SMAC 83

Proceedings of the Stockholm Music Acoustics Conference July 28—August 1, 1983

Anders Askenfelt Si Felicetti Erik Jansson Johan Sundberg
Editors

SMAC 83
CAS IAERS
STOCKHOLM

Volume I

Publications issued by the Royal Swedish Academy of Music No. 46:1
1985

Anders A.

SMAC 83

Proceedings of the Stockholm Music Acoustics Conference
July 28—August 1, 1983

Anders Askenfelt Si Felicetti Erik Jansson Johan Sundberg
Editors

Volume I

Publications issued by the Royal Swedish Academy of Music No. 46:1
1985

ISBN 91-854 28-38-8 © *Kungl. Musikaliska Akademien, Stockholm, Sweden*
ADEBE Reklam och Tryckservice AB, Sundbyberg, Sweden

VOLUME I

TABLE OF CONTENTS

VOLUME II

TABLE OF CONTENTS

PREFACE

From July 28 to August 1, 1983, the Music Acoustics Committee of the Royal Swedish Academy of Music arranged the Stockholm Music Acoustics Conference, "SMAC" as a joint meeting of the Catgut Acoustical Society and the International Association for Experimental Research in Singing as an integral part of its converence series "The International Decade of Research in Singing". SMAC was attended by about 200 participants from all over the world. Most participants presenting a paper at the Conference agreed to publish an article on their presentation in these Proceedings. As a consequence, the Proceedings appear in two volumes. Basically, all articles appear in the form in which they were submitted by the author.

The theme of SMAC was twofold: the acoustics of the singing voice and the acoustics of string instruments; all papers pertaining to the singing voice topic have been placed in the first volume and all papers on the string instruments appear in the second volume.

Only few authors used the opportunity of presenting sound illustrations on an EP gramophone record, and actually, all papers supported by such illustrations belong to the singing voice topic; for this reason, only the first volume contains a gramophone record.

A concert at the Music Museum with Carleen Hutchins' New Violin Octet was included in the SMAC program. There, the prize winning pieces of a specially arranged composition contest were performed for the first time. The concert, organized by Mr. Semmy Lazaroff, and the contest were supported by the Catgut Acoustical Society and the Henneberg Foundation. The conference was held at the Royal Institute of Technology and was supported by the Voice Foundation, the Bank of Sweden Tercentenary Foundation, the Royal Swedish Academy of Music, and the Swedish Council for Planning and Coordination of Research. Mervi Moisander-Webster kindly assisted in typing the manuscripts.

KTH, November 9, 1984

Anders Askenfelt   Si Felicetti   Erik Jansson   Johan Sundberg

Editors

ACOUSTIC PARAMETERS OF THE VOICE SOURCE*

Gunnar Fant
Dept. of Speech Communication and Music Acoustics, KTH, Stockholm

Abstract

This is a review of recent work on acoustic modeling of the human voice source to bring out general relations between time and frequency parameters of glottal flow and the constituent of a production mode. A single slope parameter is not sufficient to describe the voice source spectrum. At low voice intensities, the low frequency parts attain a dominance. Consequences for defining phonation efficiency are discussed. The analysis includes some recent findings about glottal friction and inductance as second order factors influencing vocal pulse skewness. The main emphasis is on glottal source variations with increasing voice intensity and pitch. A differentical analysis of the lung pressure, glottal maximum opening, closure speed, degree of discontinuity at closure together with supraglottal articulation as determinants of voice intensity is undertaken. This analysis is related to various modes of voice production, such as pressed voice versus a flow mode and normal voice versus breathy voice. Some experiments on analysis of the phonation at a gliding pitch are discussed to illustrate some systematic trends in parametric variations.

Introduction

The acoustics of speech production is based on the concept of a source and a filter function - in a more general sense, a raw material and a sound-shaping process. In current models the source of voiced sounds is represented by a quasiperiodic succession of pulses of air emitted through the glottis as the vocal cords open and close and the filter

---

* This paper will also be publ. in Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control.

function is assumed to be linear and time invariant. Speaker specific and contextual factors have not been given much attention. In spite of these shortcomings, speech synthesis has gained a fair quality but there remains much to improve in terms of voice quality and we still have difficulties in synthesis of female and children's voices.

In the last few years it has become apparent that in order to carry out a meaningful descriptive work we need a firmer theoretical basis of voice production including parameterization and data collection techniques, such as time- and frequency-domain inverse filtering and spectrum matching. The concept of source and filter will differ in a maximally true model of human voice production and in terminal analog synthesizers. Depending on the principle of synthesis, there exists a variety of combinations of source and filter functions that will provide one and the same, or approximately the same, output. Even in the "true" physiologically and physically oriented model there exist alternative choices of source and filter function.

In general, source and filter have a mutual dependency or rather a common dependency on underlying phonatory and articulatory events mediated by mechanical and acoustical interaction. Thus, extreme articulatory narrowing disturbs the aerodynamic flow pattern and causes significant source changes. A vocal cord abduction changes not only the voice source but also the glottal impedance as part of the overall filter system.

Source-filter decomposition of voiced sounds

The major theoretical complication in human voice production models is that in the glottal open state the sub- and supraglottal parts of the vocal tract are acoustically coupled through the time-variable and non-linear glottal impedance, whereas when glottis is closed the sub- and supraglottal systems execute approximately free and separate oscillations. Resonance frequency and especially bandwidths may differ significantly in the two states, the major effects being seen as a "truncation" of formant time-domain envelopes in the open state, typical of F1 of maximally open back vowels (Fant, 1979, 1980, 1981, 1982; Ananthapadmanabha and Fant, 1982; Fant and Ananthapadmanabha, 1982).

The physically most complete speech production model that has been developed is that of Ishizaka and Flanagan (1972) and Flanagan et al (1975). With the two-mass model of the vocal cords incorporated in a distributed parameter system, their model does not have a specific source in the linear network sense. It is a self-oscillating and self-adjusting system, the main power deriving from the expiratory force as represented by the lung pressure.

In the work of Ananthapadmanabha and Fant (1982), the acoustical modeling of voice production starts by assuming a specific glottal area function $A_g(t)$ within a fundamental period and a specific lung pressure, P_1. The flow and pressure states in other parts of the system are then calculated by techniques similar to those of Ishizaka and Flanagan (1972) leading to numerical determinations of the glottal volume velocity $U_g(t)$, the output volume velocity at the lips $U_o(t)$, and the sound pressure $P_a(t)$, at a distance of a centimeters from the speaker's mouth. By defining the filter function as the relation of $P_a(t)$ to $U_g(t)$ we define $U_g(t)$ as the source or we may go to the underlying glottal area function, $A_g(t)$ as conceptual reference which may attain the dimensionality of a flow source by being multiplied by a constant particle velocity $v_{go} = (2P_1/\varrho k)^{1/2}$. This "shortcircuit" or "no-load" source

$$U_s(t) = v_{go}A_g(t) \qquad (1)$$

differs from the true glottal source

$$U_g(t) = v_g(t)A_g(t) \qquad (2)$$

in terms of the particle velocity function $v_g(t)/v_{go}$ which carries two main features. One is a pulse skewing effect (Rothenberg, 1980, 1981; Ananthapadnamabha and Fant, 1982; Fant, 1982) accounting for a latency of the peak of $U_g(t)$ compared to $A_g(t)$, a lower rise towards the peak, and a faster fall. The second feature is the existence of oscillatory ripple, usually of a frequency close to F1, superimposed on the glottal flow pulse.

We now have a choice of two different sources in the production process. With the more basic $A_g(t)$ proportional source we have a complicated nonlinear time-variable filter function to cope with. Starting with the true glottal flow $U_g(t)$ as a source, the associated filter function is simply the supraglottal transfer function which is linear and only slowly time variable. On the other hand, the calculation of $U_g(t)$ is just as complex as calculating the output $P_a(t)$ and $U_g(t)$ contains properties of the entire sub- and supraglottal systems and is thus not invariant with the particular articulation.

The time- and frequency domain, principal representation of voice production in Fig. 1 incorporates these two alternative source functions

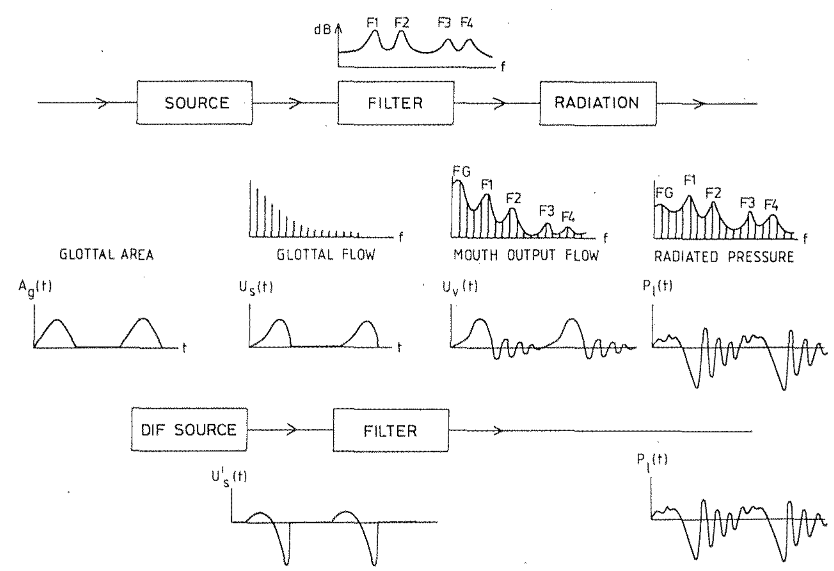


Fig. 1. Source-filter representation of voiced sounds. The radiation transfer function is included as a differentiationof the source in the lower part.

and in addition the differentiated glottal flow $U'_g(t)$ as a source with mathematically means that the differentiation inherent in the radiation transfer from lip flow $U_o(t)$ to sound pressure $P_a(t)$ is removed from the filter function and incorporated in the source function which enhances important properties of the source. The negative spike in $U'_g(t)$ representing the slope of $U_g(t)$ at closure is a proportionality parameter for all formant amplitudes (Fant, 1979; Gauffin and Sundberg, 1980). It may also be seen in the speech wave $P_a(t)$ at low F_0.

Another source characteristic apparent in the output at low F_0, high F_1, and low or medium voice intensity is a spectral maximum somewhere in the vicinity of one of the lowest harmonics, the relative amplitude of which stays rather invariant whilst F1 and higher formants gain in amplitude at higher voice effort (Fant, 1979, 1980). At a very low voice effort the source peak may dominate over F_1.

Interaction effects

The variability of the pulse skewing effect comparing different vowel articulations is illustrated by the model calculations of Fig. 2 from Ananthapadmanabha and Fant (1982). Vowels produced with large vocal tract inductance, i.e., with a narrow but not extreme constriction, such as[a] with constricted pharynx and [i] with constricted mouth have a somewhat higher $U'_g(t)$ at closure than other vowels and thus, a greater intrinsic intensity. The variations are not large, of the order of a decibel.

Superposition of formant ripple from previouas glottal periods may add to the complexity of the glottal flow waveform as illustrated by Fig. 3. This effect may be prominent at very low F_1 and relative high F_0 causing a seemingly random perturbation of glottal flow waveshape from one period to the next. The within a period induced riplle function already to be seen in Fig. 2 is especially apparent in the model calculations of Fig. 4 pertaining to a vowel [a] produced at F_0=170 Hz. A double peak is seen in the waveform of the differentiated glottal flow and the spectral correlates include a prominent antiresonance just above F_1. Such spec-

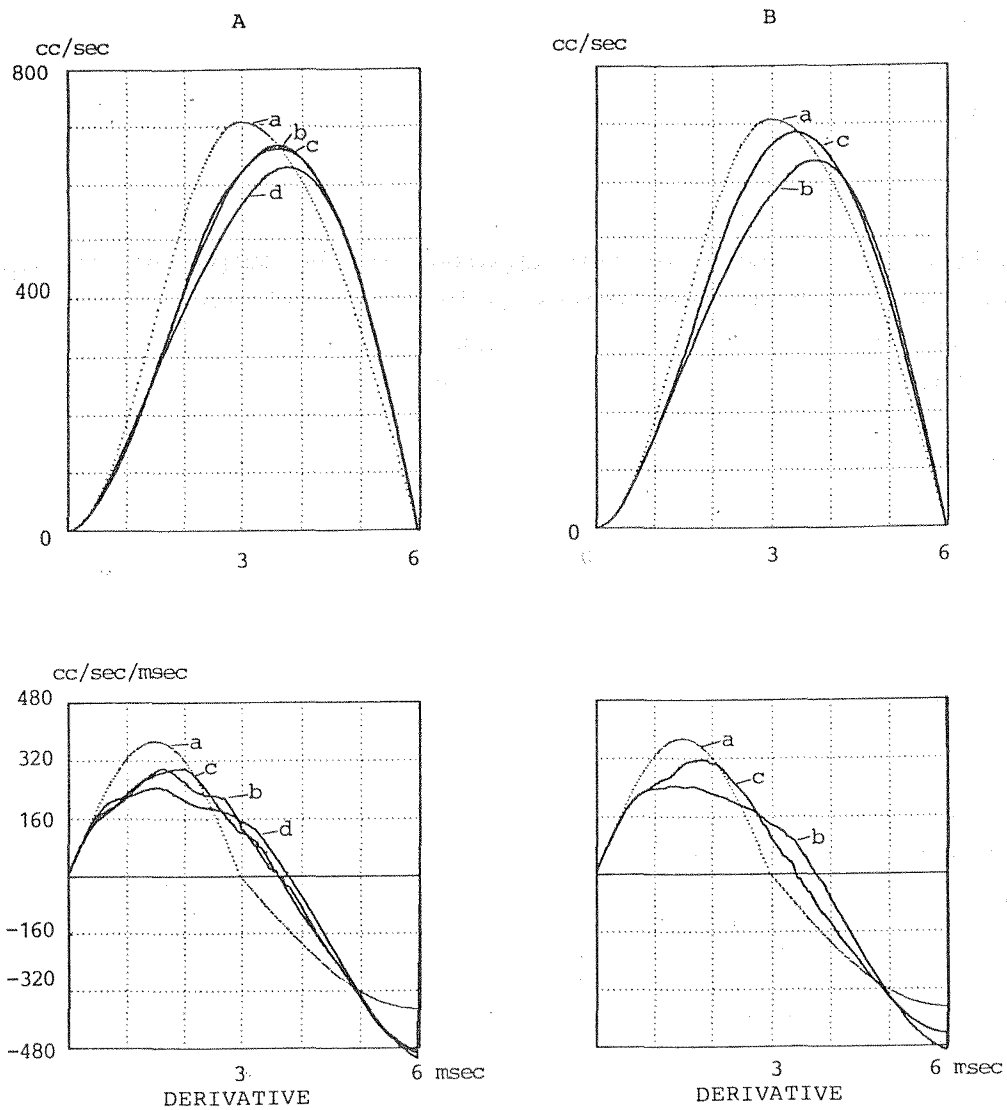


Fig. 2. Glottal source pulse and its derivative for vowels.
A: a) no load; b) vowel [a]; c) vowel [o]; d) vowel [u].

B: a) no load; b) vowel [i]; c) vowel [ɛ] (from Ananthapadmanabha and Fant (1982) Fig. I-A-6).

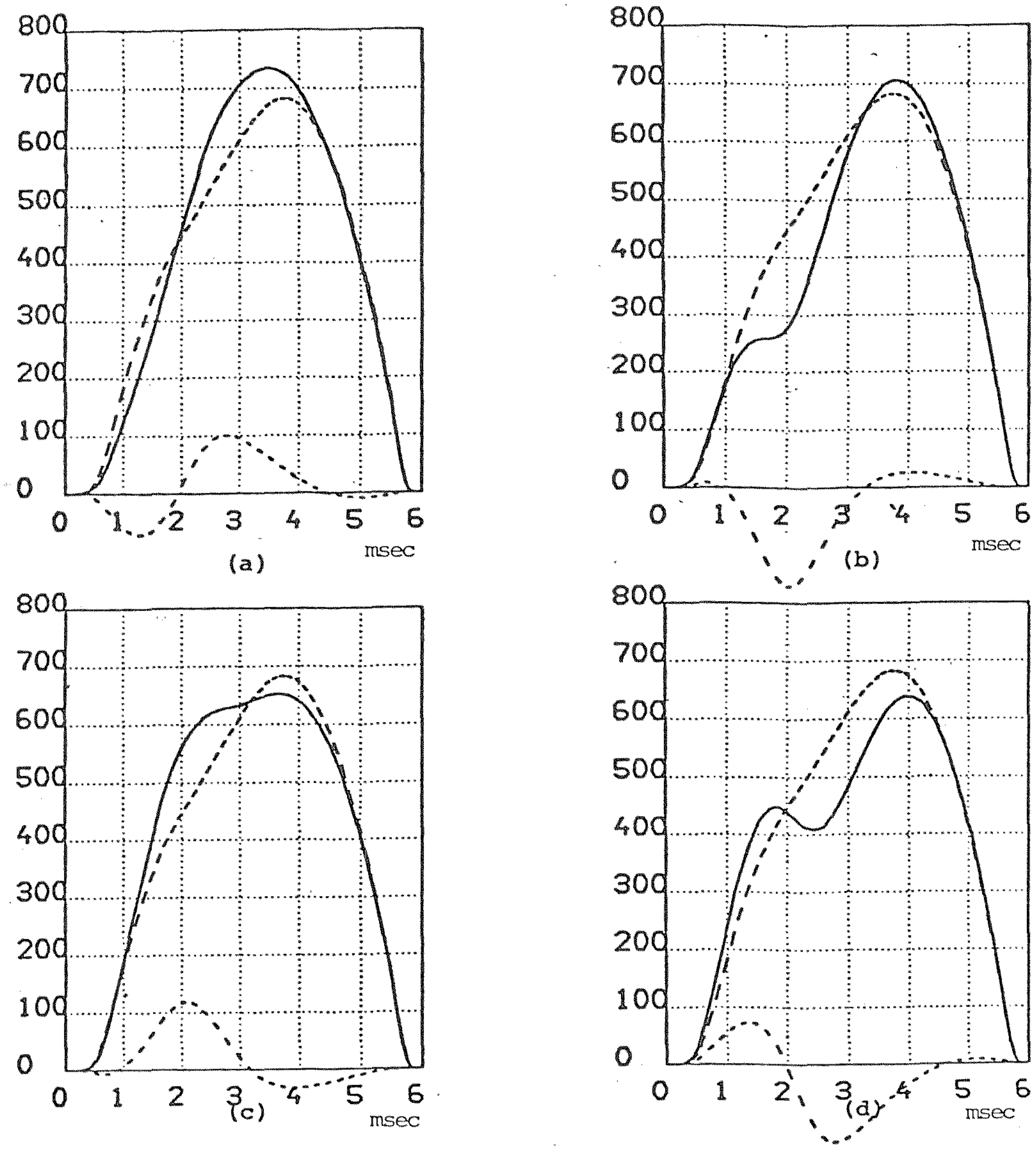


Fig. 3. Effect of superposition on the glottal flow pulse. Dotted line, without superposition; solid line, with superposition.
a) closed phase 6 ms; b) closed phase 7 ms;
c) closed phase 4.75 ms; d) closed phase 3.625 ms.
F1 VT load with F_1 = 220 Hz
(from Fant and Ananthapadmanabha (1982) Fig. 5.

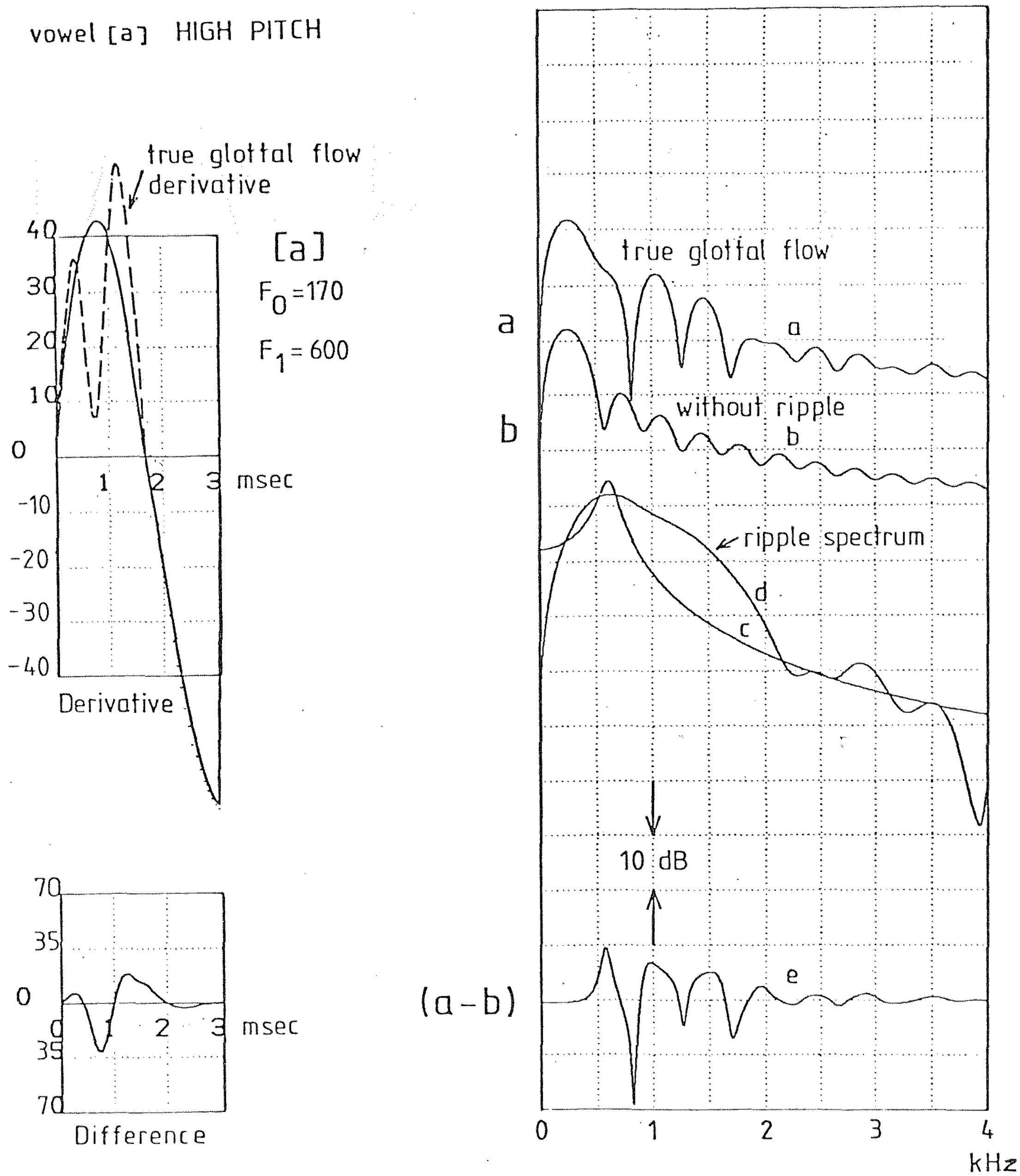


Fig. 4. Example of differentiated glottal flow showing double peak. Vowel [a], F_1 = 600 Hz, F_0 = 170 Hz (from Fant and Ananthapadmanabha (1982) Fig. 6.

tral irregularities vary with the duration of the glottal open period and thus with F_0, see Fant and Ananthapadmanabha (1982).

Eventhough the fine structure of the $U_g(t)$ source spectrum becomes complex the time domain correspondence referring to the output sound may be simpler. Part of the glottal flow ripple is a mathematical prerequisite to a detuning and truncation of a formant oscillation when it reaches next vocal open period. One can visualize this as the addition of a component in the flow which will detune and partially cancel the formant oscillation which otherwise would have proceeded with a constant rate of damping.

This view also leads us to an approximate model where the source pulse is ripple free but properly skewed versus the glottal area function $A_g(t)$. Such a source supplied as a constant current source to the vocal tract supraglottal system in parallel with the time-varying glottal- plus subglottal impedance was adopted in the modeling of Fant (1979, 1981). During the glottal open phase the shunt impedance draws current which is the negative of the ripple component in the vocal tract input flow.

By removing the time variable shunt and omitting the pulse skewing effect, we end up with the noninteractive model of a terminal analog formant synthesizer excited by constant shape smooth pulses. In order to make this synthesis model more comparible with the true model, we could adjust the pulse skewing according to the particular vocal tract inductance and the particular glottal impedance. Furthermore, we need to introduce truncation and maybe also detuning by modulating, in the first hand F_1 and B_1 within the glottal open period. An alternative close to the "true" model is to start with an $A_g(t)$ source if known and to calculate the current into a Foster reactance network representation of the vocal tract load approximated by F1 and F2 parallel resonance circuits. This procedure first suggested by Mryati et al (1976) also takes into account the main pulse skewing. The load current is then transferred as the input to a conventional formant synthesizer.

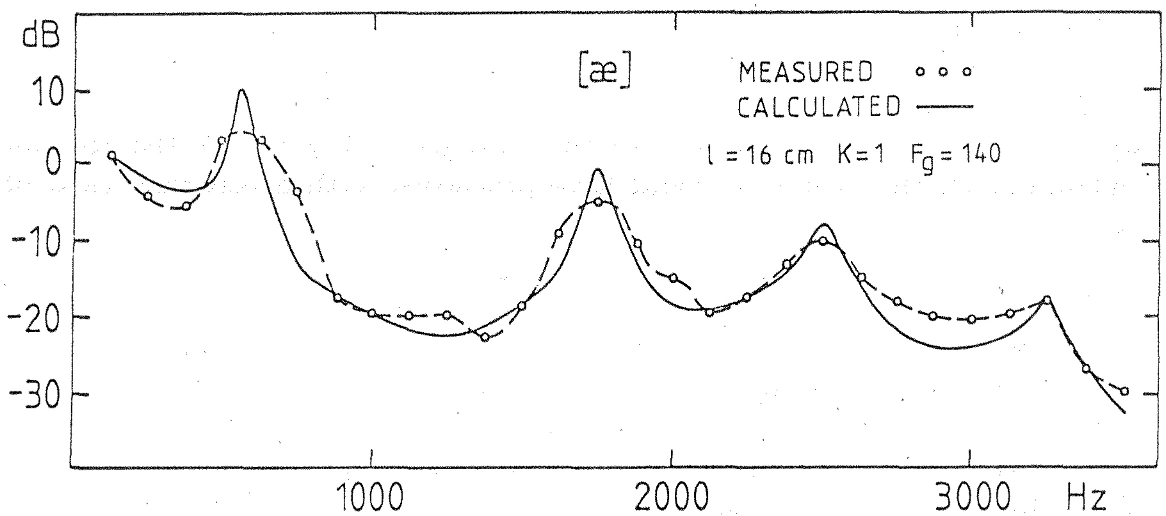


Fig. 5. Spectrum matching with noninteractive model.

We are now in a process of evaluating the perceptual importance of the interactive factors in production (Ananthapadmanabha et al, 1982). It appears that as long as relative formant amplitudes, including F_g are preserved, the perceptual differences are small. A preservation of the spectral fine structure between peaks or the corresponding time domain properties of instantaneous formant frequency and rate of envelope decay appears to be less crictical.

The difference between a measured harmonic spectrum of an[æ] vowel and a synthetic spectrum generated from a non-interactive model with a Fant (1979) source is illustrated in Fig. 5. The bandwidth of formants was set to glottal closed average data of Fant (1972) which account for the overshoot of formant amplitude levels. The tendency of dispersion of spectral energy in the F1 range towards higher frequencies is typical of the detuning effect in the glottal open interval.

A parameterization of the glottal flow pulse or of the underlying glottal area function, see Fig. 6, can be based on more or less established measures such as open quotient, speed quotient, and peak sharpness as defined by Lindqvist(-Gauffin) (1967). The measure "pulse asymmetry" which is almost the same as speed quotient is suggested as a contribution to the nomenclature.

Fig. 6, drawn from the Fant (1979) model, does not include leakage. A more general model capable of representing non-abrupt return to the zero-line after the point of maximum discontinuity is included in Fig. 7. The pulse is described in terms of parabolic consequtive parts which means that the flow derivative is a piecewise linear function and the second derivative a set of step functions and a spike at the point of maximum flow discontinuity. This model is outlined here for pedagogical purposes. Higher order power functions or connecting sinusoids would be more realistic (Ananthapadmanabha, forthcoming).

The need of the extension, Fig. 7 of the Fant (1979) model, is apparent when studying inverse filtering functions from breathy phonations, such as the intervocalic voiced [h] in Fig. 8. The differentiated glottal flow reveals the phase of continuing closing movement after the point of maximum discontinuity which perturbs the negative spike towards a slightly smoothed triangular shape, thus providing an overall picture of a fullwave rectified pulse train. The reduced efficiency in formant excitation of this pulse shape combines with a substantial F1 bandwidth increase which typically reduces the F1 amplitude in breathy phonation. This is an example of covariation of source and filter functions.

Source spectrum shapes are not adequately described by a single slope value only. An important feature derivable from the particular combination of the F_g, F0, and K parameters (Fant, 1979) is the amplitude level at some higher frequency, e.g., at 1000 Hz. In connected speech, the F_g level stays rather invariant whilst formant levels in addition to F-pattern-induced variations tend to fall off as a result of a progressive

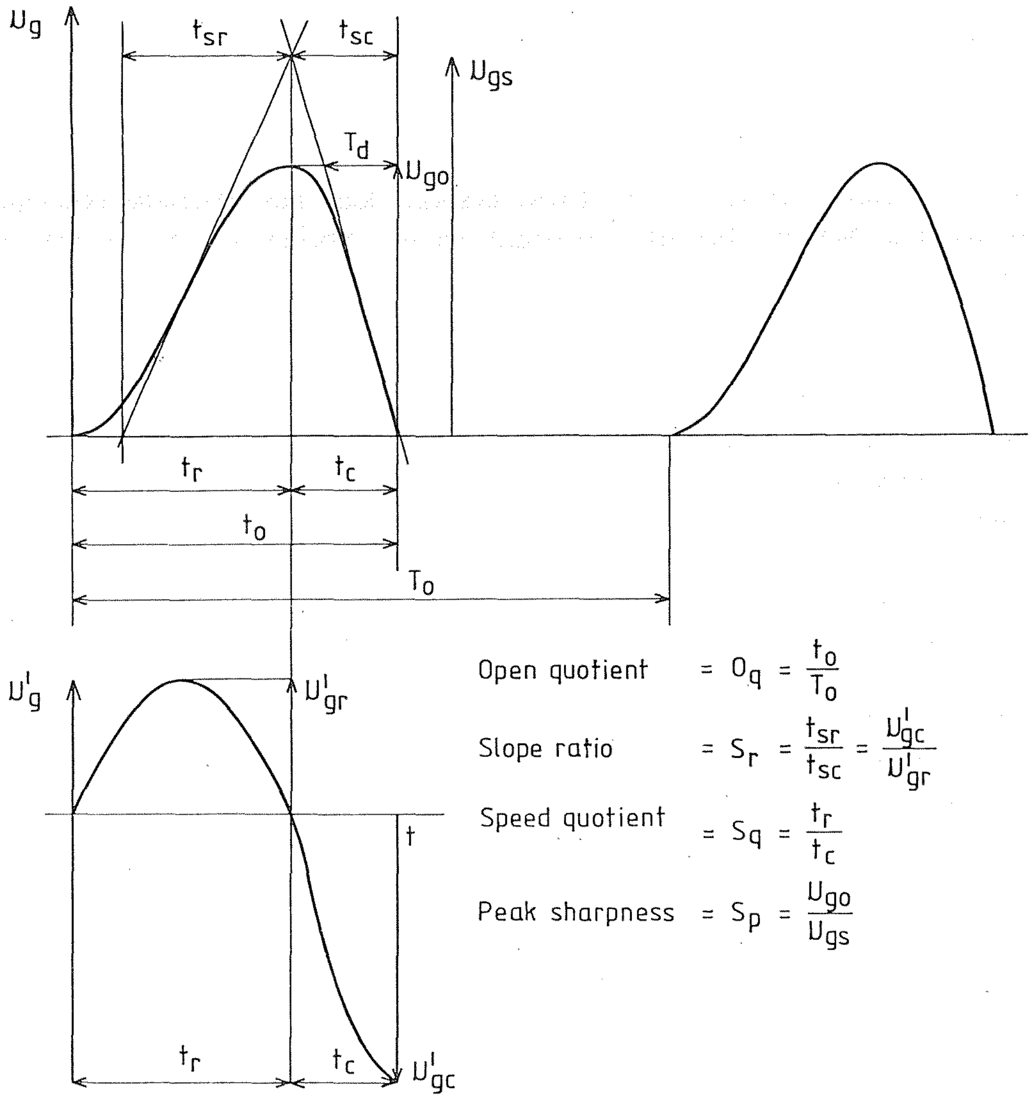


Fig. 6. Pulse parameters for idealized case with complete closure for describing glottal flow or glottal area function.

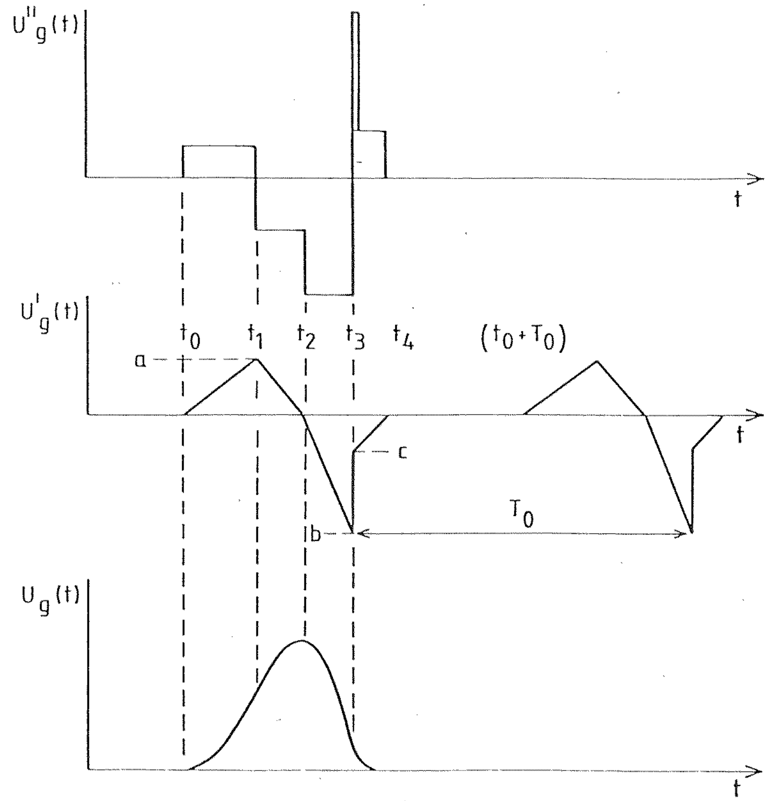


Fig. 7. Stylized, more general model of glottal flow and its first and second derivative.

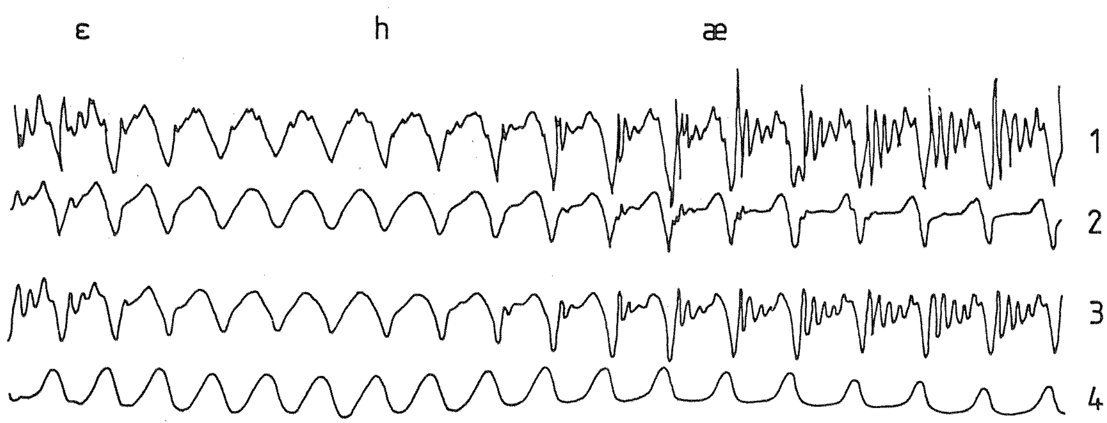


Fig. 8. Intervocalic voiced [h], 1 = oscillogram; 2 = differentiated glottal flow from inverse filtering; 3 = glottal flow; 4 = same as 3 but without F1 cancellation (from Fant and Ananthapadmanabha (1982), Fig. 2.

abduction, e.g., at the termination of voicing. Similar effects occur as a reaction from supraglottal narrowing when extended beyond that of close vowels. The amplitude level of F1 relative to that of F_g and also the absolute level of F_g are important correlates of breathy phonation (Bickley, 1982). Rules for predicting such relative amplitudes have to be based on the entire production process.

The relative role of lung pressure as a determinant of voice output intensity has been overestimated. It has been found experimentally that a doubling of subglottal pressure is associated with a raise of the overall sound pressure level by about 9 dB (Ladefoged and McKinney, 1963; Isshiki, 1964). However, if we assume that the glottal area function and F_0 remain the same, the increase in particle velocity and, thus, in glottal flow from Eq. (1) is 3 dB only. At increasing pressure the pulse skewing decreases and the net gain in flow derivative at closure would be slightly lower than 3 dB. What is then the origin of the remaining 6 decibel? As shown by Fant (1982), it may be explained by several covarying factors. Increasing subglottal pressure increases the maximum deflection of the cords and we may expect a 3 dB-increase of the overall scale factor of the glottal area function. Furthermore, the pulse duration and especially the duration of the closing phase decreases and the fundamental frequency increases somewhat. There is also a small net in the pulse skewing related to the increase in rate of change of glottal area at closure. All these factors added together account for the missing 6 decibel. This example could illustrate a "flow" type of voice. If the cords are subjected to a medial counteracting the pressure increase, the net gain in intensity would be related to a shortening of the glottal pulse rather than to an increase in maximum glottal area. In speech it is found that maximum glottal flow varies rather little with increasing voice intensity (Lindqvist-Gauffin, 1965; 1970). Apparently, there exists a large range of normal covariation of the physiological factors determining voice intensity.

These examples indicate the need of anchoring the rules of source dynamics on an insight in the underlying, more basic physiological parameters and production events. On the whole, it appears that abduction-adduction and other vocal cord activities carry the major part of the source dynamics, whilst lung pressure variations are slower and carry less information.

A promising attempt of constructing a voice source with physiologically oriented parameters is that of Rothenberg et al (1975). We are now looking forward to widening our knowledge base to meet the demands of improved and better quantified production rules and synthesis strategies.

References

Ananthapadmanabha, T.V. and Fant, G. (1982): "Calculation of true glottal flow and its components", STL-QPSR 1/1982, pp. 1-30.

Ananthapadmanabha, T.V., Nord, L., and Fant, G. (1982): "Perceptual discriminability of nonexponential/exponential damping of the first formant of vowel sounds", pp. 217-222 in (R.Carlson & B.Granström, eds.) Proc. of The Representation of Speech in the Peripheral Auditory System, Stockholm, May 17-19, 1982, Elsevier Biomedical Press, Amsterdam.

Bickley, C.(1982): "Acoustic analysis and perception of breathy vowels",MIT, Speech Communication Group, Working Papers, Vol. 1.

Fant, G. (1972): "Vocal tract wall effects, losses and resonance bandwidths", STL-QPSR 2-3/1972, pp. 28-52.

Fant, G. (1979): "Glottal source and excitation analysis", STL-QPSR 1/1979, pp. 85-107.

Fant, G. (1980): "Voice source dynamics", STL-QPSR 2-3/1980, pp. 17-37.

Fant, G. (1981): "The source-filter concept in voice production", STL-QPSR 1/1981, pp. 21-37.

Fant, G. (1982): "Preliminaries to analysis of the human voice source", to be publ. in Working Papers, Massachusetts Institute of Technology, Speech Group.

Fant, G. and Ananthapadmanabha, T.V.(1982): "Truncation and superposition", STL-QPSR 2-3/1982, pp. 1-17. 2-3/1982, 1-17.

Flanagan, J.L., Ishizaka, K., and Shipley, K.L. (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", Bell System Technical Journal 54(3), 485-506.

Gauffin, J. and Sundberg, J. (1980): "Data on the glottal voice source behavior in vowel production", STL-QPSR 2-3/1980, pp. 61-70.

Isshiki, N. (1964): "Regulatory mechanisms of voice intensity variation", J. Speech Hearing Res. 7, pp. 12-29.

Ishizaka, K. and Flanagan, J.-L. (1972): "Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell System Techn. J. 51, pp. 1233-1268.

Lindqvist-Gauffin, J. (1965): "Studies of the voice source by means of inverse filtering", STL-QPSR 2/1965, pp. 8-13.

Lindqvist-Gauffin, J. (1970): "The voice source studied by means of inverse filtering", STL-QPSR 1/1970, pp. 3-9.

Ladefoged, P. and McKinney, N.P. (1963): "Loudness, sound pressure and sub-glottal pressure in speech", J.Acoust.Soc.Am. 35, pp. 454-460.

Mrayati, M., Guerin, B., and Boe, L.J. (1976): "Etude de l'impedance d'entre du conduit vocal-couplage source - conduit vocal", Acustica, 35, pp. 330-340.

Rothenberg, M., Carlson, R., Granström, B., and Lindqvist-Gauffin, J. (1975): "A three-parameter voice source for speech synthesis", in (G.Fant, ed.) Speech Communication, Vol. 2, Almqvist & Wiksell, Stockholm.

Rothenberg, M. (1981): "An interactive model for the voice source", STL-QPSR 4/1981, pp. 1-17.

REPORT ON VOCAL REGISTERS

Harry Hollien
University of Florida, Gainesville, FL, USA

Abstract

This report emanates from one of the programs sponsored by the Collegium Medicorum Theatri; it results from the efforts of a rather substantial number of individuals. I am honored to be associated with them; indeed, it is a profound honor to be able to report their deliberations and conclusions to you.

The Collegium Medicorum Theatri (CoMeT) is a society -- by invitation -- which consists primarily of physicians who are concerned with, and treat, individuals who use their voices professionally; they are especially interested in singers and actors. Since this group fosters/encourages relevant research, the physicians are joined in their efforts by a few phoneticians (voice scientists) and engineers plus a small group of voice teachers/pedagogists.

The CoMeT committee on vocal registers was formed two years ago at the Seventh CoMeT Congress, Amsterdam. It consists of fifteen regular members plus three ex officio contributors -- over half of the group are CoMeT members. The specialty breakdown is physicians: 45%, scientists/engineers: 35%, voice pedagogists: 20%. Members are Giuseppe Bellussi (Italy), Friedrich Brodnitz (U.S.A.), W.S. Brown, Jr. (U.S.A.), Beverley Johnson (U.S.A.), Jens-Peter Köster (West Germany), Van Lawrence (U.S.A.), Richard Miller (U.S.A.), Aatto Sonninen (Finland), Johan Sundberg (Sweden), Craig Timberlake (U.S.A.), Ingo Titze (U.S.A.), J.B. van Deinse (Netherlands), Jurgen Wendler (East Germany), Fritz Winckel (West Germany), I have the honor to chair the committee. Ex officio members include CoMeT officers Hans von Veden and W.J. Gould (both U.S.A.) plus a consultant Carol Schoenhard (U.S.A.). Several other individuals have contributed to this report -- especially Suzanna Naidich (Argentina), Minoru Hirano (Japan), and Wolfram Seidner (East Germany). With only one or two exceptions, all committee members have been active in our deliberations -- many have contributed extensively.

The five "positions" to be presented represent a consensus -- often an overwhelming one -- of the committees' collective opinion. Ordinarily, I will not attribute a particular position or issue or argument to an individual by name; that is, except in those cases where they were particularly eloquent or (perhaps) when they organized the minority report. A substantial number of issues -- other than those articulated here -- have been considered. Those postulates/controversies/issues not covered in this report were considered either not relevant, too trivial for major review or incomprehensible. The positions to be reviewed are five in number.

1. Registers Exist

Voice registers have been shown to exist; hence they cannot be ignored. There is a substantial pedagogigical history in this regard, antedating (in time) even Garcia. Moreover, a substantial number of perceptual and acoustic studies have been carried out on singers with a consequence that they have demonstrated the existence of this vocal phenomenon. The minority, in this case, is led by Johnson and Winckel who argue that only untrained singers exhibit registers -- that it is only individuals not yet accomplished who distort their phonatory productions in a manner where various register qualities and breaks become apparent. They also argue that many great singers developed their voices without even being aware of the register concept; that "smoothness of scale and tone" more functionally relate to good vocal development and habits than do approaches which train "sections" of the voice (or elements within it) in piecemeal fashion.

When the two positions are contrasted, however, it becomes apparent that this controversy relates to what is desirable rather than to what exists in singers' voices. It is upon this basis that the two positions can be reconciled. In this regard, consider the relevant experiments that have been reported. Many investigators have demonstrated that accomplished singers can produce phonations in specific voice registers - even when vowel, fundamental frequency, and vocal intensity are controlled. The differences in register phonation can be heard even in those cases where the auditors are making blind judgments (i.e., those only on the basis of quality or timbre). Further, acoustic differences can be observed in the spectra of tones sung in different registers. In any case, the contrasts are even more apparent when the controls are removed from the other three vocal parameters (f0, vowel, intensity). Thus, the models, data, and descriptions provided by investigators such as Titze, Sundberg, Bellussi, and Hollien simply cannot be ignored.

But how then can these findings be made to compliment a "one register" position? Very simply. If the singer is trained to modify his or her phonation in an effort to compensate for register effects -- and is successful in doing so -- a single "register" (or no register at all) will result. However, the position of this committee is that the voice characteristic defined as register exists in the voice of singers and must be recognized as an entity. Of course, good teaching can remove or conceal its effects -- if such behavioral change is considered desirable.

2. Singing/Speaker Registers

Vocal registers in singing and voice registers in speaking (or in the untrained voice) are different and separate entities and must be treated as such. It is conceded, of course, that these two particular characteristics may overlap in function and probably have common physiological roots in the larynx. None-the-less, it must be recognized that, while vocal registers exist and are sometimes used in speaking, no attempt is made to "train" them out of the productive repertoire of the speaker. Moreover, a physiological register (vocal fry, pulse, creak) exists in speaking that is virtually nonexistent in singing. The most serious problem in this regard relates to confusions resulting from research reported in the literature. Specifically, vocal register studies carried out on non-singers cannot be extrapolated to singers on the basis of some simple relationship (Brodnitz warns of this danger). Such is often the case, however, and to the detriment of our thinking.

As we have conceded, registers also can be observed in the phonatory productions of singers. However, their uncontrolled presence apparently occurs primarily in the voices of the beginning, untrained, and/or "not-yet accomplished" singers. The artistic level or professional singer, on the other hand, works to remove or conceal the influence of registers on tone, line or timbre; that is, this approach is one taken by those singers trained in the western/classic, operatic/concert mode.

As it turned out, very little controversy occurred within the committee relative to this "position". Wendler, however, argued that because speaking registers are not essential even in artistic speaking (and do not result in problems related either to education or performance), they constitute an academic issue of no practical relevance. Perhaps this position is a little extreme but, if Sonninen's suggestion that we provide tape recorded samples of tones sung in various registers (for distribution throughout the world) is to be implemented, we probably should structure different sets to illustrate registers in speaking and in singing.

3. To "Remove" Registers

The third position taken by the committee consists of a paired set of postulates. As it turns out, this issue was, at once, the most controversial of all yet the easiest one to solve. The postulates can be stated as follows: 1) Eliminating or concealing register effects is desirable for the classical/western, concert/opera mode of singing and 2) Register effects can be fundamental to certain types of singing but individuals who utilize these effects should be trained to sing with minimum abusive effects to their voices.

Bellussi and Titze are quick to remind us that a teacher does not "remove" registers from a singer's voice; perhaps Wendler says it best. "Registers cannot be removed as they are physiologically given." Thus, it appears that it is important to conceal registers and/or their effects by training the student to equalize or balance the different sound qualities (from the different registers) until they are no longer perceptible -- that is, in those instances where such modifications are desirable and/or beneficial to the singer. We are reminded by Sundberg and Naidich, however, that there are valid types of singing where register timbre, or breaks, are used for dramatic effect -- or where the particular nature of the singing calls for production of this type of contrast.

4. The Source of Registers

The committee accepted the notion that there probably are two sources for registers -- the larynx and the vocal tract. Yet this seemingly universally agreed upon position also provoked some of the sharpest differences amongst committee members. For example, Sundberg is adamant that a voice register must originate in the larynx. On the other hand. Hollien appears to have retreated somewhat from this position. Along with Gould, he now is willing to accept the possibility of mechanical register-type operations orginating in the vocal tract (i.e., in the supra glottal oral-facial structures) and that these events may be related to "singers" registers. Further, Hollien and Schoenhard concede the possibility that the so-called middle register may illustrate this class of events. In an effort to establish a reasonable dialogue here -- and perhaps provide structure to the issue -- Wendler (and Bellussi) suggest separation of our register model into physiologic and artistic categories. Presumably, the physiological element would be equated with laryngeal registers (and speaking registers?) whereas "artistic" registers would appear related to events in the vocal tract -- or even, perhaps to the use of registers for effect.

To summarize. The dialogue relative to the source or sources of vocal registers ultimately resulted in a general agreement that there may be multiple sources for these phonatory activities. However, it should be stressed that a substantial minority of the committee argued in favor of the source (of a voice register) being only laryngeal and that the other so-called register-like phenomena actually are some sort of quality/timbre events. It is doubtful that this issue can be resolved except by further research.

5. Labels

The labeling of vocal registers is proving to be the thorniest of all problems. As it turns out, a number of committee members support the use of "the old labels." But what are these old ways and old labels? Note the confusion that would result from an attempt to use the materials presented by Fig. 1 in resolving this issue. As can be seen, the figure consists of portrayals of the efforts of four authors to name, number and specify only the frequency extent of voice registers. In any case, are Garcia's materials, the "old" ones we should use? Are his ranges accurate for all singers? Where did the labels he uses come from? What about his specification of three registers? Surely Garcia was referring to the untrained singer -- or, perhaps students -- when he structured his system. Are Appleman's and Vennard's suggestions inferior to Garcia's? Surely these gentlemen are distinguished scholars and pedagogists. Why, then, are the labels and ranges they use different from Garcia's? To confuse things yet a little more, Hollien can be seen to focus his efforts primarily upon non-singers or speakers -- but, is it possible that his estimates and terms are valid for singers also? There are many other distinguished scholars and scientists who have attempted to develop structures such as those seen in Fig. 1. Add the work of Sundberg, Bellussi, Titze, Large, Schoenhard and others to the figure and the picture becomes even more scrambled.

So, what is the answer to this problem of identifying voice registers? Before listing the proposals made by the committee, a rather startling position (perhaps one that should be articulated separately) must be presented. This opinion surfaced when the committee members began consideration of the "label" issue and commenced evaluating the sources of these labels used most commonly. Although the term "modal" may be overtaking them, the two labels that appear more often than any others (in vocal music anyway) are "chest" and "head" -- with chest referring to the lower register and head for the higher (if, indeed, there is only one higher register). As we all know, these terms are based upon singers' sensations -- i.e., on the mechanical response of the

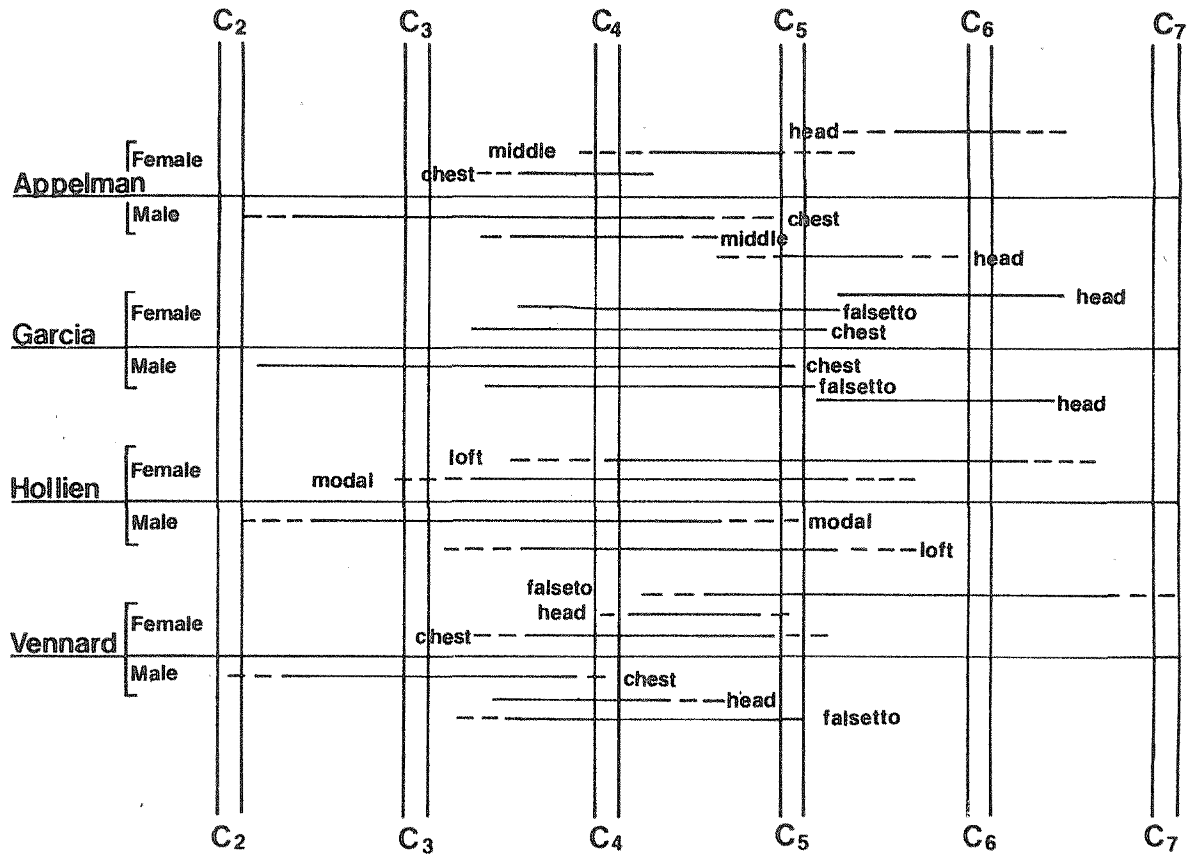


Fig. 1. Vocal registers as a function of frequency (and selected author).

bodies of singers to tones sung at certain frequencies or within certain frequency ranges. The generic connotation of these terms is such that they suggest certain relationships -- specifically that the timbre for the lower register resides in the chest; that the quality/mechanism of the upper register results from activation of the vibratory properties of the sinuses and/or cavities in the head or "mask." Once these entities are considered physiologically, and/or mechanically, their use as definitions for vocal registers is shown to be illogical if not absurd. The lower register results from operation of the larynx -- not from sympathetic vibrations of the chest to low frequency sung tones. The source of the upper register again is the larynx -- not sympathetic vibrations (to higher sung frequencies) in the face. In short, while the sensations felt by singers, of course, are valid sensations (indeed, even the non-singer can experience them), they have nothing to do with vocal registers! It is only a chance relationship that brings the two into juxtaposition; that is, voice registers are frequency related and so are the sites of the sympathetically vibrating, sensation producing, structures of the torso and head. What a classic case of misdirection this is! For three hundred years, correlary but independent operations have been viewed as related -- even casual -- yet they were not and are not. Worst yet, this seeming relationship has led scholar after scholar astray -- in the past anyway. Not so this committee; please note the following.

As stated, the majority of this committee rejects the so called "old terms." Sundberg suggests that he would be most comfortable if the registers were numbered; Wendler agrees. He further suggests that the values we used in our discussions might be appropriate. They were: # 1: the very lowest of registers, probably used only in speaking (old terms: pulse, vocal fry, creak), #2: that (low) register, which is used for most speaking and singing (old terms: modal, chest, normal, heavy), #3: a high register used primarily in singing (old terms: falsetto, light, head), #4: a very high register usually found only in some women and children and not particularly relevant to singing (old terms: flute, whistle). In addition we referred to yet another "register" as "#2A" and defined it as that "register" which is described by many voice teachers as in the middle of the frequency range; as one constituting an important problem in the training of many singers (old terms: head, mid, middle, upper). As is obvious from the research literature, this register is rather difficult to demonstrate empirically but it receives so much (subjective) support, it cannot be ignored.

An alternate approach to the suggested one would be to use, as labels, a pair of terms that kept creeping into virtually every dialogue we had: i.e., heavy, light (Vennard, Naidrich, Schoenhard) or a second set: lower, upper (relates registers to fundamental frequency). Few, if any of the committee appeared uncomfortable with either of these pairs of terms. Moreover, they do incorporate the advantage of functionally rejecting the middle register (they are laryngeally oriented) and of lending themselves to concepts relevant to the concealment of register effects. In any case, the committee (with few exceptions) appeared to favor new terms, generic terms, those that are clear and easy to understand (Köster, Miller).

In Conclusion

A final issue must be considered. Whereas logic, reasoned opinion and the available research tends to support the five positions taken by the committee, it must be remembered that, even if they were to be adopted universally, these positions can serve only to structure the concept of voice registers and to defuse several longstanding controversies. They patently do not provide the information necessary to identify and operationally describe all potential voice registers, to define the boundaries and characteristics of these phenomena, to explain their origins, sources and functions or to identify procedures useful in their modification/elimination. The materials and concepts provided by this committee simply result in a model that can be tested.

Individuals such as Titze, Sundberg, Bellussi, Hollien and Wendler all have summarized, reviewed and/or interpreted available evidence in attempts to identify and explain vocal registers. The modal articulated in this report should assist them, and others, in testing their approaches. Indeed, as nearly everyone in this committee has affirmed, the structural organization established must be testable on the basis of the perceptual, acoustic, physiologic, kinesthetic, aerodynamic and neurological approaches available. We leave many issues to be analyzed and understood. However, we hope that the work of this committee constitutes a reasonable initiative -- one that leads to the solution of the problem. We invite your input.

EPIPHYSIS VIBRATIONS OF SINGERS STUDIED BY HOLOGRAPHIC INTERFEROMETRY

Z. Pawłowski*, R. Pawluczyk**, and Z. Kraska**
*Academy of Music, Phoniatric Department, Warsaw, Poland
**Central Optical Laboratory, Department of Physical Optics, Warsaw, Poland

Abstract

The sound generated by the human vocal organ is the result of complex interactions of different parts of the human vocal organ with a flow of air. It seems that some of these parts will produce vibrations in neighboring tissues which will transmit them to the nearest area of the skin. Double-pulse holographic interferometry is applied to the present investigation of skin vibrations. Some results are presented.

Introduction

The study of human voice can be divided into two main parts:

- investigation of the sound generation mechanism,
- analysis of the generated sound.

The second of these parts does not cause any serious problems since good equipment exists both for sound registration and analysis. This apparatus produces results in a form allowing exact description of the voice in terms of Fourier spectrum. As a result, any voice can be described accurately by means of a set of formant frequencies and relative intensity.

Contrary to this, the investigation methods of the sound generation process are still unsatisfactory. The main reason of this is the complexity of the human voice generation process and, in particular, the complex structure of the human vocal organ itself.

Modern theory explaining the mechanism of voice formation considers results obtained by physiology, aerodynamics, and acoustics (Fant, 1960, especially pp. 265-272; Flanagan, 1964; Isshiki, 1964; Koike et al., 1967; v. Leden, 1968; 1969; Pawłowski et al., Źółtowski, 1982; Yanigihara and Koike, 1967). According to our present knowledge, the human vocal organ consists of two essential parts: the peripheral effector and the central analyzer interconnected by nerves into one entity. An important part of the peripheral effector is the larynx, where the sounds are generated by the air flow from the lungs. The sounds are subsequently modulated in the resonant cavities of the vocal tract. The volume of air between the surface of the glottis and the mouth opening is usually called the vocal tract (Fant, 1958; 1959; Flanagan, 1964). Some sounds are produced in this cavity connected with the nasal cavity. Apart from the vocal tract, the human resonance system consists of a series of permanently shaped and polymorphic cavities.

The development of acoustical methods during the last decades has extended our knowledge of the vocal tract function (Fant, 1960; Flanagan, 1960). As is seen from Flanagan (1960), Hirano et al. (1966), House and Stevens (1956), Kacprowski (1977; 1979), Pawłowski (1968), Pawłowski et al. (1972), Shipp (1967), Shipp and Huntington (1965), the registration of the speech wave at the mouth opening should be complemented by measurements of the acoustical pressure in individual cavities of the voice organ. Unfortunately, investigations of this kind present serious difficulties. One reason for this is that the human vocal organ in its entire complexity is hard to model (except in electronic synthesis of sound, of course), which means that the investigations must be carried out on living people. As the vocal organ is located inside the body, the introduction of an instrument is far from easy and, what is worse, it may disturb the vocal tract and, as a consequence, the sound itself.

Of course, acoustical analysis of the voice can give complete data about its Fourier spectrum, i.e., number of formants, their frequencies and relative intensity but this analysis does not reflect the complex functions of the individual parts of the vocal organ. Generally, it does not explain which ones of the numerous resonance chambers take part in the generation of particular formants or what influence the vibrations of the bones and the soft tissues have on the voice timbre, etc. The investigations of these influences are still an important problem.

It seems safe to assume that the air vibration in particular cavities as well as the vibrations of bones and soft tissues, could be transmitted through the neighboring tissues to the skin. Thus, skin vibration analysis could provide information about the localization of the stimulated resonance chambers and also show how the vibrations are transmitted. In this way analysis of the skin vibrations would carry information about the processes occurring inside the human body.

Holography seems to be the best vibration analysis method up to date. This paper presents some results obtained using this method to investigate skin vibrations of people during singing.

Experimental basis of the applied method

Holography is an optical method allowing great precision in the registration and reconstruction of any coherent wavefront. As a result, the exact reconstruction of any three-dimensional object is possible with an accuracy of a fraction of the wavelength of the light used. In a way which will be explained below, this method allows comparison of two or more wavefronts existing at different times, and as a result, the comparison of the position of the investigated object at different times with the same precision. This property of holography has been applied to vibration analysis (Collier et al., 1971; Erf, 1974; Vest, 1979).

Holography as a method of vibration analysis has some important advantages:

- contactless method of measurement,
- high sensitivity down to fractions of micrometers,
- visualization of the vibrations and displacements on a large surface of the investigated object.

On the other hand, holography is quite a demanding technique and in order to obtain a good hologram, some requirements must be met during the hologram exposure.

Coherence requirements

In the case of holographic investigations of living subjects, the coherence requirements can only be satisfied by good lasers developed for holographic needs. Without further explanation suffices it to say that the lasers for these applications must work in a single mode, i.e., in a ground transverse and a single longitudinal mode. Independently, a spectral width $\Delta\lambda$ of the emitted light must meet the following requirements:

$$\Delta\lambda < \frac{0.16\lambda^2}{l} \qquad (1)$$

where λ is the wavelength of the laser light used, l the depth of the object along the light rays. (As an example object depth, $l = 0.5$ m, the ruby laser light $\lambda = 0.7\ 10^{-6}$ m:

$$\Delta\lambda < 0.16 \cdot 10^{-9} m = 1.6 \cdot 10^{-3} Å.)$$

Stability of optical elements of the holographic system

Practically no instabilities of optical elements of the holographic system can be tolerated during the hologram exposure, as the maximal permissible translation Δl of every reflecting element during exposure is limited by:

$$\Delta l < 0.22\lambda/(n \cos\theta \cos\alpha) \qquad (2)$$

where n is the number of reflecting elements in the holographic system, θ the incident angle of the light beam at the considered reflection surface, and α the angle between the translation direction $\vec{v}$ and normal $\vec{n}$ of the reflecting surface M, see Fig. 1. If this condition not is met,

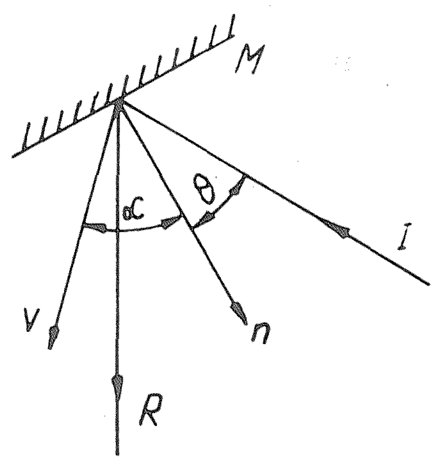


Fig. 1. Geometrical relationship between the mirror translation and changes of the pathlength: M, mirror; I, incident beam; R, reflected beam; n - normal to the mirror surface; v, velocity vector.

the image intensity drops and, in extreme cases, the image may be completely erased. Similar restrictions apply to the holographed object as well.

Holographic registration of living subjects

Physiological activities of living subjects, such as respiration, heart activity, or uncontrolled translations of different parts of the body, influence the intensity of the reconstructed image. Theoretical analysis of this influence show that the intensity of the reconstructed image depends on the kind of motion and its magnitude (Vest, 1979; Pluta, 1980). In the case of a linear motion, the image intensity I (l) depends on the translation value l as follows (see Fig. 2):

$$I(\Delta l) = I_o \sin c^2 \, [(2\Delta l \cos\alpha_d \cos\theta_d)/\lambda] \qquad (3)$$

where λ is the wavelength of the light used, Δl the total translation

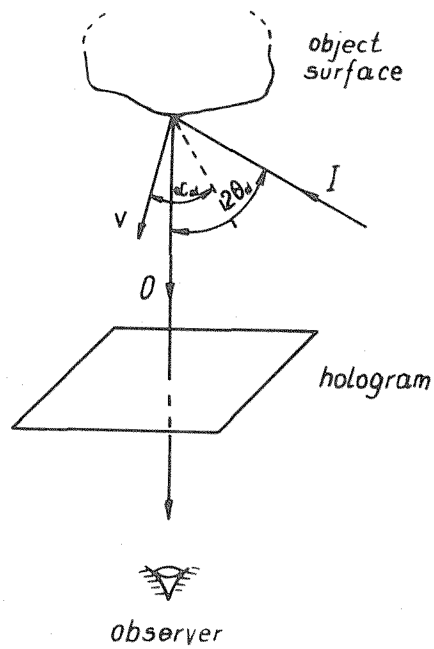


Fig. 2. Geometrical relationship between the object translation and changes of the pathlength: I, incident beam; O, observation direction; v - velocity vector.

during the exposure, α_d the angle between the velocity vector v and bisectrix the angle $2\,\theta_d$, $2\,\theta_d$ the angle between the illumination direction I and the observation direction O, $\mathrm{sinc}\,(x) = \sin(\pi x)/(\pi x)$, and I_o

is the intensity of the image obtained under the same conditions but for motionless objects. It is seen from the above equation that the image intensity drops very quickly when the translation value increases. Thus, linear translations of the holographed object are extremely undesirable during the exposition of the hologram and in the case of living persons, where different physiological motions are inevitable, the exposure time should be properly short (below 1 microsec).

Such a short exposure time makes it impossible to apply time-averaging holographic interferometry for the vibration analysis of living persons during singing, although just this method gives the best interference pattern (very clearly readable) of vibrating objects.

For this reason the double-pulse holographic interferometry is applied for skin vibration analysis.

Energy and power requirements

Energy requirements depend on many factors such as:

- sensitivity of the photosensitive material used,
- hologram size,
- size of the object and its surface properties,
- kind of object illumination: if it is diffuse or not,
- energy losses in the holographic system.

Presently, the most popular holographic plates are the 10E75 and 8E75 manufactured by Agfa-Gevaert. Their sensitivity is about 5.0 $\mu J/cm^2$ and 20.0 $\mu J/cm^2$, respectively. The energy demands are about 500 μJ and 2.0 μJ for plates of the size 9x12 cm^2. When living subjects are holographed and diffuse illumination is applied, the typical energetic efficiency of the holographic set-up is 0.1-0.01%. As a result, the total energy demand is 0.05-0.5 J for 10E75 holographic plates and 0.2-2.0 J in case of 8E75 plates. For 1 μsec exposure time, the above-mentioned energies can be delivered if the power of the laser source is equal to 0.05-0.5 MW and 0.2-2.0 MW, respectively. Until now, there are no powerful continuous lasers emitting visible light, and pulse lasers must be used. Today, there are only two lasers which provide light with desirable coherence, energy, and power:

- the ruby laser emitting light with a wavelength of λ = 694.7 nm,
- the neodymium-doped YAG laser emitting infrared radiation converted with nonlinear optics into visible light with λ = 530 nm.

At present, the most frequently used is the ruby laser which is able to provide one or more coherent laser pulses lasting 20-30nsec each, with a regulated time separation in the range of 0.01-1.0 msec. Then, to obtain the desired energy, the emitted power must be about 30-40 times higher. For these energy and power levels, the possibility of damage to the optical elements used must be taken into account.

Safety requirements

For the above-mentioned energy and power levels, the direct exposure of the human eye to a non-diffuse light is absolutely intolerable. Only diffuse light at the eye pupil is permissible, if the energy density at the diffuser does not exceed the level of 0.13 J/cm^2 for a diffuser which diffuses the light across a solid angle not smaller than $\pi/10$ steradians (Collier et al., 1971). To avoid accidental eye damage, it is advisable to shade the eyes completely.

Double-pulse holographic interferometry

As mentioned above, the most profitable time-averaging holographic interferometry could not be applied to the skin vibration analysis of living subjects. Therefore, the double-pulse interferometry is used. The method consists in making two expositions corresponding to two different phases of the object vibration on the same holographic plate in the same holographic system. During the reconstruction process, both the registered images are reconstructed simultaneously. As a result, the interference pattern appears. This pattern carries information about the relative translation of the object surface between the two exposures. The intensity distribution $I(\Delta l_t)$ of the reconstructed image depends on the relative translation Δl_t:

$$I(\Delta l_t) = I_o \cos^2(4\pi \Delta l_t \cos\alpha \cos\theta/\lambda) \quad (4)$$

As seen, the reconstructed image is covered with bright and dark fringes with constant contrast. Each fringe represents the object points with constant translations between the exposures. If the fringe number is known, it enables us to localize nodal areas as the zero-order fringes. Unfortunately, contrary to the time-averaging method, the double-exposure method does not produce any gradations of fringe visibility or contrast. This trouble may be partially overcome by applying the oblique wide-aperture diffuse illumination. In this case, the fringe contrast decreases as the relative translation increases (Kraska and Pawluczyk, 1983; Kraska et al., 1982; Matczak, 1982; 1983; Matczak et al., 1982; Pawluczyk and Kraska, 1981; Pawluczyk et al., 1982). The theoretical basis of this method has been developed by Matczak (1982, 1983) and Pawluczyk et al. (1983).

Experimental system

The experimental system used for the holographic registration of the skin vibration includes the following parts:
- holographic system,
- ruby laser system together with power supply and control circuits,
- electronic equipment for voice analysis,
- synchronization circuits,
- auxiliary control and measurement equipment.

A block diagram of the complete system is shown in Fig. 3.

Holographic system

A typical holographic system with a side reference beam has been used during the experiment. As is shown in Fig. 4, the laser light from the

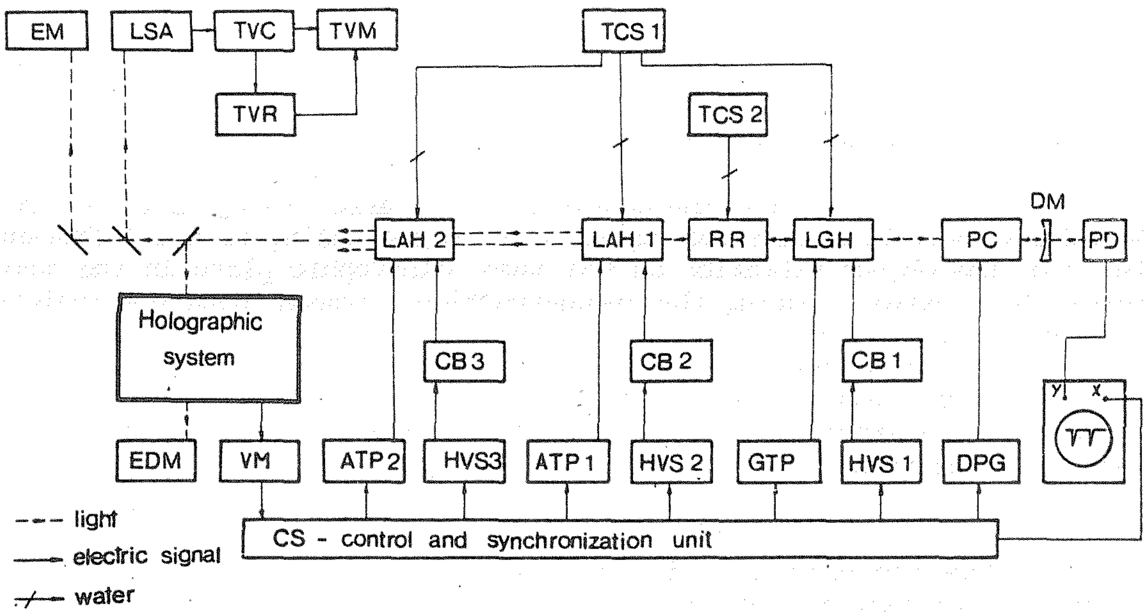


Fig. 3. Block diagram of the experimental system: LGH, laser generator head; LaH 1 and LaH 2, laser amplifier heads; DM, dielectric concave mirror; PC, Pockels cell; RR, resonant reflector; TCS1, temperature control system of laser heads; TCS2, temperature control system of resonant reflector; CB1, CB2- and CB3, capacitor banks; HVS1, HVS2 and HVS3, high-voltage supplies; GTP, trigger pulser of generator; PD, photodetector; O, oscilloscope; LSA, laser spectrum analyzer; TVC, TV camera; TVR, TV recorder; TVM, TV monitor; EM, energy meter; VM, vibration meter; EDM, energy density meter.

ruby laser RL is reflected within the prism Pr into the holographic system. The beam-splitter BS divides the incident laser beam into two parts: the object beam and the reference beam. The beam-splitter is made of a wedge glass plate coated with an anti-reflection layer on one side. The other side divides the incident laser beam. The transmitted part carries about 95% of the total energy and is the object beam. This beam is reflected with the auxiliary mirrors M1 and M2, expanded with the negative lens L1 and directed to the diffuser GP (glass ground plate).

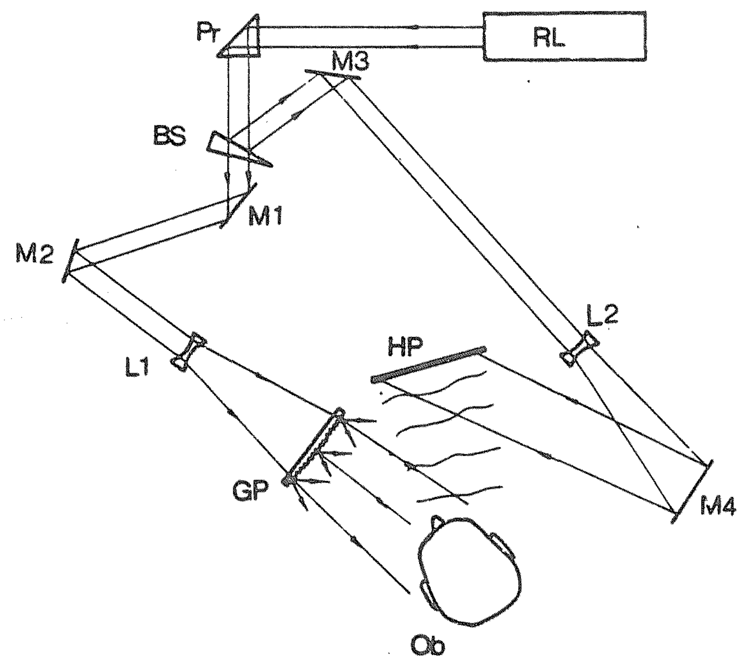


Fig. 4. Schematic diagram of the holographic system: RL, ruby laser; Pr, prism; BS, beam splitter; M1, M2, M3 and M4, mirrors; L1 and L2, negative lenses; GP, ground glass plate; Ob, object; HP, photosensitive plate.

The diameter of the light spot at the diffuser surface is equal to about 15 cm, which ensures that the security requirements are filled and that a good diffuse illumination for the fringe contrast differentiation is pronounced. A diffuse light from the diffuser illuminates a holographed object Ob. Part of the light diffracted at the object reaches the photosensitive material HP.

Part of the light reflected by the beam splitter makes a reference beam which is expanded by the second negative lens L2 and directed by the auxiliary mirrors M3 and M4 to the photosensitive material. As a result of the superposition of the two beams, an interreference pattern occurs which is registered on photosensitive material as a hologram. The holographic plates of 10E75-type, manufactured by Agfa-Gevaert, were used. The plates were processed with the chemical reagents recommended by the producer.

A simple optical set-up has been used for the hologram reconstruction, as is shown in Fig. 5.

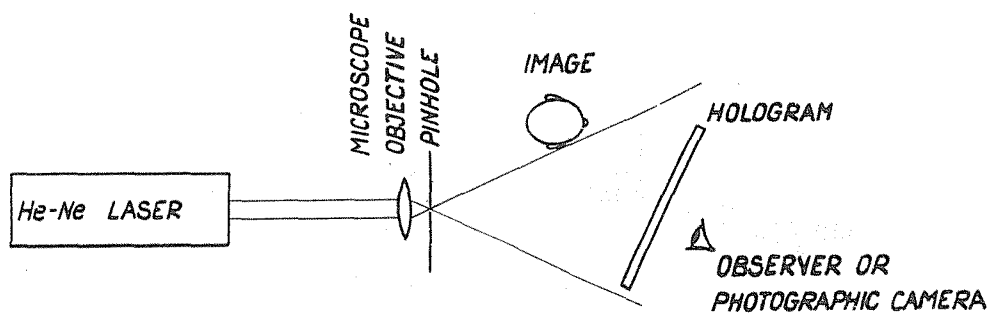


Fig. 5. Schematic diagram of the reconstruction system.

Ruby laser system

The system used in the experiment is presented in part in Fig. 3 with its optical part in Fig. 6. The ruby laser system consists of a low-

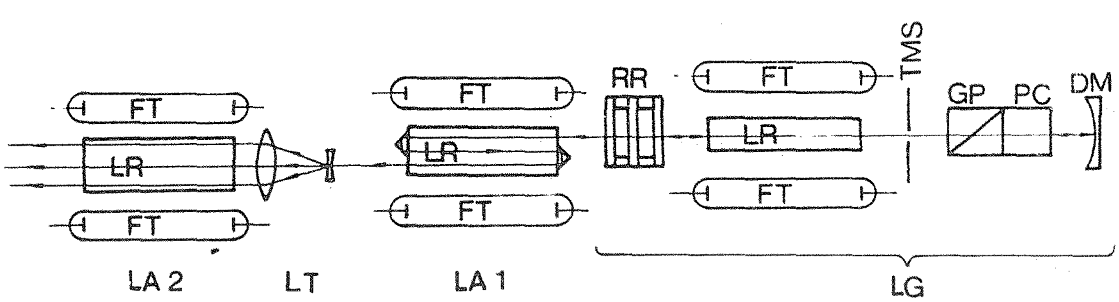


Fig. 6. Schematic diagram of double-pulse ruby laser system: LG, laser generator; LR, ruby rods; FT, flashtubes; PC, Pockels cell; GP, Glan polarizer; TMS, transverse mode selector; RR, resonant reflector, i.e., longitudinal mode selector; LA1 and LA2, laser amplifiers; LT, telescope.

energy but high-coherent electro-optically Q-switched laser generator, LG, and one or two laser amplifiers, LA1 and LA2. The laser generator radiates two light pulses lasting approx. 30 nsec each, with a time interval between them regulated to 0.15-0.25 msec. The interval chosen is 0.2 msec. The transverse TMS and the longitudinal RR mode selectors are used to achieve higher coherence. The concave high-reflecting dielectric mirror, DM, and the resonant reflector, RR, which plays a role of a longitudinal mode selector used as a partially transmitted mirror, form the resonator of the laser generator. To obtain short laser pulses at the desired time, the laser is electro-optically Q-switched by means of a Pockels cell, PC, cooperating with the air-spaced Glan polarizer, GP. The Pockels cell is controlled by a double-pulse generator, DPG, synchronized with the investigated object by the control and synchronization unit, CS.

The laser generator provides several millijoules of energy which could be increased by using the one-, two-, or three-pass laser amplifier (LA1) and the second one-pass (LA2) as required. To avoid damage of the laser rod of the second amplifier, the laser beam emerging from the first amplifier is expanded by means of a telescope (LT). Such a system allows optimization of the hologram exposure conditions for the energy level selection up to approx. 1J.

To secure repeatable laser operations and good mode selection, the temperature of the laser generator (LGH) and laser amplifiers LAH1 and LAH2 heads are controlled by the temperature control system TCS1 with an accuracy of 0.1°C, and the temperature of the resonance mirror is controlled by the second control system TCS2 with an accuercy of 0.01°C. The temperature values are somewhat higher than the temperature of the environment and are adjusted for the best single-mode operation of the laser system.

The laser heads of the generator LGH and the two amplifiers LAH1 and LAH2 are fed from the capacitor banks CB1, CB2, and CB3 charged by the individual high voltage systems HVS1, HVS2, and HVS3, respectively. The high-voltage triggers GTP and ATP1 and ATP2 produce, given the appropriate signals, high-voltage pulses of approx. 0.5 sec duration, sufficient for the preionization of the gas in the flashtubes (FT) of the heads. The control signals for the initiation of the high-volatage triggers arrive at appropriate times from the control and synchronization unit (CS).

Control and synchronization system

An important problem in the double-pulse holographic interferometry is the production of laser actions at precise times, i.e., to synchronize the laser pulses with changes occurring in the object.

This problem may be solved using different methods depending on the object under examination and the changes occurring in it. To control the laser action, the fundamental frequency of the voice of the person examined has been used. The whole process of the laser action has been synchronized with the phase of that component. The whole control of laser actions is accomplished by the control and synchronization system which consists of two parts:

- the sound analysis system,
- the control and synchronization unit.

The functional block diagram of the sound analysis system, the synchronization, and control unit are shown in Fig. 7. The time sequence of the control of the pulses together with the fundamental frequency of the acoustic vibrations and light phenomena in the laser are presented in Fig. 8.

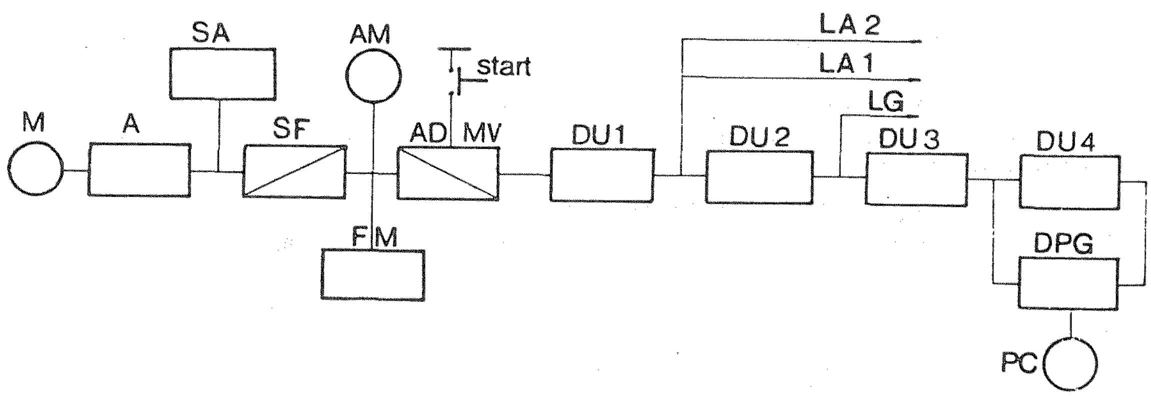


Fig. 7. Block diagram of sound analysis, synchronization, and control systems: M, microphone; A, acoustical amplifiers; SA, signal analyzer; SF, selective filter; FM, frequency meter of fundamental acoustic component; AM, amplitude meter of this component; AD, amplitude discriminator; MV, monovibrator; DU1, DU2, DU3 and DU4, delay units; DPG, double-pulse generator; PC, Pockels cell; LA1, LA2 and LG, pulses to trigger circuits of laser amplifiers and generator.

Until now holographic registrations of skin vibrations have been made for three different groups of subjects: singers, persons with untrained voices, and some pathological cases. To obtain the controlled results in the first two cases, the subject sings a vowel at a certain pitch, the fundamental frequency of which is demonstrated with a tuning fork. The sound produced is received by a microphone (M), see Fig. 7. Its elec-

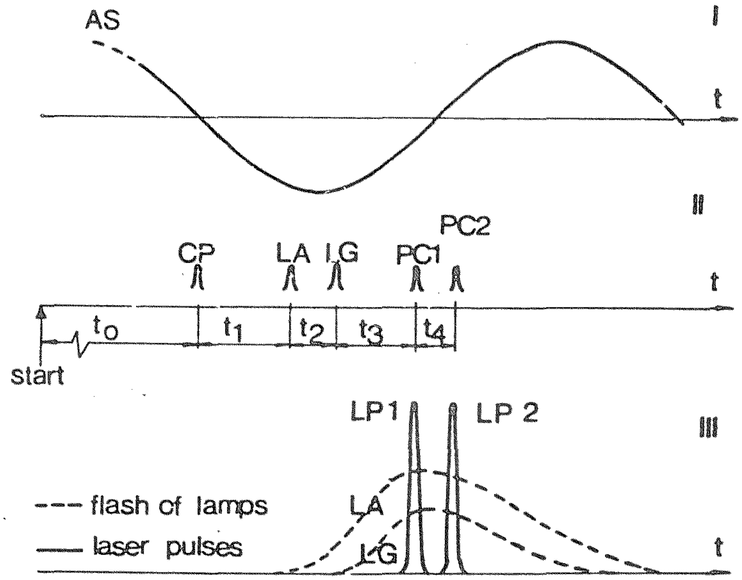


Fig. 8. Time sequence of phenomena in holographic and laser systems: /I/, fundamental acoustic component of the voice; /II/, control pulses; CP, first electrical control pulse; LA, control pulse triggering laser amplifiers; LG, control pulse triggering laser generator; PC1 and PC2, control pulses to Pockels cell; t_0, t_1, t_2, t_3 and t_4, time intervals; /III/, light processes; LG, in generator; LA, in amplifier; LP1 and LP2, laser pulses.

trical signal is transmitted to the spectrum analyzer (SA) after amplification by the calibrated amplifier (A). The analyzer detects the fundamental frequency of the voice and this frequency is filtered by a selective filter (SF), the output of which is the sinusoidal signal, see Fig. 8.1. The frequency of this signal is checked by the frequency meter (FM), and its amplitude by the amplitude meter (AM), Fig. 7. If the frequency is in agreement with the frequency produced by the tuning fork, an interferogram is made. The amplitude discriminator (AD) is activated (by pushing the START button) and the sinusoidal signal arrives from the selective filter. The amplitude discriminator waits until the sinusoid reaches the preset value at a given (increasing or decreasing) slope of the sinusoidal curve and at this moment, the discriminator activates the monovibrator (MV) forming a first control pulse (CP), see Fig. 8.2. This pulse starts a chain of delay units: DU1, DU2, DU3, and DU4. These units produce a sequence of control pulses: LA, LG, PC1, and PC2, which activate the triggers of the laser amplifiers, the trigger of the laser generator, and the control pulses of the Pockels cell, respectively. The Pockels cell pulses cause light generation in the laser system. This control system allows generation of laser pulses at the desired times.

In the pathological cases, the same system is used the only difference being that the basic frequency is not forced by means of a tuning fork.

The exact time of occurrence of the laser pulses is registered by means of the photodetector (PD), see Fig. 3, with the oscilloscope (O) permitting observation of laser pulses only, each of them separately or both at the same time, as well as in the background of the curve showing acoustical vibrations received by a microphone. The energy of the two laser pulses can be measured by means of the accumulating energy meter (EM), which may be used as the ratiometer of energy (or energy density) of the object to the reference beams. The laser energy density meter (EDM), located in the plane of the holographic plate, allows energy density to be measured, whereas the laser spectrum analyzer (LSA) with a TV-camera is used for the spectrum recording. The obtained spectrum could be recorded on a TV-recorder and monitored on a TV-monitor.

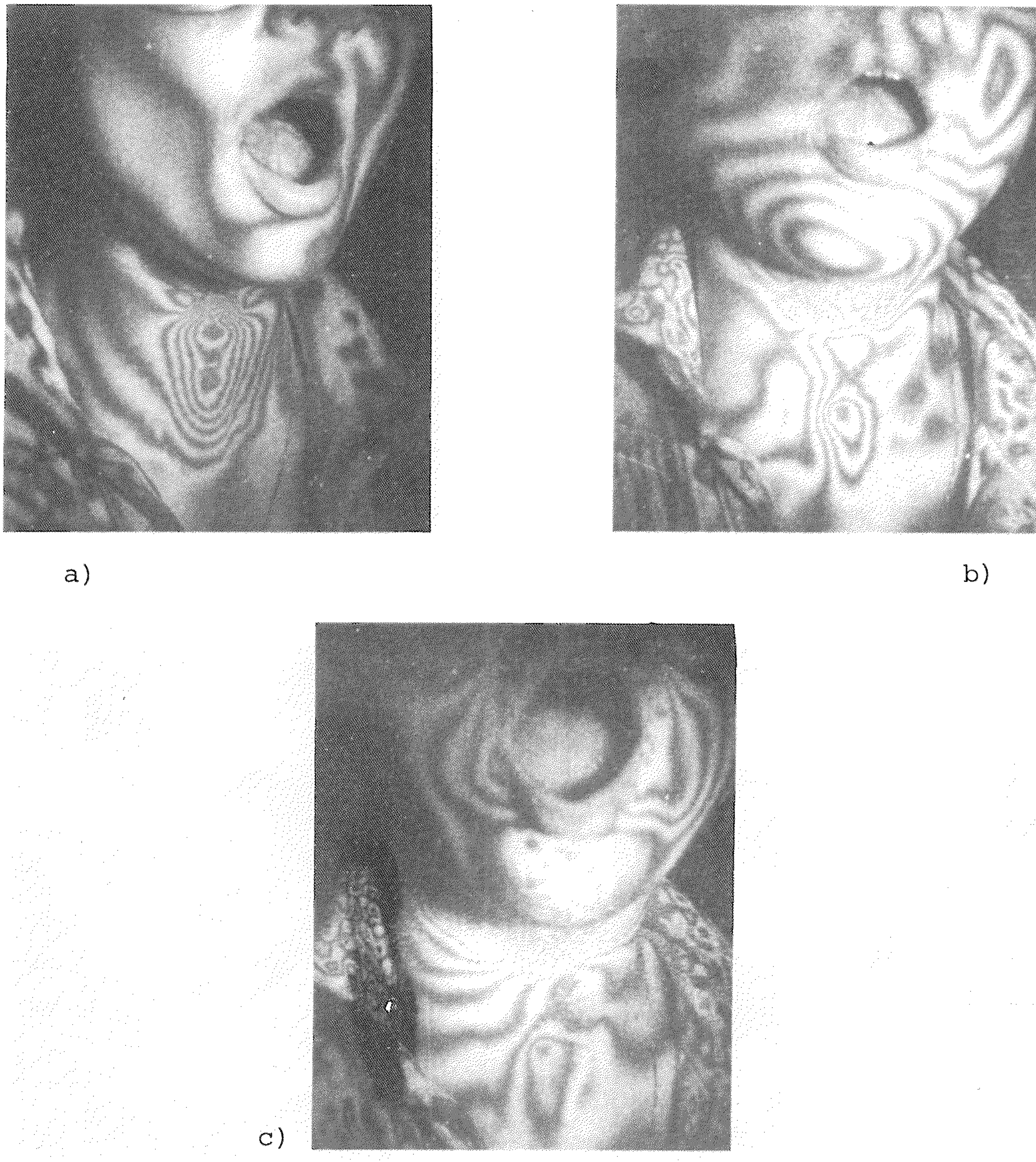
All this auxiliary equipment secures an exact control of the laser action.

For the interested reader more details regarding the holographic system may be found in Pawluczyk et al. (1982).

Results

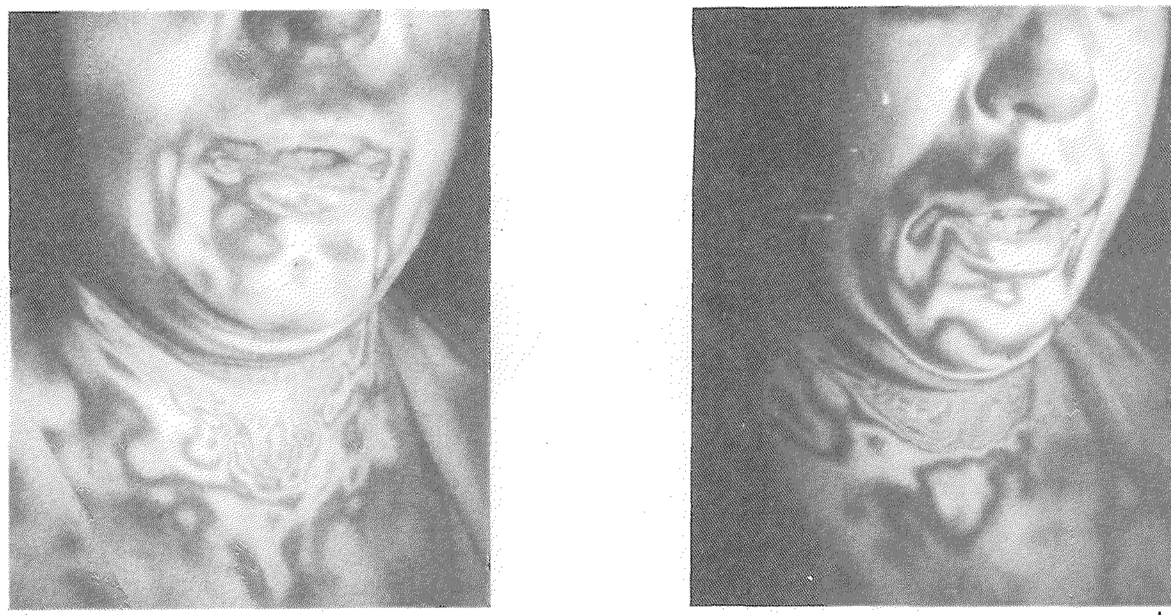
The main purpose of the investigations was to ascertain the usefulness of the holographic method for the analysis of skin vibrations during singing in those parts of the human body which have any connection with the human vocal organ, i.e., epiphysis, larynx, neck, and manubrium of the sternum. The intention of these investigations was to test the usefulness of the holographic double-pulse interferometry for voice diagnosis as well as the possibilities of applying this method for medical diagnosis. As was mentioned above, three groups of people have been holographed. Some exemples of such interferograms have been published previously (Pawluczyk et al., 1982a; 1982b), and a further set is presented in this paper (Figs. 9-14).

Fig. 9 presents three interferograms of the vibrations of the epiphysis of a woman (EJ) during singing of the vowel "a" in Polish pronunciation at the following frequencies: 446 Hz, 668 Hz, and 697 Hz. The next series of interferograms present images of the vibrations of another



a) b) c)

Fig. 9. Interferograms of vibrations of the epiphysis of woman E.J.
Vowel a: a - 446 Hz; b - 668 Hz; c - 697 Hz.



a) b)

Vowel y: a - 96 Hz; b - 128 Hz;

c) d)

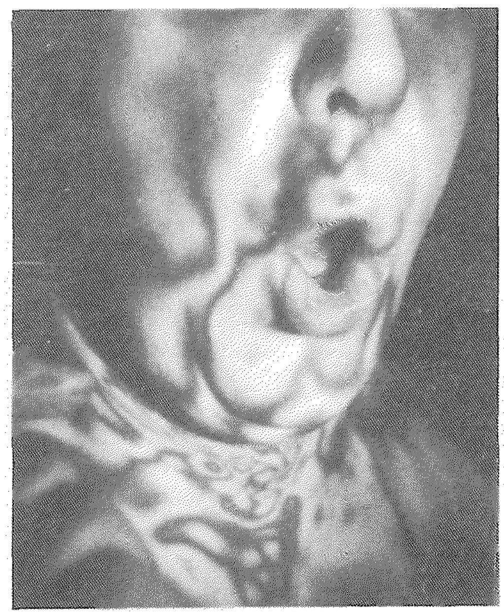
Vowel u: c - 96 Hz; d - 128 Hz;

Fig.lo. Interferograms of vibrations of the man M.W.

e)

f)

Vowel a: e - 96 Hz; f - 128 Hz;



g)

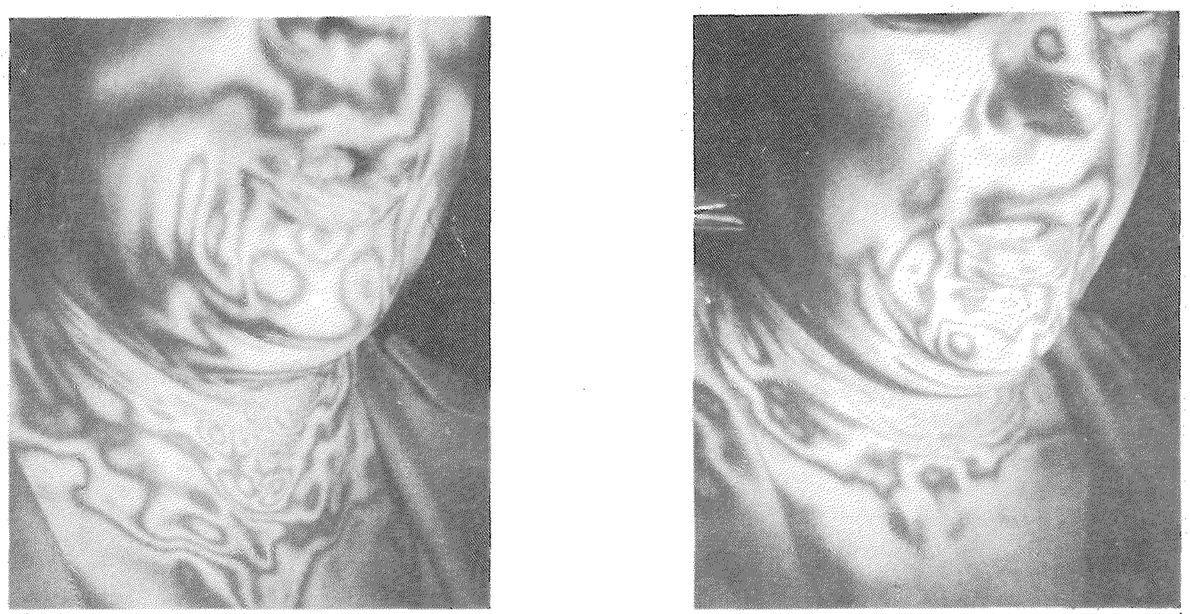
h)

Vowel o: g - 98 Hz; h - 129 Hz;

Fig. 10 (cont.)

i) j)

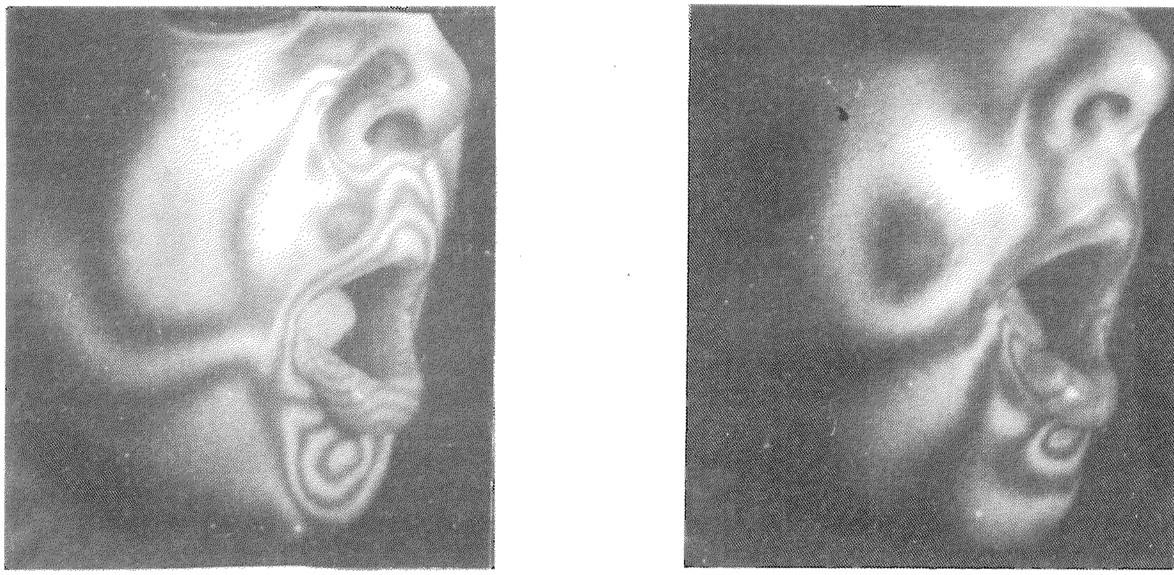
Vowel e: i - 96 Hz; j - 129 Hz;



k) l)

Vowel i: k - 97 Hz; l - 128 Hz.

Fig. 10 (cont.)

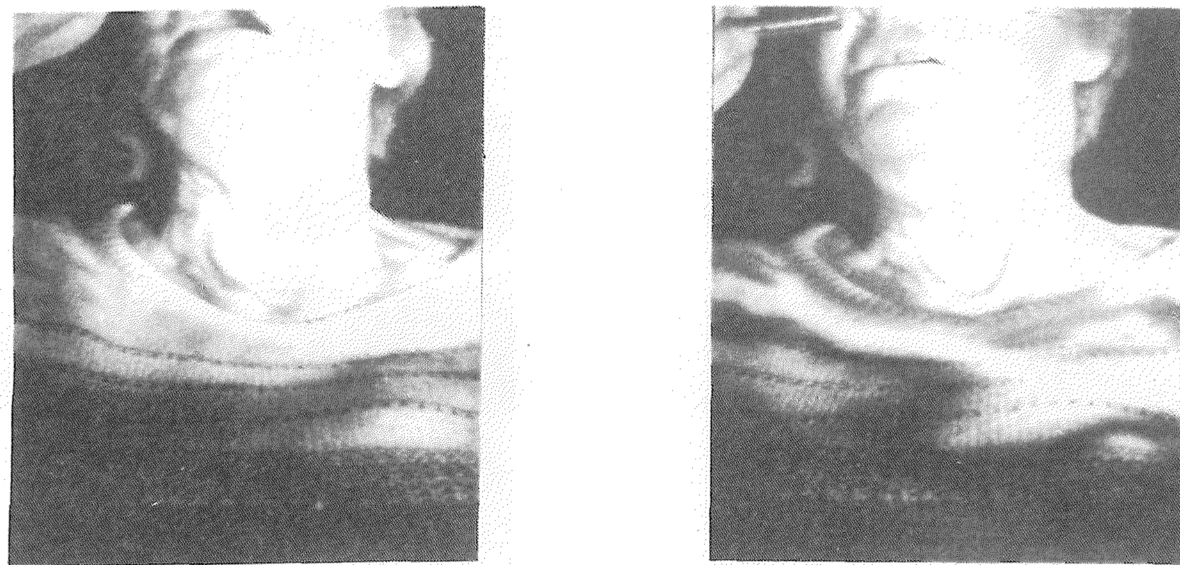


a) b)

Fig. 11. Interferograms of lip vibrations during singing the vowel "a" at frequencies: a - 519 Hz; b - 524 Hz.

a) b)

Fig. 12. Interferograms of skin vibrations of the woman P.E. (pathological case): a - vowel "o" - 173 Hz; b - vowel "i" - 209 Hz.



a) b)

Fig. 13. Interferograms of skin vibrations of the woman O.Z. after surgical operation and voice rehabilitation: a - vowel "a" - 232 Hz; b - vowel "e" - 257 Hz.

a) b)

Fig. 14. Interferograms of skin vibration and voice rehabilitation: a - vowel "o" - 242 Hz; b - vowel "e" - 286 Hz.

person, the man (MW) during singing of the vowels: "y", "u", "a", "o", "e", "i" at two different basic frequencies: 96 98 Hz and 128 129 Hz. The interferograms show that the increase of the fundamental frequency causes a decrease of the vibration amplitude in the lower parts, i.e., in the region of the manubrium of the sternum and the larynx; simultaneously an increase of the vibration amplitude in the region of microgenia and on some fragments of the surface of the face occur. This series of interferograms illustrates well the changes of the vibrations, especially on the surface of the face when a different vowel is sung. Such changes take place both on the surface of the larynx and on the surface of the face. On account of the very complex character of these changes on the surface of the larynx, the analysis presents some difficulties. It is much easier to study the changes on the surface of the face. The changes in the distributions of the vibrations on the face take place also as the the pitch of the sung vowel is changed. In this case even a small variation of the fundamental frequency causes the distributions of the changes of the vibrations to be very great. It demonstrates a great selectivity of the individual portions of the face. The interferograms in Fig. 11 serve as examples of the changes of the vibrations of the lips with a change of fundamental frequency of the vowel "a" by 5 Hz. The next three sets of interferograms show vibrations of the skin of different pathological cases.

The first set (seen in Fig. 12) presents two interferograms of a woman (PE) during singing of the vowel "o" at a basic frequency of 173 Hz and the vowel "i" at a frequency of 209 Hz. The medical diagnosis was laringitis chronica subformam in form OEDEMA REINCKE.

The next two sets (Figs. 13 and 14) present the interferograms of two women (OZ and WM) with the same medical diagnosis: bilateral paralysis of the vocal cords. The interferograms present the state of the post-surgical operation consisting in dilatation of the glottis - the left arytenoidectomi and rehabilitation of the voice. Fig. 13 presents the vibrations during singing of the vowels "a" and "e" at the basic frequencies 232 Hz and 257 Hz, respectively, and Fig. 14 shows the interferograms during singing the vowels "o" at the frequency 242 Hz and "e" at 286 Hz. These interferograms have been obtained quite recently and no analysis of these cases is discussed in this paper.

Conclusion

The results obtained during the experiments make it possible to draw the following conclusions:

- the method gives reproducible interferograms for the same persons singing the same sound at the same pitch, i.e., the same fundamental frequency of the voice.
- in all cases the holographic method allows visualization of the vibrations of the surface of the skin of different parts of the human body,
- depending on the kind of sound sung and the basic frequency, the distribution of the vibrations changes,
- a comparison of the interference pictures shows that for all persons increasing fundamental frequency causes decreasing vibration amplitude in the lower parts, i.e., in the region of the manubrium sternum and the larynx, simultaneously with increasing the vibration amplitude in the region of microgenia and the segments of the surface of the face,
- it seems that after getting the appropriate statistics and applying the quantitative analysis in connection with a Fourier analysis of the voice, it will be possible to use this method for the qualification of the timbre of a vocalist's voice,
- it seems that this method will become useful as a diagnosis method in the examination of pathological voices.
- the applied holographic system needs some further improvements allowing the complex quantitative correlation analysis of the produced voice and the obtained interferograms,
- a full appreciation of the usefulness of this method in phoniatry, voice research, and medicine requires further studies and quantitative analysis applied.

References

Collier, R.J., Burckhardt, Ch.B., and Lin, L.H. (1971): Optical Holography, Academic Press, New York and London.

Erf, R.K. (ed.)(1974): Holographic Nondestructive Testing, Academic Press, New York and London.

Fant, G. (1958); "Modern instruments and methods for acoustics studies of speech", Proc. VIII Int.Congr.Linguistics, Oslo.

Fant, G. (1959): "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics, No. 1.

Fant, G. (1960): Acoustic Theory of Speech Production, Mouton, The Hague.

Flanagan, J.L. (1960): "Analog measurements of sound radiation from the mouth", J.Acoust.Soc.Am., 32, pp. 1613-1620.

Flanagan, J.L. (1964): Speech Analysis, Synthesis and Perception, Springer Verlag, Heidelberg.

Hirano, M., Takouchi, Y., and Hiroto, I. (1966): "Intranasal sound pressure during utterance of speech sounds", Folia Phoniat., 18, pp. 369-381.

House, A.S. and Stevens, K.N. (1956): "Analog studies of the nasalization of vowels", J. Speech and Hearing Disorders, 21, pp. 218-232.

Isshiki, N. (1964): "Regulatory mechanism of voice intensity variation", J. Speech and Hearing Research, 7, pp. 17-29.

Kacprowski, J. (1977): "Physical models of the larynx source", Arch. Acoust., 2, pp. 47-70.

Kacprowski, J. (1979): "Objective acoustic methods in phoniatric diagnosis of speech organ disorders", Arch. Acoust., 4, pp. 289-304.

Koike, Y., Hirano, M., and v. Leden, H. (1967): "Vocal initiation: Acoustic and aerodynamic investigations of normal subjects", Folia Phoniat., 19, pp. 173-182.

Kraska, Z. and Pawluczyk, R. (1983): "Holographic interferometry with diffuse illumination", European Optical Conf., Rydzyna, Poland, May 30-June 4.

Kraska, Z., Pawluczyk, R., and Matczak, M. (1982): "Identification of the local minima of fringers order in holographic interferometry", Proc. Int.Conf. on Holographic Data non-Destructive Testing, Dubrovnik, Yugoslavia, Oct. 5-8.

v. Leden, H. (1969): "Objective measures of laryngeal function and phonation", Ann. New York Acad.Sci., 155, pp. 56-67.

v. Leden, H. (1969): "Objective methods of vocal analysis in singers and actors", Exc. Medica Int.Congr. Reports, Series No. 206.

Matczak. M.J. (1982): "Fringe visibility method as a new method of holographic interferograms interpretation", Proc.Int.Conf. on Holographic Data non-Destructive Testing, Dubrovnik, Yugoslavia, Oct. 5-8.

Matczak, M.J. (1983): "Diffractional description of interference image formation and its consequences in holographic interferometry", European Optical Conf., Rydzyna, Poland, May 30-June 4.

Matczak, M.J., Pawluczyk, R., and Kraska, Z. (1982): "Diffuse illumination in holographic interferometry", Proc.Int.Conf. on Holographic Data non-Destructive Testing, Dubrovnik, Yugoslavia, Oct. 5-8.

Pawłowski, Z. (1968): "Risultati di ricerche acustiche sul sistema di risonanza di alcuni difetti congeniti del palato", Annali di Laring., Otol., Rinol., Faring., 67, pp. 733-755.

Pawłowski, Z., Letowski, T., and Rakowski, A. (1972): "Comparative analysis of microphone recordings registered at various points of the vocal tract", Folia Phoniat., 24, pp. 360-370.

Pawłowski, Z. and Żółtowski, M. (1982): "Fundamental aspects of aeroacoustics in singing", Arch. Acoust., 7, pp. 207-224.

Pawluczyk, R. and Kraska, Z. (1981): "Diffuse illumination in holographic interferometry of living objects", ICO-12 Satellite Meeting, Optics in Biomedical Sciences, Sept. 7-11, Graz, Austria.

Pawluczyk R., Kraska, Z., and Pawłowski, Z. (1982a): "Holographic vibration analysis of the frontal part of the human neck during singing", Proc. Int.Conf. Optics in Biomedical Sciences, (eds. G. v. Bally, & P. Greguss), Springer-Verlag, Heidelberg.

Pawluczyk, Z., Kraska, Z., and Pawłowski, Z. (1982b): "Holographic investigations of skin vibrations", Appl. Optics, 21, pp. 759-765.

Pluta, M. (ed.)(1980): Hologriafia Optyczna, PWN, Warsaw.

Shipp, T. (1967): "Frequency, duration, and perceptual measures in relations to judgments of alaryngeal speech and acceptability", J. Speech and Hearing Research, 10, pp. 417-427.

Shipp, T. and Huntington, D.A. (1965): "Some acoustic and perceptual factors in acute-laryngitic hoarseness", J. Speech and Hearing Disorders, 30, pp. 350-359.

Vest, Ch.M. (1979): Holographic Interferometry, John Wiley & Sons, New York.

Yanagihara, N. and Koike, Y. (1967): "The regulation of sustained phonation", Folia Phoniat., 19, pp. 1-18.

THE IMPORTANCE OF VOCAL TRACT LOADING IN MAINTAINING VOCAL FOLD OSCILLATION

Ingo R. Titze
Dept. of Speech Pathology and Audiology, University of Iowa
Iowa City, IA, USA

Abstract

The mechanism of flow-induced oscillation of the vocal folds has been explained on the basis of negative dynamic stiffness and negative dynamic resistance. For small-amplitude oscillation, a negative aerodynamic coupling stiffness exists between upper and lower portions of the vocal folds as shown by Ishizaka and Matsudaira (SCRL Monography No. 8, 1972). For a larger-amplitude oscillation, this takes the form of an asymmetric driving force that depends on both displacement and velocity of the center of the medial surface of the vocal folds. The asymmetry results from a phase-delay in vertical propagation of surface waves in the mucosal covering.

Introduction

Much has been written recently about the importance of acoustic interaction between the glottal flow and the vocal tract system. It has been demonstrated, for example, that the inertance of the vocal tract is responsible for the asymmetry in the glottal flow waveform (Rothenberg, 1983). It has also been shown that formant ripple in the glottal flow waveform is the result of superposition of acoustic energy from the vocal tract onto the glottal airstream (Ishizaka and Flanagan, 1972; Ananthapadmanabha and Fant, 1982). Considerable work is in progress to relate temporal and spectral features of the glottal source function to physiologic variables, such as subglottal pressure, maximum glottal area, and vocal tract configuration (Fant, 1982; Titze, 1984).

The more recent flow calculations, however, do not include the energy exchange between glottal flow and tissue movement. This energy exchange appears to be implicit in the simulations of Ishizaka and Flanagan (1972) and Titze and Talkin (1979), but we are becoming increasingly concerned about the validity of the glottal flow circuit used in these simulations. In particular, it has been recognized recently that the displacement flow resulting from movement of the medial surface of the folds is not simply a small correction to the glottal flow, as suggested by Ishizaka and Flanagan (1976), but plays a major role in the establishment of asymmetry in driving pressure and flow between the opening phase and the closing phase. The glottal flow circuit used repeatedly in all previous flow calculations (except Conrad's, 1980, to be discussed later) does not include the yielding-wall parallel branch that is necessary to have a proper energy exchange between glottal airflow and tissue movement. The absence of this branch is particulary noticeable in the simulations of Ishizaka and Flanagan (1976), who use the proper yielding-wall circuit throughout the vocal tract, but not in the vocal fold region where wall-yielding seems to be most evident. Rather the effective parallel branch is modeled with flow sources that introduce displacement flow in a somewhat artificial manner. According to Cooper and Titze (1983) the energy dissipation in the vocal folds can be a significant portion of the energy available from the lungs and on the same order of magnitude as the acoustic energy radiated. Proper representation of tissue losses in the glottal flow circuit is therefore important.

In contrast to the single-loop glottal flow circuits used by most investigators, Conrad (1980) has proposed a flow circuit that makes the parallel branch a key element in self-oscillation. Although we are basically in agreement with Conrad's analogy between flow-induced vibrations in the vocal folds and collapsible tubes, we do not believe that his choice of a resistive vocal tract load is appropriate. It is known that a sizable reactive component is necessary to represent the resonant nature of the vocal tract. In the following sections we will demonstrate how a self-oscillating vocal fold model can be interfaced with a multi-section transmission-line-analog of the vocal tract, such that three-way interactions between glottal airflow, elastic wave propagation in vocal fold tissue, and acoustic wave propagation in the vocal tract can be quantified by relatively simple formulas.

I. Coupled Equations for Glottal Flow and Driving Pressure in Terms of Air and Tissue Waves

Consider the glottal flow circuit of Fig. 1, where a single T-section is used to represent the glottis. Let P_1 be the subglottal pressure (prior to glottal entry), P_2 the supraglottal pressure (after glottal exit), and P the glottal midpoint pressure (net driving pressure). Using transmission line theory, P_1 and P_2 can be written as

$$P_1 = 2P_1^+ - \frac{\rho c}{A_1}(u + u_d) \qquad (1)$$

$$P_2 = 2P_2^- + \frac{\rho c}{A_2} u , \qquad (2)$$

where P_1^+ is the forward-travelling wave incident upon the glottis from the lungs, P_2^- is the backward-travelling wave incident upon the glottis from the pharynx, ρ is the density of air, c is the velocity of sound, A_1 is the subglottal area (before glottal entry), A_2 is the supraglottal area (after glottal exit), u is the exit flow, and u_d is the displacement flow at the center of the glottis.

The transmission line equation (1) and (2) follow directly from a simple D'Alembert solution of the one-demensional wave equation for acoustic disturbance in uniform ducts. A varying vocal tract configuration can be approximated by using an arbitrary number of sections of different cross-sectional areas if continuity laws for pressure and flow are applied at the junctions. In addition, losses can be incorporated by introducing appropriate side branches (Guerin, et al., 1983; Degryse, 1981).

The one-dimensional wave concept can also be applied to surface waves propagating on vocal fold tissue in the direction of glottal airflow. If we assume c' to be the phase velocity of the surface waves and z to be the direction of propagation, then

$$\xi(z,t) = \xi(z-c't) \tag{3}$$

is a general displacement disturbance on the medial surface of the vocal folds (the yielding wall) that travels from glottal entry to glottal exit. To first order, we assure the displacement to be a linear function of z. This allows us to d scribe rectangular, converging, and diverging glottal ducts. We also assume, for simplicity, that the equilibrium glottal duct is rectangular, i.e., that the glottal width is constant from bottom to top prior to phonation. If we now let a_1 be the glottal area at entry, $a=2L\xi$ the midpoint glottal area, and a_2 the glottal area at exit, then

$$a_1 = a - \frac{\partial a}{\partial z}\frac{T}{2} = a - \frac{\partial \xi}{\partial z} LT = a + \frac{\partial \xi}{\partial t}\frac{LT}{c'} = a + u_d/2c' \tag{4}$$

$$a_2 = a + \frac{\partial a}{\partial z}\frac{T}{2} = a + \frac{\partial \xi}{\partial z} LT = a - \frac{\partial \xi}{\partial t}\frac{LT}{c'} = a - u_d/2c' \tag{5}$$

where L and T are the length and thickness of the vocal folds, respectively, and u_d is the midpoint displacement flow previously defined.

Equations (4) and (5) are important relationships between glottal areas, displacement flow, and phase velocity of the travelling wave in the mucosal region. Note that as the phase velocity becomes very large, that is, as mucosal waves travel very quickly, the exit and entry areas approach the midpoint area and the folds behave like a one-mass lumped oscillator. For moderate to slow-moving surface waves, a net outward (lateral) movement at the midpoint demands a convergent glottis, since u_d is positive and a_1 is larger than a_2. On the other hand, inward (medial) movement corresponds to a divergent glottis, a_1 being less than a_2 for negative u_d. Thus, the typical vertical phase relationships in tissue movement of the vocal folds are a necessary consequence of surface waves moving in the +z direction.

Returning now to Fig. 1, the glottal exit flow u is obtained by solving the outside loop equation

$$R_1(u + u_d) + R_2 u - (P_1 - P_2) = 0, \tag{6}$$

which upon substitution of Eqs. (1) and (2) for P_1 and P_2 explicit formulation of kinetic resistances R_1 and R_2 becomes

$$\frac{1}{2}\rho k_1 \frac{|u + u_d|}{(a + \frac{u_d}{2c'})^2}(u + u_d) + \frac{1}{2}\rho k_2 \frac{|u|}{(a - \frac{u_d}{2c'})^2} u$$
$$+ \rho c(\frac{1}{A_1} + \frac{1}{A_2})u + \frac{\rho c}{A_1} u_d - 2(P_1^+ - P_2^-) = 0 \tag{7}$$

Here k_1 and k_2 are pressure coefficients for kinetic pressure losses at glottal entry and exit, respectively. Note that the glottal areas as defined by Eqs. (4) and (5) have been introduced in the kinetic resistances. Viscous losses and inertance within the glottis are neglected at this point because their effect is secondary to the kinetic pressure drops. Eq. (7) is quadratic in both u and u_d. If one variable is known, the other can be solved by formula. Typically, u_d and u are found numerically by solving two simultaneous loop equations according to Fig. 1.

Of primary importance in the discussion of self-oscillation is the effective driving pressure at the midpoint of the glottis. It is informative to compare this pressure, side-by-side, with the flow equation to determine how vocal tract loading shapes the flow and maintains optimum conditions for oscillation. The midpoint driving pressure P in Fig. 1 can be written in a symmetrical form by adding the relationships $P = P_1 - R_1(u + u_d)$ and $P = P_2 + R_2 u$ to obtain

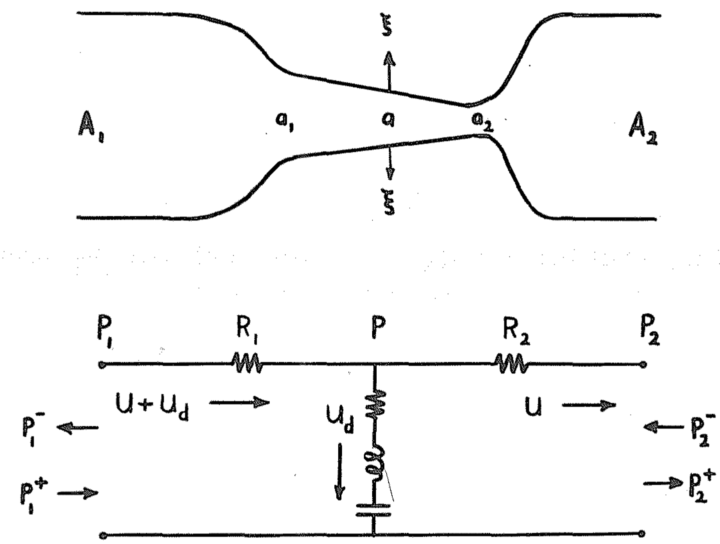


Fig. 1. Glottal airway and glottal flow circuit.

$$2P = -R_1(u + u_d) + R_2 u + (P_1 + P_2), \tag{8}$$

or more explicitely,

$$2P = -\frac{1}{2}\rho k_1 \frac{|u + u_d|}{(a + \frac{u_d}{2c'})^2}(u + u_d) + \frac{1}{2}\rho k_2 \frac{|u|}{(a - \frac{u_d}{2c'})^2} u$$

$$- \rho c(\frac{1}{A_1} - \frac{1}{A_2})u - \frac{\rho c}{A_1} u_d + 2(P_1^{\ +} + P_2^{\ -}). \tag{9}$$

The similarities and differences between Eqs. (6) and (8) and (7) and (9) are noteworthy. First of all, note that the flow depends upon P_1-P_2, the transglottal pressure, whereas the driving pressure depends upon the sum $P_1 + P_2$. Also note that the effects of subglottal and supraglottal acoustic impedance $\rho c/A_1$ and $\rho c/A_2$ are additive for the flow, but differential for the driving pressure. The same can be said about entry and exit resistances. We will now embark on a detailed discussion of how various terms in Eq. (9) influence vocal fold oscillation.

II. Discussion

In previous explanations of the self-oscillatory characteristics of the vocal folds (Stevens, 1979; Titze, 1980) it was pointed out that asymmetry in the net driving force with respect to the opening phase and closing phase can overcome the next energy losses per cycle and thereby maintain oscillation. The discussions were only semi-quantitative in that no derivation for the driving pressure was given, but the importance of vertical phase differences in movement was recognized. This phase difference, along with an aerodynamic coupling stiffness, was also the key to oscillation in the original small-amplitude oscillation theory posited by Ishizaka & Matsudaira (1972).

Eqs. (8) and (9) represent a compact way of visualizing not only the effects of asymmetry in vocal fold movement, but also of vocal tract loading. Consider first the asymmetry in movement. In the opening phase u_d is positive. In magnitude it can be estimated to be

$$u_d = (2LT)\left(\frac{\partial \xi}{\partial t}\right) = (1\ \text{cm}^2)\left(\frac{0.1\ \text{cm}}{0.0025\ \text{s}}\right) = 40\ \text{cm}^3/\text{s} \qquad (10)$$

for a 1.5 cm vocal fold length, a 1/3 cm vocal fold thickness, a 0.1 cm vibrational amplitude, and 100 Hz fundamental frequency. This flow is negligible only when the overall glottal flow is near its maximum (about 500 cm^3/s). At glottal opening and closing it can in fact be a major component of the glottal flow.

Hirano et al. (1981) has measured the approximate phase velocity of mucosal waves on the top surface of the folds. If we assume their 50 cm/sec value to apply also to surface waves within the glottis, $u_d/2c' = 0.4\ \text{cm}^2$ is a component of glottal area which corresponds to the vertical "out-of-phase" movement. The maximum midpoint glottal area is $a = 2L\xi = (2)(1.5\ \text{cm})(0.1\ \text{cm}) = 0.3\ \text{cm}^2$. This suggests that during glottal opening the entry area can be large and the entry pressure drop can be very small (first term in Eq. (9)), whereas the glottal exit area can be small and the exit pressure drop large (second term). Both of these conditions assure a favorable positive driving pressure. During glottal closing, on the other hand, u_d is negative and the flow resistances adjust themselves in the opposite direction. A large resistance at entry and a small resistance at exit reduce the midpoint driving pressure. Negative pressures can, in fact, be experienced if the first term is sufficiently large. This is, of course, the well-known and often mentioned Bernoulli Effect in vocal fold vibration.

The asymmetry in the driving pressure is the key to maintaining oscillation in the face of energy dissipation in the tissue and the glottis. A large positive pressure during outward movement (positive tissue velocity) and a lesser or even negative pressure during inward movement (negative tissue velocity) guarantees a force that is in phase with velocity and opposes the natural dissipative forces. If the net energy gained per cycle is greater than the net energy lost, oscillation is maintained.

Consider now the effect of vocal tract loading in generating asymmetric driving pressures. If either P_1^+ or P_2^-, the incident pressures from above and below the glottis, are phased such that they are positive during glottal opening and negative during closing, the proper conditions for enhancement of oscillation are met. This is illustrated in Fig. 2. At the instant of glottal closure, the subglottal pressure is maximally positive and the supraglottal pressure is maximally negative. This is a consequence of the inertia of the air stream, which at the instant of closure maintains air particles in motion to create a compression below the glottis and a rarefaction (suction) above the glottis. The travelling pressures P_1^- and P_2^+ leaving the glottis at the instant of closure are therefore of opposite phase. If the first formant of the subglottal system or the first formant of the supraglottal system (or both) are now of such a frequency that $P_1 + P_2$ is positive during opening of the next cycle and negative during closing (Fig. 2), the vocal fold driving pressure is assisted by these acoustic pressures to maintain its asymmetry, i.e., its phase-lock with tissue velocity.

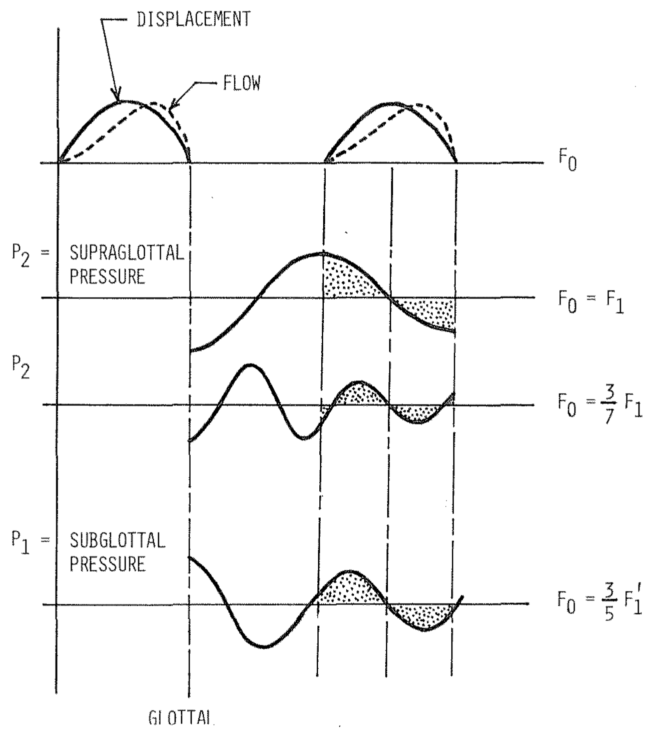


Fig. 2. Phase relationships between vocal fold displacement, flow, supraglottal and subglottal pressures.

It is clear that the condition $F_0 = F_1$ for the supraglottal system provides one such positive phase enhancement. The supraglottal pressure P_2 is maximum at the instant of glottal opening and minimum at the instant of glottal closure. Thus, adjusting the fundamental frequency to lie on the first formant frequency tends to improve conditions for self-oscillation. There are other possibilities, however. If F_0 = (3.7) F_1, or slightly less than half of the first formant frequency, a positive force is experienced over nearly all of the opening portion of the cycle and a negative force over the closing portion, although the peak force comes later during opening and earlier during closing (Fig. 2).

A similar condition is found when the fundamental frequency is slightly greater than half of the first formant frequency of the subglottal system. Specifically F_0 = (3.5) F_1' is an optimal case, where F_1' is the first subglottal resonance frequency. According to measurements made by Ishizaka et al. (1976), this resonance is approximately 600 Hz, which would imply that a fundamental frequency of 360 Hz would interact favorable with the subglottal system to sustain self-oscillation.

Note that if F_0 is larger than F_1 or (3/5) F_1', the second cycle of glottal opening is shifted forward in time (Fig. 2, top trace), the negative portions of the acoustic pressures being delayed with respect to the closing movement of the vocal folds. On the average, then, the pressure ($P_1 + P_2$) in Eq. (8) has a positive bias. This will tend to force the vocal folds toward a more open position, with a reduced sharpness of closure. It is possible that this interaction is responsible for the major chest-voice to falsetto-voice transition experience in male and female voices in the region of 250-400 Hz. It would explain why, for a given voice, the transition is rather abrupt if no supraglottal (vowel) adjustments are made to compensate for the effect of the subglottal interaction, which is rather fixed in frequency.

The 360 Hz optimal value of F_0 for subglottal interaction falls between F and F# above middle C on the keyboard. This is the "bottle-neck" or passaggio region for higher voices, for lower voices it is a few semitones lower, possibly because the trachea is a bit longer. The measurements of F_1' were based on four male and one female Japanese subjects. The average frequency was reduced from 640 Hz to 600 Hz for Westeners because they are slightly taller, on the average. Aside from the one female, whose F_1' was not much different from the males, no data contracting male and female subglottal resonances are available. It is clear that more work is needed to substantiate this claim about chest register transitions being related to subglottal interactions.

Conclusion

Vocal fold oscillation is produced and maintained by velocity-dependent driving forces. The primary mechanism for producing a positive net driving force during glottal opening and a lesser (or negative) driving force during glottal closing is the surface-wave motion on the mucosa of the vocal folds, which is set up by an interaction between fluid flow and the yielding glottal wall.

A secondary mechanism for enhancement of self-oscillation is found in the interaction between subglottal and supraglottal acoustic pressures and the velocity-dependent driving force. A fundamental frequency of approximately one-half of the first subglottal formant frequency, or a fundamental frequency equal to the first supraglottal formant frequency (or both), appears to provide maximum positive interaction. This has led to the hypothesis that the primary chest-falsetto register transition is tied to the somewhat fixed tracheal (chest) resonance, but can be modified by supraglottal (vowel) adjustments.

At low fundamental frequencies (100-200 Hz) the interaction would tend to be less strong because the formant energy has died down during the closing portion of the cycle, and there are too many phase reversals in the formant-related driving pressures to cause much synchrony with the tissue movement. The acoustic pressures may contribute to pitch irregularities (jitter), however. More research is needed to investigate these lesser interactions, including secondary register shifts that seem to exist at very high F_0's in many voices.

References

Ananthapadmanabha, T. and Fant, G. (1982): "Calculation of the true glottal flow and its components", STL-QPSR 1/1982, 1-30.

Conrad, W. (1980): "A new model of the vocal cords based on the collapsible tube analogy", Med.Res.Eng. 13, No. 2, 7-10.

Cooper, D. and Titze, I. (1983): "Generation and dissipation of heat in vocal fold tissue", J. Speech Hear.Res., in review.

Degryse, D. (1981): "Temporal simulation of wave propagation in the lossy vocal tract", The Fourth F.A.S.E. Symposium, Venezia, 193-196.

Fant, G. (1982): "Preliminaries to the analysis of the human voice source", STL-QPSR 4/1982, 1-27.

Guerin, B., Degryse, D., and Lherm, B. (1983): "Vocal tract analog synthesizer with lumped losses", J.Acoust.Soc.Am. 73, Suppl. 1. S5.

Hirano, M., Kahita, Y., Kawasaki, H., Gould, W.J., and Lambiase, A. (1981): "Data from high-speed motion picture studies", pp. 85-91 in (Stevens & Hirano, eds.) Vocal Fold Physiology, Univ. of Tokyo Press.

Ishizaka, K. and Flanagan, J. (1972): "Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell Syst.Tech.J. 51, 1233-1268.

Ishizaka, K. and Flanagan, J. (1976): "Acoustic characteristics of the air volume displaced by the vibrating vocal cords", Proc. (British) Inst.Acoust.

Ishizaka, K. and Matsudaira, M. (1972): "Fluid mechanical considerations of vocal cord vibration", SCRL Monography No. 8, Santa Barbara, CA, USA.

Ishizaka, K., Matsudaira, M., and Kaneko, T. (1976): "Input acoustic impedance measurement of the subglottal system", J.Acoust.Soc.Am. 60, No. 1, 190-197.

Rothenberg, M. (1981): "Acoustic interactive model for the voice source", in (Abbs & Bless, eds.) Vocal Fold Physiology: Current Research and Clinical Issues, College Hill Press.

Stevens, K. (1979): "Physics of laryngeal behavior and larynx modes", Phonetica 34, 264-279.

Titze, I. (1980): "Comments on the myoelastic-aerodynamic theory of phonation", J. Speech Hear.Res.23, No. 3, 495-510.

Titze, O. (1984): "Parameterization of the glottal area, glottal flow, and vocal fold contact area", J.Acoust.Soc.Am., 75, 570-580.

Titze, I. and Talkin, D. (1980): "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation", J.Acoust. Soc.Am. 66, 60-74.

THE QUALITY OF CHILDREN'S SINGING VOICE

M. C. Ametrano Jackson
Department of Audiology and Speech Sciences
Michigan State University
Michigan, USA

Abstract
The purpose of this paper is to study the quality of a normal child's singing voice. This analysis provides objective elements to distinguish the voice of a child who has been professionally trained from the voice of a child who hasn't.

Introduction

We expect in this study to give an objective distinction between good and mediocre vocal technique.

The infantile voice has specific characteristics different from the adult male and adult female voices. These specific characteristics are related to the anatomy, the physiology, and the development of the child's vocal structures.

Bibliography on children's voice characteristics is scarce. Generally, research on this topic concerns pathological aspects and almost never studies of the quality of "normal" children's voices.

In the world of music there are more adult choirs than children choirs because the adult voice is more stabilized than the child's voice.

The infantile voice is stabile enough for professional singing only at the age of four years. During these four years of active professional singing, professionally trained children learn faster and more intensely, due maybe to their natural talent than children who might sing occasionally.

The author thinks that it is not convenient to wait until adult age to begin training the voice. Rather, this training can start at an early age, providing that proper techniques which would not damage the vocal structure are utilized.

The laws of nature have to be followed for this purpose, e.g., if a child is endowed with a soprano voice, then it is not possible to force this child to sing as a contralto. Such a change may produce serious problems in the vocal structures. Clearly, if the "tessitura" is between SOL3 (G3) and FA4 (F4) the child would be a soprano, if the "tessitura" is between DO3 (C3) at SI3 (B3), the child could be considered a contralto.

The voice quality is more apparent for the "tessitura" range than for other notes. The soprano and the contralto can have a range from 14 to 16 notes.

The best period for singing a sufficiently large musical repertorie is approximately from the ages of 9 to 14.

Cornut (1971) states that the frequency range of the children's speaking voices is approximately from RE3 (D3) to LA3 (294-440 Hz).

The extension and the tessitura of the singing voice of children are:

<u>Extension</u>

Soprano - - - - - - - - DO3 (C3) 262 Hz to RE5 (D5) 1175 Hz
Mezzo soprano - - - - - SOL2 (G2) 196 Hz to LA4 (A4) 880 Hz
Contralto - - - - - - - FA2 (F2) 174 Hz to SOL4 (G4) 784 Hz

Tessitura

Soprano - - - - - - - - SOL3 (G3) 392 Hz to FA4 (F4) 698 Hz
Mezzo soprano - - - - - RE3 (D3) 294 Hz to DO4 (C4) 523 Hz
Contralto - - - - - - - DO3 (C3) 262 Hz to SI3 (B3) 494 Hz

The range of the singing voice is progressively enlarged as the children become more experienced in choir practice; also their musical knowledge becomes more significant.

The author thinks that a correct singing voice has to be based on:

(1) Good physical health
(2) A vocal tract with no functional or organic pathologies
(3) Good resonance cavities
(4) No hearing disabilities
(5) Talent and musical temperament

Clearly, it is necessary to teach the singing child according to phonological rules, considering proper methods of relaxation, phonorespiration, resonance and acoustic emission. It is possible to train children's voices in order to obtain full solid, round, vibrant sound uttered with no hesitation, jitter, or shimmer within the whole musical scale.

Subjects

Eight soloist children, belonging to 3 choirs of Paris, France, were utilized for this measurement. The subjects were from 9 to 12 years old. All subjects were natives of France, speaking the Parisian dialect of French.

According to the subjective evaluation of the author two groups were differentiated:

Group = 1 ; A - B - C - D -, Good vocal technique
Group = 2 ; E - F - G - H -, Mediocre vocal technique

Phonetic-materials

(a) Each subject sang the vowels /a/, /i/, /u/, using a diatonic scale of 8 notes; the scales started at SOL3 (G3) for sopranos and DO3 (C3) for contraltos.

(b) A song.

(c) The spoken text of the song.

Procedure

The acoustical analysis of the voice involved consisted of comparing the spectra of the voices by three independent methods, namely (1) aural short term memory, (2) spectrography, and (3) integrator of spectral density (I.D.S.). Spectra of children with good vocal technique were compared with spectra of the mediocre vocal technique.

Subjects produced the recordings of the phonetic materials in the rehearsal rooms utilized for the 3 choirs involved; these rooms had comparable acoustics and resonance. Subjects stood at one meter apart from the microphone (Sennheiser) connected to a tape recorder (Nagra, E mono 1 speed CCNR 50 Hz), tape utilized was Agfa PEM 468 open reel, 7 inches diameter, thinnest 1/1000 inches. Recordings were monophonic at the speed 7.5 i.p.s. and at the SPL of 80 dB.

Recordings were play-back into:

(1) Aural short term memory method of comparing voices. This method (Tosi, 1979) consists of listening loop cartridge recordings using special equipment that includes two tape recorders connected to a dual channel amplifier and a system of fast switches that allows the listener to pass rapidly from one voice to another voice being compared with it.

A listener then is able to listen to short segments of the two singing voices (good vocal technique, mediocre vocal technique) on a continuous loop fashion.

The evaluation of similarities or differences between the samples compared was established by a subjective scale from 0 (zero) to 10 (ten), 0 meaning completely different and 10 absolutely similar.

(2) A spectrograph (Kay Electric, model 7029A) to obtain spectrograms utilizing 150 Hz band filters and a linear scale up to 8000 Hz. Five spectrograms per subject were obtained, i.e., 3 spectrograms corresponding to the scale, 1 spectrogram for the singing voice, and 1 spectrogram for the spoken text of the song. Spectrograms of children were compared with same vowels sung. These comparisons were done by using the same subjective scale as that used in item (1).

(3) An integrator of spectral density (IDS integrateur de densite spectrale) model 03 Sodiment, France was used to obtain the acoustic energy distribution within the following frequency bands: 50-200 Hz; 200-400 Hz; 400-800 Hz; 800-1200 Hz; 1200-1800 Hz; 1800-3000 Hz; 3000-6000 Hz; and 6000-15000 Hz.

The output graph yields the acoustic energy within each frequency band as a percentage of the total energy of 10 seconds of input.

Results

(1) From aural term memory method of comparing voices

Results from this test are displayed in Table 1.

(2) From spectrograms

Spectrograms yielded by group 1 with good vocal technique were compared with those yielded by group 2 with mediocre vocal technique. It was observed that consistently group 1 spectrograms presented much stronger intensities in the harmonic between 3000 and 4000 Hz than group

Table 1. Results from aural term and spectrographic comparison of voice "good vocal technique, mediocre vocal technique".

Results from aural short term memory analysis.

Scale 0 to 10

Mean score of comparison good vocal technique vs. mediocre vocal technique (singing voice)

Subjects

	E	F	G	H	X	D = 10-X
A	5	4	4	7	5	5
B	4	4	5	5	4.5	5.5
C	6	5	5	4	5	5
D	4	4	5	6	4.7	5.2
Grand mean =					4.8	5.2

Results from spectrographic analysis.

Scale 0 to 10

Mean score of comparison spectrogram, good vocal technique and mediocre vocal technique (singing voice)

Subjects

	E	F	G	H	X	D = 10-X
A	5	5	5	6	5.2	4.7
B	5	5	4	6	5	5
C	6	6	5	6	5.7	4.2
D	5	5	6	4	5	5
Grand mean =					5.2	4.7

2. This is possible due to a good vocal technique.

Results from this test are displayed in Table 1.

(3) From IDS (integrator of spectral density)

The distribution of energy within the bands is presented in Tables 2 and 3.

Discussion

(2) From spectrograms

The harmonics of the contralto are closer in comparison with those of the sopranos.

Table 2. Distribution of energy from IDS (integrator of spectral density). Diatonic scale, vocalization of the vowels.

Groups			Good technique				Mediocre technique			
	Subjects		A	B	C	D	E	F	G	H
Band 1	50 Hz	a	5	6	5	7	7.5	11	5	7
	to	i	6	11	8	8	7	12.5	8.5	11
	200 Hz	u	7.5	13.5	7	8	8	12.5	7.5	9.5
Band 2	200 Hz	a	7	8.5	6.5	10	10.5	20.5	6.5	12
	to	i	9	22	13	12.5	11.5	27.5	14	20
		u	12	26	10.5	13	12	25	11.5	16.5
Band 3	400 Hz	a	19.5	18.5	18.5	24	30.5	40	20	27.5
	to	i	24.5	16	31	35	28.5	36	40	40
	800 Hz	u	29	22.5	30	32.5	34.5	40	35	35
Band 4	800 Hz	a	27	21	30	22.5	25	11	27.5	21
	to	i	24	3.5	21.5	12.5	25.5	7	13	11.5
	1200 Hz	u	23	4	25	17	22	8.5	18	13.5
Bands 2-3-4	200 Hz	a	53.5	48	54.5	56.5	66	71.5	54	60.5
	to	i	64	41.5	65.5	60	65.5	70.5	81.5	81
	1200 Hz	u	57.5	52.5	65.5	62.5	68.5	73.5	64.5	65
Band 5	1200 Hz	a	20.5	11	21.5	15	11	6	20	16.5
	to	i	9	11.5	7.5	13	11.5	3	10.5	5.5
	1800 Hz	u	14	3.5	13	8	10	4	18	13
Band 6	1800 Hz	a	12	8	10	8	8	4.5	10	7.5
	to	i	13.5	17.5	7.5	10	8	5	7	5.5
	3000 Hz	u	8	13	7.5	6	7	4	7	6
Band 7	3000 Hz	a	7	21	5.5	6	4.5	2.5	5	3.5
	to	i	7	18.5	4.5	5	4.5	3.5	3.5	3.5
	6000	u	5	12	4	3.5	4	2.5	4	3.5
Bands 6-7	1800 Hz	a	22	35	18.5	16.5	15	8	17.5	13
	to	i	23	42.5	14	17	15	9.5	12.5	10.5
	6000 Hz	u	15.5	30	13.5	11.5	13	7.5	13.5	11.5
Band 8	6000 Hz	a	3	6	3	2.5	2.5	1	2.5	2
	to	i	2.5	6.5	2	2	2.5	1	2	1.5
	15000 Hz	u	2.5	5	2	2	2	1	2.5	2

Group 2 has the low frequency harmonics more reinforced than the high frequency harmonics, because of a technical vocal problem. On the contrary, group 1 has the low frequency harmonics less reinforced than the high frequency harmonics around 3000 Hz to 4000 Hz; in this region often a formant appears for the singer's voice.

The spectrograph analysis of the subjects A, B, C, and D showed that the area from 3000 to 4000 Hz displays a large amount of acoustic energy. According to the subjective evaluation of the author, these subjects have a voice of a "professional" quality, the conclusion is that concentration

Table 3. Distribution of energy from IDS (integrator of spectral density). Singing and speaking voice of groups 1 and 2.

Groups			Good technique				Mediocre technique			
Subjects			A	B	C	D	E	F	G	H
Band 1	50 Hz to 200 Hz	Singing	6	8.5	8.5	10	15	14	13	10
		Speaking	12	12.5	15.5	23.5	20	16	35	19
Band 2	200 Hz to 400 Hz	Singing	10	16	15.5	17	33.5	26.5	27.5	18
		Speaking	22	26	32	31	32	30	28.5	37.5
Band 3	400 Hz to 800 Hz	Singing	27	22	31.5	30.5	22.5	32	35	40
		Speaking	18	21.5	27.5	24	25	30.5	18	20.5
Band 4	800 Hz to 1200 Hz	Singing	16	10.5	16.5	17.5	8	10	9	9
		Speaking	10	9.5	6.5	6	6	6.5	4	4.5
Bands 2-3-4	200 Hz to 1200 Hz	Singing	54	49.5	63.5	65	64	68.5	71.5	67
		Speaking	56	56	66	61	63	67	50.5	62
Band 5	1200 Hz to 1800 Hz	Singing	19	12	9	8	6	7	5	7.5
		Speaking	11	7.5	3.5	3.5	4	4	2.5	2.5
Band 6	1800 Hz to 3000 Hz	Singing	11	11.5	9	8	6	5	4	6
		Speaking	12	8	5	4	5	5	4.5	4.5
Band 7	3000 Hz to 6000 Hz	Singing	8	14.5	6.5	6	5	3.5	3	4.5
		Speaking	10	9.5	5	3.5	4	5	4	4.5
Bands 7-8	1800 Hz to 6000 Hz	Singing	19	26	15.5	14	11	8.5	7	10.5
		Speaking	21	17.5	10	7.5	9	10	7.5	8.5
Band 8	6000 Hz to 15000 Hz	Singing	3	5	2.5	3	3.5	1.5	2	3
		Speaking	5	5	5	3.5	3.5	3	2	4

of energy on this region of frequency is an indicator of such quality. This agrees with the conclusion of Winckel (1973) and Sundberg (1973) in a similar study on adult singers' voices.

In addition, spectrograms for subjects A, B, C, D from group 1 showed "vibrato" technique, apparent by listening, that is not present for subjects E, F, G, and H from group 2.

Also the singing and the speaking vowels present similar formant mean frequencies ratio. However, the absolute value increases with the increasing of the fundamental frequency.

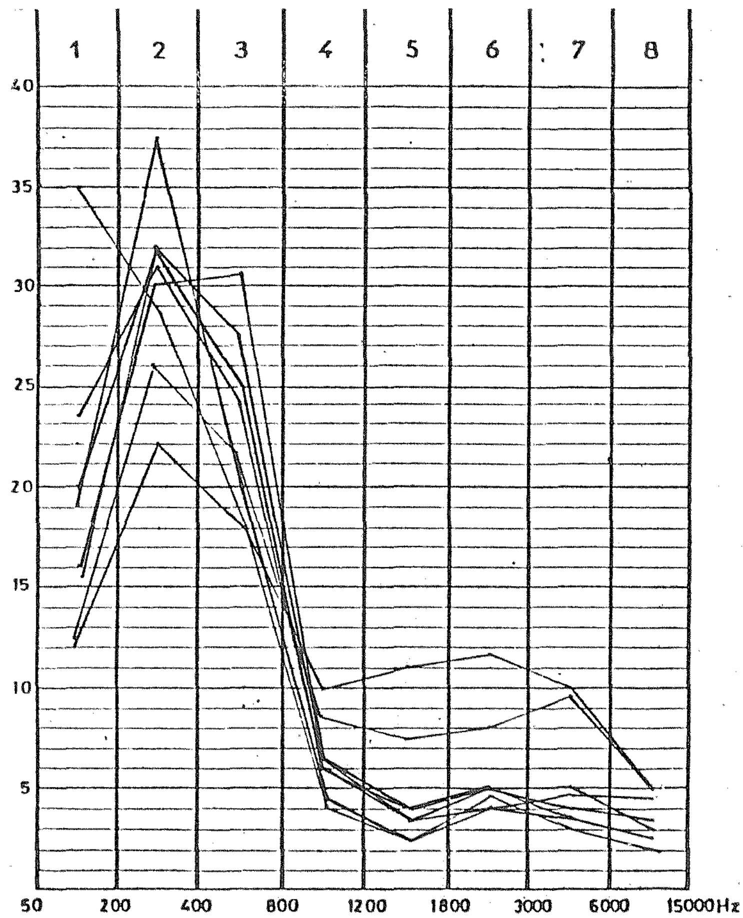


Fig. 1. Distribution of energy from IDS (integrator of spectral density). Speaking voice of groups 1 and 2.

(3) From the integrator of spectral density I.D.S.

Subjects from group 2 presented a large percentage of energy distribution in bands 2, 3, and 4 for all vowels, possibly due to their poor singing techniques. See Table 2.

For this reason the intensity of the singers with the poor singing technique lowers in bands 6 and 7 (1800 Hz - 6000 Hz).

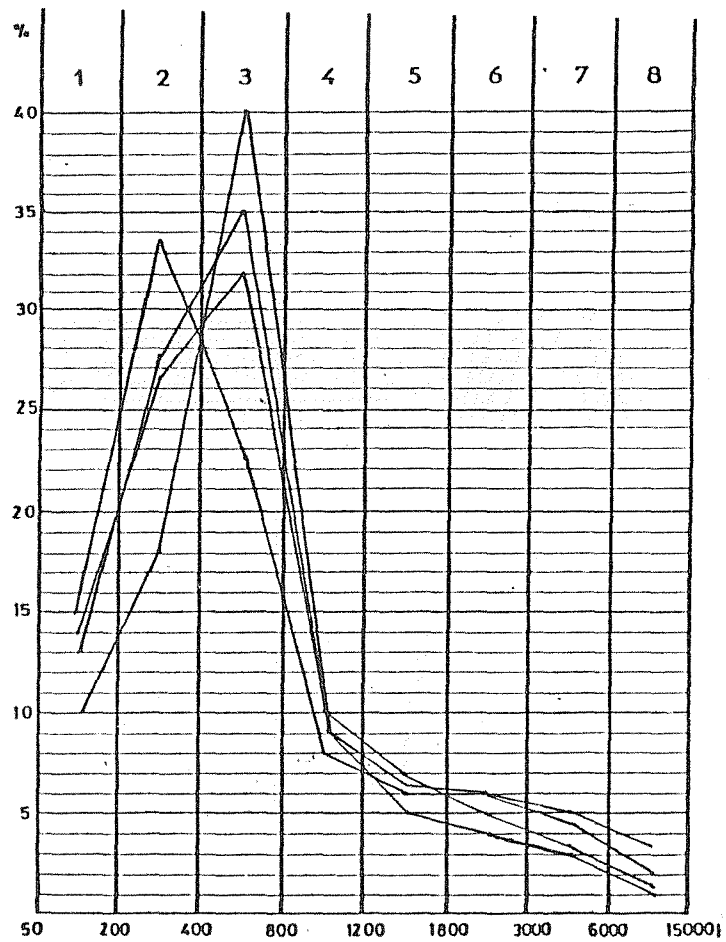


Fig. 2. Distribution of energy from IDS (integrator of spectral density). Singing voice of group 2 (mediocre technique).

The acoustic energy is concentrated in bands 6 and 7 for subjects A, B, C, and D from group 1, those with the better singing techniques. This is a general characteristic of the professional singer's voice. See Fig. 1.

This differential distribution of energy apparent through the I.D.S. instruments allows the evaluation of a "professional" and a "non professional" child's singing voice. See Table 3.

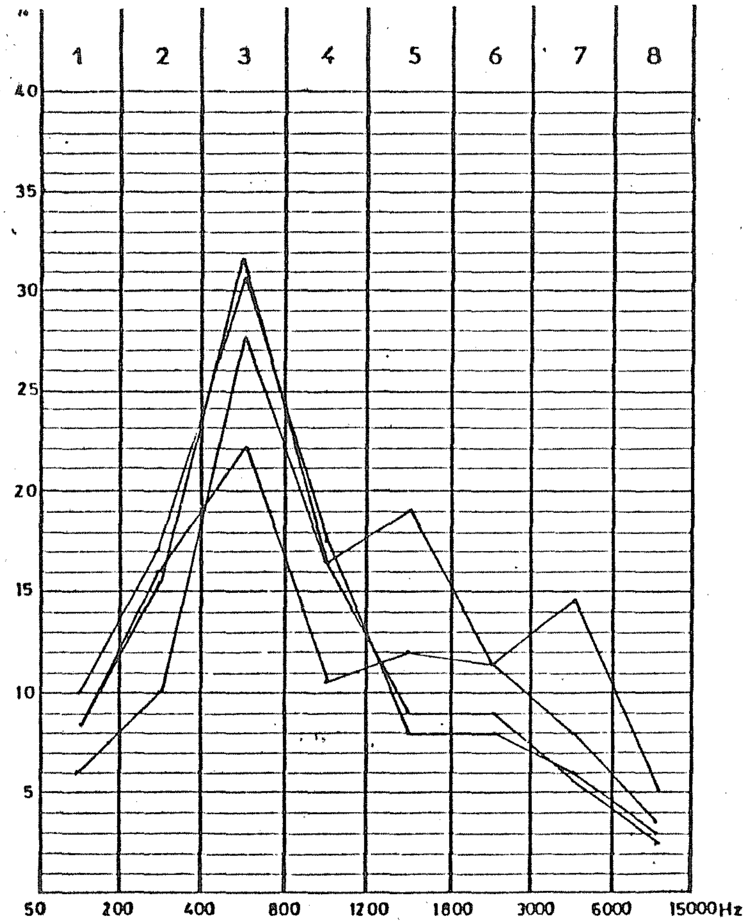


Fig. 3. Distribution of energy from IDS (integrator of spectral density). Singing voice of group 1 (good technique).

The speaking voice has a concentration of energy in bands 2 and 3 for all the subjects from groups 1 and 2, see Fig. 1. The voices from group 2 have a concentration of energy in bands 2 and 3 and do not have a differentiation from the singing voice. See Fig. 2.

Nevertheless, group 1 has a concentration of energy in bands 3 and 4 and 6 and 7 (the peak that appears in it between 3000-4000 Hz is called the 'singing formant'), see Fig. 3.

Conclusions

This study provided objective elements to distinguish the voice of a child who has been professionally trained from the voice of a child who has not.

The I.D.S. (integrator of spectral density) output provides objective elements to explain the general quality of the voices of singing children. It can be stated that there is a strong correlation between the results of the I.D.S. and the spectrograms.

It could be stated that the spectral region between 3000 and 4000 Hz defines the quality of the voice of the singing children.

The singing and the speaking vowels present a similar formant mean frequency ratio. However, the absolute value increases with the increasing of the fundamental frequency.

The distribution of energy from group 2 (mediocre vocal technique) is close to the speaking voice. Nevertheless, group 1 (good vocal technique) has more concentration of energy in bands 6 and 7 (the peak called the singing formant).

The comparison of voices from the subjective analysis gives a difference of 47% between good and mediocre vocal technique.

Acknowledgements

This study was initiated at the "Laboratoire d'Acoustique", University of Paris VI, France, and finished at the Department of Audiology and Speech Sciences, Michigan State University.

The author thanks: Mlle M. Castellengo, J. Sapaly, and Dr. O. Tosi for their technical collaboration; and the choir conductors Mr. F. Bardot, Mr. R. Thirot, and Mr. J.P. Loré for making possible the recordings of the children. Thanks are also due to Professor J. Sundberg for his interesting advices in the preliminary version of this paper.

References

Ametrano Jackson, M.C. (1976): "Clasificación de la voz cantada", Fonoaudiologica 22, 3, 128-130.

Ametrano Jackson, M.C. (1982): "La voix du jeune chanteur", Folia Phon. 34, 276-280.

Cornut, G. (1971): "Contribution a l'étude de la voix parlée et chantée de l'enfant normal de 5 à 9 ans", Folia Phon. 23, 381-389.

Sundberg, J. (1973): "The source spectrum in professional singing", Folia Phon. 25, 71-90.

Tosi, O. (1979): Voice Identification. Theory and Legal Applications, University Park Press, Baltimore.

Winckel, F. (1973): "Comment mésurer l'êfficacité de la voix chantée professionnelle", Bulletin du Groupe d'Acoustique Musicale, Université Paris VI. No 67.

SPECTRAL AND TEMPORAL DIFFERENCES BETWEEN GLOTTAL FLOW PULSES AS A FUNCTION OF SPEAKER AND SUPRA-GLOTTAL LOAD

L. Boves and B. Cranen
Institute of Phonetics
Nijmegen University, The Netherlands

Abstract

Using a carefully designed and extensively tested inverse filtering technique glottal flow pulses are reconstructed from the acoustic speech wave. This is done for the vowels /a,e,i,o,u/ produced by a singer and by un untrained adult male speaker with various phonation types. The resulting flow pulses are analysed and described in both the spectral and the temporal domain. The present study focusses on the influence which the vowel formants have on the characteristics of the flow pulses. Preliminary results tend to confirm Rothenberg's contention that the supra-glottal system acts as an inductive load on the source generator, and that this effect manifests itself in a vowel dependent skewing of the pulses.

Introduction

Despite of much research and long standing traditions of teaching singing it is not quite clear what the detailed differences are between trained singers' voices and untrained, but normal, speakers' voices. A similar remark applies to differences between voice qualities in singing. Voice and voice quality can be described in a number of ways. In our approach the description will be in terms of the temporal and spectral properties of glottal flow pulses as they can be reconstructed from the acoustic speech wave by means of inverse filtering.

The customary theory of acoustic speech production underlying inverse filtering is based on the assumption that the functioning of the sound generator at the glottis is not normally affected by the resonances of the vocal tract (Fant, 1960; Flanagan, 1972). Recent research has, however, yielded convincing evidence of interactions between the form of the glottal flow pulses and the shape of the vocal tract (Cranen & Boves, 1984; Fant & Ananthapadmanabha, 1982; Rothenberg, 1981). Therefore, it might be worthwhile to try and see if differences between speakers and singers can, at least to some extent, be accounted for by the fact that singers exploit this interaction differently - and notably more effectively - than untrained speakers. In a similar vein, different voice qualities might derive from different forms of interaction between the glottal sound source and the vocal tract load.

In Rothenberg's model the glottis is represented by a time-varying conductance and the vocal tract load by an inductance. Increasing the load inductance gives rise to an increased skewing of the flow pulses with a consequent enhancement of the level of the higher harmonics. Fant & Ananthapadmamabha distinguish two further interactions, viz. truncation, i.e., an increased damping of the formant oscillations in the open glottis interval and superposition of the formant oscillations on what remains from the oscillations excited by the previous glottal pulse. Truncation results in energy disappearing into the subglottal system; thus it is not likely to be a desirable attribute of a singer's voice. Superposition might enhance the formant oscillations, but only if the newly excited oscillations are in phase with those carried over from the previous period. Again, singers might (probably unconsciously) exploit this effect in order to reenforce their voices. The idea that different voice qualities can be brought about by coupling of the voice source to the vocal tract is supported by the finding (from flexible laryngoscope filming) that articulations of the lower pharynx are heavily involved in imitations of speakers and in changing one's voice.

Method

The material on which this study is based comprises vocalisations of two speakers, a professional bass-baritone singer and an untrained speaker whose habitual speaking pitch is in the same region. The singer produced swelltones on 3 vowels (/a, i, o/) at a nominal fundamental frequency of 98 Hz in four different voice qualities, viz. neutral, light, dark and pressed or strained. Due, perhaps, to the fact he had to maintain a constant pitch, not all attempts to produce a target voice quality were equally successful. This will, of course, tamper with the results of a comparison of pulse characteristics between voice qualities. Recordings were made in a small auditorium using normal audio equipment. They were made available to us by Gerrit Bloothooft, who at the time of making the recordings was with the Dept. of Experimental Audiology of the Free University in Amsterdam. We have the recordings in a digital form, sampled at a 10 kHz rate and quantized to 12 bits.

An attempt to have the untrained speaker imitate the vocalisations of the singer failed completely. Therefore we decided to content ourselves with existing recordings that had been made by feeding the output of a B&K type 4134 condenser microphone directly into a computer where it was sampled at 8 kHz, quantized to 12 bits and stored on disc and tape. The speech material used for this study consisted of vowels produced in isolation and vowels taken from monosyllabic words that were also produced in isolation. Unfortunately, it appeared that the fundamental frequency in these vowels was slightly higher than 100 Hz.

Glottal flow pulses were reconstructed by means of acoustic inverse filtering. To that end the instant of major excitation of the vocal tract was determined from the acoustic speech wave. Normally, this 'epoch' coincides with the moment of glottal closure and thus it is followed by the closed glottis interval, the time period during which the vocal system approaches the idealized linear source-system model most closely. No attempt was made to establish the moment of glottal opening. It simply was assumed that the glottis remained effectively closed during approximately 4 ms in the singer (corresponding to 40 samples) and slightly shorter in the speaker (3.75 ms or 30 samples). Covariance Linear Prediction analysis was used to establish the transfer function of the vocal tract during each closed glottis interval and the result used to construct an optimal inverse filter for each period. The number of predictor coefficients was otimized for each vowel token. This was done by visually monitoring the inverse filter output that was displayed on an oscilloscope connected to the computer.

The output of the inverse filter was stored on disc for subsequent hard copy plotting and further spectral analysis. Moreover, formant frequencies and bandwidths were determined by solving the predictor polynomial for its zeros. This information was also written to disc, along with the RMS signal energy in each glottal period, the period duration, the maximum amplitude value of each reconstructed glottal pulse and the maximum (negative) value of the slope of the pulse during the closing gesture of the vocal folds.

We strongly believe that our understanding of what really is going on at the level of the larynx in singing and in speech would greatly benefit from an automatic processing of a large number of vocalisations of a wide variety of singers and speakers. Such an undertaking is, however, senseless if one does not know where to look at in order to find what. Specific research hypotheses can only be formulated on the basis of a detailed study of a limited number of examples. The present paper obviously belongs to the class of exemplary studies, if only since it does not involve more than one speaker and one singer.

The 4 (voice qualities) times 3 (vowels) vocalisations of the singer were analyzed at two points, viz., at the point of maximal amplitude and at a point close to the end where the amplitude was much lower. The high amplitude segments analyzed comprised at least two complete vibrato cycles as judged from the RMS level. Visual inspection of the glottal flow pulses did, however, not reveal a dependency of the pulse forms on their position in a vibrato cycle. Therefore the vibrato will not be a separate factor in the remainder of this study.

From the speaker's utterances detailed analyses were made of three tokens of the vowels (/a, i, o/), two of which were vowels produced in isolation and the remaining one taken from a monosyllable. Since even the isolated vowels had very short onsets and offsets only segments from the central portion of the vowels were analyzed. The vocalisations of the speaker were, of course, vibratoless, but they contained a small amount of pitch perturbation instead. Finally, the pulses reconstructed from vowels produced by human subjects are compared with waveforms produced by means of the two- mass model of the vocal folds developed by Ishizaka and Flanagan (1972). All signals generated by the model were produced with standard settings of the vocal fold parameters.

The spectral analysis of the flow pulses was done by means of cepstrally smoothed FFT spectra. A single period was taken from the signal, differentiated in order to emphasize the higher frequency components and padded with enough zero samples to fill a 1024 point record. The cepstrum window used to smooth the spectra was 25 samples wide, 15 of which were unity, the remaining 10 forming a quarter period of a cosine. Spectra were plotted for at least five consecutive periods of each vocalisation. This was done in order to get an impression of the variability of the results. The spectra were checked to see whether or not they contain zeros attributable to interactions of the source and the load.

The temporal analysis of the reconstructed glottal flow pulses focussed on the steepness of the falling slopes and on the duration of the open glottis interval.

Results

The single most salient experience made in carrying out this research is how much easier it is to obtain smooth and essentially ripple free glottal flow waveforms from the vocalisations of the singer than those of the speaker. Also, in general the order of the inverse filter needed to cancel all formant oscillations was much higher in the singer. A second observation is that the singer did not tend to tune his formant frequencies to harmonics of the fundamental more successfully than the speaker. This is true for all formants, including the higher ones that mostly are absent from the speaker's signals.

All smoothed spectra contain clear ripples, as can be seen from Figs.

1 and 2. The figures should be interpreted against the background of the corresponding formant frequencies and bandwidths given in Table I. Although it is tempting to attribute these ripples to source-load interactions, it should be borne in mind that they may very well be artefacts of the analysis procedure. This is especially the case in the spectra of the speaker's pulses and - to a slightly smaller degree - in the spectra of the pulses from the low amplitude portions of the sung vowels, where the pulse shapes vary quite a bit from period to period, which gives rise to discontinuities at the edges of the time window. The probability that the ripples are indeed an artefact is highest in the pulses reconstructed from the singer's /i/ vowels, where neither the peaks nor the dips bear any systematic relation to the formant frequencies. It is equally tempting, however, to note that they do tend to coincide with the frequencies of the poles of the subglottal system as found by Boves (1983). In any case, neither the pulse spectra of the singer nor those of the speaker exhibit clear traces of a spectral zero at the frequency of the first formant of the /i/ vowels, with one exception, viz. the low amplitude portion of the dark /i/ (not shown in Fig. 1).

The spectra of the pulses derived from /a/ and /o/ vowels, on the other hand, show unmistakable signs of source system interactions in the form of spectral dips at the frequencies of the lower formants of the /a/ and the second formant of the /o/.

Taking due account of all uncertainties in the details of the spectra it is not possible to assess the differences in source-system interaction between the speaker and the singer. Thus we have to confine ourselves to noting that the two subjects show similar effects with respect to the different behavior of the pulses derived from /i/ vowels. The most obvious difference between the spectra of the speaker and the singer is, of course, the slope towards the higher frequencies. The pulse spectra of the speaker's voice fall off much more steeply. The slopes are even steeper than in the pulses of the low amplitude portions in the vowels of the singer.

It is hard to avoid the impression that the pulses produced by the

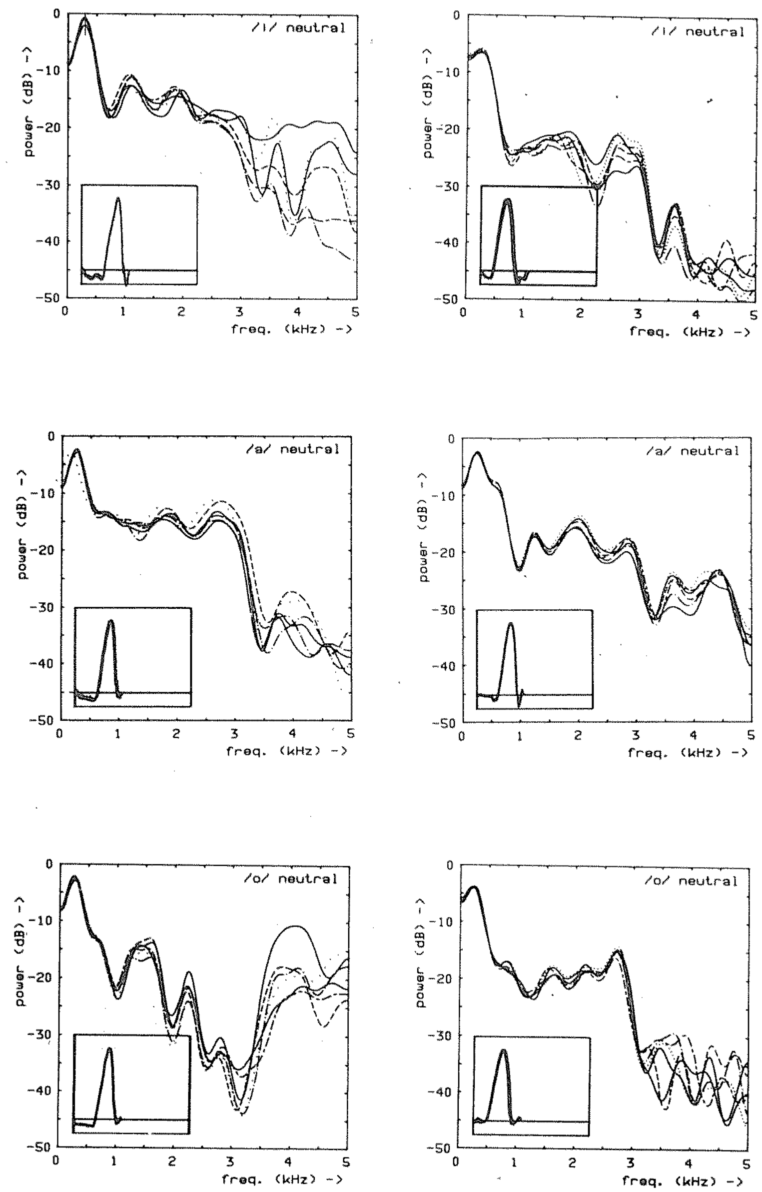


Fig. 1. Spectra of the glottal flow pulses derived from the neutrally sung vowels. The left column contains spectra pertaining to pulses from the centra, high amplitude part of the vocalisations. The right hand column contains spectra derived from the trailing, low amplitude part of the vowels.

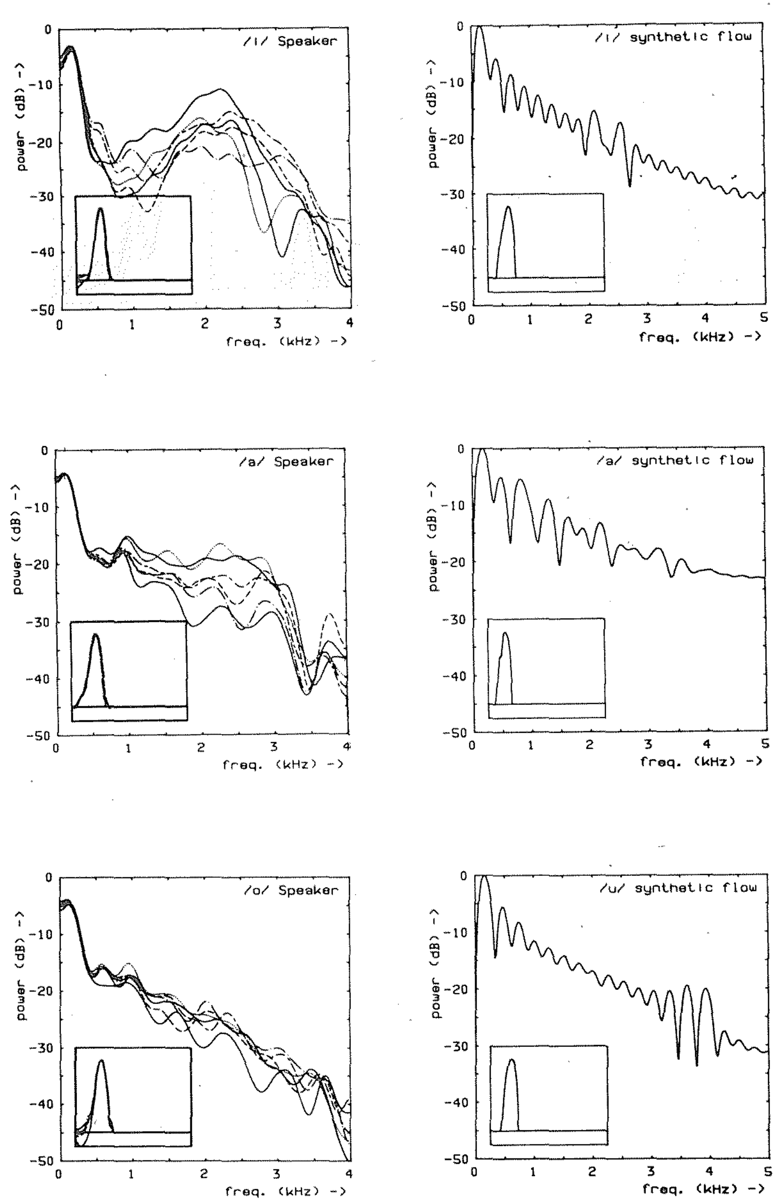


Fig. 2. Spectra of glottal flow pulses derived from vocalisations of an untrained speaker (left column) and of flow pulses generated by means of the two-mass model of the vocal folds, loaded by different vowel tubes (right column).

Table I. Formant frequencies and bandwidths of the vowels of which pulse spectra are shown in Figs. 1 - 3.

Singer		F1	F2	F3	F4	B1	B2	B3	B4
/a/	Light	749	1139	2428	2954	49	90	159	216
		386	1195	2176	--	42	193	126	-
	Dark	652	967	2217	2965	34	44	103	78
		478	905	2068	3046	53	63	117	220
	Pressed	874	1335	2343	2970	44	78	208	107
		764	1158	2300	2798	225	131	469	479
	Neutral	676	1172	2275	2924	46	51	88	120
		833	1061	2321	3172	152	146	286	178
/i/	Light	293	1838	2453	3018	54	54	231	187
		242	1819	2230	3016	56	87	420	192
	Dark	305	1712	2609	2975	54	40	72	69
		229	1627	2711	3081	29	44	81	99
	Pressed	371	1992	2512	3069	43	59	276	188
		249	1923	2427	3198	66	40	192	139
	Neutral	332	1793	2726	3244	60	48	86	147
		294	1748	2340	3243	39	49	132	61
/o/	Light	599	819	2439	3090	87	53	129	111
		440	797	2443	3271	48	58	81	98
	Dark	570	797	2518	3025	55	50	36	179
		445	736	2201	3300	29	65	82	158
	Pressed	619	917	2473	2964	49	67	71	229
		521	795	2057	--	64	76	117	-
	Neutral	588	915	2598	3015	38	53	57	159
		459	767	2299	3051	99	163	91	70

Speaker		F1	F2	F3	F4	B1	B2	B3	B4
/a/		657	1173	2019	3280	89	96	209	115
		658	1167	2169	3250	67	61	201	178
		699	1232	2274	3457	107	110	150	256
/i/		253	1946	2569	3064	42	80	200	280
		303	2022	2320	3034	70	151	206	302
/o/		390	772	2201	3007	16	28	153	62
		381	754	2249	3059	16	32	101	37
		394	868	2293	3114	52	48	210	169

two-mass model show a ripple structure of an amplitude much higher than in the pulses derived from human subjects. Also, the spectral slopes, especially in the upper frequencies, seem to be less steep.

Comparing spectral slopes across vowels it appears that, irrespective of voice quality, the /o/ pulse spectra of the singer tend to fall off more rapidly than the spectra of the /a/ and /i/ pulses. As yet, we

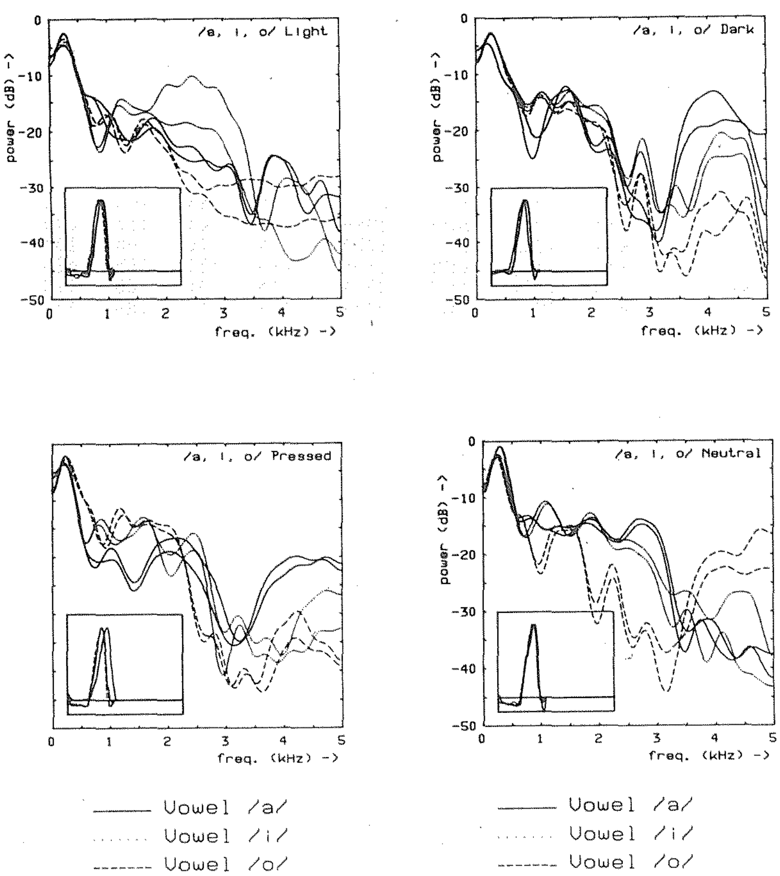


Fig. 3. Pulse spectra derived from the high amplitude portions of the sung vowels compared for vowel and voice quality.

cannot explain this finding. In the speaker's voice the variability of the slopes within the vowel classes seems to be so great as to prevent any conclusion. A similar remark applies to a comparison of the four voice qualities produced by the singer. As can be seen from Fig. 3 the relative spectral slopes associated with voice quality seem to be vowel dependent. It should, however, be remembered that the vocalisations differ in the closeness with which the intended voice quality was approached. Apparently a much larger material should be processed in order to obtain insight into the degree to which these results are systematic.

Figs. 4-6 contain plots of the acoustic pressure waves and the corresponding reconstructed glottal pulses. Except for the pulses derived from the low amplitude portion of the dark /i/ vowel, all glottal flow pulses produced by the singer appear to contain a clear closed interval. There is some evidence of truncation in the singer's /a/ vowels, even in the high amplitude portions; in the light /a/ the truncation seems to affect only every second period. In the speech of the untrained speaker the truncation effect seems to be outweighed by secondary excitations of the vocal tract at the moment of glottal opening.

In the pulses reconstructed from the speaker's utterances flat closed intervals are the exception rather than the rule. His pulses seem to consist of an opening phase, a closing phase, that is interrupted before the glottis reaches a complete closure and a phase during which the remaining glottal opening varies slowly and most probably only slightly.

Upon further inspection a checking of the closing movement and a consequent transition to a much less steep slope just before the moment of complete glottal closure is also evident in the pulses of the singer. It is, however, not clear whether this 'knee' in the pulses of the singer has the same origin as in the speaker's pulses. It is quite conceivable that in the singer the effect is mainly an aerodynamic one, whereas in the speaker it is most probably due to a slight habitual abduction of the vocal folds which gives rise to a never closing chink in the cartilageneous glottis.

The waveforms in Figs. 4-6 are all plotted to the same relative scale. Thus it is clear that the speaker produces pulses of a much larger amplitude than the singer. Also, his open ratio seems to be much higher. Therefore, we are forced to assume that the speaker spends considerably more air per unit phonation time than the singer, which is in accord with the assumption of a slight habitual abduction of the folds in the speaker.

The most important time domain characteristics of the pulses are summarised in Table II. The Table contains information on the pulse amplitude, the (negative) amplitude of the derivative of the flow pulses and - mainly to have some reference as to the effect of the pulses - the

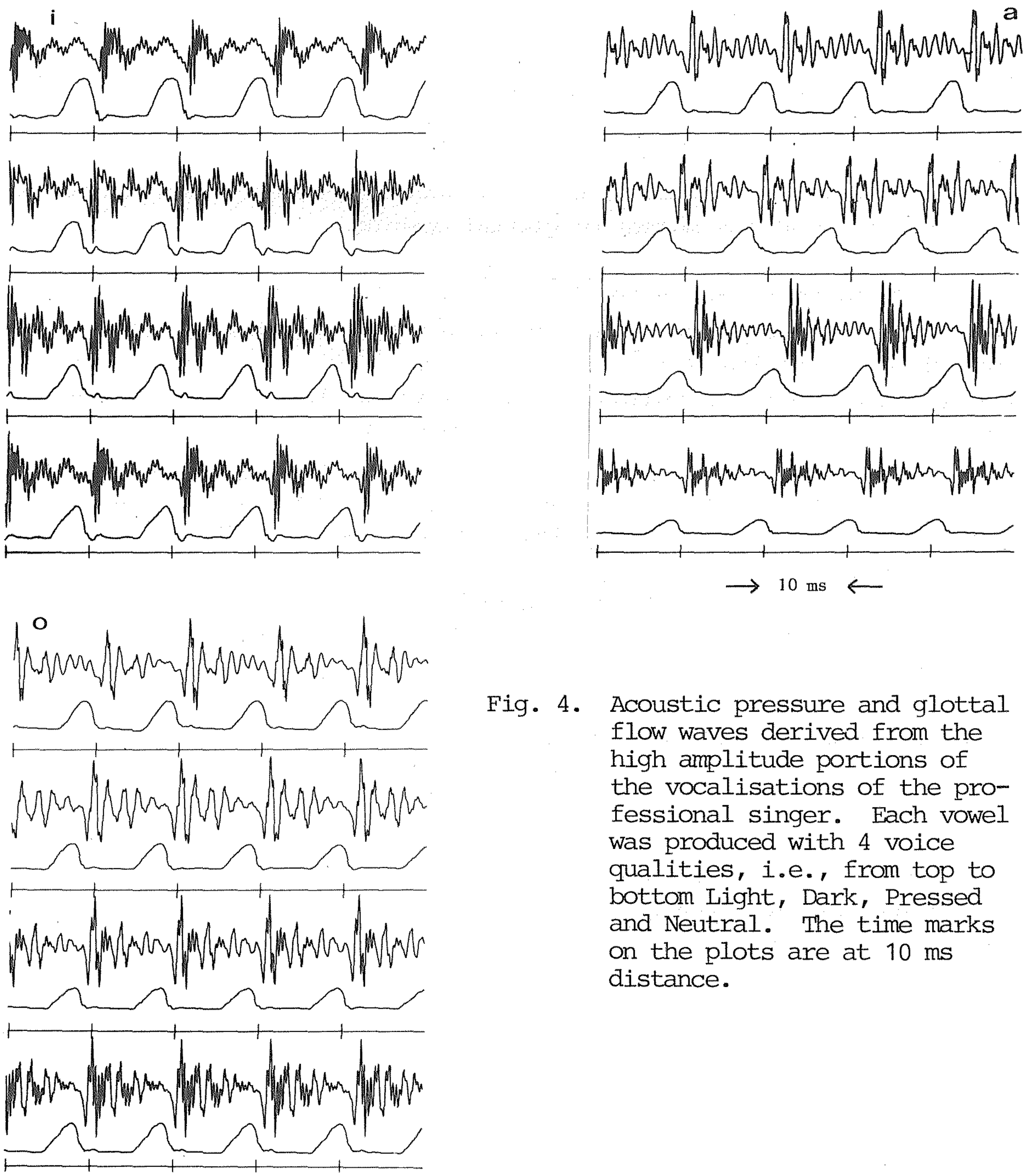
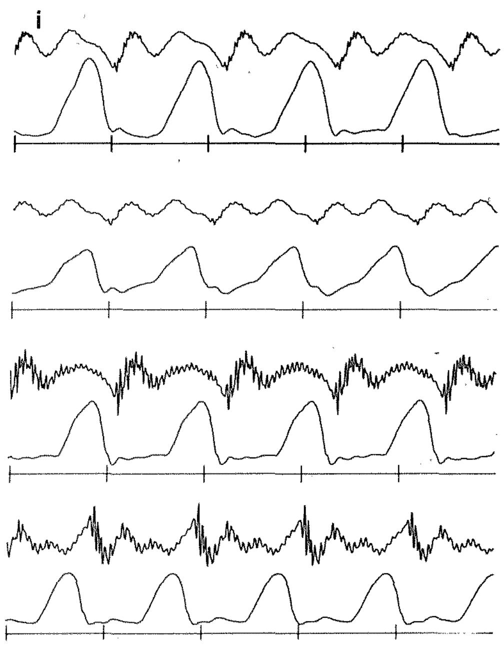
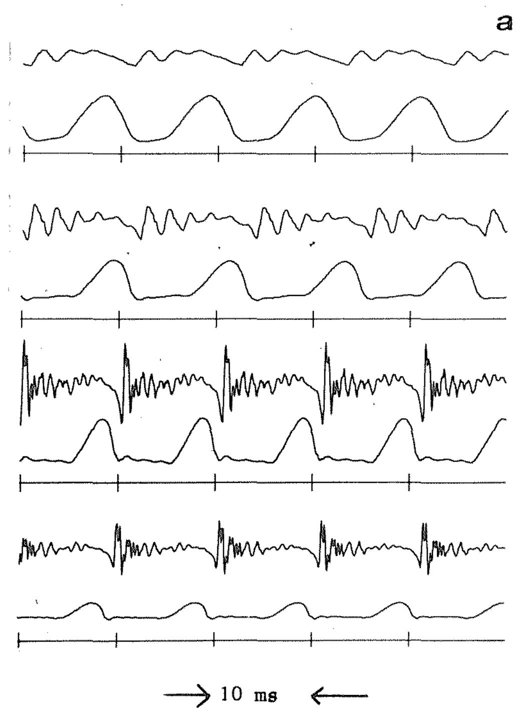


Fig. 4. Acoustic pressure and glottal flow waves derived from the high amplitude portions of the vocalisations of the professional singer. Each vowel was produced with 4 voice qualities, i.e., from top to bottom Light, Dark, Pressed and Neutral. The time marks on the plots are at 10 ms distance.





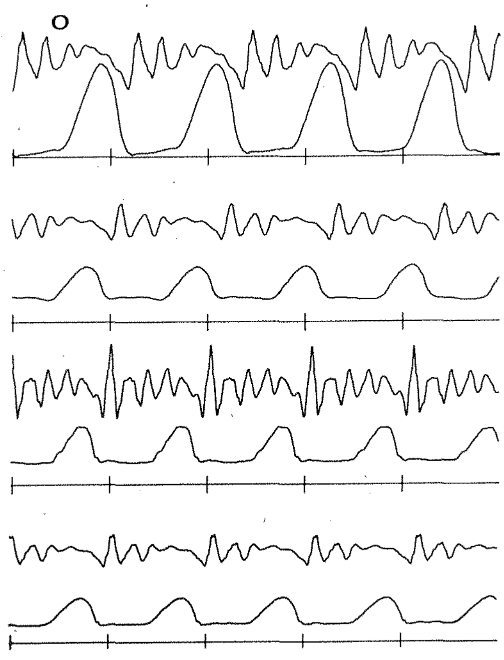
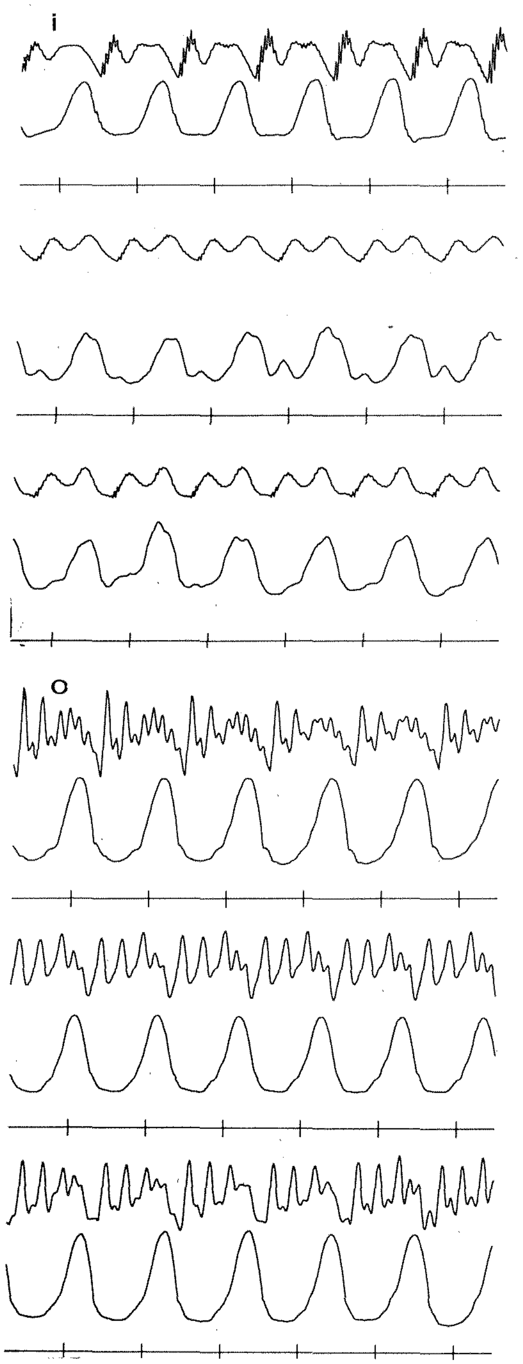


Fig. 5. Acoustic pressure and glottal flow waves derived from the low amplitude portions of the vocalisations of the professional singer. Each vowel was produced with 4 voice qualities, i.e., from top to bottom Light, Dark, Pressed and Neutral.



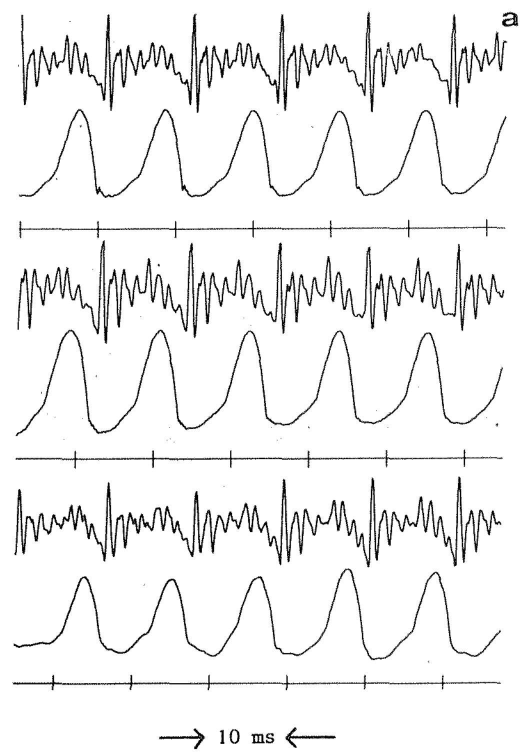


Fig. 6. Acoustic pressure and glottal flow waves derived from three realisations of the vowels /a, i, o/ by an untrained male speaker.

Table II. Time domain characteristics of the glottal flow pulses. The open ratio is given in terms of the frequency and bandwidth of the 'Source Maximum', in the columns labelled Fg and Bg respectively.

		Pulse amplitude	Deriv-ative	Signal amplitude	Fg	Bg
Singer						
/a/	Light	2931	-587	5872	270	260
	Dark	2972	-589	8131	225	310
	Pressed	2918	-510	6661	220	290
	Neutral	2451	-589	6751	265	260
/i/	Light	2700	-726	4447	245	360
	Dark	2039	-505	5731	300	230
	Pressed	2162	-608	7046	250	220
	Neutral	2743	-667	5369	300	240
/o/	Light	2462	-478	4754	265	360
	Dark	2024	-470	5532	265	280
	Pressed	1728	-422	5532	265	280
	Neutral	2595	-561	6801	250	255
Speaker						
/a/		5192	-828	5320	125	220
		5579	-860	5996	110	200
		5099	-860	5121	140	215
/i/		2994	-557	2643	160	250
		5221	-385	2313	150	260
		5207	-551	2798	170	220
/o/		2916	-446	5358	160	250
		5701	-670	5452	150	260
		5443	-820	4874	170	220

RMS value of the speech. Since it proved to be difficult to determine the open ratio of the pulses from the speaker's utterances accurately, we decided to approach this aspect via the detour suggested in a number of papers by Fant (1982). Using his model of the glottal pulse the open time can be estimated from the frequency of the spectral peak corresponding to the 'Source maximum'. For the low amplitude segments of the sung vowels, however, this approach appeared to fail, since they almost invariably have their Source maximum at DC. Therefore, the low amplitude portions of the sung vowels are not included in the time domain analysis.

The data pertaining to the singer display some evidence of a dependence on the falling slope on the supraglottal load, which is highest in /i/ vowels. This is in accord with model predictions by Rothenberg and

Fant. The data derived from the speaker's pulses, on the other hand, show much less systematic behavior. This may be due to chance problems in the reconstruction of the pulses from the particular phonation analyses for this paper. A coarse spectral analysis of a great number of glottal flow pulses recovered from running speech produced by 5 male speakers does indeed suggest that the spectral slope in closed (high) vowels is less steep than in the remaining (half open and open) vowels (Boves, 1982) The data on the frequency (and bandwidth) of the glottal 'Source maximum' do not show any clear vowel dependence. As could have been anticipated from a glance at the pulses depicted in the figures, the frequency of the 'Source maximum' in the speaker is approximately half as high as in the singer. In the pulses of the speaker the maximum is always between the first and second harmonic, whereas it is close to the third harmonic in the singer.

Contrary to our expectation the pulses produced by the two-mass model did not comply with Rothenberg's and Fant's model predictions. It appears that the derivative at the moment of glottal closure is greatest in the case of the /a/ vowel and smallest in the /i/ vowel, the /o/ taking an intermedicate position.

Conclusion

The single most important conclusion that can - or perhaps must - be drawn from this reseach is that the results of studies in which a small number of vocalisations of an even smaller number of subjects are analysed should be looked upon with great caution. Model studies are, by their nature, not subjected to chance fluctuations in their results, but the ecological validity of existing models of the voice source leaves as yet much to be improved.

Acknowledgement

Part of this research was suppoorted by the Foundation for Linguistic Research which is funded by the Netherlands Organization for the Advancement of Pure Research, ZWO.

References

Ananthapadmanabha, T.V. and Fant. G. (1982): "Calculation of true glottal flow and its components", Speech Communication 1, pp. 167-184.

Boves, L. (1982): "Spectral properties of glottal flow pulses as a function of speakers, vowel and stress level", Proc.Inst. of Acoustics, paper A3.

Boves, L. (1983): "Frequency analysis of broad-band subglottal pressure signals", Proc.Inst. of Phonetics Nijmegen, 7, pp. 9-18.

Cranen, B. and Boves, L. (1984): "Spectral consequences of a time-varying glottal impedance", pp. 361-366 in (M.P.R. Van den Broecke & A. Cohen, eds.) Proc.Tenth Int.Congr. of Phonetic Sciences, Utrecht, Vol. IIB, Foris Publ., Dordrecht.

Fant, G. (1960): Acoustic Theory of Speech Production, Mouton, The Hague (2nd edition 1970).

Fant, G. (1982): "The voice source - acoustic modeling", STL-QPSR 4/1982, pp. 28-48.

Flanagan, J.L. (1972): Speech Analysis, Synthesis and Perception, Springer, Berlin.

Ishizaka, K. and Flanagan, J.L. (1972): "Synthesis of voiced sounds from a two mass model of the vocal cords", Bell System Techn.J. 51, pp. 1233-1268.

Rothenberg, M. (1982): "An interactive model for the voice source", STL-QPSR 1/1982, pp. 1-17.

THE MASKING OF CHARACTERISTIC SPECTRA BY THE ACOUSTICO-MECHANICAL RECORDING PROCESS

G. Brock-Nannestad
Gentofte, Denmark

Abstract

From the conception of the idea of sound recording (on paper) until 1925, a horn was used exclusively for the collection of the sound input. Since the turn of the century gramophone records were made by pressing from masters, which had been engraved by using the mechanical force generated by the sound pressure. The desired mechanical efficiency necessitated the use of resonances in the recording system. Many musically important recordings of the voice and instruments were made in this period, and the paper demonstrates the distortions in the spectra of these sound sources caused by the recording process. Two variables carry the greatest importance, viz. recording speed and recording system transfer function. Replay with a speed different from the recording speed causes a transposition of the recorded spectrum, upsetting the absolute relations between fundamental and formants corresponding to a given register. The second variable influences mainly the amplitude of the fundamental and lower formants and the transient properties.

The paper introduces a novel method for extracting recording system parameters from the actual recordings and compares the results with those obtained by simulation of the recording process by means of a surviving recording horn of the period and electrical filters for simulating the parameters of the recording soundbox. Finally, the problem of multiple recording horns is discussed with reference to the large number of recordings made with ´soloist´ and ´ensemble´ horns.

Before 1925 the only process available for recording performances was the acoustico-mechanical (Brock-Nannestad, 1981). Three parameters influence the possibility of correctly replaying such recordings today: speed of revolution, transfer function (linear) distortion, and the variations in tuning pitch with time and place. The present paper will concentrate on the two former.

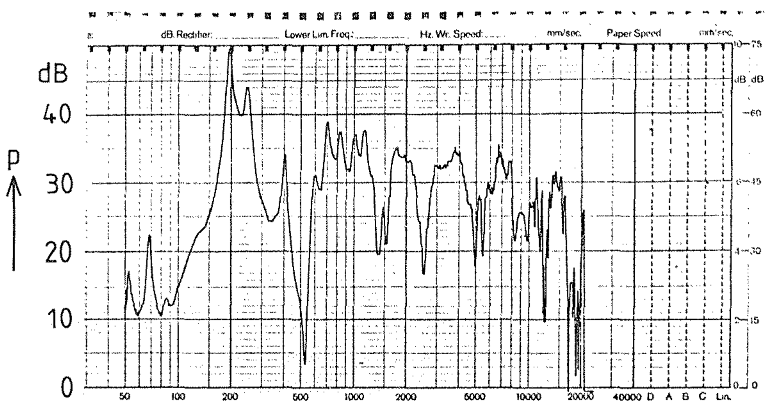


Fig. 1. Transfer function of horn (closed, long tube + sound box) using measuring setup of Fig. 2.

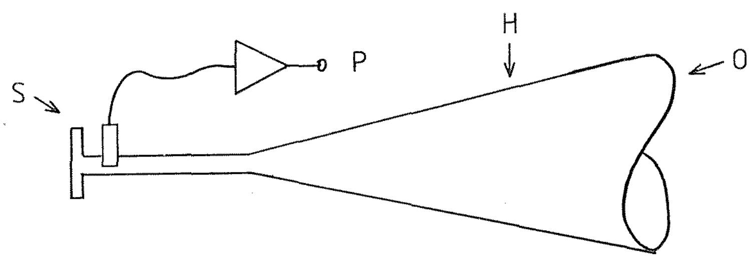


Fig. 2. Setup for measurement of transfer function of a horn. The horn (H) with soundbox (S) is placed in an anechoic chamber and excited through its open end (O) with constant sound pressure and the resulting sound pressure at the throat (P) is measured.

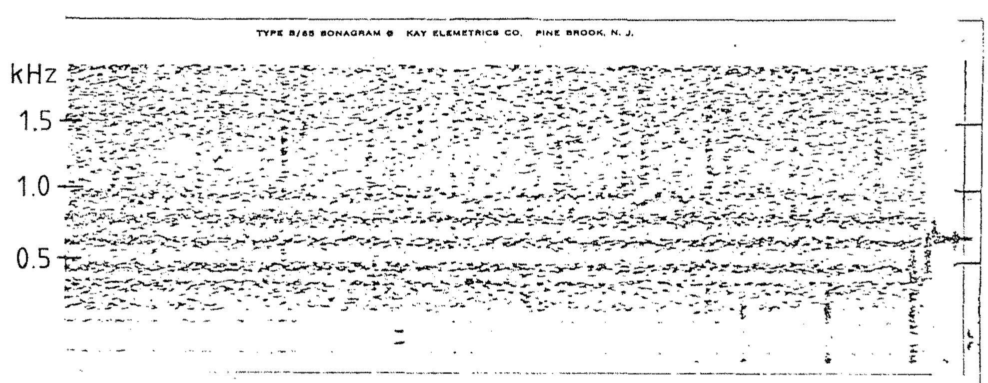


Fig. 3. Spectral distribution of hiss from silent grooves.

The speed problem actually exists for recordings as late as 1940, however, it is the interplay of speed uncertainty with transfer function uncertainty which causes substantial problems in recordings before 1925. It is my contention that the transfer function problem may be solved first - objectively, and that one may (in particular where the voice is concerned) adjust speed afterwards according to a mixture of objective and subjective criteria.

Horns were used for collecting the sound output from performers in the acoustico-mechanical recording process. High efficiency walked hand in hand with sharp resonances which were actually quite unrealistic in the mechanical recording process, since replay could not cope with the high stylus velocities that ensued. Practical experiments with surviving recording horns, terminated with a volume of air corresponding to that in use in the acoustico-mechanical recording process, have shown that dynamic differences of 40 dB exist in the transfer function under such circumstances, Figs. 1 and 2. This range would not be tolerated for two reasons: the record would wear out after just a few replays and the sound quality would be noticeably "tubby". It is interesting to note that Dayton C. Miller who was the first to note the inadequacies of the recording horn (1913) did not have the same dynamic limitations in his Phonodeik. Indeed, he devised a graphical method for compensating his measurements for the horn resonances.

A new method has been developed in order to determine the actual transfer function without recourse to (intelligent!) guesswork as to resonances and damping. The method consists in measuring the spectral distribution of the background noise present on all records made by the acoustico-mechanical recording process. In this way the spectral distribution of the noise signal preceding a recording by Adelina Patti ("la Calasera", 1906) was determined, Figs. 3 and 4. The spectrum was inverted and used in digital filtering of the performance. Sonagrams were made of some characteristic examples of the voice production both in the "straight" and in the correct version, Figs. 5 and 6 (sound examples 1 and 2). Differences in the levels of the harmonics are distinct, however, it is in the amplitude distribution (envelope) that the real, dynamic difference is most clearly observed: the corrected version does

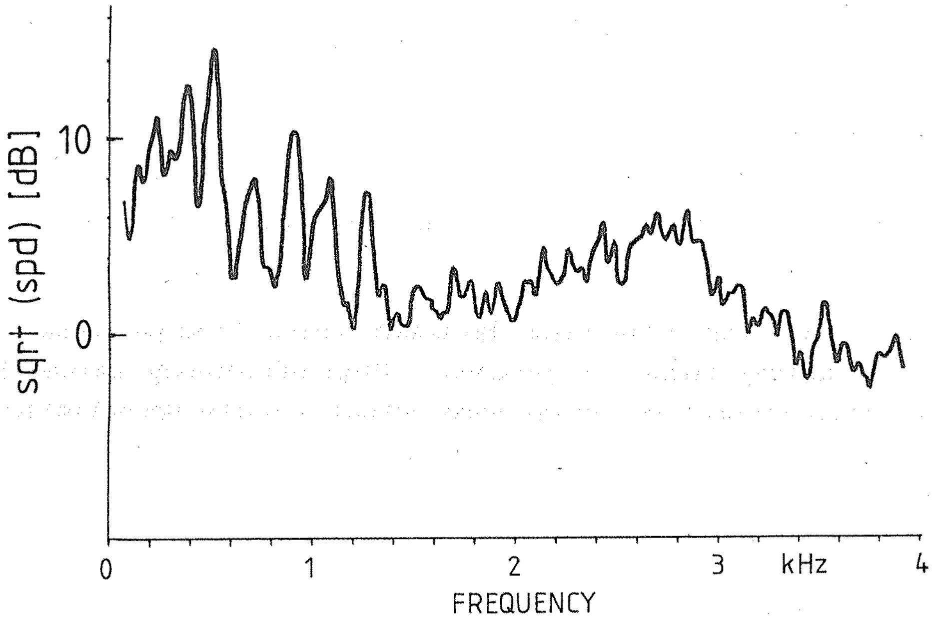


Fig. 4. Integration of spectral distribution of hiss equals horn transfer function.

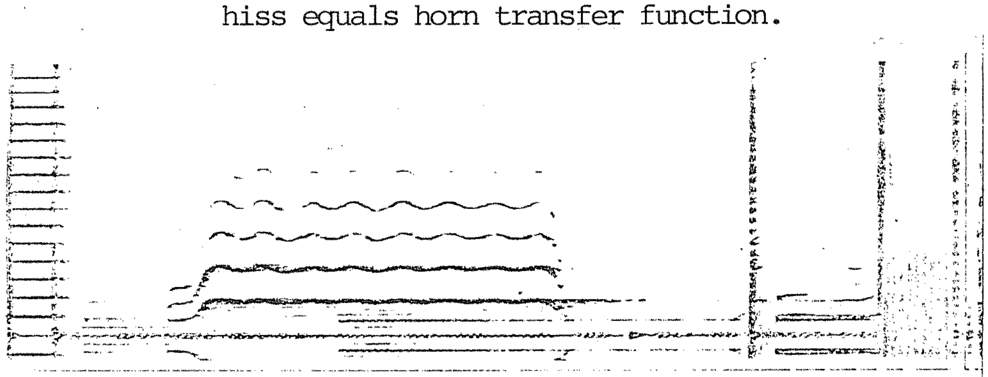


Fig. 5. High note on record "la Calasera" - as found.

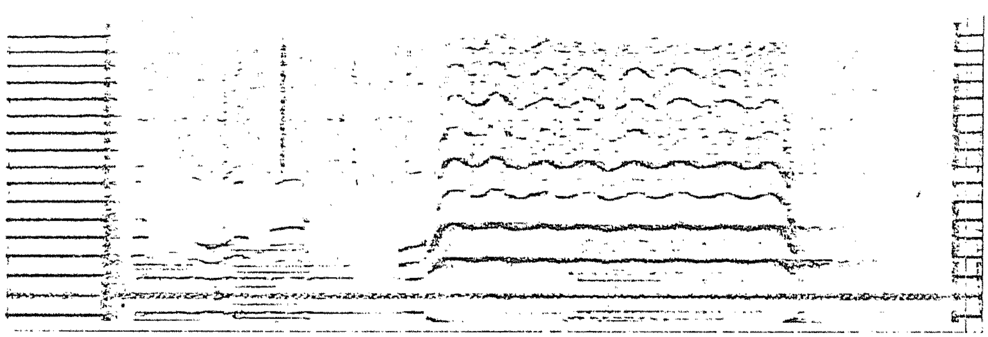


Fig. 6. High note on record "la Calasera" - corrected as described.

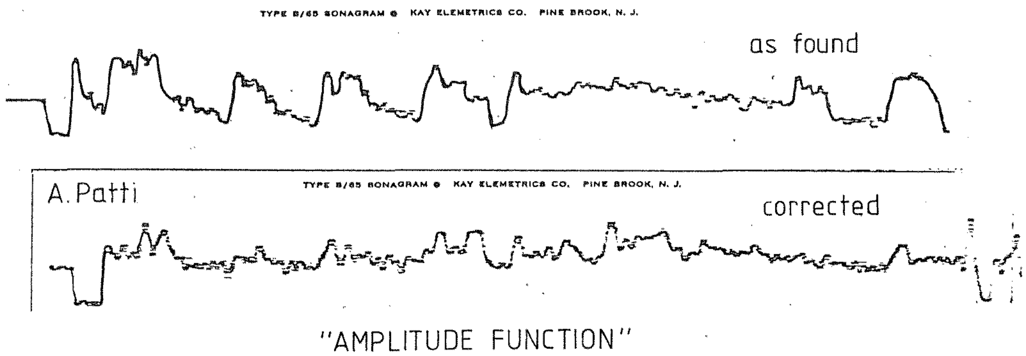


Fig. 7. Envelope or "amplitude" Sonagram of register change.

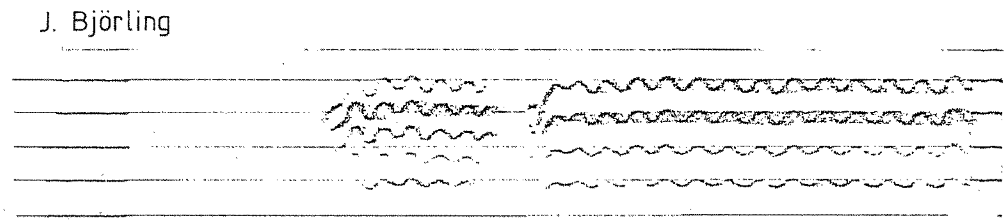


Fig. 8. Last notes of "la donna è mobile" in singers' formant range sung by Jussi Björling (F-sharp and B-natural).

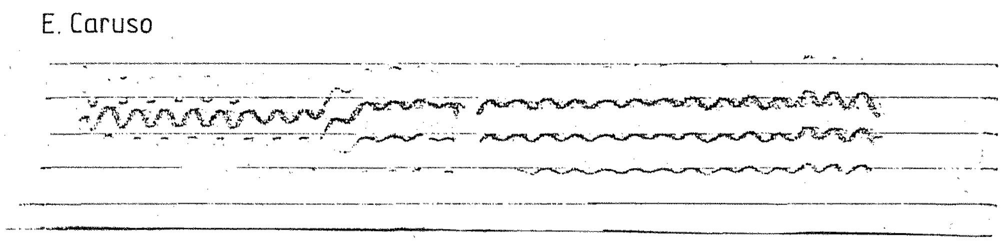


Fig. 9. Last notes of "la donna è mobile" in singers' formant range sung by Enrico Caruso (F-sharp and B-natural).

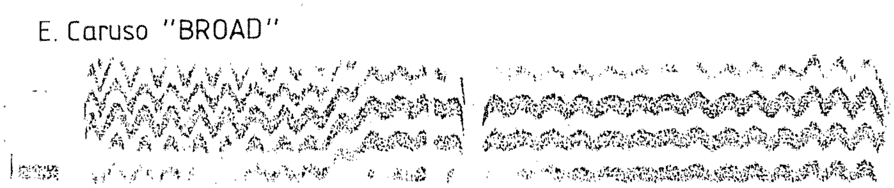


Fig. 10. Same as Fig. 9, however, "broad" Sonagram.

not display the dynamic variation caused by the exciting of the various resonances, Fig. 7.

Among the objective approaches to obtaining absolute speed calibration of voice recordings, the distribution of vibrato has been looked into. Fundamental work in this area - albeit with another objective in mind - was made under the supervision of Carl E. Seashore (1932). Apart from utilizing live test subjects, early gramophone recordings were also used. Unfortunately there is no mention of the conditions of utilization of these; in particular the speed information is lacking. It is expected that the research outlines in the present paper will allow one to deduce the conditions used in these early experiments.

The utilization of recordings for vibrato studies requires very much of the fundamental frequency analyzer. In the case of tenors good results have been obtained on a Kay Sonagraph using the following approach: using the range 40 Hz to 4 kHz the input is sharply filtered (18 dB/oct) in order to obtain a pass band of 2.5 kHz to 3.5 kHz, see Figs. 8-10 (sound examples 3 and 4). This corresponds to the frequency range in which one finds the "singing formant" and it is a range in which instrumental acocompaniment contributes only little. In this way the signal-to-disturbance ratio is very much increased, and graphic analysis may be performed on the Sonagram. It is a logical extension of this concept to look in on one of the harmonics in the same range of frequencies, when working in real-time. Incidentally, the filtering which avoids overload of the recording system by the lower frequencies will also make replay at one quarter speed very listenable.

Acknowledgments

This research is funded by the Danish Research council for the Humanities under grant no. 15-1709.

Valuable assistance from Ove Christensen and Werner Deutsch is gratefully acknowledged.

References

Brock-Nannestad, G. (1981): "Zur Entwicklung einer Quellenkritik bei Schallaufnahmen", Musica 1/1981, pp. 76-80 (Jan.-Febr.).

Miller, Dayton C. (1913): "The graphical recording of sound waves; effect of free vibrations of the recording apparatus", Proc.Fifth Int. Congr. of Mathematics 1912, Vol. II, pp. 245-249 (Cambridge).

Seashore, Carl E. (ed.): The Vibrato, University of Iowa Studies in the Psychology of Music, Vol. 1; in particular pp. 170-175.

Brock-Nannestad sound illustrations (record side)

1) Adelina Patti as found: includes both high note and change of register and piano interlude 20 s

2) Adelina Patti corrected: identical selection as 1) 20 s

3) Jussi Björling: last notes of "la donna è mobile", filtered, played twice 10 s

4) Enrico Caruso: last notes of "la donna è mobile", filtered, played twice 10 s

For practical purposes the speed of reproduction of the original Patti record was 78 rpm.

STUDY OF THE ACOUSTICAL PHENOMENA CHARACTERISTIC OF THE TRANSITION BETWEEN CHEST VOICE AND FALSETTO

Michele Castellengo, Bernard Roubeau, and Claude Valette
University of Paris VI, France

Abstract

The two main registers of the human voice, usually referred to as chest voice and falsetto, correspond to two different emission mechanisms of the vocal cords; thick cords with long closed phase or thin cords with no closed phase. During the transition from one type of emission to the other, a loss of control of the pitch and of the vocal color can be observed. The vocal color of falsetto is generally different from chest voice, in particular, significant differences in the distribution of energy between the first harmonics may be observed. We have carried out the study of this break with different subjects, singers or non-singers of both sexes. Whether the subject is a singer or not, according to his vocal technique and according to the cultural context, this break can be so masked as to become imperceptible, or on the contrary, it can be used as a musical effect (yodel).

Introduction

A simple vocal test is to sweep the whole vocal extent with a glissando. When we ask a singer (or a non-singer) to execute such a glissando through the whole range of his voice, from the lower limit to the very upper limit (not only the trained part of the voice), we find at least one or two breaks or failures in the continuous sound.

Figs. 1a and b shows two breaks for a male voice as well as for a female voice. This paper concerns only the lower one of the two breaks shown in Fig. 1.

One of us (Roubeau, 1982) previously used a different techniques such as radiography and electroglottography to study this phenomenon; we present here results of an acoustical investigation.

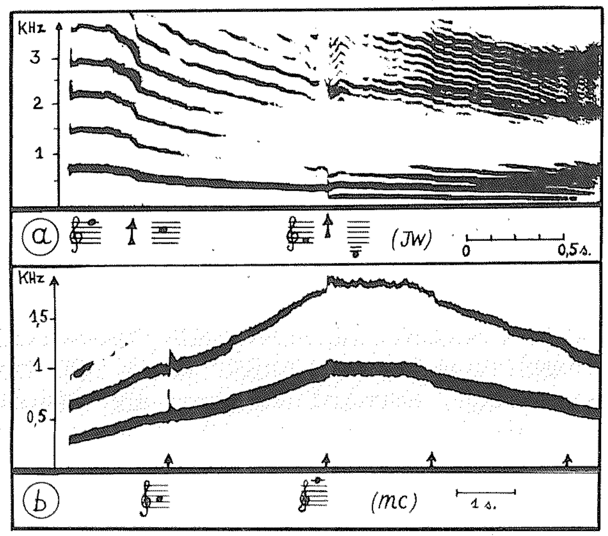


Fig. 1. Glissando over the whole range of the voice. (a) male voice: a bass showing two breaks near 600 and 300 Hz. (b) female voice: a soprano first singing an ascending, then a descending glissando. The sonagram shows two breaks near 880 and 440 Hz. The breaks, also two (marked with arrows) delimit three registers.

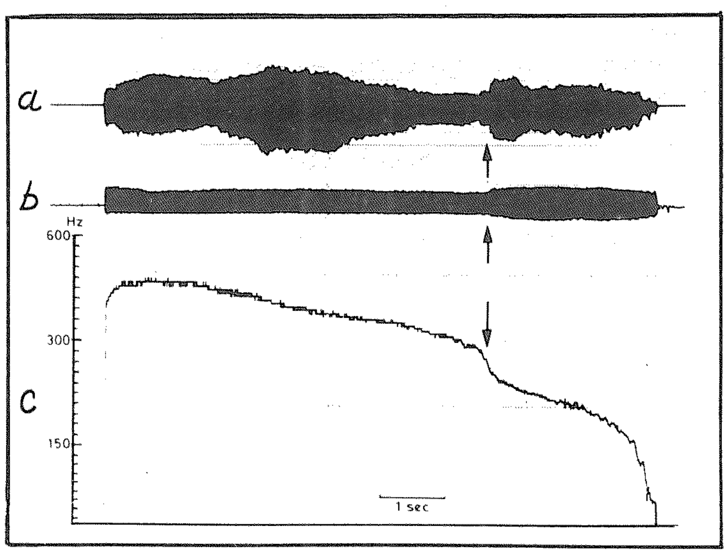


Fig. 2. Study of a descending glissando from a male voice. (a) Sound wave. (b) Electroglottographic curve. (c) Fundamental frequency. One can see a simultaneous alteration of the three curves when the change between the two registers occurs (arrows).

Fig. 2 shows clearly that the break (here not exactly a break but a sudden change of the slope) is related to a change in the glottographic curve. We agree with Garnault (1895), Hirano, et al. (1970), and Hollien (1974) that there are two main registers for men as well as for women.

There remains the difficult problem of choosing terms to designate the registers: first and second, low and high, modal and loft... etc. Arbitrarily we use "PLAIN" voice for all the sounds produced below the break, and "FALSETTO" voice for sounds above this break.

Acoustical characteristics of a break found in a glissando

Let us now look at the acoustical characteristics of such a change between Plain voice and Falsetto voice during a glissando. They can be different depending on whether the singer is trained or not.

Fig. 3 shows two ascending glissandos on the same vowel /a/ from untrained singers, a male and a female, with a jump in the frequency and an important change in the spectra.

Fig. 4 is an ascending and descending glissando from a student singer. At the transition between the registers there is no break but a kink, and a drop in the brightness.

Among the professional singers, those whose voice normally use the two registers, for instance counter-tenor, practice to rub out the transition. Nevertheless, in an example given by a singer from "Les Arts florissants" singing an ascending and a descending scale, we hear clearly the transition as a break and we can see it on the sonagram (Fig. 5). Another example from R. Jacobs is more subtle. When the pitch is descending, we hear clearly a break but we can notice in Fig. 6 that probably consciously the singer changes the colour of his voice one note before crossing the transition. Ascending, the transition is perceivable as a slight change in the legato between the two notes E and F#(MI_3 and

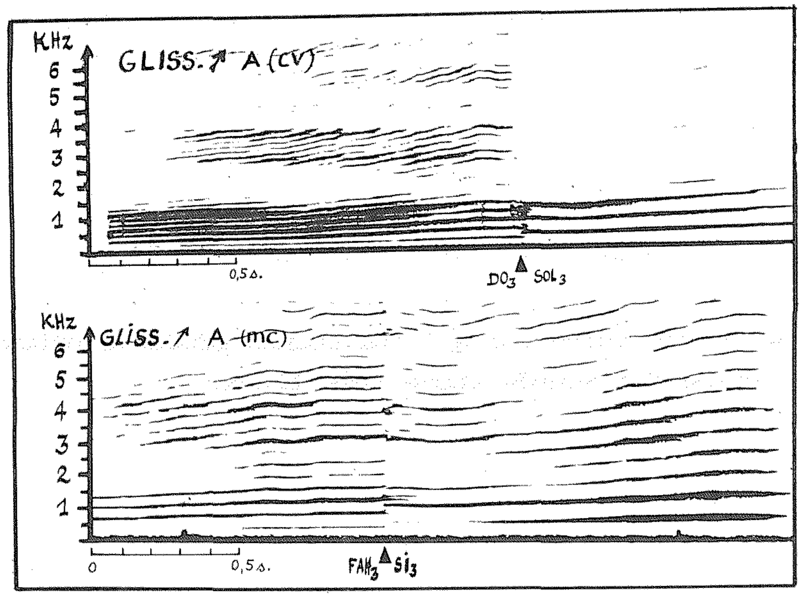


Fig. 3. Sonagrams of ascending glissandos produced by two non-trained singers, a male (cv) and a female (mc). The break is clearly visible (arrows): it appears as a frequency jump and a change in the spectrum.

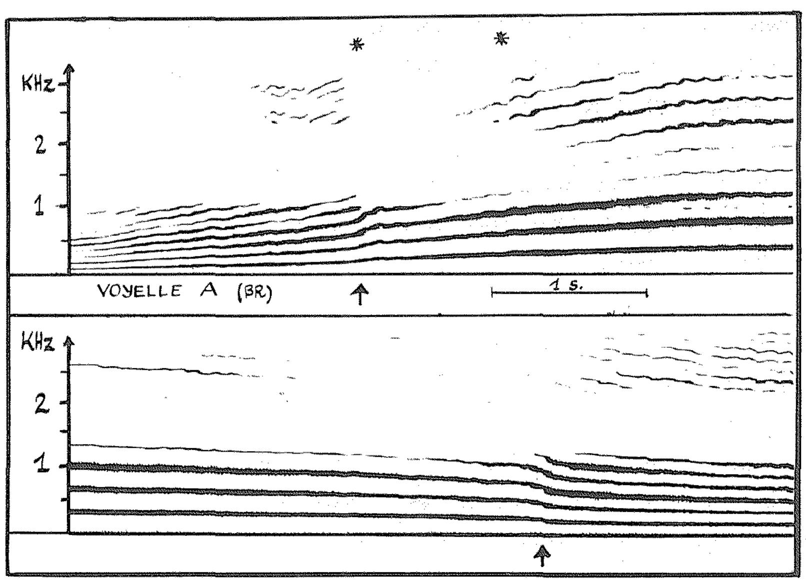


Fig. 4. Sonagram of an ascending and a descending glissando from a student singer: at the passages (arrows) there is no break but a kink and then spectral change perceived as a drop in the brightness.

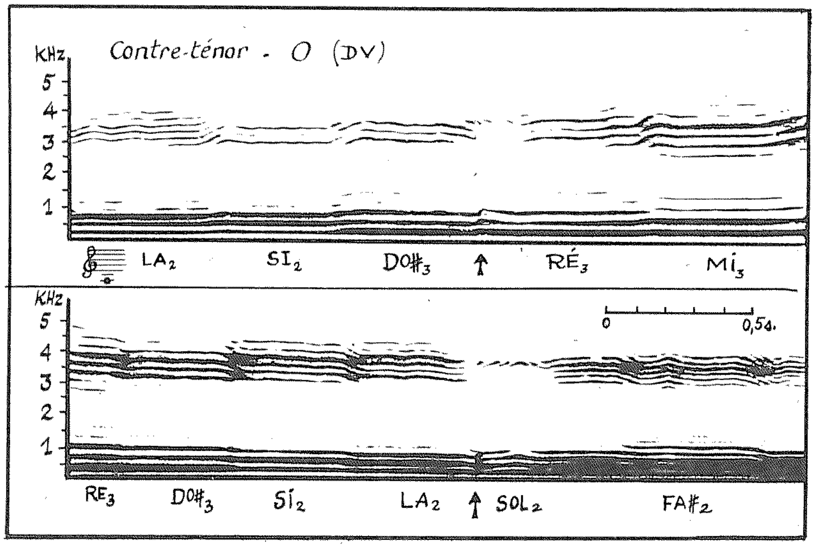


Fig. 5. Sonagram of an ascending and a descending scale from a professional counter-tenor: the spectrum is similar below and above the passage (arrows), but we notice a short drop in the intensity of the harmonics at the passage.

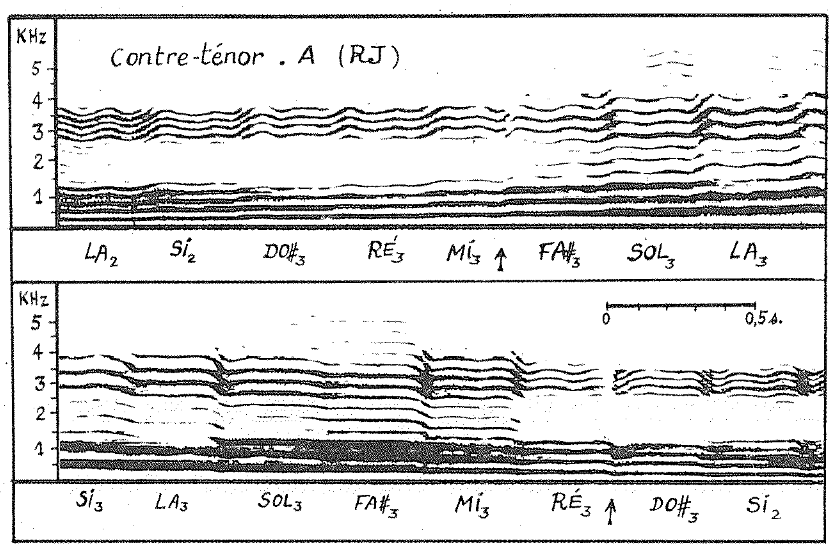


Fig. 6. Sonagram of an ascending and a descending scale by a professional counter-tenor: René Jacobs. At the passage in the descending scale there is an extremely short loss of coherence in the voice: notice that the singer consciously changes the color of his voice one note before the passage. The upper part of the falsetto voice has a different spectrum than the lower part but the change is continuous. The passages are marked by arrows.

FA_3 in the figure): a short drop of intensity near 3 kHz and a modification of the frequency slope of the legato.

Let us also remark that for both singers the transition occurs on a higher note, when the pitch is ascending than when it is descending.

Fig. 7 shows another example sung by a professional singer in which the transition is marked by a change of intensity and a loss of vibrato.

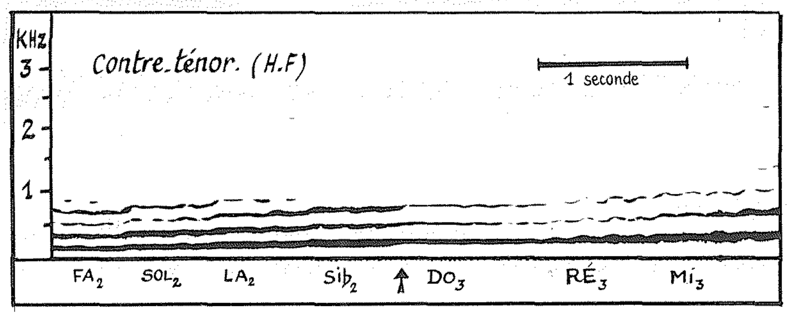


Fig. 7. Sonagram of an ascending scale sung by a professional singer: there is no break at the passage (arrow) but a short drop of intensity and a loss of vibrato.

Acoustical characteristics of the break found in isoparametric tones

Another good test for studying the acoustical change between two registers is to pass from one register to another while singing the same pitch with the same intensity.

We asked three professional singers: a Tenor, a Baritone and a Countertenor, to make this test (Fig. 8).

With the tenor, there is no break but a loss of harmonics in the falsetto voice. With the baritone there is a break and also a loss of

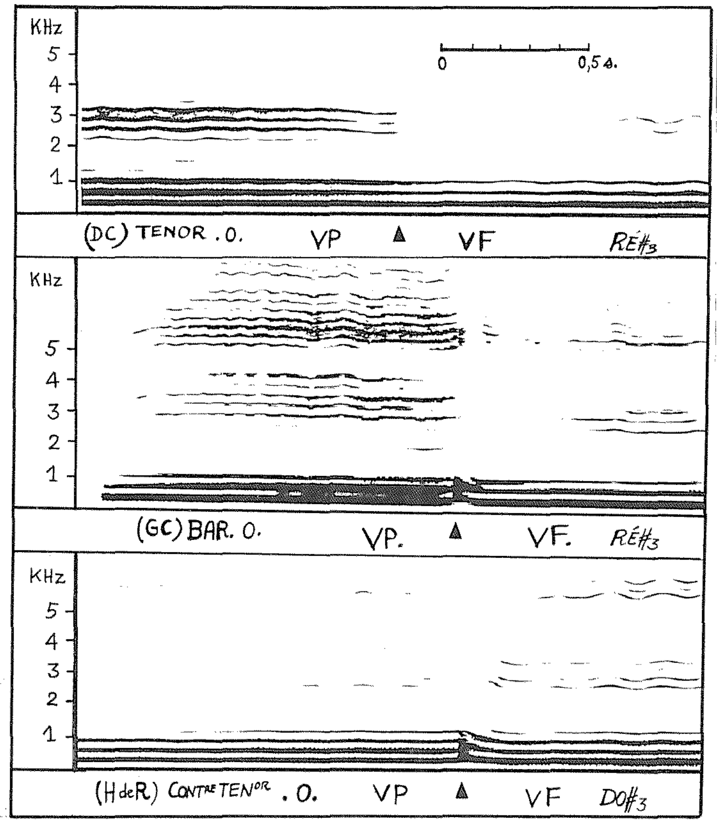


Fig. 8. Sonagram of isoparametric tones sung by three professional singers. Various features can be seen at the passage (arrows): a loss of amplitude in the higher harmonics for the tenor; break and loss of amplitude in the higher harmonics for the baritone; break and gain in the amplitude of the higher harmonics for the counter-tenor.

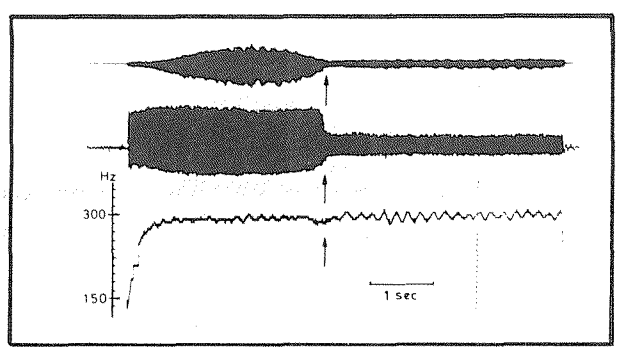


Fig. 9. Sound wave, electroglottographic curve and melodic curve of the isoparametric sound of the tenor (see Fig. 8). There is no break in fundamental frequency but the glottographic curve shows a sudden change, probably related to different laryngeal adjustments.

amplitude of the higher harmonics. With the countertenor there is a break, but on the contrary, the voice is richer in the falsetto register. We have verified that this does not depend on the vowel.

Even when there is no break in fundamental frequency, the glottogram does show a sudden change, suggesting the existence of two different mechanisms (Fig. 9).

<u>Some comments upon the frequency jump during the transition</u>

Let us look carefully at the jump of frequency when it does occur. In Fig. 10, three subjects, two males and a female, sing the same note on

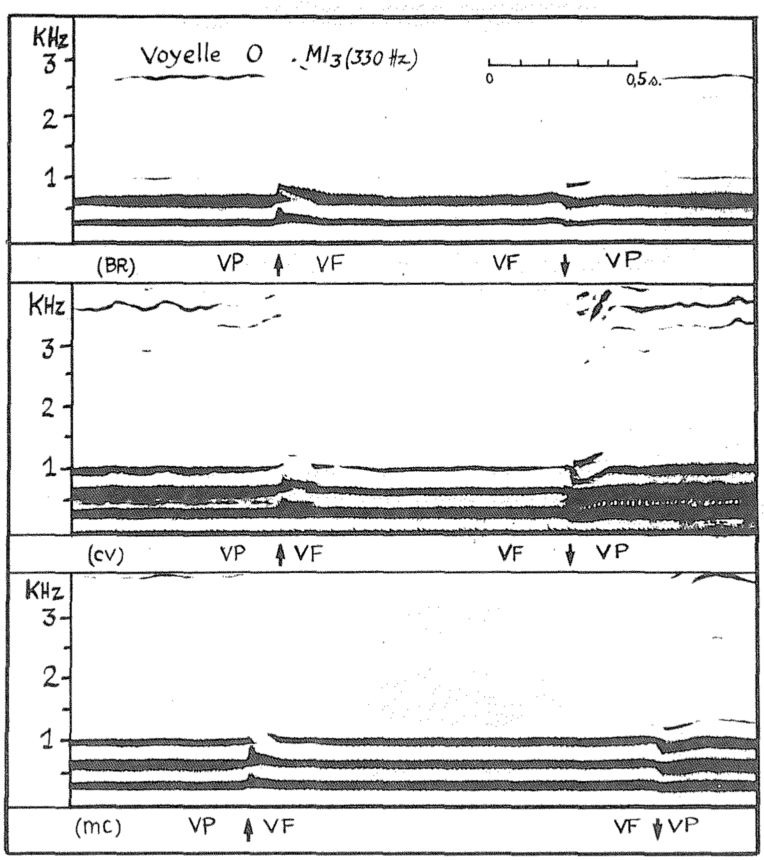


Fig. 10. Isoparametric sounds from three untrained singers (2 males and a female) singing the same note on the same vowel: the jump in fundamental frequency at the transition (arrow) occurs upward when the change is from the plain voice (VP) to the falsetto voice (VF) and downward when the change is from the falsetto voice to the plain voice.

the same vowel: the frequency jump during the transition occurs UPWARD when the change is from the plain voice (VP) to the falsetto voice (VF), and DOWNWARD when the change is from the falsetto voice to the plain voice.

This phenomenon seems to be quite general and supposedly related to the physiological properties of the larynx.

Musical use of the transition between two registers

Contrary to the classical technique of singing, there are examples of the musical use of the break such as the yodel. Sonagrams of African yodel as well as Tyrolian yodel are similar to the one shown in Fig. 11, which was obtained from a folk singer.

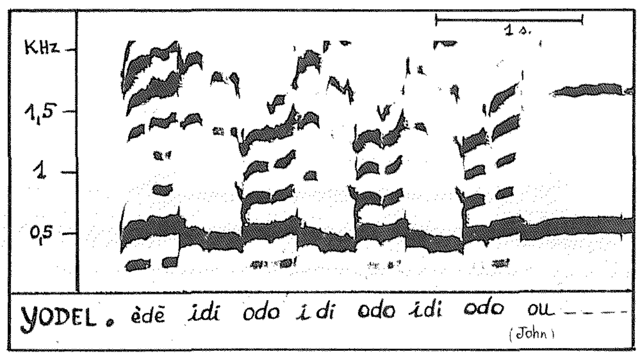


Fig. 11. Sonagram of yodel obtained from a folk singer: the second harmonic of the plain voice comes just above the fundamental of the falsetto voice: one gets the visual impression that the plain voice is higher than the falsetto voice.

We remark that the second harmonic of the plain voice is just above the fundamental of the falsetto voice in such a way that, at first sight, you could have the visual impression that the falsetto is at a lower pitch than the plain voice. The weakness of the fundamental in a plain voice compared to the dominance of the fundamental in a falsetto voice is emphasized by the change of vowel, which probably helps the voice to jump across the transition.

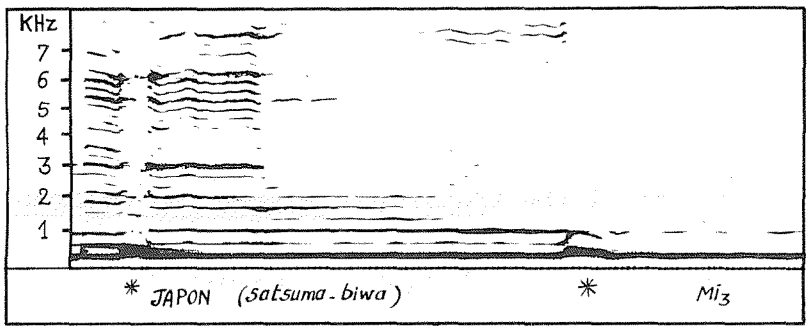


Fig. 12. Sonagram from a Japanese singer using as a musical effect the change of register on the same note: one can see a sudden change of color simultaneously with a short controlled upward jump of pitch.

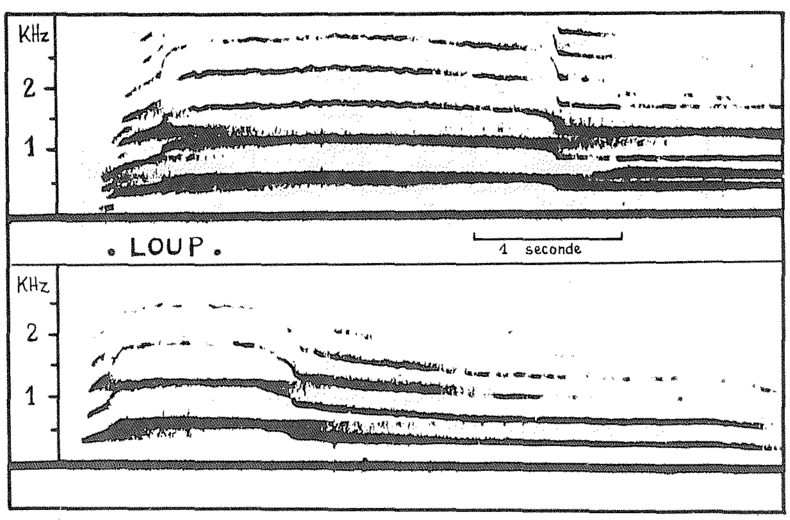


Fig. 13. Sonagram of wolves' cries including examples of breaks.

A Japanese singer was found to use a sudden change in voice colour as a musical effect simultaneously with a jump in fundamental frequency (Fig. 12).

Conclusion

The passage between the two main registers is a quite general feature which can be assumed to be related to a mechanical change in the vocal folds. The register break does not seem to be limited to the human voice: similar breaks can be found in the cry of wolves, for instance (Fig. 13).

References

Roubeau, B. (1982): "Exploration instrumentale des mécanismes laryngés en voix de fausset chez l'homme; contribution a l'étude du passage", Memoire d'Orthophonie. Université Paris VI-Salpetriére.

Garnault, Dr. (1895): Physiologie, Hygiène et Thérapeutique de la Voix Parlée et chantée, Maloine Ed., Paris. A study of registers can be found in this book with direct observation support of vocal chords activity in professional singers, including thickness observations.

Hirano, M., Vennard, N., and Ohala, J. (1970): "Regulation of register, pitch and intensity of voice", Folia Phon. 22, 1-20. (1970).

Hollien, H. (1974): "On vocal registers", J.Phonet. 2, 125-144.

SOURCE/VOCAL-TRACT INTERACTION IN SPEECH AND SINGING SYNTHESIS

D.G. Childers, J.J. Yea, and E.L. Boccheri
Dept. of Electrical Engineering, University of Florida
Gainesville, FL, USA

Abstract

We are using two models to study the influence of the glottal source on the production of synthesized speech and singing.

One model employs experimentally and theoretically derived source waveforms as the excitation for a serial/parallel Klatt formant synthesizer to produce speech and singing.

The other model is articulatory and based on Flanagan's and Ishizaka's model, which we have modified and extended to study source/tract interaction in speech and singing synthesis. This model also uses a computer graphics terminal to display a cross-section of the vocal tract area function as the synthesis is performed.

We evaluate the naturalness of our synthetic speech and singing by formal listening tests.

Our goal is to assess the contribution of various source parameters to the production of natural sounding synthetic speech and singing. We report our results to date which illustrates the effect of source / tract interaction.

Introduction

Two factors affecting the quality of synthetic speech are the improper modeling of the glottal excitation function, e.g., using an impulse or stylized waveform as the glottal excitation, and neglecting source/tract interaction.

One aspect of our research improves the quality of synthetic speech by using a formant synthesis scheme, which adopts a more accurate model of the glottal excitation. We do this by employing a "glottal area" excitation function which controls the time varying glottal impedance in an equivalent circuit of the vocal system. The output of this circuit is a time-varying "glottal volume velocity" function which includes the effect of source/tract interaction. This glottal volume velocity function provides the excitation for the formant synthesizer. We have used this procedure to synthesize sentences which sounded as if they have been spoken by a male, female and child. The other glottal excitation functions (an impulse and a stylized waveform) were used as well for comparison purposes, but the vocal tract transfer functions were kept the same regardless of the form of excitation. A formal listening test using the natural sentence as well as the synthesized sentences was conducted. This test showed that the glottal volume velocity excitation derived from the glottal area waveform and using source/tract interaction can produce more natural sounding speech than the other two forms of excitation and this synthetic speech is nearly as good as the original.

We have also used our procedure to synthesize a singing voice producing three scales of /a /.

Our other synthesis model uses an animated computer graphic's system to derive articulatory parameters for a transmission line (wave propagation) type speech synthesizer. This system is being used to investigate the physiology of speech production, to test linguistic theories, and to model the effects of surgical intervention of the vocal cords or tract on the production of speech or singing. This model has also been used to study the effects of source/tract interaction and to model Sundberg's (1970) singing formant.

Formant Speech Synthesis

The speech is synthesized using parameters extracted from speaker generated sentences. The parameters are those required for Klatt's formant synthesizer, e.g., formant locations, amplitudes, bandwidth, pitch contour, transitions, etc. (Klatt, 1980). But we have modified this synthesizer to include source/ tract interaction via a simple Guérin

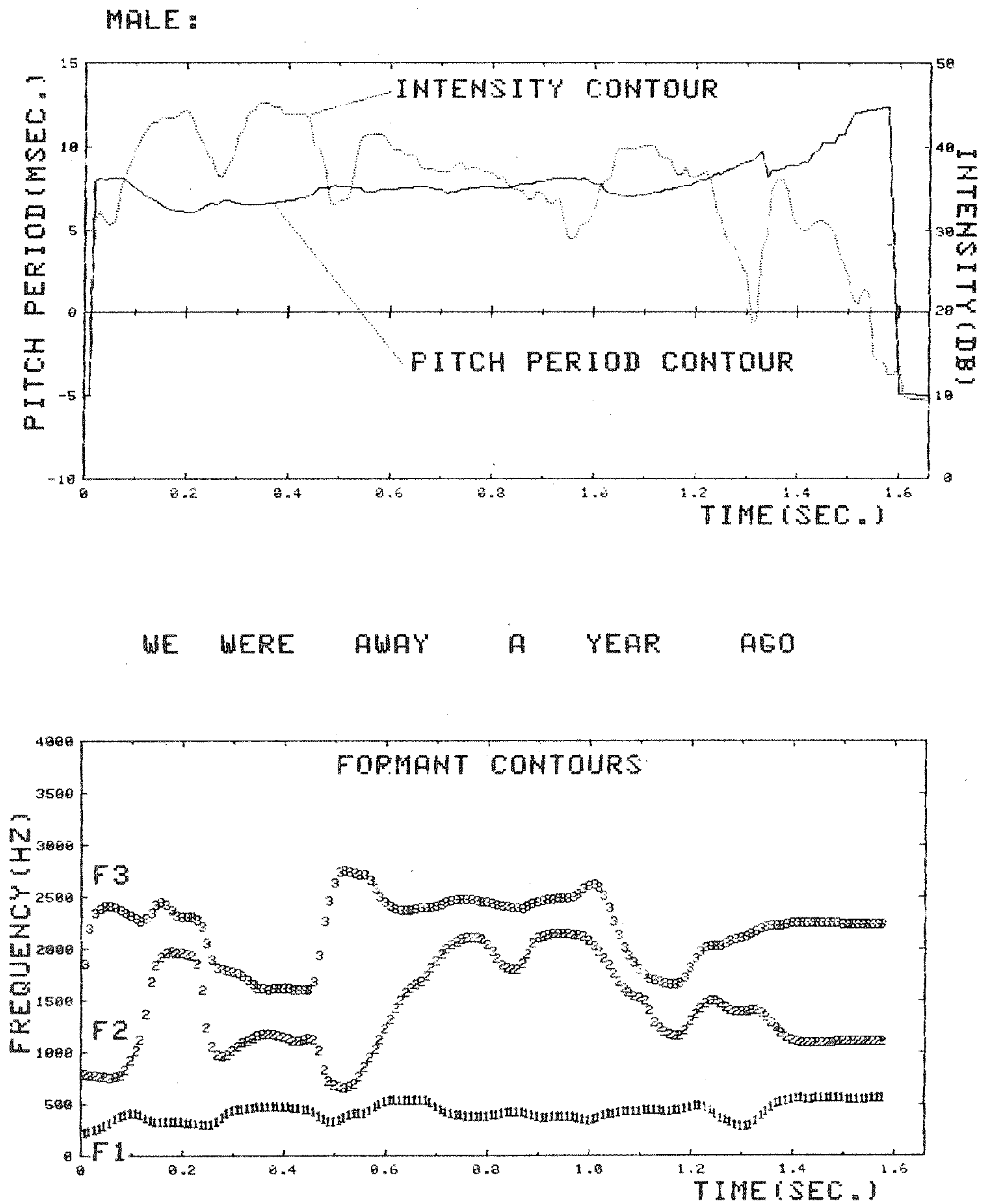


Fig. 1. Intensity, pitch, and formant contours for the same sentence spoken by two different speakers, male and female child.

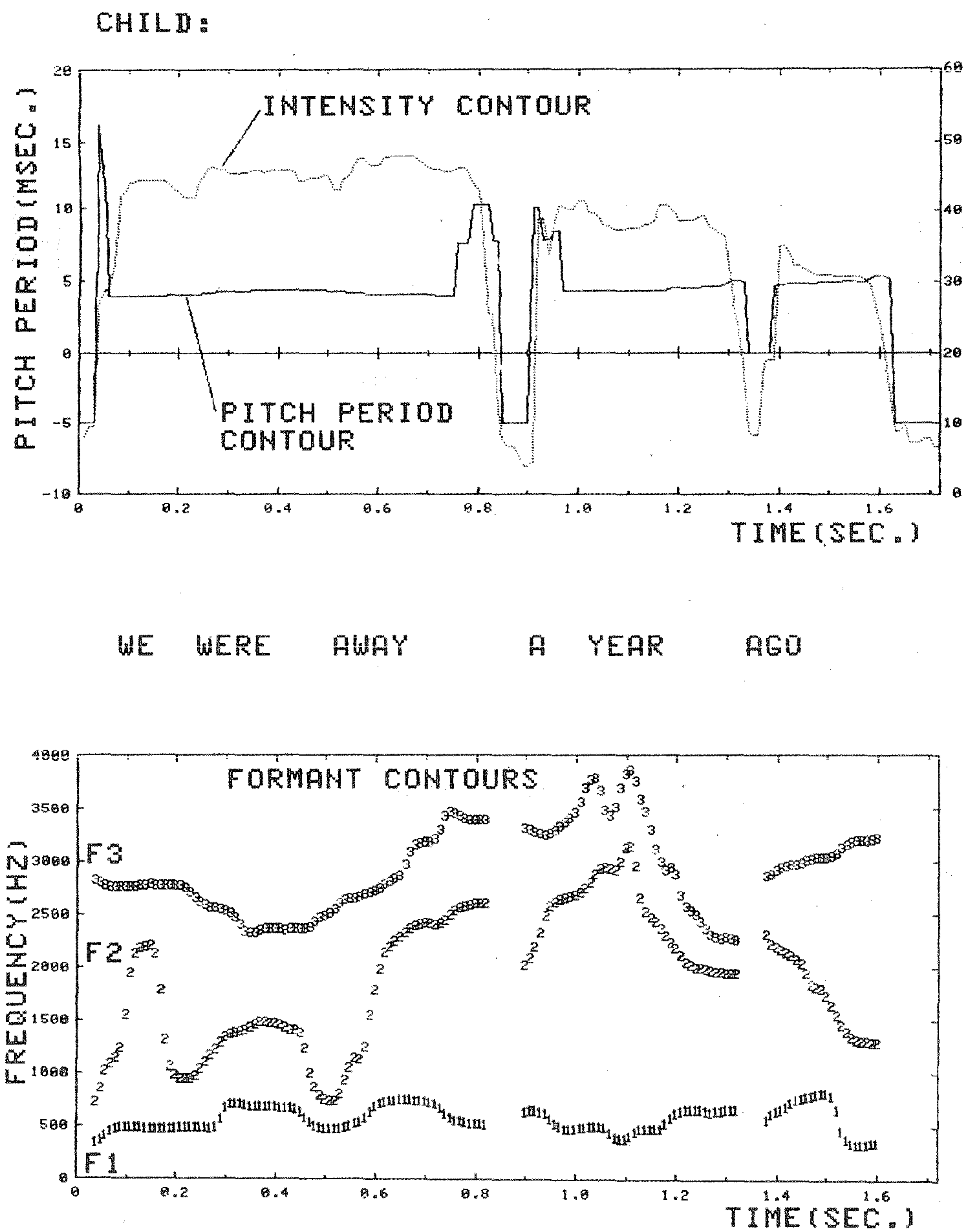


Fig. 1. Continued.

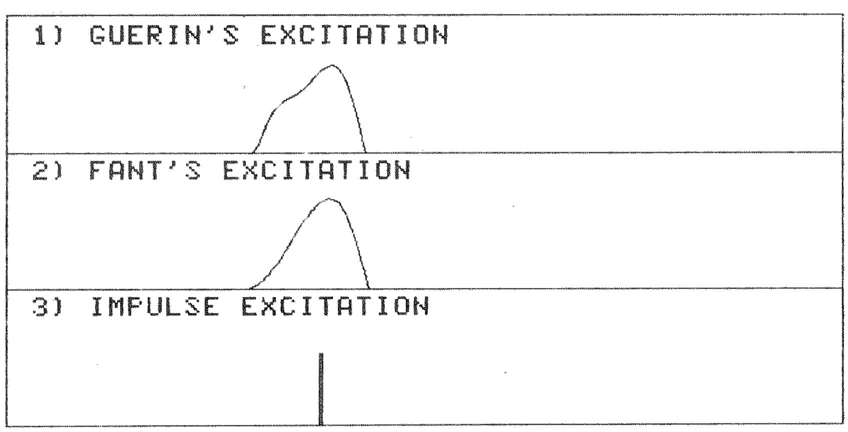
impedance network (Guérin et al., 1976). This allows us to use various glottal waveforms as excitation for voiced sounds. These glottal waveforms may be experimentally derived, as in our experiments, or generated by a model, as in Fant's work (Fant, 1979; Ananthapadmanabha and Fant, 1982). We use Fant's source model without source/tract interaction.

Fig. 1 shows the formant contour, pitch contour and voiced, unvoiced and silence regions for two sentences. The information from Fig. 1 is used for setting the parameters of the formant synthesizer.

The three types of glottal excitation used for speech synthesis appear in Fig. 2.

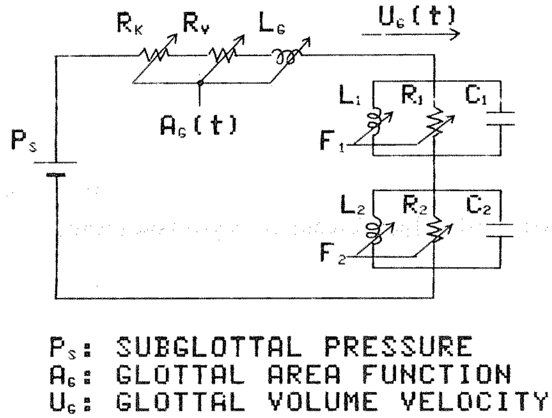
Guérin's (Guérin et al., 1976) circuit for including coupling between the source and tract is shown in Fig. 3.

We use the actual glottal area function measured from our films (or the volume-velocity waveform), which in turn excites the formant synthesizer. Since there are several parameters to adjust in Guérin's circuit we established an error criterion which minimized the mean-square error between the speech inverse filtered volume-velocity waveform and the Guérin generated volume-velocity waveform. The error function is shown



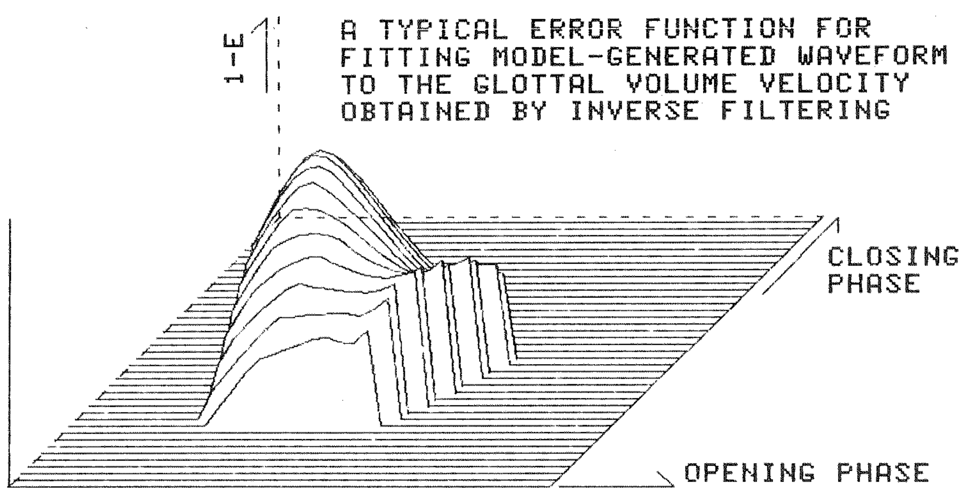
DIFFERENT MODELS OF GLOTTAL EXCITATION

Fig. 2.



GUERIN'S CIRCUIT FOR STUDYING SOURCE-TRACT INTERACTION

Fig. 3.



A TYPICAL ERROR FUNCTION FOR FITTING MODEL-GENERATED WAVEFORM TO THE GLOTTAL VOLUME VELOCITY OBTAINED BY INVERSE FILTERING

Fig. 4. Error function.

in Fig. 4 plotted against the open duration (total open duration as measured from the glottal area) and the closing phase (duration from the point of maximum glottal opening to glottal closure). The best Guérin circuit waveform is shown in Fig. 5 where it is compared with the inverse filter derived waveform.

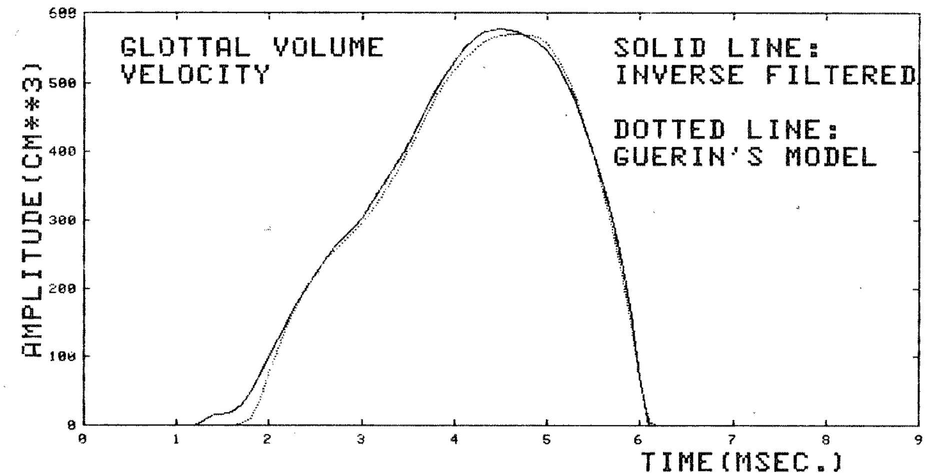


Fig. 5. Best excitation waveform.

When we put all our analysis results together we obtain the three synthesized vowels /I/ (as in it) shown in Fig. 6. The best result was obtained using Guérin's circuit. Shown in Fig. 7 is the spectrogram of the complete original sentence compared with the synthesized version. The two are nearly identical and sound that way too.

One of our main themes is that source/tract interaction is important in synthesis, but one can have too much as well as too little interaction. The amount of source/tract interaction can be controlled by the following two methods:

a) varying the vocal tract input impedance,
b) varying the glottal impedance.

Not much is known about how the vocal tract impedance should change. On the other hand, the glottal impedance is known to be inversely propor-

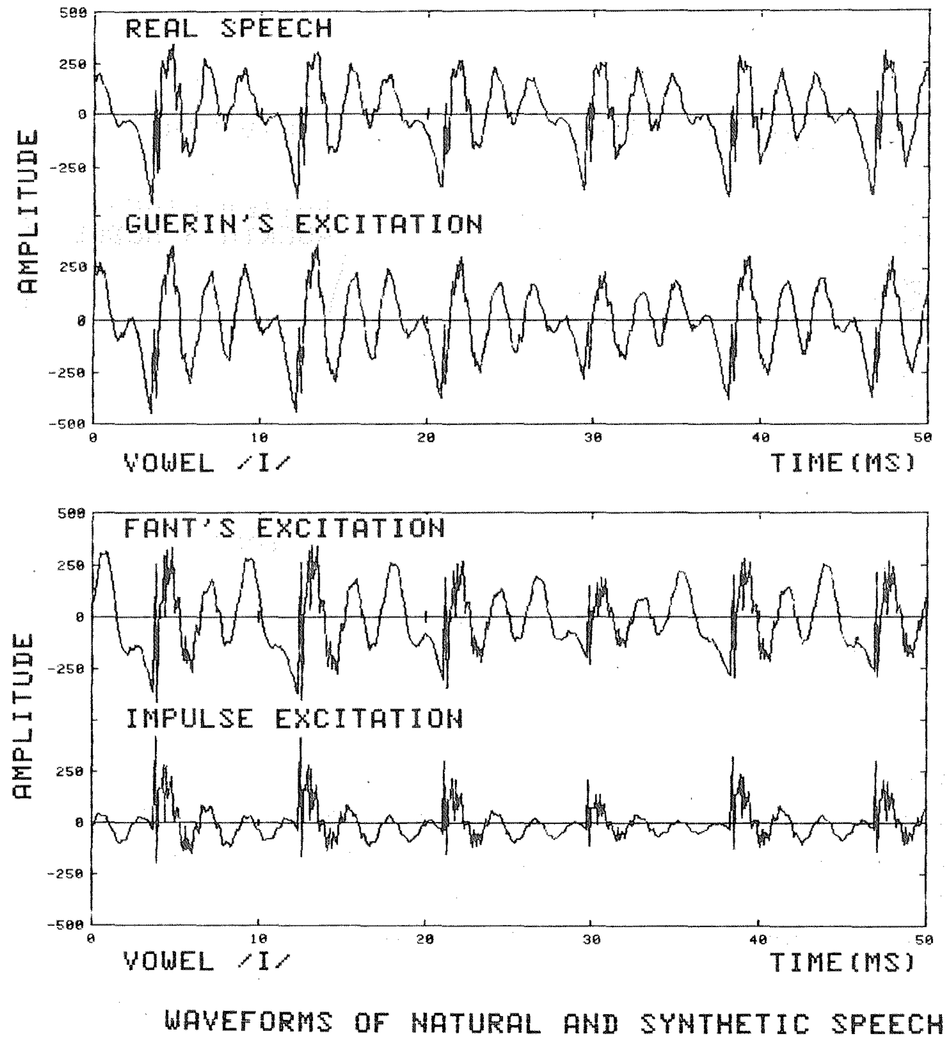


Fig. 6. Waveforms of natural and synthetic speech.

tional to the size of the glottis. Thus, if the vocal tract input impedance remains the same, the source/tract interaction effect should be most significant for the adult male voice, because the glottal area is the largest in this case. On the other hand, the source/tract interaction should be least for a child.

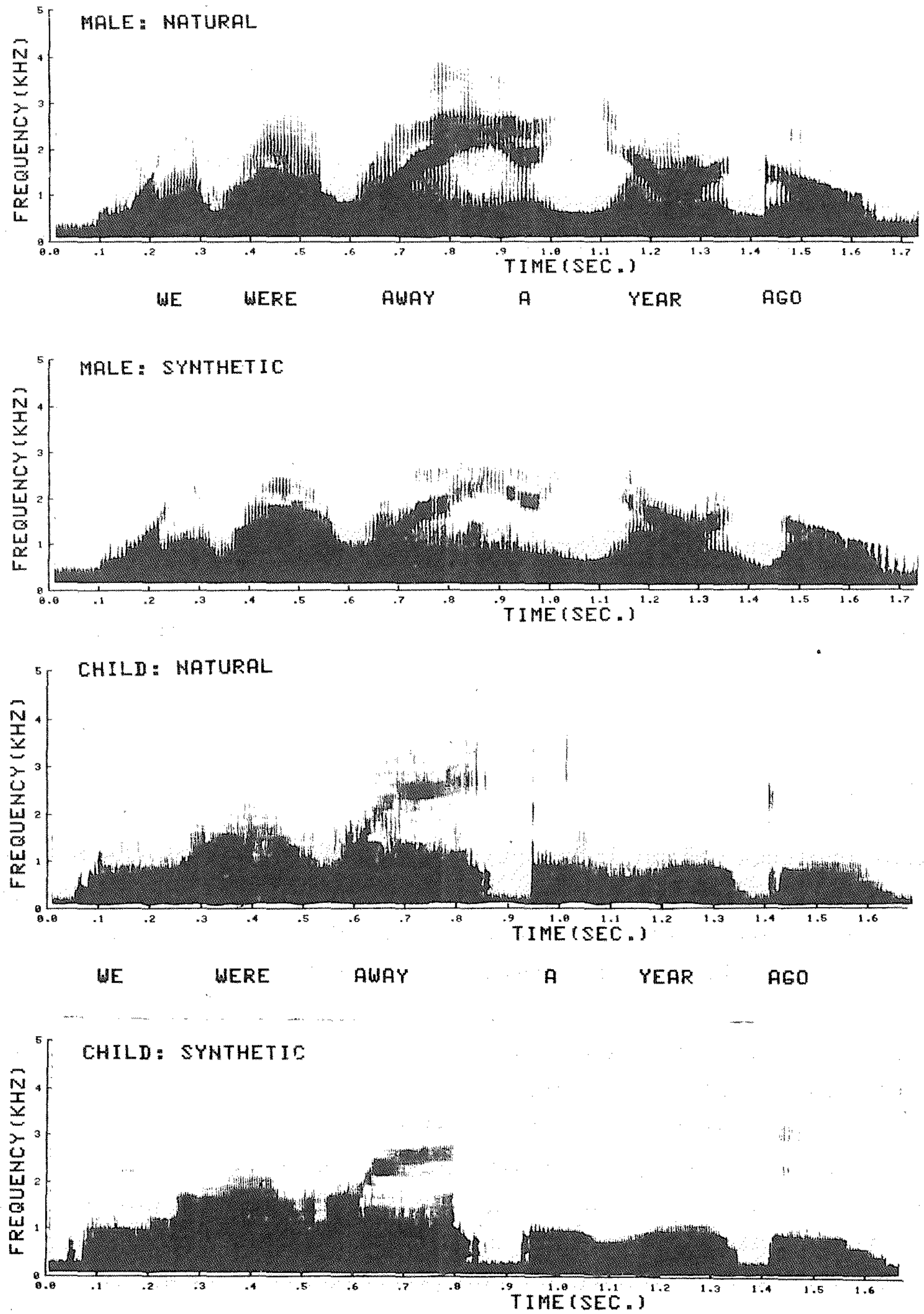


Fig. 7. Spectrograms of a sentence spoken by a male and female child and their synthetic counterparts.

For illustrative purposes, we have synthesized a sentence using three levels of source/tract interaction, i.e., no interaction, optimal interaction, and too much interaction.

We have also applied our method to synthesizing a singer producing three scales of /a/. Fig. 8 shows the inverse filtered glottal volume velocity waveform for /a/ for one of these scales. This waveform was the excitation for the formant synthesizer. A closed phase linear prediction 12 pole model was used for inverse filtering. A spectrogram showing the natural and the synthesized note is presented in Fig. 9. The quality is good because there is little change in the source/ tract interaction.

Articulatory/transmission line synthesis

The acoustic and perceptual properties of speech depend on the position of the lips, tongue, jaws and other vocal organs. The physiological mechanism of speech production (articulation) involves precisely timed movements of the vocal organs to produce the acoustic wave we perceive as continuous speech or singing.

We have implemented an interactive graphic's model of the human articulatory system, which we use to generate an animated representation of the vocal organ movements during speech production. This system is modeled after Coker (1976) and Mermelstein (1973).

The articulatory or "muscular" model of the human vocal cavities is connected (off-line) to a computer model of the vocal cavity acoustics (Flanagan et al., 1975), which calculates the radiated speech, air pressure, and volume velocity distribution in the vocal cavities and models the motion of the vocal cords. All this information can be displayed together in an animated manner. Fig. 10 shows the articulation model, with cross-hairs for adjusting the position of the articulators, the vocal tract area function, and formant frequencies. Fig. 11 is a typical animation frame including the vocal tract, vacal cords, and various waveforms calculated by the model.

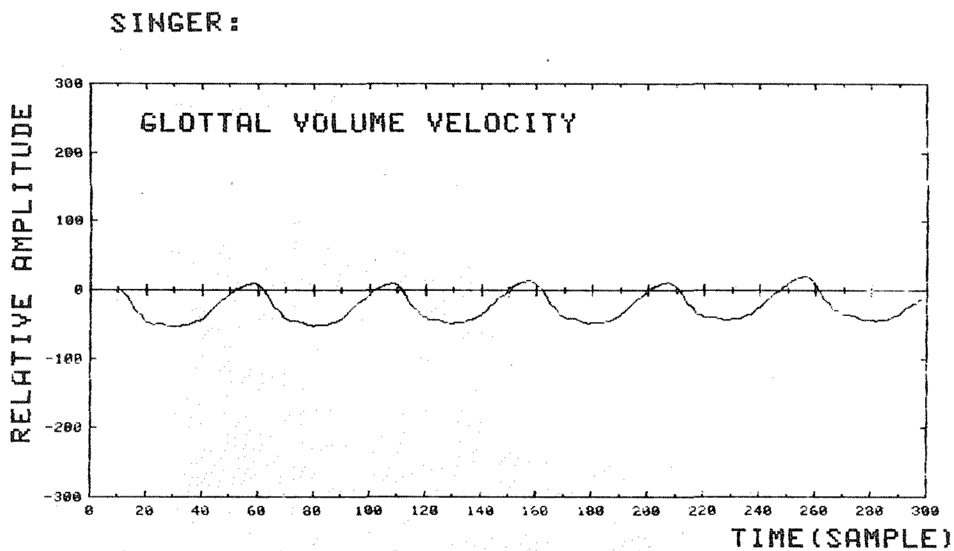


Fig. 8. Singer glottal volume-velocity.

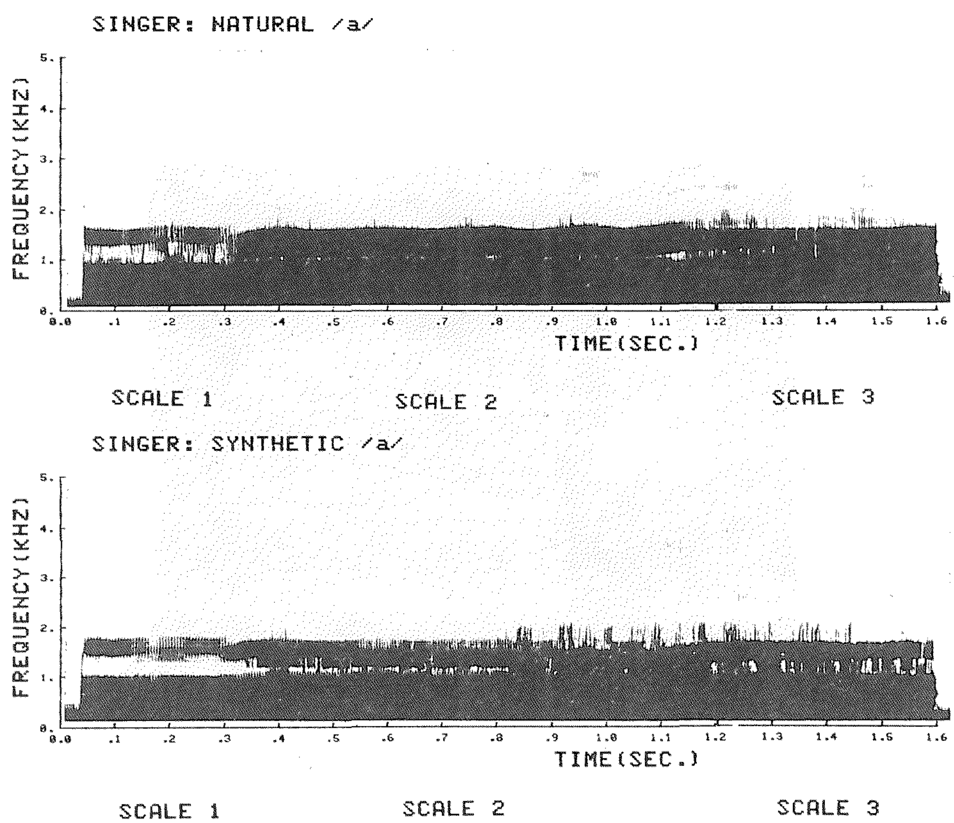


Fig. 9. Spectrograms of male singer singing three scales of /a/, natural and synthetic.

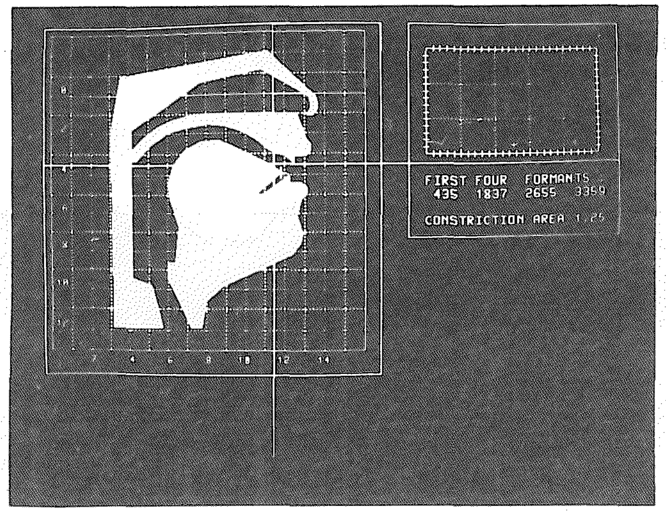


Fig. 10. Spatial articulation model, vocal tract area function, and formant frequencies.

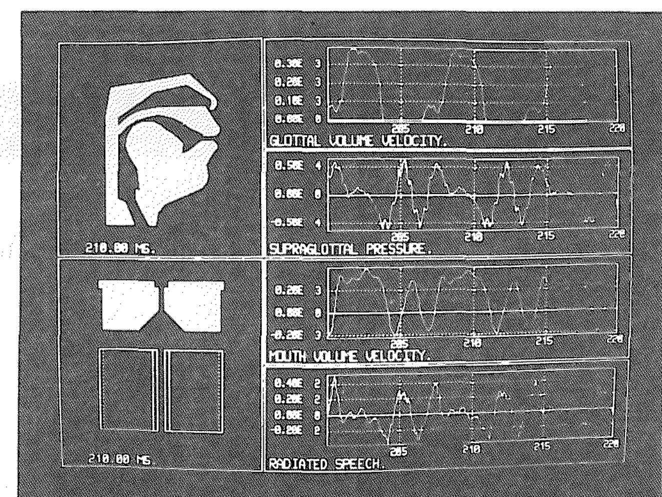


Fig. 11. A typical animation frame showing vocal tract and vocal cords along with various data waveforms calculated by the model.

This model has been used to study source/tract interaction by varying the amount of glottal area for various synthesized utterances. The word "good-bye" has been synthesized with too "little", too "much" and "just right" source/tract interaction. This synthesis was accomplished using three synthetic glottal volume velocity waveforms with a varying degree of source/tract interaction, shown in Fig. 12. The glottal pulse with the greatest source/tract interaction is characterized by a more pronounced 1st formant ripple during the rising slope and by a steeper final slope (Rothenberg, 1983) which increases the higher frequency components of the glottal excitation.

Spectra of the glottal waveforms show that the greatest source/tract interaction causes an increase of about 8 dB around 3 kHz. This demonstrates that source/tract interaction can be used by singers to modify their voice quality (Rothenberg, 1981). (Note that our synthesizer models source/tract interaction over the entire frequency range.)

The spectrogram for "good-bye" with the "just right" glottal excitation appears in Fig. 13.

The voice quality during singing is characterized by the "singer's formant" (Sundberg, 1970) which appears as a concentration of the acoustic energy around 3 kHz, making the singer's voice more easily discerned against the background of the orchestra (Sundberg, 1981). The singer's formant is articulated by clustering the 3rd, 4th and sometimes 5th formants around 3 kHz so that the sound transmission of the vocal tract is enhanced at these frequencies.

The spectra of synthetically "sung" and "spoken" /u/'s is shown in Fig. 14. The articulatory information to drive the synthesizer has been obtained from Sundberg (1970). The articulation of the "sung" vowel is characterized by a more abrupt enlargement of the pharynx with respect to the larynx tube, desreasing the 4th formant frequency. The 3rd formant frequency is increased by an enlargement of the vocal tract volume in the region of the soft palate and by a reduction of the mouth volume behind the incisors.

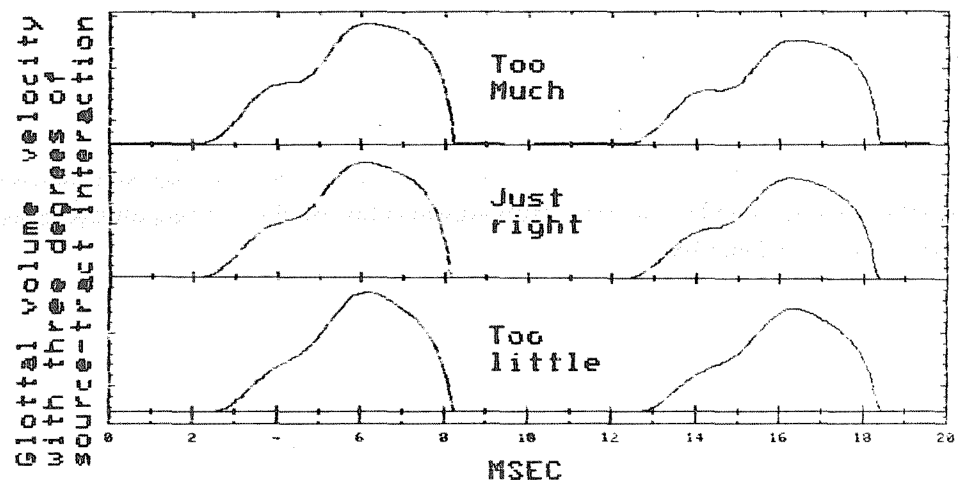


Fig. 12. Three glottal volume-velocity waveforms used to synthesize speech in the articulation model.

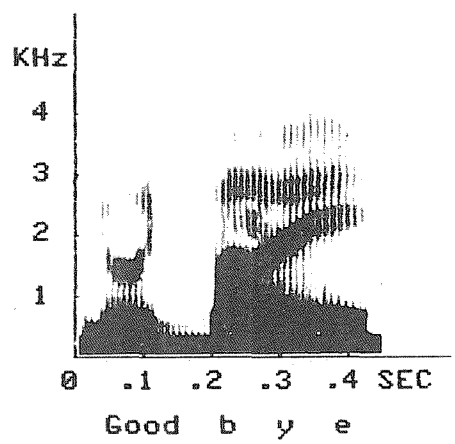


Fig. 13. Spectrogram of "good-bye" synthesized with the "just right" glottal excitation.

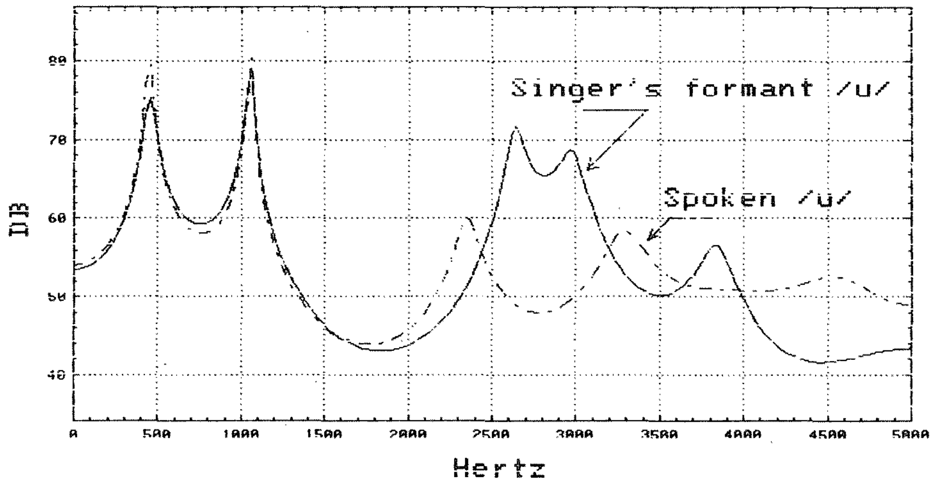


Fig. 14. Illustration of singer's formant for /u/ compared with spoken /u/. Both waveforms were obtained by synthesis using the articulation model.

Conclusions

The synthesis by analysis technique using a formant synthesizer provides very good quality, better than the articulatory synthesis model. Why is this so? We know little about the proper articulator model positions. The modeling to date is two-dimensional, i.e., cross-section based, and is not three-dimensional. Finally, the articulator based model requires much more computation than the formant based synthesizer. The computations are interpolated frequently in the articulator model and this has prevented us from synthesizing longer utterances with the articulator model. However, we have successfully modeled different acoustic events which are related to the physiology of speech articulation. Here we have briefly discussed source/ tract interaction and the "singer's" formant but we have also modeled other effects like the reduction of speech intensity during nasalization, the onset spectra of voiced stops, and the vibration of the vocal cords during vocal tract closure using the articulator model.

Acknowledgments

This research was supported in part by grant NIH NS17078 from the National Institute of Health, grant NSF ECS 8116341 from the National Science Foundation, and the University of Florida Canter of Excellence Program which provides funding to the Mind-Machine Interaction Research Center.

References

Ananthapadmanabha, T.V. and Fant, G. (1982): "Calculation of true glottal flow and its components", STL-QPSR 1/1982, pp. 1-30.

Coker, C.H. (1976): "A model of articulatory dynamics and control", Proc. of the IEEE, 64, 4, April, pp. 412-460.

Fant, G. (1979): "Glottal source and excitation analysis", STL-QPSR 1/1979, pp. 85-107.

Flanagan, J.L., Ishizaka, K., and Shipley, K.L. (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", The Bell System Techn.J. 54, 3, March, 485-506.

Guérin, B., Mryati, M., and Carré, R. (1976): "A voice source taking account of coupling with the supraglottal cavities", IEEE Int.Conf. on Acoustics, Speech and Signal Processing, pp. 47-50.

Klatt, D.H. (1980): "Sopftware for a cascade/parallel formant synthesizer", J.Acoust.Soc.Am. 67, March, pp. 971-995.

Mermelstein. P. (1973): "Articulatory model for the study of speech production", J.Acoust.Soc.Am. 53, April, pp. 1070-1083.

Rothenberg, M. (1981): "The voice source in singing", pp. 15-33 in Research Aspects on Singing, Royal Swedish Academy of Music, publ. no. 33.

Rothenberg, M. (1983): "The effect of flow-dependence on source-tract acoustic interaction", presented at and to appear in Proc. of Vocal Fold Physiology and Biophysics of Voice, May 4-7, University of Iowa.

Sundberg, J. (1970): "Formant structure and articulation of spoken and sung vowels", Folia Phoniat. 22, pp. 28-48.

Sundberg, J. (1981): "The voice as a sound generator", Research Aspects on Singing, Royal Swedish Academy of Music, publ. no. 33, 1981, pp. 6-14.

ACOUSTIC ANALYSIS OF THREE MALE VOICES OF DIFFERENT QUALITY

T. Cleveland* and J. Sundberg**
*University of Southern California, Los Angeles, CA, USA
**Dept. of Speech Communication and Music Acoustics, KTH

Abstract

We know that formants typically differ between voice categories. It is reasonable to assume that differences exist also regarding the glottal voice source. The purpose of the present investigation was to map such voice source differences in representatives of three different male voice categories. Data are presented on subglottic pressure (as estimated from oral pressure during /p/-occlusion) on the overall SPL as well as on the SPL of the "singer's formant", on the amplitude of the fundamental, and on the duration of the closed phase of the glottal vibratory cycle as measured from acoustic glottograms derived by means of inverse filtering. The results show intersubject differences, the possible relevance of which to voice classification is discussed.

Introduction

Voice classification in singers has been the subject of two previous investigations by the present authors (Cleveland, 1977; Ågren and Sundberg, 1978). The results showed that in isolated, sustained vowels the pitch was the most important cue to perceptual voice classification. These results are similar to those obtained in a study on maleness or femaleness in voice timbre (Coleman, 1976).

Vowel formant frequencies have also been shown to be another cue to perceptual voice classification. Cleveland (1977) found that the higher the formant frequencies in a vowel frame, the more tenor-like the voice quality appeared to a group of trained listeners. From a study of two representatives of each of the alto and the tenor categories, Ågren and Sundberg (1978) inferred that the frequency of the fourth formant was probably typically higher in altos than in tenors. These findings are supported by data on the vocal tract length in singers (Dmitriev and Kiselev, 1979). These results support the conclusion that formant frequencies and vocal tract morphology is significant to voice classification.

Both the Cleveland and the Ågren and Sundberg articles also considered the relevance of the voice source to perceptual voice classification. The results were limited to the finding that, within an individual singer, the source spectrum slope seemed to differ between low and high pitches. For example, it was observed that on pitch C3, which is a low pitch for a tenor and a high pitch for a bass, the tenor source spectrum slope was greater than the bass source spectrum.

Since the aforementioned studies, a great deal of research relative to the voice source has been presented which has direct implications concerning our further understanding of the voice source characteristics also in professional singers (Rothenberg, 1981; Sundberg and Gauffin, 1979; Sundberg and Gauffin, 1981). These studies have improved our understanding of how characteristics in the voice source waveform are manifested in the radiated vowel spectrum; therefore, it now seems appropriate to complement previous studies of voice categories with voice source data. The present investigation is a pilot study of voice source characteristics versus pitch in three subjects with differing voice classifications. The study includes data on subglottic pressure, sound pressure level (SPL), amplitudes of the fundamental and of the singer's formant, and glottogram characteristics.

Method

The singer subjects were a tenor, a baritone, and a bass (the two last mentioned being the authors). They all have a considerable background in solo singing. They sang a chromatic scale on the vowel /a / from the pitch E3 (fundamental frequency approximately 165 Hz) to the pitch E4 in the three different dynamic levels, forte, mezzoforte, and piano. The /a/-vowel was preceded by the consonant /p/ on each pitch in the chromatic scale. In this way the singer's subglottic pressure could be estimated from the oral pressure during the p-occlusion (see, e.g,. Bouhuys, 1968; Rothenberg, 1973; Löfqvist, Carlborg, and Kitzing, 1982). Henceforth these oral pressure values will be referred to as the subglottic pressure.

The oral pressure was measured by means of a pressure transducer connected to a 50 cm long plastic tube of 1.5 mm inner diameter. The subjects held this tube in the mouth corner and the resulting signal was recorded on one track of the FM tape recorder. The oral acoustic output was picked up by means of a mask of the type described by Rothenberg (1973), which the subjects pressed against the face. This signal was recorded on a second track of the tape recorder. Finally, the signal from a B&K condensor microphone at 50 cm distance from the subject's mouth was recorded on a third track of the tape recorder. All recordings were made in an anechoic chamber.

Analysis

The oral pressure during the p-occlusion was analyzed by means of an oscillograph. The sound pressure level at 50 cm distance, henceforth the SPL, was recorded using a B & K level recorder. In addition to this, the vowel spectra were analyzed by means of an FFT computer program with a time window corresponding to at least one fundamental frequency period. In selecting the part of the vowel sounds to be analyzed, no attention was paid to the phase of the vibrato undulations. Given the spectra and the SPL of the individual notes, the absolute amplitudes of the fundamental and of the singer's formant could be determined. Additionally, each note in each scale as recorded by the mask microphone was inverse filtered by means of a computer program, the output waveform of which represented an acoustic glottogram.

Results

Fig. 1 shows the subglottic pressure as function of pitch. As expected, all three subjects use the pressure as a tool for regulating loudness; the piano scale gives the lowest pressure values and the forte

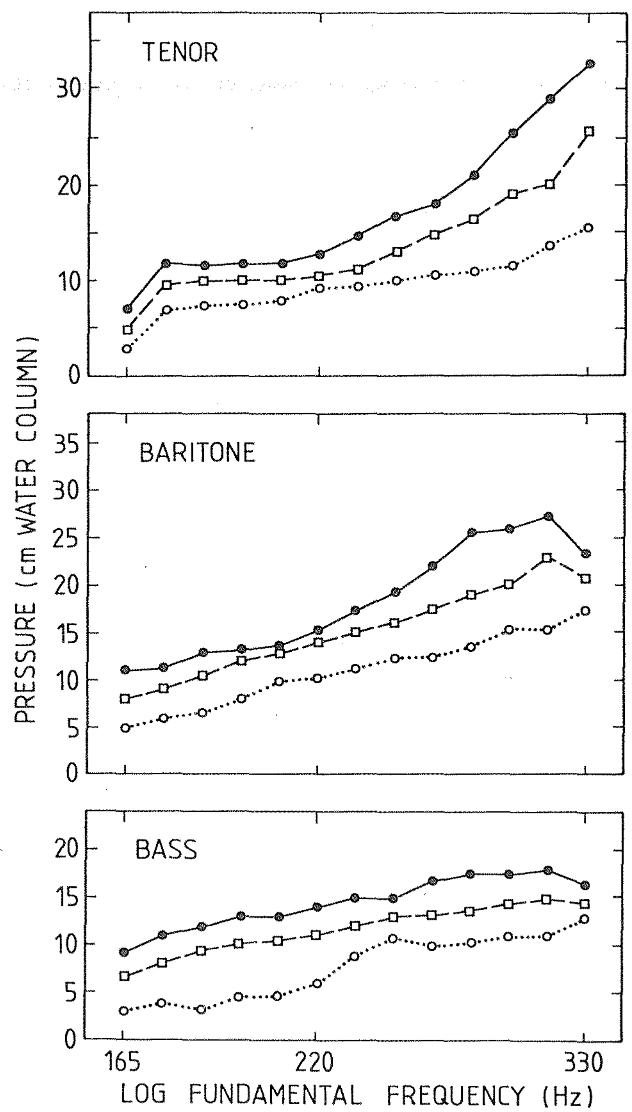


Fig. 1. Subglottic pressure as function of pitch in the three singers. Filled circles, squares and open circuits refer to high, medium, and low degree of vocal effort.

scale the highest. Also, all subjects increase pressure with rising fundamental frequency. This increase is greatest in the tenor and smallest in the bass. The pressure values pertaining to the mezzoforte scale is situated approximately midway between the forte and piano curves.

Subglottic pressure is known as the main agent in the control of phonatory loudness; however, the particular pattern of spectrum partials and formants complicates the quantification of this relationship. For this reason, the influence of the frequencies of the formants and harmonic partials on the SPL was estimated for each note in each scale. Such estimates were obtained from measurements of the mean amplitude produced by a terminal analogue, which was adjusted to the formant frequencies and excited by a standard source having the average fundamental frequency of each note. In these measurements the subjects' vibrato characteristics were taken into account. The SPL readings of the sung scale tones were then corrected accordingly. The SPL values thus obtained are directly related to the peak amplitude of the differentiated glottogram (see Fant 1979; Sundberg and Gauffin, 1980), or, in other words, they reflect a voice source characteristic.

Fig. 2 shows such corrected SPL values as a function of subglottic pressure plotted on a logarithmic scale. For a given pressure the SPL-values scatter about +4 dB within the subjects. Comparing the absolute SPL values of the different subjects, we observe that the bass tends to produce the highest SPL while the tenor tends to produce the lowest SPL. This difference is probably due to the fact that the maximum SPL for a voice rises as function of the pitch position along a subject's individual range, which, of course, is higher in tenors than in basses and baritones. Thus, we may assume that these SPL differences would disappear if we compared values pertaining to a comparable position along the singers' individual ranges rather than values pertaining to the same absolute fundamental frequencies. Fig. 2 also indicates that the tenor used the highest subglottic pressure of the three subjects, and achieved the lowest SPL values while the bass used the lowest pressure and reached the highest SPL values. This relationship shows that different subjects pay differing prices in terms of subglottic pressure for a given SPL.

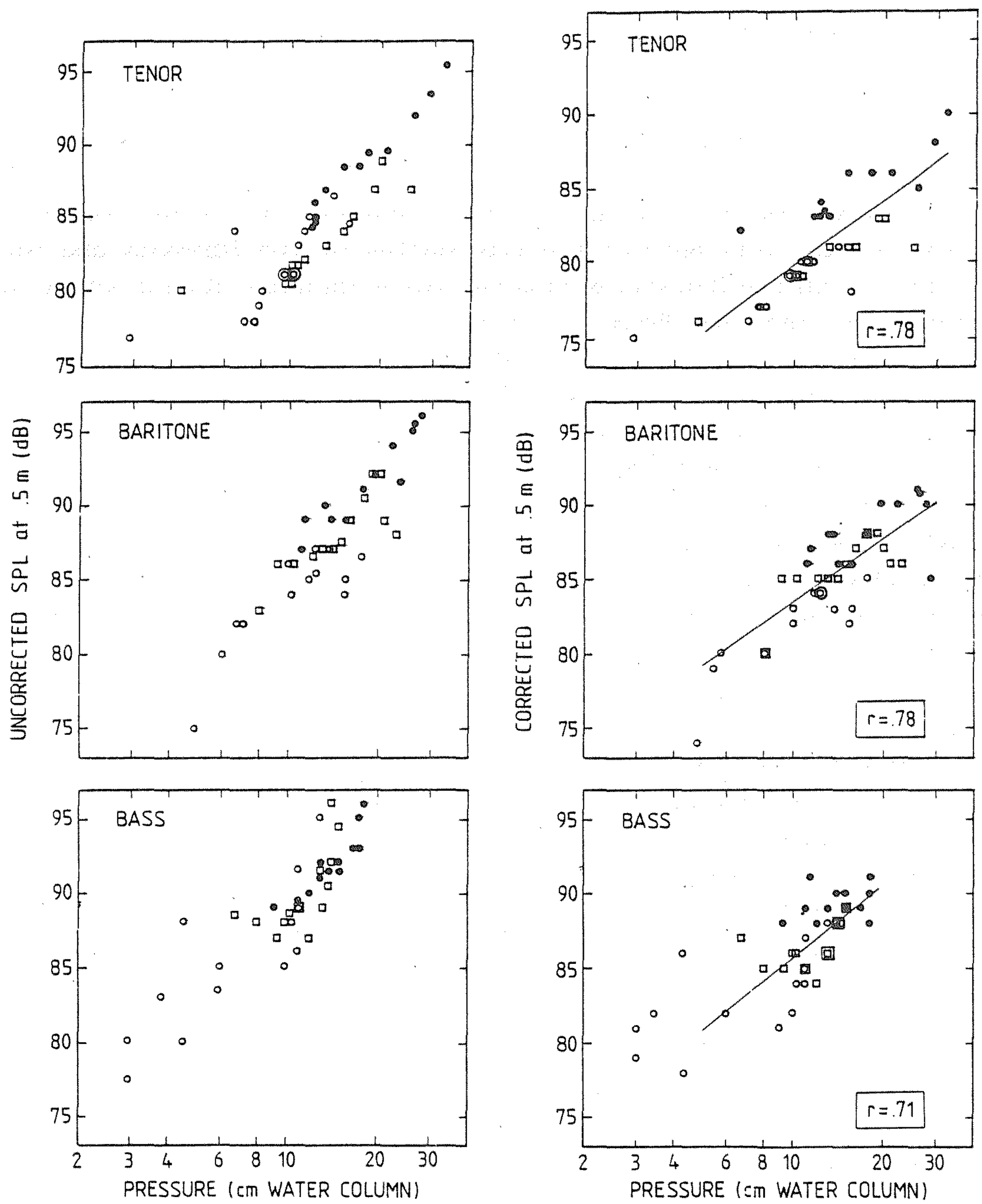


Fig. 2. Uncorrected and corrected SPL values as function of the logarithm of the subglottic pressure. Filled circles, squares, and open circuits refer to high medium, and low vocal effort.

These intersubject differences may reflect differences in vocal fold dimensions or in singing technique.

The level of the singer's formant is given as a function of the (uncorrected) SPL values in Fig. 3. The level of the singer's formant

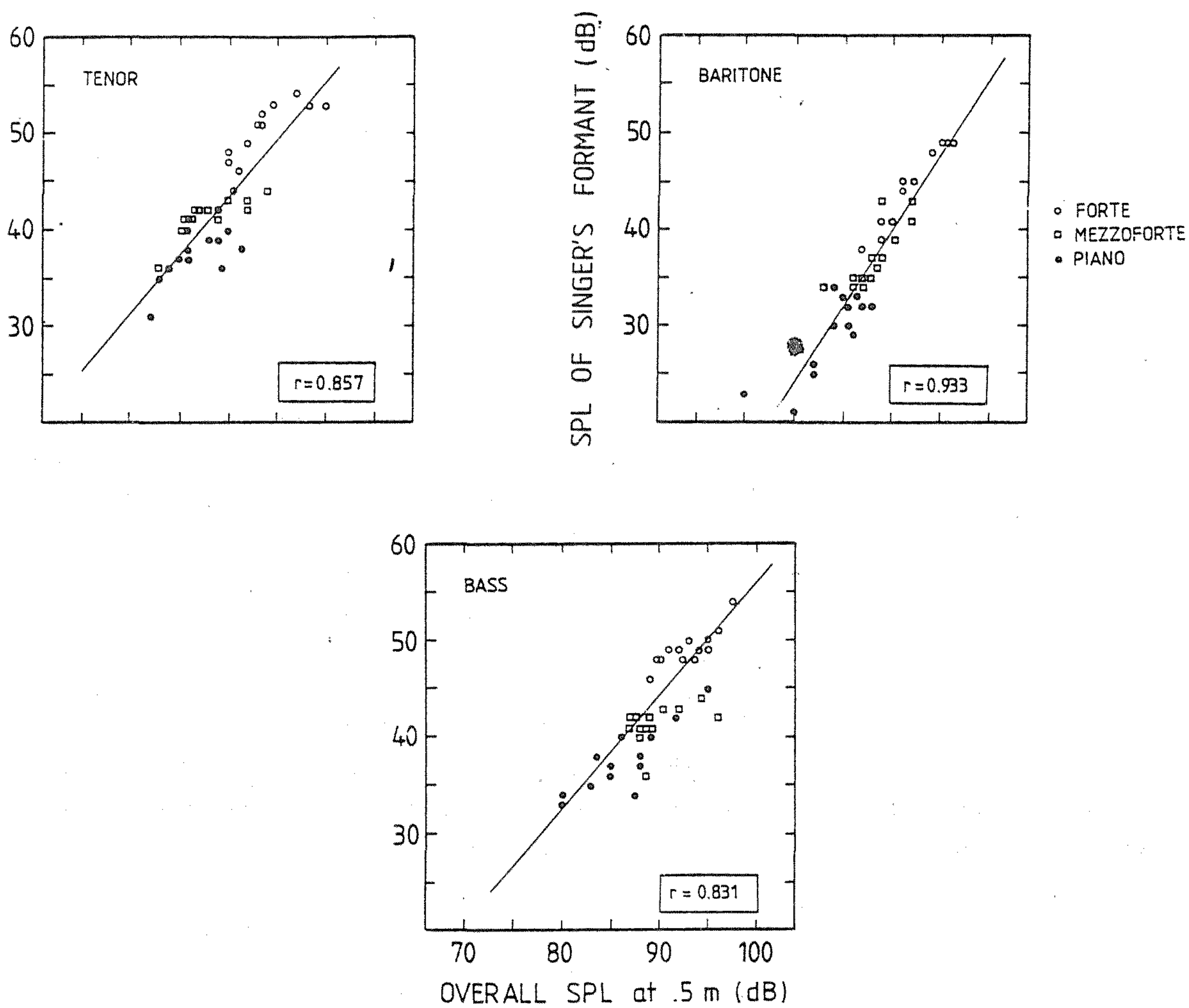


Fig. 3. The level of the singer's formant as function of the uncorrected SPL values given in Fig. 3. Filled circles, squares, and open circuits refer to high, medium, and low degree of vocal effort.

rises with both pitch and vocal effort. The baritone shows the lowest singer's formant level, and the tenor shows the highest levels. The correlation between SPL and the amplitude of the singer's formant is quite high in all subjects. In fact, 75% or more of the variation in the level of the singer's formant can be accounted for by the SPL variations. The slopes of the correlation lines are higher than 1.0 in all subjects indicating that the singer's formant amplitude increases more rapidly than the amplitudes of the lower overtones, which are responsible for the SPL readings. The slope is 1.2 for the tenor and the bass and 1.6 for the baritone implying that the intersubject differences in SPL are boosted in the frequency region of the singer's formant.

The amplitude of the source spectrum fundamental is closely related to the peak-to-peak amplitude of the glottogram, thereby revealing an aspect of the operation of the vocal folds (see Sundberg and Gauffin, 1979). In the radiated spectrum the amplitude of the fundamental is influenced also by the formants, although in a predictable way. Thus, given the formant frequencies and the fundamental frequency, the influence on the amplitude of the fundamental of the radiated spectrum can be compensated. Using the terminal analogue mentioned above, such compensations were made for each note in each scale as sung by each subject. The resulting estimated amplitudes of the voice source fundamental, shown in Fig. 4, demonstrate that the amplitude of the source spectrum fundamental remains essentially unaffected by pitch in these three subjects. In the case of the tenor and the bass, it tends to grow more or less systematically with rising vocal effort. In the case of the baritone, on the other hand, the amplitude is similar for all three degrees of vocal effort. This means that the baritone varies the amplitudes of his source spectrum overtones only, when he changes his vocal effort. The tenor subject shows the weakest fundamental, and the bass shows the strongest.

According to previous findings (Sundberg and Gauffin, 1979) the amplitude of the source spectrum fundamental is low in the lower part of a subject's fundamental frequency range, which agrees with the finding that the tenor shows the lowest values. In the same investigation, however,

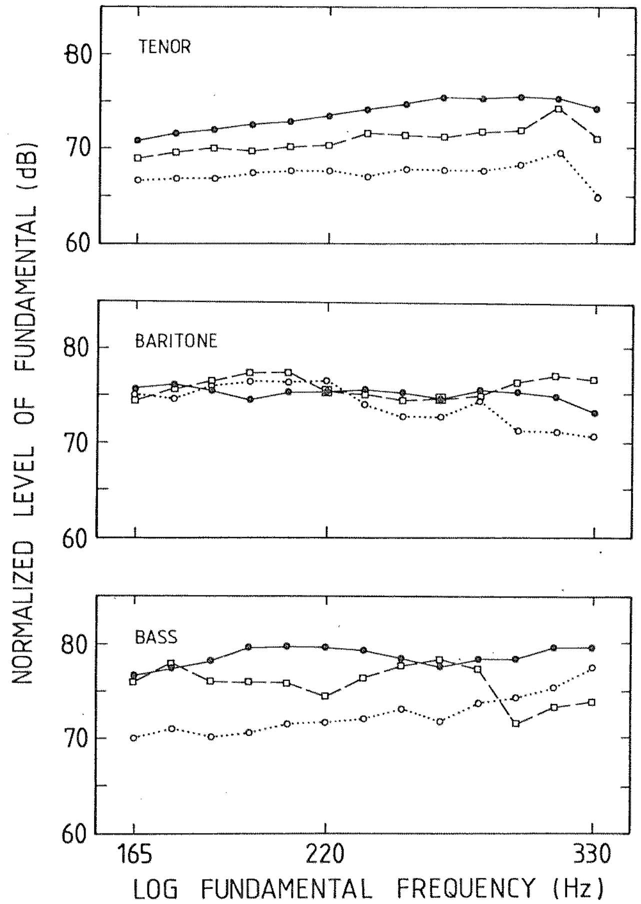


Fig. 4. Amplitude of the voice source fundamental as function of the logarithm of the fundamental frequency. Filled circles, squares, and open circuits refer to high, medium, and low degree of vocal effort.

the amplitude was found to drop towards the upper end of a professional singer subject's range, which is contrary to the present data. Similar results have been reported for speech (Fant, 1982). This discrepance might be the result of differing singing techniques used.

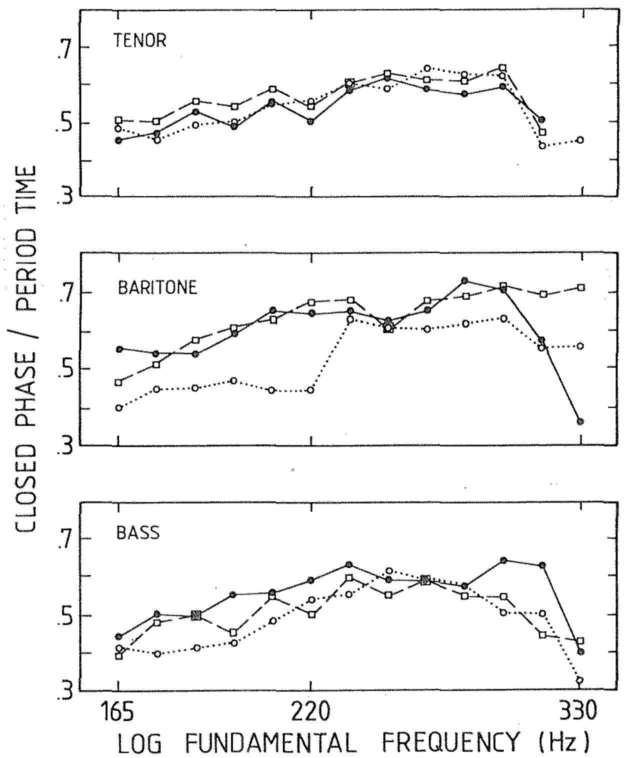


Fig. 5. Ratio between the durations of the closed phase and the period time in acoustic glottograms for the three singers plotted as function of the logarithm of the fundamental frequency. Filled circles, squares, and open circuits refer to high, medium, and low degree of vocal effort.

Apart from the corrected SPL and the corrected amplitude of the fundamental, there are other voice source characteristics that can be studied from acoustic glottograms. The glottograms were obtained by means of computerized inverse filtering of the voice signal picked up by the mask microphone. The inverse filtering gave reasonably convincing waveforms in most cases. The top note of the tenor's scales showed a distorted waveform, and was, therefore, excluded from analysis. In most scales, at least a few notes gave a tilted closed phase; however, most notes yielded normal looking glottograms. The formant frequencies were found to be essentially constant in all three singers within one scale. In other words, the subjects did not adjust their articulation for each individual tone in the scale.

A physiologically relevant parameter, that can be studied from acoustic glottograms, is the length of the closed phase in the glottal vibration cycle. The closed phase is important to two acoustic voice parameters, namely the amplitude of the source spectrum fundamental, and the overall SPL. The reason for this is that a lengthening of the closed phase shortens the glottal pulse time, and tends to shorten the closing time. The duration of the closed phase, normalized with respect to the period time, is shown in Fig. 5. Vocal effort does not appear to affect this parameter to any great extent. Pitch, on the other hand, seems to have a small effect; the closed phase is shorter for the lowest notes of the scales than for the higher notes and is again decreased towards the top notes of most of the scales. Thus, the curves are arch-shaped for all three subjects. The maximum seems to fall on lower pitches in the case of the bass than in the other two singers.

Discussion and conclusions

In previous investigations, vocal loudness has been found to be closely correlated with subglottic pressure (see, e.g., Proctor, 1968). The relationship is close to a +9-dB loudness increase per doubling of subglottic pressure. Recently, Fant discussed systematic SPL effects that can be expected from a change of the subglottic pressure (Fant, 1982). An SPL increase of 9 dB per doubling of subglottic pressure would result because of the concomitant increase of mean glottal particle velocity, mean glottal area, and other glottal waveform consequences of an altered subglottic pressure. In the present study we found correlation coefficients of about .75 and slopes in the vicinity of 7 implying a 4-dB loudness increase per doubling of subglottic pressure, thus only half of what has been reported before. There are, probably, several reasons for this discrepancy. One is that in the present study, we compensated for the influence of the frequencies of the formants and of the fundamental on SPL; this reduces the comparability between our results and previously published measurements. Another reason certainly is that our measurements pertain to scale singing. During conditions of changing pitch, a singer is likely to adjust his mode of phonation more or less. If the mode of phonation is changed, a perturbation of the correlation between subglottic pressure and SPL can be expected. For this reason, the data published here on this correlation would be less typical for the subglottic-pressure/SPL correlation than those collected under more constant phonatory conditions, e.g., constant fundamental frequency.

The level of the singer's formant was found to increase 1.2 to 1.6 times more rapidly than the overall SPL in our singer subjects. This means that the higher spectrum partials become increasingly dominant in the vowel spectrum as the phonatory loudness is increased. Similar values have been observed for untrained voices under the same conditions (see, e.g., Fant, 1959). This confirms previous assumptions that vocal-effort-dependent level variations of the singer's formant do not depend on a supernormal functioning of the voice source, but is mainly a resonatory phenomenon in the vocal tract.

One aim of the present investigation was to explore the possible voice source differences that might exist between the male voice categories. As we have confined this study to only three subjects, our data reflect both interindividual differences and differences that are typical for voice categories. We may assume that a difference typical for the voice categories changes gradually from tenor to baritone to bass. If we accept this criterion, we arrive at the following possible voice source differences that might merit future study.

All subjects increased subglottic pressure slightly with rising fundamental frequency. The increase was found to be greatest in the tenor and smallest in the bass. Moreover, the tenor used the highest subglottic pressure of the three subjects, and arrived at the lowest SPL values, while the bass used the lowest pressure and reached the highest SPL values. Thus, the three subjects pay different prices in terms of subglottic pressure for the same SPL values. Presumably, such differences can be heard both in the voice timbre and in the onset and release of phonation, and it is possible that these differences in the subglottic pressure are typical for the voice categories studied.

Another difference that might be typical concerns the amplitude of the voice source fundamental. We found that our tenor subject had the weakest fundamental, and the bass had the strongest fundamental. Similar findings were made previously in a study of the voice source spectrum of male singers, where a voice with a "dark" timbre showed stronger low frequency component in the source spectrum than a voice with a "lighter" timbre (Sundberg, 1973). Our results from the present study lend further support to the assumption that tenor voices typically have weaker source spectrum fundamental than bass voices.

Our study has shown that some intersubject voice source differences can be observed, if the phonatory behavior is studied as a function of pitch and loudness. Perhaps these acoustic parameters have perceptual correlates used by experts to classify voices into traditional categories. Nevertheless, it seems important to take into consideration such dynamic aspects of the voice source in future research.

Acknowledgements

The present investigation was made in the spring of 1981 during Dr. Cleveland's visit which was supported by a grant from The Voice Foundation.

References

Bouhuys, A. (1968): "Pressure-flow events during wind instrument playing", in Sound Production in Man, Annals of the New York Academy of Sciences 155, Art 1, 264-275.

Bouhuys, A., Mead, J., Proctor, D., and Stevens, K. (1968): "Pressure-flow events during singing", in Sound Production in Man, Annals of the New York Academy of Sciences 155, Art 1, 165-176.

Cleveland, T. (1977): "Acoustic properties of voice timbre types and their influence on voice classification", J.Acoust.Soc.Am. 61, 1622-1629.

Coleman, R. (1976): "A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice", J. Speech and Hearing Research 19, 168-180.

Dmitriev, L. and Kiselev, A. (1979): "Relationship between the formant structure of different types of singing voices and the dimensions of supraglottal cavities", Fol. Phoniat. 31, 238-241.

Fant, G. (1959): "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics No. 1, 1-108.

Fant, G. (1979): "Voice source analysis. A progress report", STL-QPSR 3-4/1979, 31-53.

Fant, G. (1982): "Preliminaries to analysis of the human voice source", STL-QPSR 4/1979, 1-27.

Löfqvist, A., Carlborg, B., and Kitzing, P. (1982): "Initial validation of an indirect measure of subglottal pressure during vowels", J.Acoust. Soc.Am. 72, 633-635.

Rothenberg, M. (1973): "A new inverse filtering technique for deriving the glottal air-flow waveform durning voicing", J.Acoust.Soc.Am. 53, 1632-1645

Rothenberg, M. (1981): "The voice source in singing", in Research Aspect of Singing, Publ. issued by the Royal Swedish Academy of Music, No 33, Stockholm, 15-33.

Sundberg, J. (1973): "The source spectrum in professional singing", Folia Phoniat. 25, 71-90.

Sundberg, J. and Gauffin, J. (1979): "Waveform and spectrum of the glottal voice source", in (B.Lindblom & S.Öhman, eds.) Frontiers of Speech Communication Research, Festschrift for Gunnar Fant, Academic Press, London, 301-320

Sundberg, J. and Gauffin, J. (1980): "Data on the glottal voice source behavior in vowel production", STL-QPSR 2-3/1980, 61-70.

Ågren, K. and Sundberg, J. (1978): "An acoustic comparison of alto and tenor voices", J.Research in Singing 1:3, 26-32.

SUPRALARYNGEAL ACTIVITY IN A STUDY OF SIX VOICE QUALITIES [1)]

Jo Estill, Thomas Baer, Kiyoshi Honda, and Katherine S. Harris
Haskins Laboratories, New Haven, CT, USA
and
The City University of New York, USA

Abstract

In a study, a trained singer held voice quality constant over a 2-octave range while EMG recordings were made of four supralaryngeal muscles at five frequencies (196, 294, 392, 587, and 784 Hz). At each frequency, six different qualities were produced, which on the basis of previous research can be described as follows: 1) Speech (neutral vocal tract/massed vocal folds), 2) Falsetto (relaxed vocal tract/thin vocal folds), 3) Sob (expanded vocal tract/thin folds), 4) Twang (constricted vocal tract/open velopharyngeal port/thick folds), 5) Opera (expanded vocal tract/ thick folds/constricted aryepiglottic sphincter), and 6) Belting (constricted vocal tract/thick folds).

Electrodes were inserted into the levator palatini, palatopharyngeus, middle constrictor, and geniohyoid. The laryngeal signal was recorded using a laryngograph and photoglottography. EMG activity of the levator palatini and the palatopharyngeus showed major increases in four of the six qualities, between 392 and 587 Hz, while in Sob (low larynx) and Opera, there was a more gradual increase in activity as frequency was increased. The middle constrictor was more active at 587 Hz for all qualities except Falsetto, which showed an increase in activity at 784 Hz.

From a preliminary analysis, there appear to be major shifts in these supralaryngeal muscles in some qualities, perhaps to maintain the quality beyond a critical frequency.

It is usual whenever "voice quality" is announced as the topic of a research paper, for the audience to assume that the reseach is aimed at clarifying one or another aspect of the "vocal register" problem. However, the relationship between "vocal register" and "voice quality" is

1) This paper is illustrated by sound examples.

terminologically confused. For some, a register is a perceived quality, i.e., chest, head, falsetto, etc. For others register is a description of that mechanism which may be used for only some part of the frequency range as when the untrained singer, in negotiating a scale will suddenly break into a falsetto quality if he is ascending, or into a chest quality if he is descending the scale. Implied in this instance is a change in physiology associated with the change in the frequency. Over the years, this confusion between the perceptual definition of registers and physiological registers has contributed to a proliferation of terms and meanings (Mörner, Fransson, and Fant, 1963).

Adding to the confusion are the conventions of the classical singers who are trained to proceed into the upper frequencies by changing to a "head" quality, which is deemed to be more beautiful in Western culture than extending the "chest" quality upward. Chest quality is believed to be not only more difficult to produce in the higher frequencies, but it is considered by many to be "dangerous" there as well.

The question which has not been asked nor investigated is: If the trained subject were to use a single perceived quality such as chest, head, falsetto, etc., with no frequency limitations, would there be some physiological evidence that the production mechanism changes at a critical frequency, either to increase frequency or to maintain that quality in the upper frequencies? Stated another way: Are there some set of perceived qualities which can be produced by a trained singer over the whole frequency range? If quality is maintained constant, is there some physiological change at crucial frequencies?

Our study of registers, therefore, has proceeded from this slightly different perspective. That is, we have held quality constant and looked for evidence of physiological differences as a function of frequency changing. From that perspective and using a single voice on which to make the comparisons, we began a series of experiments in 1973 of four selected voice qualities, using fiberoptic recordings, acoustic analysis, x-rays, cinefluorography, high-speed films, inverse filtering, and an extensive perception study (Colton and Estill, 1981).

Essentially, what we found was that singers can maintain perceptually identifiable qualities over wide frequency ranges. Perceptually discriminable qualities have characteristic spectra and overall SPL. Preliminary evidence suggests that qualities have both different laryngeal and supralaryngeal correlates.

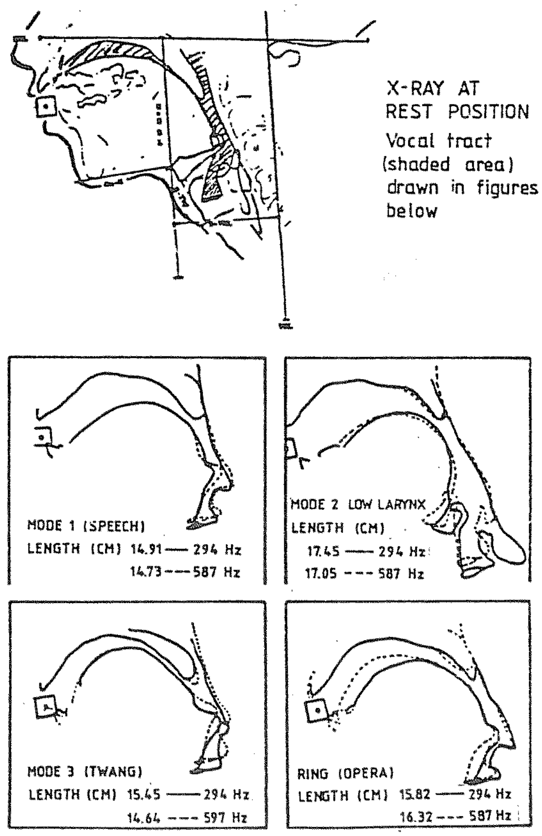


Fig. 1. X-rays of vocal tract at rest and for four voice qualities at 2 different frequencies for Speech, Low Larynx, Twang (nasalized), and Opera. The solid line is for the lower frequency, 294 Hz; the dotted line is for the upper frequency, 587 Hz.

Fig. 1 is taken from one of those studies (Colton and Estill, 1977; 1981) and is a lateral x-ray view of two frequencies in four different qualities. At the top is a mid-sagittal section of the head in rest position and the shaded area is the vocal tract. Only the vocal tract outlines are displayed in the four lower plates. The vocal tract shown in the upper left is for Speech quality; in the upper right, for Low Larynx quality; in the lower left is the vocal tract for Twang quality (nasalized), and in the lower right, that for the "ringing" Opera quality. The solid line contour is for the lower frequency, 294 Hz, while the dotted line contour is for the upper frequency, one octave higher, 587 Hz.

There were measured differences in length of the vocal tract and in length of the vocal folds among these qualities at the same frequencies (Colton and Estill, 1981). But one of the most tantalizing observations made of this figure was that in three of the qualities, the larynx was raised for the higher frequency, but in Opera quality, the larynx remained at approximately the same level as that for Rest or for Speech, at least for this octave, 294 to 587 Hz. With differences in vocal tract shape and laryngeal behavior for each quality, could there also be different muscular adjustments for frequency raising depending on quality? We formed hypotheses about the muscular control of frequency given the differences in vocal tract shape required for each quality. But clearly, what was needed were EMG studies to determine the roles of various muscles in quality maintenance and frequency change.

It was not possible to study all the muscles of interest in one EMG recording session. Therefore, the study was divided into two parts: The first, to record EMG activity from certain supralaryngeal muscles and the second, to record intrinsic and several extrinsic laryngeal muscles. What follows is a report from the first part of this two-part study.

In the first EMG experiment, hooked wire electrodes were inserted into four muscles according to current techniques used at Haskins Laboratories (Hirose, 1971). As indicated schematically in Fig. 2, those muscles were: (1) the Levator Palatini, which upon contraction will pull the soft palate up and back; (2) the Palatopharyngeus, which when contracting will pull the palate down, or if the palate is fixed, may help pull the thyroid wings and the larynx up. Its contraction may also result in medial movements of the lateral pharyngeal walls. (3) the Middle Con-

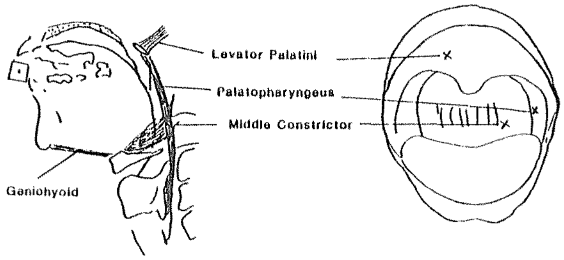


Fig. 2. Insertions of EMG electrodes into the supralaryngeal muscles: the Levator Palatini, the Palatopharyngeus, the Middle Constrictor, and the Geniohyoid.

strictor, which upon contracting will narrow the pharynx and pull the hyoid bone posteriorly. (4) the Geniohyoid which when contracting pulls the hyoid anteriorly and may help lift the larynx.

In addition to the EMG and the audio signals, glottographic signals were recorded using a Synchrovoice Electroglottograph. A nasopharyngeal fiberscope delivered light to the laryngopharynx and photoglottographic recordings were attempted.

Protocol

In an extensive perceptual study of four qualities, Speech, Low Larynx Twang, and Opera, Colton and Estill (1981) found that these qualities could be produced throughout the frequency range. They were also perceived as different throughout the range, although perceptual distinctiveness was reduced at the highest and lowest frequencies. Besides the first four qualities examined in the previous studies, two other qualities were added to this one: Falsetto and Belting. Belting is the quality used in some American musical theatre and heard around the world in much ethnic music. Informal listening to the tokens produced in the present study appeared to show a similar consistency across the frequency range, which suggests that the qualities in this study were like those of the earlier study. The vowel used was /i/, because it is produced with a fronted tongue root allowing the most favorable photoglottography.

The task here required that the subject phonate over a two-octave range in six qualities in each of the two protocols: first, using discrete pitches (196, 294, 392, 587, and 784 Hz) and then, gliding from one frequency to the next. The following results are from the first of these two protocols using discrete pitches.

Results: Spectral Analysis

Was each quality maintained throughout the range? Using a Kay Digital Spectrograph, a spectral analysis was made of all the tokens. If the spectra across frequencies within a single quality showed similar characteristics, then we might assume the quality distinctions were maintained from low to high range. In Fig. 3, each plot represents one quality and all the frequencies in that quality are numbered from 1-5, from low to high frequency.

Although the fundamental frequencies are different (196-784 Hz) and one might expect the difference in partials to make a difference in envelope, especially in the upper frequencies, within each quality the spectral envelopes are quite similar. This is confirmed by Plomp (1976) who notes that timbre (quality) is determined by the absolute frequency position of the spectral envelope rather than the position of the envelope relative to the fundamental. With the possible exception of Frequency Levels 4 and 5 in Falsetto and Low Larynx, these envelopes suggest that each quality was maintained across the frequency range. Especially interesting are the envelopes for Opera where all frequencies exhibit an identical spectral peak in the 2-3 kHz range.

Were the qualities distinct from each other? To compare the differences among the qualities, it would be useful if we could compute an average envelope for each quality. By dividing each of the 30 spectra (6

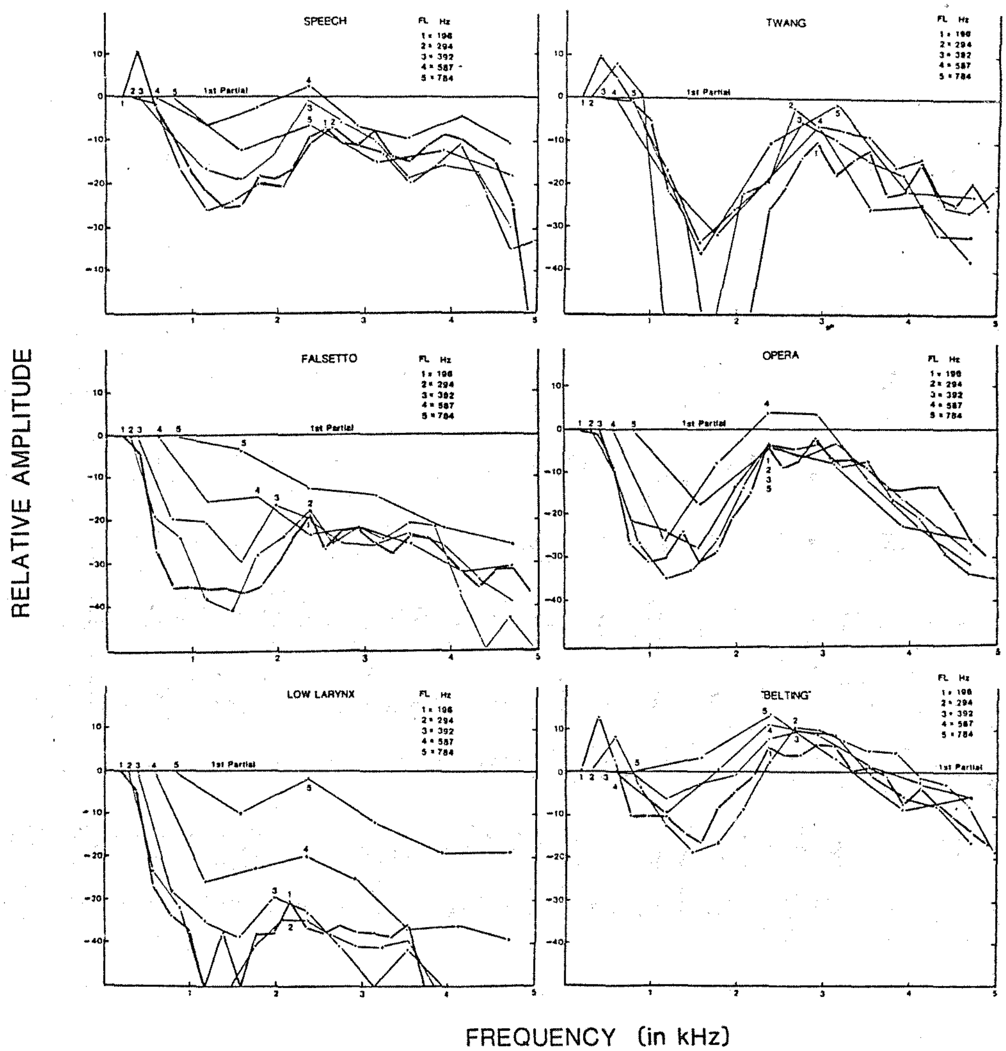


Fig. 3. Spectral envelopes for each voice quality displayed by the frequency levels. All partials were normalized to the fundamental frequency and each envelope represents the average of all the tokens for that condition.

qualities x 5 frequencies) into 500 Hz bins, and averaging all the energy in each bin, it was possible to extract an envelope that represents the total acoustic energy for all five frequencies for each quality. These are displayed in Fig. 4 and a comparison of quality by spectra is possible.

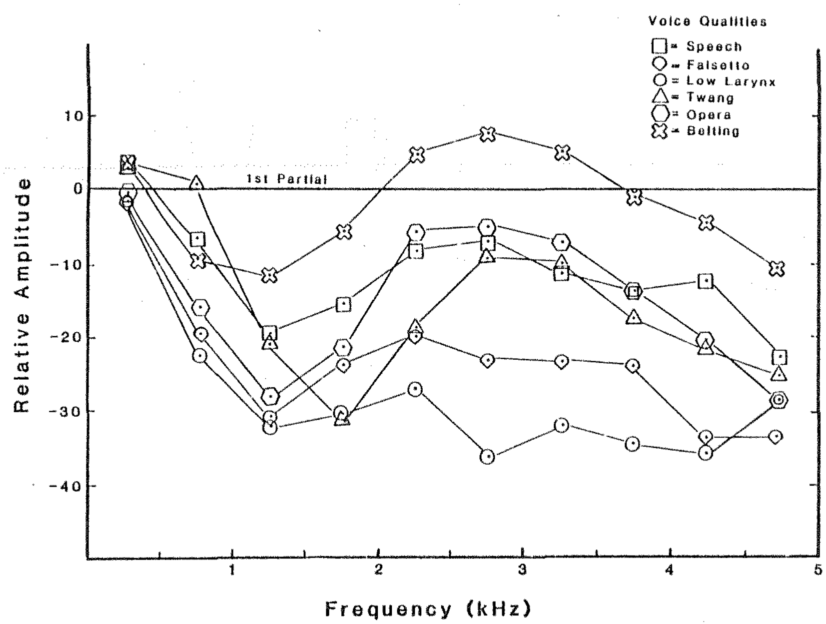


Fig. 4. A comparison of average spectral envelopes by quality. The spectra were divided into 500 Hz bins and all partials for all frequencies in each bin were averaged and plotted.

There are differences in spectral energy by the second 500 Hz bin, which become larger in the 2-4 kHz range, diminishing gradually in the bins above 4 kHz.

The analyses in Figs. 3 and 4 show in general, that each quality was maintained from low to high and that the qualities were different from each other. There did not appear to be any obvious markers suggesting a change due to frequency raising in either analysis. Did the physiology change as a function of frequency?

Results: Electroglottographic Analysis

With this question in mind, we analyzed the electroglottographic (EGG) data to look for the maintenance of laryngeal behavior across frequencies within a single quality or evidence of any change in laryngeal behavior due to frequency changing. In Fig. 5, electroglottograms are shown for each quality.

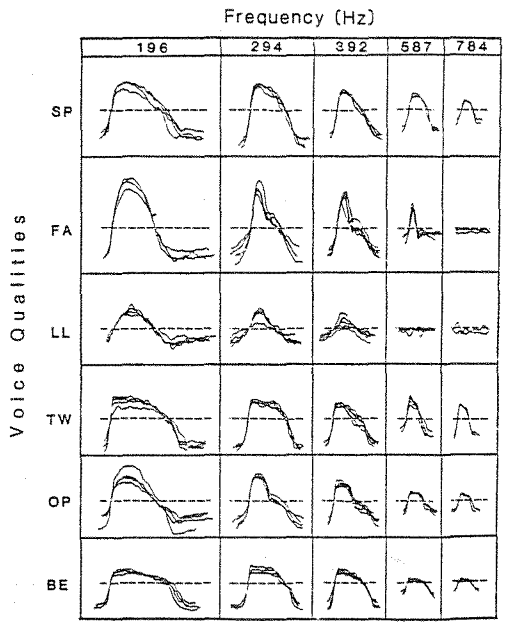


Fig. 5. Electroglottograms of six qualities at five frequencies. Each column is a frequency; each row is a quality. Each cell contains the waveform for one vibratory cycle. All tokens for each condition are displayed. Vocal fold closing is represented by the upward slope and all tokens were matched at the first intersection of the dotted line.

Each column represents one frequency and each row represents one quality. Within each cell, all tokens for that condition have been overlaid matching the first crossing at the dotted line. Closure occurs with the initial upward deflection. In Falsetto, most of the tokens showed intermittent closure, and no closure was recorded for the highest frequency. The electroglottograms for the lower four frequencies are taken from the part of the sample where closure occurred.

In general there appear to be similar patterns within a quality and distinctly different patterns across qualities. There does not seem to be strong evidence for a quantal change in source behavior for a given quality as a function of frequency. However, we are continuing to study both the spectral data and this laryngeal data, correlating them with the electromyographic data below. The EMG data which follow are the major focus of this report.

Results: EMG

Fig. 6 is an example of the muscle activity data for two qualities, Speech and Belting, at 392 Hz.

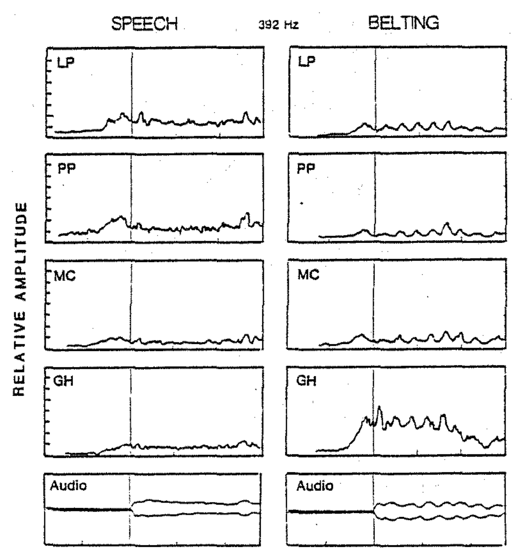


Fig. 6. Electromyographic activity for Speech and Belting voice qualities at 392 Hz for LP (Levator Palatini), PP (Palato-pharyngeus), MC (Middle Constrictor), and GH (Geniohyoid). Relative amplitude is displayed on the vertical axis and time in 500 msec is on the horizontal axis. Onset of tone is marked by the vertical line.

The top panels are for the Levator Palatini (LP), below that are the Palatopharyngeus (PP), followed by the Middle Constrictor (MC), and finally, the Geniohyoid (GH). The audio signals are in the lowermost panel and the vertical lines indicate voice onset. The level of the rectified and integrated EMG signal in μV is displayed on the vertical axis and is the average for all the tokens for a given condition. Time in 500 msec units is on the horizontal axis. A template was devised to average the activity for each muscle for each condition and these averaged values are plotted on the next figure. The level of activity was scaled relative to the average level observed for any condition.

Before proceeding to Fig. 7, we wish to call attention to the presence of amplitude fluctuations at 5-6 cycle/sec rate in the LP, PP, and MC muscles in Belting which accompanies acoustic vibrato. Vibrato was also a feature of the Opera and Low Larynx qualities, and occasionally, the Twang quality, as well. Vibrato seems to be a natural consequence of the production of some qualities, which a singer can voluntarily remove. In other qualities, it is naturally absent and singers can voluntarily put it into the tone. In this instance, this subject voluntarily used vibrato to "sweeten" the sound and to add some poignancy into what would otherwise be heard was yelling. But vibrato will be the topic of a future study and no more will be said of it in this paper.

In Fig. 7, the activity for each muscle is displayed separately by quality and by frequency.

At first glance, one notes that the activity patterns for three of the muscles look somewhat similar, in that the activity is low at the three lower frequencies and generally increases with the higher frequencies. The GH is distinguished by a much higher activity for three of the sex qualities: Twang, Opera, and Belting.

Since on contraction, the GH will pull the hyoid anteriorly, this higher level of GH activity may contribute to an enlarged vocal tract in the A-P dimension in the louder qualities. There is reason to believe,

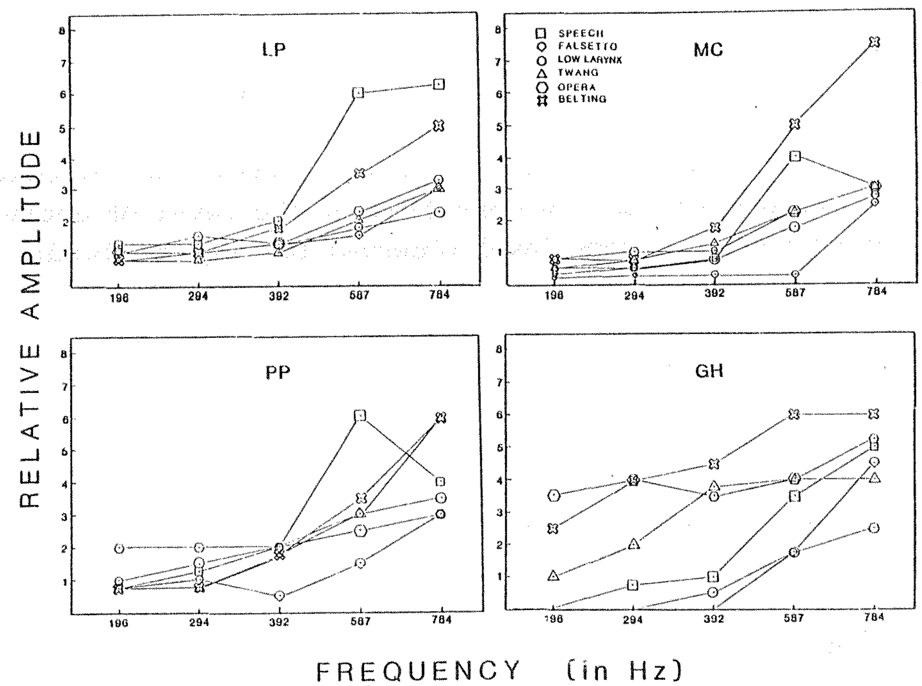


Fig. 7. Comparison of averaged muscle activity over 5 frequencies for each of 4 supralaryngeal muscles, active in the production of six voice qualities: Speech, Low Larynx, Twang (nasalized), Opera, and Belting.

given the work of Sundberg (1977) and others, that an expanded pharynx may contribute to a spectral peak at about 3 kHz, one of the features of these louder qualities (Figs. 3 and 4).

Returning to the other plates in this figure, all of the muscles including the three remaining qualities in the GH, have low amplitudes at the three lower frequencies with significant increases at or before 587 Hz. Is this the evidence for physiological registers? It is in this area of the range where we have come to expect that a change is necessary. Is this phenomenon observable for all qualities?

When we replot the activity by qualities, Fig. 8, a slightly clear picture emerges.

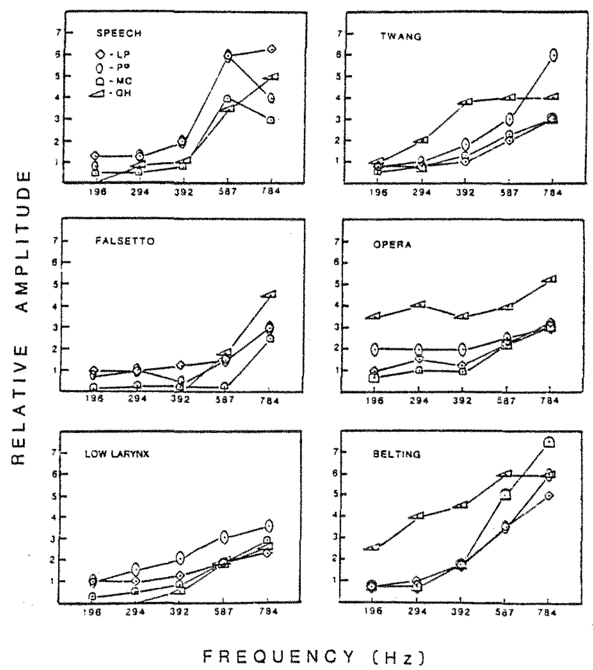


Fig. 8. The averaged amplitude of four supralaryngeal muscles at each of five frequencies displayed by quality: LP (Levator Palatini), PP (Palatopharyngeus), MC (Middle Constrictor), amd GH (Geniohyoid).

In Speech, there is a decided change in EMG amplitude between 392 and 587 Hz. In the Opera and Low Larynx qualities, the increase in activity is more gradual and the change is not so dramatic. In Falsetto, there appears to be an abrupt shift between 587 and 784 Hz. In Twang and Belting, the activity for each muscle is so different as to make any decision debatable. But there is another way we can look at these data.

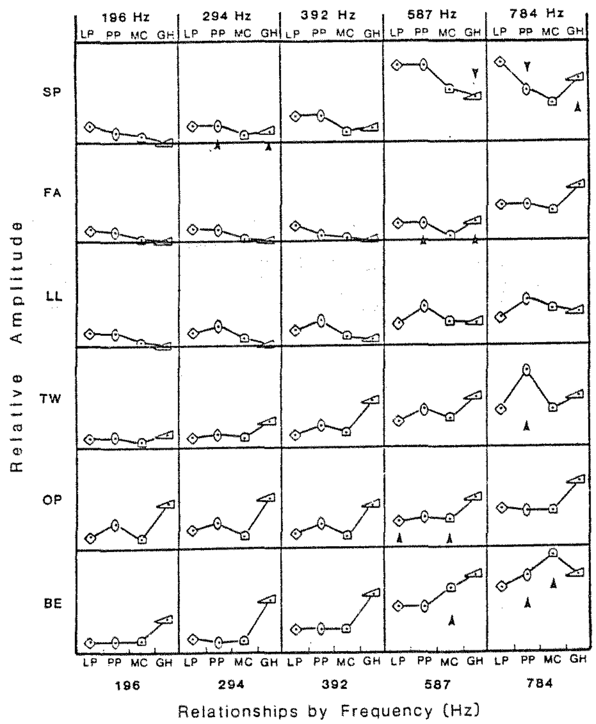


Fig. 9. The relationship of four supralaryngeal muscles is displayed by frequency (columns) and quality (rows). Within each cell is the averaged muscle activity of LP (Levator Palatini), PP (Palatopharyngeus), MC (Middle Constrictor) and GH (Geniohyoid).

In Fig. 9, each column represents a frequency and each row, a quality. Within each cell, the relative amplitude increases along the vertical axis, and different muscles are given on the horizontal axis. Strictly speaking, one cannot compare the amplitude of the EMG activity of one muscle with that of another, but that is not what we are doing here. In this figure, we are looking at relationships of these muscles to each other for evidence of a relative change in activity as frequency increases.

Looking at the columns, for any given frequency, there are significant differences across qualities with three exceptions. At 196 Hz, the cells for Speech, Falsetto, and Low Larynx are similar. At 392 Hz, Twang and Opera are similar. At 784 Hz, Opera looks like Belting at 196, 294, and 392 Hz. But again, these comparisons omit the infrahyoid activity. More productive at the moment may be the examination of frequencies within a single quality.

In Opera, there is a change in the relationship among muscles at 587 Hz which seems to have continued in the same direction at 784 Hz. In the X-rays of this quality (Fig. 1) there was very little difference in laryngeal height between 294 and 587 Hz. In the acoustic domain (Fig. 3) there was no difference in the spectral peaks in the 2-3 kHz range at any frequency. Yet, by 784 Hz, the relationships and the amplitude of the EMG activity have changed from that at the lower frequencies. This will tend to coincide with what many singers describe as their "passaggio" around 698 Hz. i.e., between 587 and 784 Hz. Does this phenomenon qualify as a physiological register definition?

In Belting, the relationship begins to change at or before 587 Hz, continuing to 784 Hz. The spectra and laryngeal behavior for this quality were also consistent at all frequencies (Figs. 3-5). Again, is this evidence for a physiological register definition? If so, then change occurs where it occurs as in Speech, between 392 and 587 Hz. But the pattern is not the same. The pattern appears to be a mirror image of that in Speech.

Discussion

In discussions on voice registers, there are often disagreements about the number of physiological registers in the voice. Here, two qualities, Low Larynx, and Twang (up to 587 Hz), display a consistency in the relationships of the muscles to each other across the range. Increase in frequency is marked by a consistent increase in muscle activity in all the muscles. It would appear in these qualities, one might find support for a "one-register" argument.

In Speech, there is a change in amplitude between 392 and 587 Hz (with only slight change in the relationship of the muscles to each other) at the place in the range where we have come to expect a change in physiological registers. Between 587 and 784 Hz, there is a change in the relationship (but slight changes in amplitude). Could we say this may be evidence for the existence of two or several physiological registers in this quality?

Taking an overall look at Fig. 9 again, one wonders: If the relationships in Speech, that quality common to the untrained voice, had been held constant beyond 392 Hz, would the subject have shifted automatically to Falsetto or, perhaps to a light voice quality similar to Low Larynx? Similarly in Belting, if the relationship had been held constant beyond 392 Hz, would the singer have shifted automatically into Opera quality.

This phenomenon is often heard when a singer cannot make this quality at these higher frequencies. These observations speak to the issue of effort in vocalizing and where the effort might be located that effects changes in quality. Does the fact that at the highest frequency in Opera, the pattern is similar to Belting (although in Opera, the pattern is higher in amplitude), indicate something about the work or effort involved in making this ringing tone in Opera? Belting is considered by many to be a hyperfunctioning use of the voice, possibly damaging and Opera is considered by most to be "good" use of the voice. In this figure, the relationships for the high Operatic G (784 Hz) are slightly higher in amplitude, but comparable to that of the Belted G, an octave lower (392 Hz). According to this observation, both qualities would appear to be hyperfunctioning. Again, such speculations are premature, no matter how tempting they may be at the moment. They must wait for the data from the muscles below the hyoid.

There is one final observation we can make about Belting. We can imagine the height of the larynx from the degree of activity in the MC and GH with almost as much activity in the LP and the PP. Johnson, Sundberg, and Wilbrand (1983) studying Swedish "kölning," found using X-rays techniques that the pharynx was increasingly constricted and the larynx higher with increasing frequency. This would seem to be the case with Belting as well. The SPL of "kölning" was reported to be "almost as high as has been reported for operatic singing." In this study, the SPL of Belting was found to be higher than Opera at all frequencies (Figs. 3-4).

Conclusion

In the two-octave range examined in this protocol, two kinds of action seem to be at work in these four muscles, the LP, the PP, and the MC, and the GH:

(1) A CHANGE OF RELATIONSHIP OR POSTURAL SET: Some qualities at some critical frequency changed markedly, as in Belting at or before 587 Hz. Other qualities changed posture more gradually, as in Opera, between 587 and 784 Hz.

(2) INCREASE IN AMPLITUDE: As frequency is increased, the relationship between the muscles is maintained, but the level of activity is increased. Presumably, the relationship or postural set of these muscles maintained the quality, while the increase in amplitude was necessary for frequency changing in Low Larynx and Twang qaulities. Again, without the activity from the muscles below the hyoid, this is at best only an observation.

Where the change in action occurs appears to be quality dependent. In four of the six qualities, Speech, Falsetto. Opera, and Belting, changes in posture were observed and of these four qualities, two of them, Speech and Belting, also doubled in amplitude. Two qualities, Twang and Low Larynx, retained their postural relationship throughout the range, with only gradual increases in amplitude and no evidence of a critical frequency.

While the purpose of this study was to understand the mechanism for frequency changing given different voice qualities, what we may indeed be defining is the difference between registers when they are perceived as qualities, and registers when they are meant as physiological adjustments associated with frequency change.

It is premature to identify physiological register shifts on the basis of these four supralaryngeal muscles without the analysis from the muscles below the hyoid. However, when that information is availabe, it may be possible to understand more clearly the mechanical basis for frequency changing, given the physiological constraints imposed on the vocal folds and on the vocal tract by different voice qualities.

Acknowledgment

This work was supported in part by NINCDS Grant NS 13870 and NS 13617 to Haskins Laboratories.

References

Colton, R.H. and Estill, J. (1981): "Elements of voice quality: Perceptal, acoustic, and physiologic aspects", in Speech and Language: Advances in Basic Research and Practice, Vol. V (ed. N. Lass), Academic Press, New York.

Colton, R.H., Estill, J., and Gould, L.V. (1977): "Physiology of voice modes: Vocal tract characteristic", paper presented to the 6th Symposium: Care of the Professional Voice, The Juilliard School, New York City.

Hirose, H. (1971): "Electromyography of the articulatory msucles: Current instrumentation and technique", Haskins Laboratories, SR-25/26, 73-86.

Johnson, A., Sundberg, J., and Wilbrand, H. (1982): "Study of phonation and articulation in a type of Swedish herding song", STL-QPSR 4/1982, 99-116.

Mörner, M., Fransson, F., and Fant, G. (1964): "Voice registers", STL-QPSR 4/1963, 17-23.

Plomp, R. (1976): Aspects of Tone Sensation, Academic Press, London.

Sundberg, J. (1977): "Acoustics of the singing voice", Scientific American 236, March, 82-191.

MEASURING INSUFFICIENT VOCAL FOLD CLOSURE DURING PHONATION

B. Fritzell*, J. Gauffin**, B. Hammarberg*,
I. Karlsson**, and J. Sundberg**
*Huddinge University Hospital, Sweden
**Royal Institute of Technology, Stockholm, Sweden

ABSTRACT

There is a clear relationship between breathy voice quality and insufficient vocal fold closure during phonation. In patients with dysphonia, a "leak" between the vocal folds is often observed during laryngoscopy. Treatment aims at eliminating or at least diminishing this insufficiency. Hence, an objective measure of this insufficiency is needed.

By means of inverse filtering according to Rothenberg, the flow of air through the glottis can be determined at any point during the vibratory cycle. If the subglottic pressure is measured simultaneously, the area of the glottis during the approximation phase can be calculated and, thus also the degree of vocal fold closure insufficiency. A series of examples will be shown.

Introduction

There is a close relationship between voice quality and vocal fold closure during the glottal vibratory cycle in phonation. In patients with dysphonia, an insufficient vocal fold closure is often observed during laryngoscopy. Usually the voice quality sounds breathy; there is a "leakage", i.e., a waste of flow. Treatment is aimed at diminishing or possibly eliminating this insufficiency of closure.

For this reason, objective measurement of the insufficiency of vocal fold closure during phonation is highly desirable. This can be realized by analysis of high speed motion picture films, which, however, is a very difficult, expensive, and time-consuming procedure (Hirano, 1981). Moreover, it requires that the subject phonates with a laryngeal mirror in his throat, which undoubtedly is a most uncomfortable and unnatural condition for phonation.

In some previous attempts to evaluate quantitatively the glottal efficiency, the so called glottal resistance has been used (e.g., Rubin & al., 1967; cf., also Schutte, 1980). This is defined as the ratio between the sound pressure level of the sound produced and the overall airflow used for producing this sound. However, this is not a proper measure of the glottal leakage; rather it is a measure based on the average airflow over the entire glottal pulse. The peak value of the glottal pulse depends on the type of phonation (Sundberg & Gauffin, 1979). Thus, if the phonation is "pressed" (high subglottic pressure, high degree of medial compression) the peak value of the airflow is low, and when the phonation is of the "flow" type, the peak value of the airflow is high. This variation of the peak value of the glottal airflow will obscure the relationship between glottal resistance and glottal leakage. In order to measure this leakage, information on the waveform of the glottal airflow is required, so that the closed phase can be identified in the glottal vibratory cycle.

Acoustic inverse filtering ad modum Rothenberg (1973) represents the best method for this purpose. It is a non-invasive method which interferes little with the subject's normal phonatory behavior. The present paper is a report on our first clinical experiences with this method.

Material

For this study, a number of patients were selected who all suffered from insufficient vocal fold closure during phonation. The causes or types of vocal dysfunction are specified in Table I.

Table I. Phoniatric diagnoses of the patients.

Recurrent nerve palsy	5
Bowed vocal folds (sulcus glottidis 2)	3
Nodules	3
Vocal fold polyp	1
Habitual diplophonia	1

The material consisted of 22 recordings from these 13 patients. In some patients, recordings were made both before and after therapy.

Methods

The equipment used for acoustic inverse filtering has previously been described by Rothenberg (1973) and Sundberg & Gauffin (1979). In the present study, the signal from a Rothenberg mask was lowpass filtered at 1.2 kHz and stored in an Eclipse S 140 minicomputer. The inverse filtering of the recorded vowel signal was carried out by an interactive filtering program, INA, written by J. Liljencrantz and P. Branderud. With this program antifilters were adjusted so as to cancel the first and second formants whereby a maximally even spectrum slope was obtained for the filtered signal.

The subglottic pressure was assumed to equal the intraoral pressure during the occlusion for the /p/ which followed the inverse filtered vowel (cf., Rothenberg, 1973; Löfqvist & al, 1982). The intraoral pressure was measured via a narrow plastic tube, which the patients held in the mouth corner and which was connected to a pressure transducer (Fig. 1).

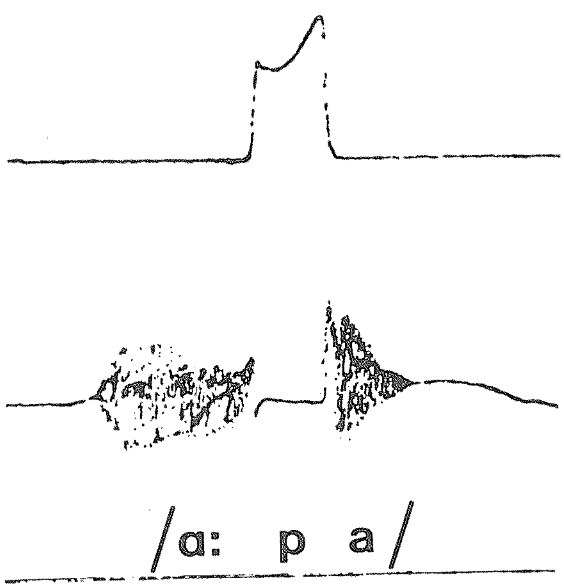


Fig. 1. Recordings of intra-oral pressure (upper trace) and microphone signal (lower trace) with the subject saying "apa". In the present study, the initial pressure during the production of /p/ is used as an equivalent of the subglottic pressure.

The overall sound pressure level was determined in each recording as the RMS amplitude of the differentiated air flow signal recorded by the mask microphone corrected to represent the SPL at 0.5 m from the mouth.

The patients were instructed to repeatedly pronounce the word /ɑ:pa/ (Swedish word for "monkey") first at a normal voice level, then with a loud voice, and, finally, with a soft voice. For each of these three conditions, one utterance was recorded and the initial vowel was subjected to inverse filter analysis. Fig. 2 shows an example of the results. The figure also illustrates that the measure of the minimum flow (mf) was obtained as the flow which corresponds to that phase of the vibratory cycle, during which the folds should ideally meet and close the glottis.

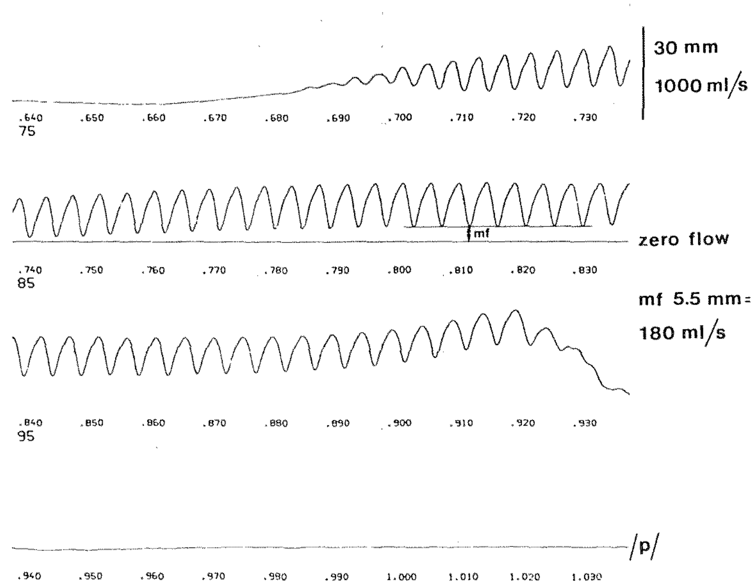


Fig. 2. The voice source representation after inverse filtering from a patient with vocal nodules phonating at normal level. The first vowel and the lip closure phase of the /p/ (in "apa") is seen. During /p/ there is no air flow. From this level the distance to the lowest part of the voice source is measured at a point chosen during the stable phase of the vowel, which corresponds to the minimum flow (mf).

Given air flow and pressure drop across the glottis, we can get an estimate of the area of the glottis. We know from model experiments that, except for very small openings, the pressure drop across the glottis is a function of the kinetic energy in the glottis:

$$P = k \frac{\varrho U^2}{2A^2} \qquad \text{or} \quad A = \sqrt{\frac{k \cdot \varrho}{2}} \cdot \frac{U}{\sqrt{P}}$$

where U is the volume velocity, P is the pressure drop across the glottis, ϱ is the density of air, and k is a correction factor. The factor k is usually assumed to be constant but is, in reality, both dependent on vocal fold cross-section shape and on the flow and may vary as much as from 0.9 to 1.5, approximately (Gauffin & al., 1983). It is, accordingly, not possible to make a very exact estimate of the glottis area but it may still be convenient to use one measure, which is closely connected with the area, instead of using two, U and P, to specify the glottal leakage. The value we chose for k is, thus, not very important as long as it is within the normal range found in model experiments. However, if results are to be compared among different investigators, it is convenient if the same k-value is used. Therefore, we suggest that, for simplicity, a k-value of 1 is used. In this article, the calculated value A is referred to as the projected glottis area.

Results

As seen in Table II, the minimum flow for a normal intensity varied from 1100 ml/s in a patient with recurrent nerve palsy to 10 ml/s in a patient, who was recovering from "bowed vocal folds" following a laryngitis. These extreme values correspond to a variation in the calculated glottal insufficiency area from 27 mm^2 to 0.3 mm^2.

Three of the patients with recurrent nerve palsy had teflon injected into the paralytic and atrophic vocal fold. Their voices improved con-

Table II. Results of measurements and calculations for normal intensity of voice at the first recording for each patient.

	Minimum flow, ml/s	Psubglott cm H_2O	Glottal insuff, mm^2	SPL dB
Recurrent nerve palsy				
OW	1100	9.2	27	65
EW	780	10	18	64
PL	630	10	15	68
KO	106	4.4	3.7	62
LA	140	4.2	5.0	58
Bowed vocal folds				
ST	200	12.5	4.1	75
LR	220	6.4	6.3	60
LT	10	4.0	0.3	62
Nodules				
MB	50	9.0	1.2	73
SM	70	12.0	1.4	66
LD	170	6.0	5.1	74
Vocal fold polyp				
LL	220	5.0	7.2	55
Habitual diplophonia				
GJ	240	11.6	5.2	70

siderably, and the measurements for minimum flow and glottal insufficiency decreased accordingly, as shown in Table III.

Table III. Results of measurements and calculations for 3 patients who had teflon injections. Normal intensity of voice. Post-operative recording approximately 3 months after the teflon injection.

		Minimum flow, ml/s	Psubglott cm H_2O	Glottal insuff, mm^2	SPL dB
OW	pre-op	1100	9.2	27	65
	post-op	300	12.5	6.2	70
EW	pre-op	780	10	18	64
	post-op	230	7	6.4	60
PL	pre-op	630	10	15	68
	post-op	200	7	5.5	62

Table IV. Results of measurements and calculations on a patient with bowed vocal folds and glottal sulsi. Post-op. recording 1 month after operation.

ST

Pre-op	Minimum flow, ml/s	Psubglott cm H_2O	Glottal insuff, mm^2	SPL dB
Normal intensity	200	12.5	4.1	75
Loud	180	16	3.3	82
Soft	40	7.5	0.9	68
Post-op				
Normal	300	15	5.7	65
Loud	230	21.5	3.7	78
Soft	70	7.5	1.8	68

One of the two patients who suffered from glottal sulci (furrows) underwent an operation involving excision of the sulcus and injection of teflon into the vocal fold on the same side. For a long time her voice was in poor condition (Table IV) in spite of regular voice therapy.

Four months after the operation an improvement started, and recently she has occasionally experienced that her voice is nearly normal. The most recent inverse filtering recording could not be used for measurements, however, presumably because of a leak between the mask and the skin, which was not observed during the recording.

Patients with vocal nodules often have a glottal insufficiency at the posterior end of the glottis, as revealed by indirect laryngoscopy. With increased loudness of voice this insufficiency tends to decrease. This is clearly exemplified in the two examples of Table V.

Both patients were treated with voice therapy, when the recordings were made. The first patient was about to finish therapy, while the other patient had just started.

Table V. Results of measurements and calculations on two patients with vocal nodules.

MB	Minimum flow, ml/s	Psubglott cm H_2O	Glottal insuff, mm^2	SPL dB
normal	50	9	1.2	73
loud	40	15	0.7	87
soft	110	2.8	4.8	54
LD				
normal	170	6	5.1	74
loud	90	6.5	2.6	83
soft	170	4.5	5.8	59

One of the patients had a large polyp on the left vocal fold, and it was removed surgically. Her rough voice was normalized by the operation. In Fig. 3 the pre- and post-operative inverse filter representations of this patient is shown.

Finally, one patient was studied who had habitual diplophonia for more than 10 years. He was given voice therapy and quickly learned to use a more normal voice (Fig. 3).

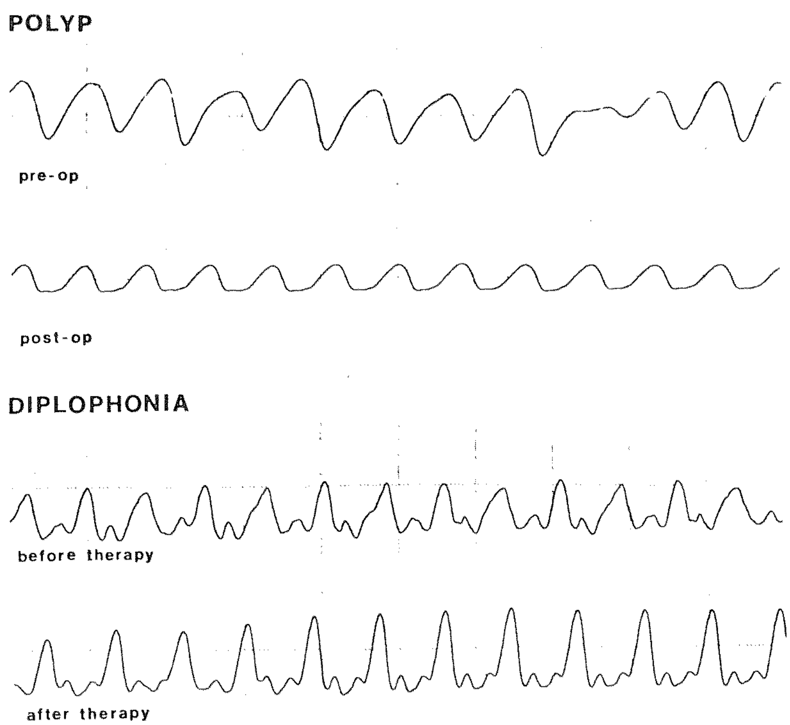


Fig. 3. Examples of inverse filtering recordings from two patients before and after therapy. The upper two recordings are from a female patient with a large vocal fold polyp, which was removed surgically. The glottogram became normalized. The lower two recordings are from a male patient with habitual diplophonia. After voice therapy, his voice improved considerably, and the glottogram became more regular.

Discussion and conclusions

In the beginning stage of phoniatrics, a great deal of attention was directed towards insufficient closure of the vocal folds, as observed by indirect laryngoscopy. Various deviant configurations of the glottis during phonation were described and attributed to the dysfunction of single pairs of muscles of the larynx. In German, the term "Internusparese" is still in use, designating the spindle-shaped insufficiency of the glottis, which in English usually is referred to as "bowed" vocal folds. The German term mirrors the assumption that it was a failing function of the internal thyroarytenoid muscles, which caused this kind of glottal insufficiency.

In modern literature on voice pathology, comparatively little is found about this matter. Schutte (1980), however, discusses the insufficient vocal fold closure at the posterior end of the glottis in patients with functional dysphonia. He points out that we do not know, whether this chink has organic causes or if it should be regarded as a habitual muscular maladjustment.

In most children with habitual, hyperfunctionl dysphonia, vocal nodules or more diffuse swellings of the anterior parts of the vocal folds are seen, provided that the examination can be carried out during respiration. During phonation one can sometimes observe insufficient vocal fold closure in the posterior half of the glottis, but more often this part of the vocal folds cannot be seen. Because of the hyperfunction, the entrance to the larynx is narrowed and the posterior part of the glottis is concealed by the aryepiglottic folds and the arytenoids. Thus, indirect laryngoscopy cannot tell whether or not these children suffer from an insufficient vocal fold closure.

Indirect laryngoscopy is an excellent, relatively simple way to examine the structure and the abduction of the vocal folds. More subtle aspects of their functions, however, cannot be reliably examined with this method because of the strange situation and the highly abnormal conditions for phonation; the patient has his tongue extended and a laryngeal mirror in his throat. Conclusions from observations made by laryngeal photography or by high speed motion pictures under such conditions must be drawn with great care, in particular if the patient is phonetically unsophisticated.

It must also be pointed out that vocal fold closure, as observed during indirect laryngoscopy, may vary a great deal because of various factors. One such factor influencing the closure is vocal intensity. With increasing voice effort, an insufficiency of closure usually diminishes and may disappear completely.

With these limitations of mirror laryngoscopy in mind, acoustic inverse filtering appears to be a very promising tool for clinical work. It was first described by Miller (1959). Rothenberg (1973) combined it with air flow recording and made it much easier to use. The present report of our first clinical experiences demonstrates the value of the method. It enables us to follow changes in vocal fold closure as a result of our treatment, and it will allow us to measure jitter and shimmer from the recording. It gives us a quantitative method to assess voice therapy.

It remains to test the accuracy and reliability of our procedures and to systematically investigate the problems involved in working with patients. Two sources of error are obvious; the method requires velopharyngeal closure during phonation, and the mask must be tight to the skin all around.

As pointed out in the beginning of this paper, insufficient vocal fold closure during phonation is apparently closely related to the perceptual quality of breathiness. Which are these relations? How does the degree of insufficiency influence the breathiness? Schönhärl (1960) as well as Koike & Hirano (1973) have observed a posterior glottic chink in subjects with normal voices. Maybe an insufficient vocal fold closure is not always heard? Presumably, acoustic inverse filtering will help us to answer these questions.

Acknowledgments

This work was supported by the Swedish Medical Research Council project No. 04783.

References

Gauffin, J., Binh, N., Ananthapadmanabha, T., & Fant, G. (1983): "Glottal geometry and volume velocity waveform", pp. 194-201 in the proc. of the Research Conf. on Voice Physiology, USA, 1981: Vocal Fold Physiology (eds. D.M. Bless & J.H. Abbs), College-Hill Press, San Diego, CA.

Hirano, M. (1981): Clinical Examination of Voice, Springer-Verlag, Wien.

Koike, Y. & Hirano, M. (1973): "Glottal-area time function and subglottal pressure variation", J.Acoust.Soc.Am., 54, pp. 1618-1627.

Löfqvist, A., Carlborg, B., & Kitzing, P. (1982): "Initial validation of an indirect measure of subglottal pressure during vowels", J.Acoust.Soc. Am., 72, pp. 633-635

Miller, R.L. (1959): "Nature of the vocal cord wave", J.Acoust.Soc.Am., 31, pp. 667-677.

Rothenberg, M. (1973): "A new inverse-filtering technique for deriving the glottal airflow waveform during voicing", J.Acoust.Soc.Am., 53, pp. 1632-1645.

Rubin, H.J., LeCover, M., & Vennard, W. (1967): "Vocal intensity, subglottal pressure and air flow relationships in singers", Folia Phon., 19, pp. 393-413.

Schutte, H.K. (1980): "The efficiency of voice production", Diss. University of Groningen.

Schönhärl, E. (1960): Die Stroboskopie in der praktischen Laryngologie, George Thieme Verlag, Stuttgart.

Sundberg, J. & Gauffin, J. (1979): "Waveform and spectrum of the glottal voice source", pp. 301-320 in Frontiers of Speech Communication Research (eds. B. Lindblom & S. Öhman), Academic Press, London.

"KÖLNING" 1)

STUDY OF PHONATION AND ARTICULATION IN A TYPE OF SWEDISH HERDING SONG

Anna Johnson*, Johan Sundberg**, and Hermann Wilbrand***
*Dept. of Musicology, The University of Uppsala
**Dept. of Speech Communication and Music Acoustics, KTH, Stockholm
***Dept. of Diagnostic Radiology, The University Hospital, Uppsala

Abstract

Within the Nordic herding culture a particular way of using the voice is practised for calling on the cattle. It is called "kölning". The physiological and acoustical characteristics of this type of voice use are examined and compared with a more normal type of singing in one female subject. The articulatory characteristics are studied by means of conventional radiographs of the vocal tract in a lateral projection. The subglottic pressure was estimated from the oral pressure during p-occlusions which the subject inserted between the notes. The formant frequencies associated with the vocal tract configurations as documented by the radiographs were estimated from area functions of the vocal tract. These area functions were derived from the radiographs. The results show that in kölning (1) the SPL is in the range of 80-100 dB at 50 cm distance in an anechoic room; (2) the subglottic pressure rises with fundamental frequency and may be as high as 60 cm of water column; (3) with rising fundamental frequency the vertical position of the larynx is raised considerably, the jaw and lip openings are increased, and the tongue shape is changed systematically; (4) the first formant is tuned to a frequency close to the fundamental while the second, third and fourth formant frequencies seem to remain in the vicinity of 1.7, 3, and 4 kHz, respectively, regardless of the pitch sung.

Introduction

In Scandinavian forest and high mountain pastures, herding is a woman's work. Here a special type of singing is used which in one of many dialects is called "kölning". This term is a derivative of the verb "kalla" (=to call), and kölning serves the function of calling the cattle from the forest, keeping it together or leading it in the right direc-

---

1) This paper is illustrated by sound examples.

tion. Thus, obviously this particular herding song can be heard over long distances. For a more detailed description of kölning the reader is referred to Johnson (1980) and Johnson et al. (1978). As compared with other types of folk song in Scandinavia, the singing style of kölning is unique in that it has a very high pitch range, similar to that of an operatic soprano. From the point of view of voice physiology, kölning is an interesting example of a manner in which the voice can be used so that very loud sounds can be generated at very high pitches, apparently without causing damage to the voice organ. The purpose of the present investigation was to examine the characteristics of this type of voice use from a physiological as well as from an acoustical point of view.

Description

The musical structure of this Scandinavian herding song is flexible and adaped to its function. It consists of phrases of varying length and vocal style: speech-song, calls, or song phrases. These song phrases are based on some long skeleton notes with more or less rich melismatic ornaments, i.e., textless melodic elaborations on one or several vowels, and are mostly free from consonants. These three different types of phrases are combined by means of free addition into chains, the length of which reflects the demands of the actual herding situation. For instance, if the cattle is close, speech-song phrases may be preferred while loud and high-pitched song phrases dominate when the cattle is far away. The present investigation is focused on the melismatic type of singing. Henceforth, the word kölning will be used for this melismatic type of high-pitched singing.

Method

A woman, born 1909, was selected as subject. She has learned this singing technique from unbroken oral tradition, and she still uses kölning for its original purpose each summer. In addition to this, she also has a large repertory of folk songs. According to her own expertise, she applies radically different singing techniques for kölning and for the folk song repertory. Also, she is a member of the local church choir.

The physiological characteristics of kölning were determined by means of conventional radiography of the entire vocal tract including the contours of the lips, the hard palate, the back pharynx wall, and the glottis.

The radiographs were taken in the standing position in the lateral view with a remote control x-ray apparatus (Siemens, Siregraph E), using fluoroscopy. The radiographic system includes a magnification factor of 1.2-1.3 for this type of examination. To reduce the radiation dose to a minimal lèvel, the x-ray beam was suitably coned and the imaging system was made up of cassettes with high speed screens (MR 600) and a universal x-ray film, Medichrome (Dahlin et al, 1978). The monochrome bleu color film has the advantage of reproducing structural details with high resolution because of the low granularity of the silver halogenide in the emulsion. The low granularity reduces the influences of the quantum noise of fast screens in the cassettes on the image quality, and thereby preserves the high resulution capacity of the film. According to the wide latitude of its characteristic curve, the blue color film has an inherent wide density range. This permits - on a normal negatoscope - visual perception of structures of low density such as, e.g., the lips, and at the same time of structural detail of high density such as bone tissue in the vocal tract walls. The remote control TV-fluoroscopy facilitated managements of the procedure.

Fourteen radiographs were taken at different moments while the subject was performing kölning; thereby, imaging was carried out so as to offer information on a representative choice of pitches within the relevant pitch range. For comparison, five radiographs were taken when the subject sang a melismatic part of a folk song, where the subject used different vowels. A tape recording was made of the entire session. As soon as the experimenter made a radiographic exposure, he simultaneously tapped a microphone connected to one track of the dual track tape recorder. The tape recording thus obtained allowed determination of the pitch pertaining to each radiograph. Additional radiographs were taken during rest with the subject holding a ruler in front of her lips, so as to also obtain the scale factor. Two radiographs pertaining to kölning can be seen in Fig. 1. From such radiographs a number of articulatory parameters were measured. Also area functions were derived, and then the formant frequencies associated with these area functions were determined.

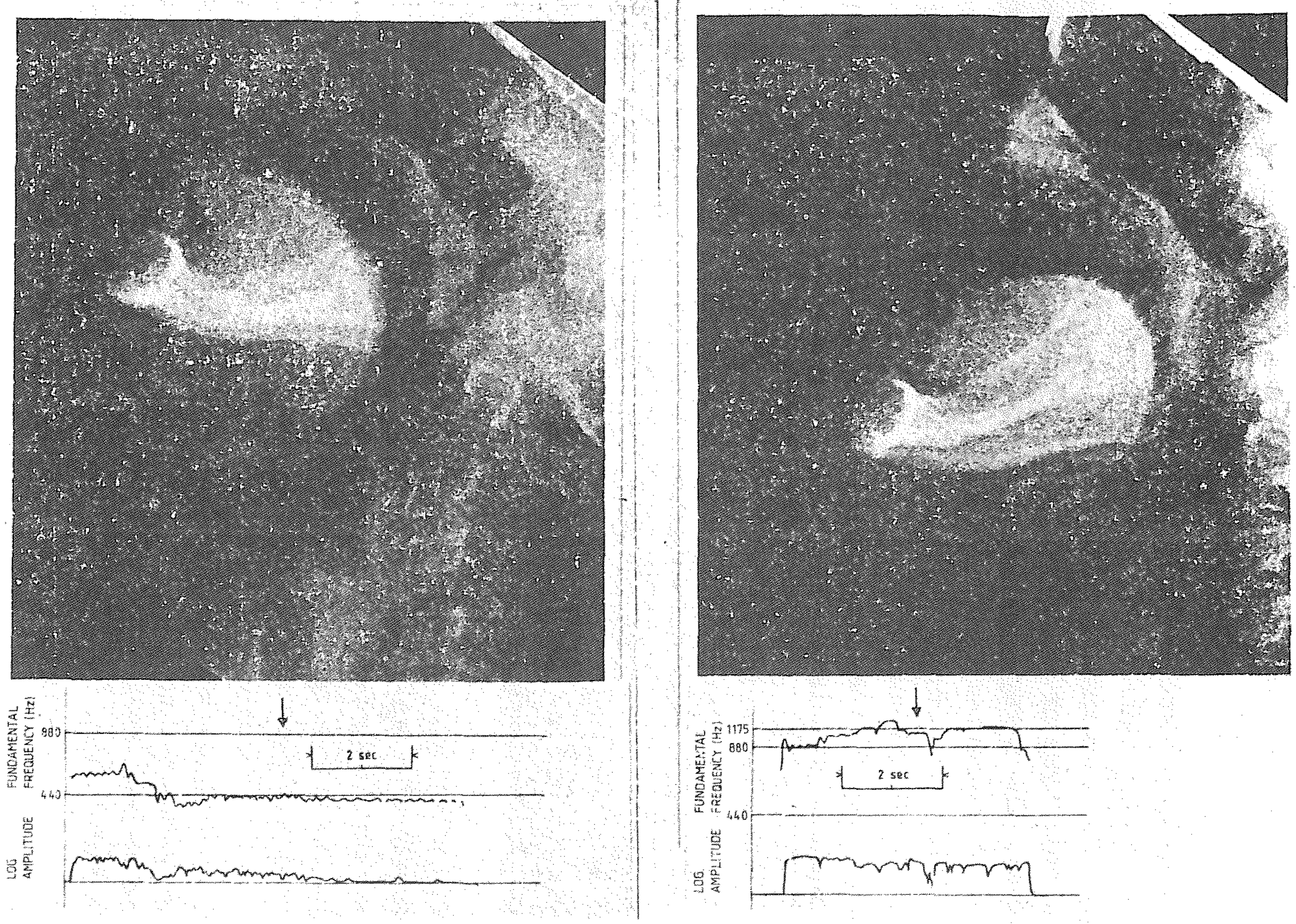


Fig. 1. Radiograms of the subject taken during kölning at the fundamental frequencies shown. Below are shown the Corresponding graphs of fundamental frequency versus time. The arrows show the instances at which the radiographs were taken.

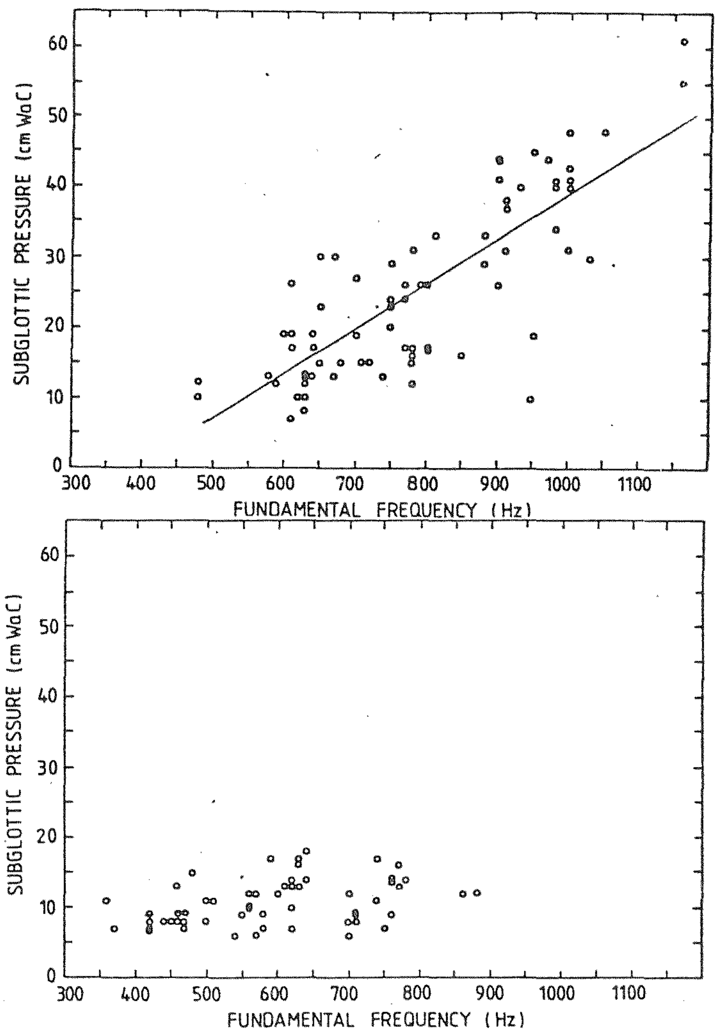


Fig. 2. Subglottic pressure as determined from the oral pressure during p-occlusion as function of fundamental frequency in the subject's phonation in singing a folksong (upper) and in kölning (lower). Also shown in the linear regression line for kölning (r=.805).

In a second experimental session the subject performed kölning and sang a folk song in an anechoic room. The sound was recorded on a calibrated tape recorder equipment so that the SPL could be determined.

A third experiment was made in order to estimate the subglottic pressure in kölning and folk song. In this session the subject inserted a p-sound between each note while she was holding a thin plastic tube in her mouth corner. This tube was connected to a pressure transducer, so that the oral pressure during the p-occlusions could be measured, and hence, an estimate of the sublottic pressure could be obtained. This experiment was made in an ordinary room and was recorded on tape.

Phonation characteristics

A: Subglottic pressure

The estimates of the subglottic pressure in singing and kölning are shown as function of fundamental frequency in Fig. 2. During kölning the subglottic pressure values up to a maximum of no less than 60 cm water

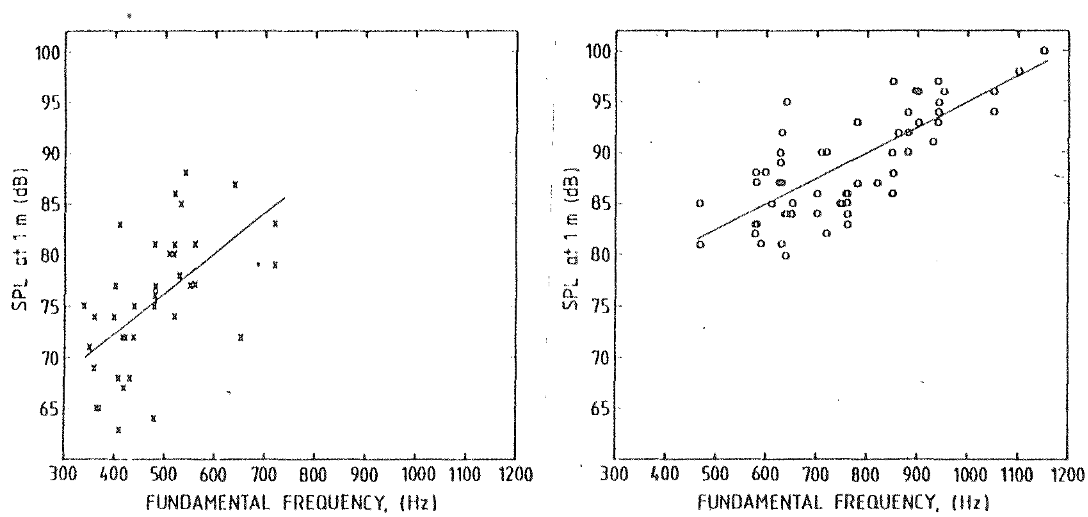


Fig. 3. Sound pressure level versus fundamental frequency and linear regression lines for data collected when the subject sang a folksong (left) and in kölning (right). For singing and kölning r=.567 and r=.760.

column (WaC) occur. During the same subject's normal singing the higest value was 18 cm WaC, which is in good agreement with previously data observed for operatic singing (see, e.g., Bouhuys et al., 1968). A typical value for normal conversational speech is about 6 cm WaC. From this we conclude that kölning requires unusually high subglottic pressures.

Normally the subglottic pressure is the main tool for regulating vocal loudness, other factors being equal: the higher the pressure, the higher the SPL (see, e.g., Bouhuys et al., 1968). SPL values for kölning and a folk song as recorded in the anechoic room are shown as function of fundamental frequency in Figs. 3a and 3b. It can be seen that the dependence on fundamental frequency is greater in the case of kölning. Probably this simply is a consequence of the fact that kölning is performed at much higher SPL values. The maximum SPL that can be produced by a voice is known to rise with fundamental frequency. Consequently, the louder one sings, the closer to this pitch dependent maximum value are the SPL values. For this reason a louder rendering of a song should always be expected to show a stronger correlation between SPL and fundamental frequency. Still, we may conclude that the SPL is more dependent in the subject's kölning than in her singing of folk songs.

Table I. Correlation coefficients (linear regression) for subglottic pressure (P), sound level in a normal room (SL), fundamental frequency (F0), and SPL in anechoic room in the subject's normal singing and kölning.

Type of phonation	P & SL	P & F0	F0 & SPL
Singing	.168	.345	.567
Kölning	.582	.805	.760

The correlation coefficients (linear regression) obtained from the subject's normal singing and from kölning are shown in Table I. It is clear that neither the SPL, nor the fundamental frequency is significantly dependent upon the subglottic pressure in the subject's normal singing. In kölning, on the other hand, the strongest correlation is found between fundamental frequency and subglottic pressure, while the correlation between SPL and subglottic pressure is less clearly pronounced. The lack of correlation between subglottic pressure and SPL would be due to a variation of one or more additional factors which also affect the SPL. As will be shown later, the formant frequencies are chosen with regard to the fundamental frequency, both in the subject's normal singing and in her kölning. As the formant frequencies influence the SPL, this explains why there is no high correlation between subglottic pressure and SPL in kölning. The high correlation between subglottic pressure and fundamental frequency in kölning suggests that the subglottic pressure is involved in the pitch regulating system in kölning.

B: Articulation

As mentioned, several articulatory parameters were examined in the radiographs. The larynx height was determined in relation to its position during rest and using the contour of the cervical vertebrae as the reference. Similarly, the jaw opening was measured using the rest position as reference.

Figs. 4 and 5 show larynx height and jaw opening as functions of fundamental frequency. In kölning in particular the vertical positioning of the larynx is considerably higher than during rest, and it shows a

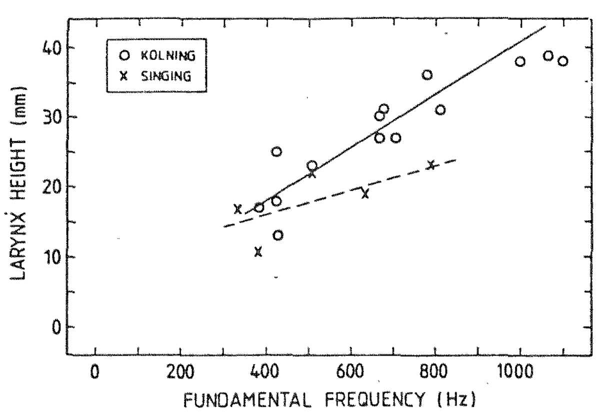


Fig. 4.

The vertical position of the larynx relative to the rest position as determined from the radiograms. Also shown are the linear repression lines. In singing and kölning r=.709 and r=.921.

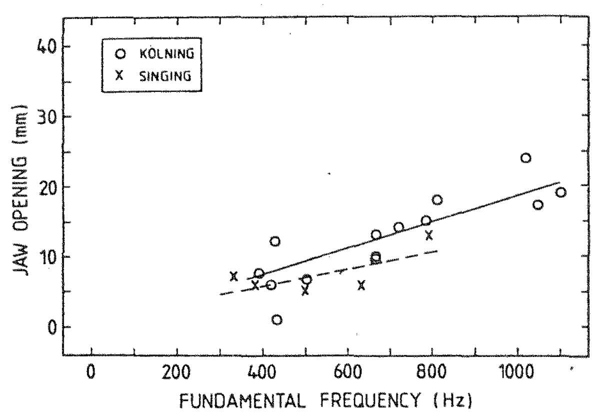


Fig. 5.

The jaw opening measured as the distance between the upper and lower incisors relative to the rest position value and determined from the radiograms. Also shown are the linear regression lines. In singing and kölning r=.693 and r=.854.

stronger dependence on fundamental frequency in kölning than in singing. Thus, in kölning the larynx height seems to be more closely related to fundamental frequency than in normal singing. At the highest fundamental frequencies the larynx rise is substantial, almost 4 cm in kölning.

Two measures were derived from the contours of the lips, see Fig. 6. The vertical distance between the contours of the upper and lower lip is

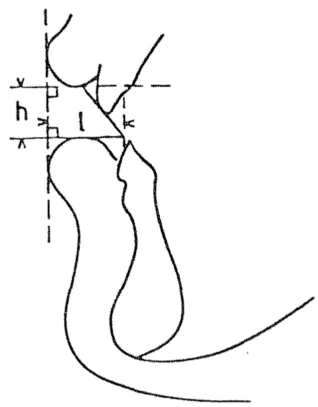


Fig. 6. Definition of the two lip measures derived from the radiograms: the vertical distance between the contours of the upper and lower lips, h, and the retraction of the mouth corner l. In cases where two contours were seen from the mouth corners because of parallax errors, the average was taken of the two resulting l values.

shown as function of the jaw opening in Fig. 7. Fig. 8 shows the parameter which we call retraction of mouth corners (cf., Fig. 6) as function of the same vertical distance between the lips. The graph reveals that there are no conspicuous differences between singing and kölning in neither of these two parameters. The jaw opening determines the vertical upper-lower-lip distance almost entirely, and this distance in turn determines the retraction of the mouth corners.

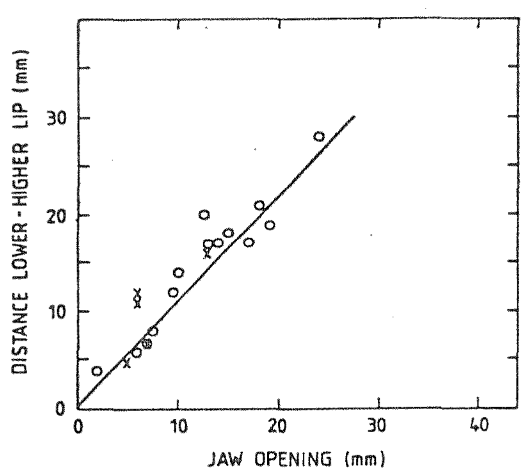


Fig. 7. Distance between the upper and lower lip versus jaw opening as determined from the radiographs and the corresponding linear regression line. In kölning r=.947.

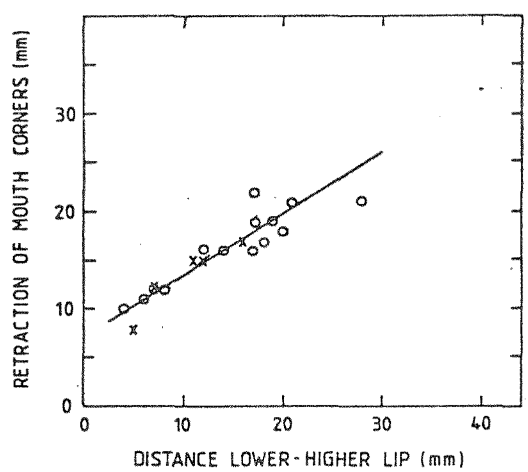


Fig. 8. Retraction of the mouth corners (l in Fig. 6) versus jaw opening as determined from the radiographs and the corresponding linear regression line for kölning (r=.902).

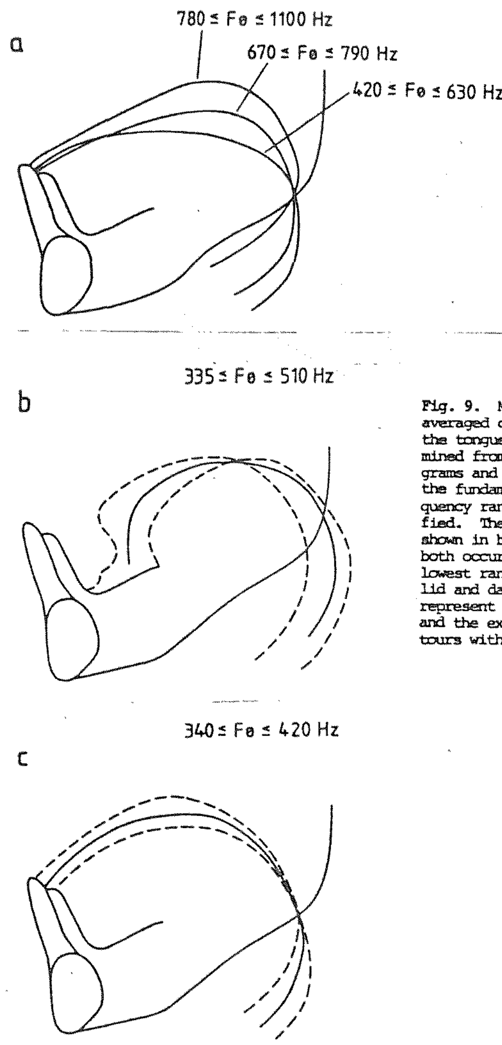


Fig. 9. Midsagittal averaged contours of the tongue as determined from the radiograms and observed in the fundamental frequency ranges specified. The contours shown in b) and c) both occurred in the lowest range; the solid and dashed curves represent the average and the extreme contours within the groups.

The tongue contours for the radiographs in kölning were examined using the contour of the lower jaw as the frame of reference. The tongue contour was found to vary systematically with fundamental frequency, so that similar tongue contours were observed for notes that were similar in fun-damental frequency. The averaged contours for three groups of tongue shapes are shown in Fig. 9a. They can be described as pharyngeal with a frontal tongue tip. With rising fundamental frequency, involving, as we have seen above, a rising of the larynx and a lowering of the jaw, the tongue root is increasingly raised. Seemingly as a result of this, the tongue hump rises with fundamental frequency. Slightly different tongue shapes were observed in the lowest fundamental frequency range, as seen in Figs. 9b and 9c. One of these (Fig. 9b) is similar to those observed for the higher notes, while the other (Fig. 9c) can be described as velar with a retracted tongue tip.

C: Formant frequencies

Acoustical data in terms of formant frequencies were derived from the radiographs. The processing of the radiographs involved the following steps. First, the mid line of the vocal tract in the mid-sagittal plane was estimated by means of a semipolar coordinate system which was anchored on the contours of the hard palate. Second, the dimensions of the vocal tract in the mid-sagittal plane were determined at each .5 cm along the midline mentioned. Third, each of these distances was converted into a cross-sectional area by means of data, which for the buccal region were derived from a plaster cast made of the subject's hard palate; in the case of the pharyngeal region previously published data were used (Sundberg, 1969). In this way an area function was derived for each radiograph.

The formant frequencies for the area functions were then estimated by means of acoustic models. Thus, each area function was constructed by piling .5 cm thick plexiglass plates with a center hole of well defined size. The sizes of these holes were chosen according to the values listed in the area function modelled. This model was then excited by means of a sinewave of variable frequency and generated by the STL ionophone, so that the four lowest formant frequencies could be determined (Fransson and Jansson, 1973).

Fig. 10 shows the resulting formant frequencies for the notes analyzed in kölning and in singing. The frequency values are seen to form a rather regular pattern, particularly in the case of the kölning. The variability that can be seen in the data pertaining to singing is expected, as the

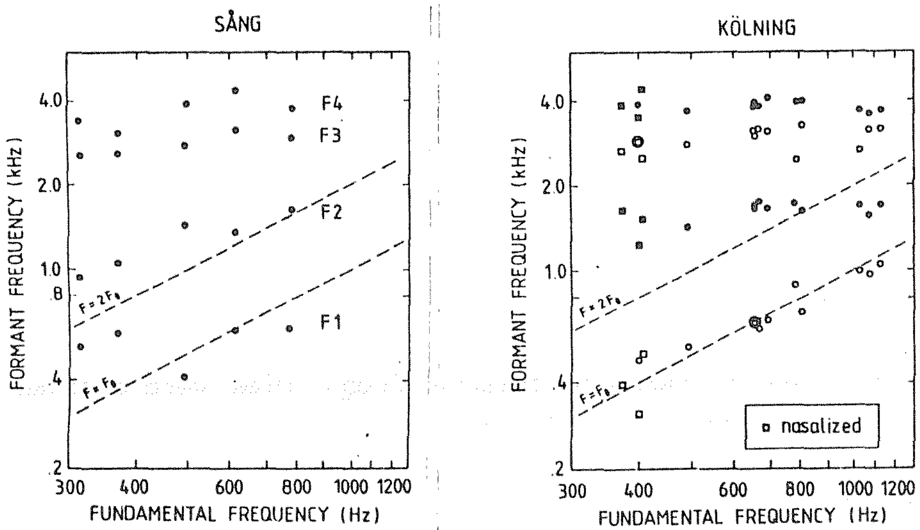


Fig. 10. Frequency estimates of the four lowest formants derived from the radiograms by means of the method described in the text.

radiographs were taken when the subject sang on different vowels. These observations support the assumption that the formant frequency data shown in Fig. 10 are reasonably reliable.

The formant techniques seem to be systematically different in kölning and singing. In kölning the first formant seems to be tuned to a frequency close to the fundamental throughout the pitch range of relevance. Also, regardless of fundamental frequency, the frequencies of the second, third, and fourth formants remain close to about 1700 Hz, 3000 Hz, and 4000 Hz, respectively. In the case of singing, on the other hand, the subject seems to tune the frequency of her first formant more freely, even though she appears to apply the principle of not allowing the fundamental to raise beyond the first formant frequency also in this case.

Apparently, this does not apply to the fundamental frequencies of 390 Hz and 790 Hz, where the frequency of the first formant is lower than that of the fundamental. However, this is probably because of an error in the estimation of the first formant frequency. It has been shown that in a real vocal tract, the first formant frequency is raised by yielding vocal tract walls, and this effect increases with frequency (Fant, 1960). Thus, our estimates of the first formant frequency must be too low at high frequencies, so that, probably, the correct values of the first formant are very close to the frequency of the fundamental for the three highest sung notes.

Discussion

We have seen in most plots that data on notes in kölning, which are similar in fundamental frequency, also display a similarity regarding articulatory as well as acoustic characterstics. This can actually be taken as a support for two assumptions; one is that kölning is a type of vocalization where the fundamental frequency is the key to articulation, so that other factors are of little relevance; the other is that our data are representative of kölning.

Considerable differences in subglottic pressure were found between singing and kölning. These great differences should be viewed against the background that the SPL was much lower in the case of singing. Certainly, our subject is able to produce a higher SPL by using a higher subglottic pressure in singing than she did when she sang the folksong in the present experiment, e.g., when she sings in her choir. However, the data collected would be typical for the voice use in singing the folk song repertory. Thus, the great differences in subglottic pressure between kölning and folk song singing are probably typical. Kölning may be less extreme in this respect when compared with, e.g., operatic singing, even though subglottic pressures as high as those we observed during kölning have been reported to be vary rare in operatic singing (see Proctor, 1974). We conclude from this that kölning probably requires higher subglottic pressures than other types of singing, and that the differences between kölning and operatic singing would be smaller than those seen in Fig. 2.

The articulation for the top pitches in kölning is extreme with respect to larynx height. In this regard, kölning seems to differ considerably from what has been recently found in operatic female singing

(Johansson et al., 1983). For one thing, in all the kölning notes examined, our subject operates her larynx in a position which is higher than the rest position. The professional soprano studied in the investigation mentioned used a larynx position below rest position for her lowest notes; then, she raised it with rising fundamental frequency, so that she reached a larynx level close to her rest position for her highest pitches. It is interesting that such high larynx positions may be habitually combined with the extremely high subglottic pressures mentioned before without causing damage to the voice organ. On the other hand, since kölning is reported to be efficient for its function, it would not be necessary to execute kölning for very long periods of time (simply because the cattle obeys the call) and from a voice hygienic point of view this may be important.

There is another interesting difference between the professional soprano's data and the data presented here. The contour of the back pharynx wall is almost blurred in the radiographs for the highest pitches in kölning, as can be seen in Fig. 1. This would be due to two factors. One factor is the very high subglottic pressure used at these extreme pitches, see Fig. 2, which decreases the x-ray transillumination of the tissues. A second factor may be a contraction of the medium constrictor muscle, which originates at the hyoid bone, courses posteriorly, and inserts in the mid-line raphe of the pharynx. Possibly, this muscle is engaged in the elevation of the larynx.

In our material on kölning, the pitch dependent choice of formant frequencies seems to apply more or less throughout the relevant pitch range. Moreover, the frequencies of the second, third, and fourth formants remain rather constant, regardless of fundamental frequency. In these two respects, kölning differs from what has been found in operatic soprano singing (Sundberg, 1975). On the other hand, there is an interesting similarity between operatic soprano singing and kölning as well; in both these types of singing the jaw opening is increased with increasing fundamental frequency. The underlying acoustical goal certainly is to tune the frequency of the first formant to the vicinity of the fundamental frequency. An acoustic result of this is a considerable gain in SPL obtained irrespective of vocal effort (Sundberg, 1982). Thus, vocal economy seems to be a common denominator in kölning and operatic soprano singing. This is a natural consequence of the fact that both kölning and operatic singing requires very loud sounds in order to serve its functions.

Conclusions

From the data presented and discussed above we want to infer the following. Kölning seems distinct from other types of singing regarding phonatory, articulatory as well as acoustical characteristics. Throughout the pitch range of relevance, fundamental frequency is a decisive factor in determining the subglottic pressure, jaw and lip opening, tongue shape, and larynx height. Also, the pharynx is increasingly constricted with rising fundamental frequency. Acoustically, the formant frequencies show characteristics distinct from other types of singing studied previously. Regardless of the fundamental frequency, the first formant seems to be tuned to a frequency close to the fundamental frequency, while the second, third, and fourth formants appear to remain in the vicinity of 1700 Hz, 3000 Hz, and 4000 Hz. The resulting SPL is almost as high as has been reported for operatic soprano singing. Since kölning is reported to cause no harm to the voice organ, it seems that it represents a way to vocal economy, which is different from that developed in opera singing.

Acknowledgments

This investigation was supported by the Bank of Sweden Tercentenary Foundation.

References

Dahlin, H., Welander, U., and Wilbrand, H. (1978): "Clinical comparison between monochrome colour film and black-and-white film", Acta Radiologica Suppl. 356.

Fant, G. (1960): Acoustic Theory of Speech Production, s'Gravenhage, The Haague.

Fransson, F. and Jansson, E. (1973): "The STL-Ionophone: Transducer properties and construction", J Acoust.Soc.Am. 58, 910-915.

Johansson, C., Sundberg, J., and Wilbrand, H. (1983): "X ray study of two professional female singers", (forthcoming)

Johnson, A. et al. (1978): "Fäbodmusik i förvandling", Svensk Tidskrift för Musikforskning 60:2, 5-39.

Johnson, A. (1980): "Sången i skogen - skogen i sången", pp. 66-103 in Folkmusikboken (J.Ling, M.Ramsten, & G.Ternhag eds.), Stockholm.

Lindblom, B. and Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw, and larynx movement", J.Acoust.Soc.Am. 50, 1166-1179.

Proctor, D. (1974): "Statement in a discussion of a symposium paper", p. 476 in Ventilatory and Phonatory Control Systems (B.Wyke, ed.), Oxford.

Sundberg, J. (1975): "Formant technique in a professional female singer", Acustica 32, 89-96.

Sundberg, J. (1982): "Perception of singing", pp. 59-98 in Psychology of Music, (D.Deutsch, ed.) Academic Press, New York.

X-RAY STUDY OF ARTICULATION AND FORMANT FREQUENCIES IN TWO FEMALE SINGERS

C. Johansson*, J. Sundberg** and H. Wilbrand***
*Dept. of Musicology, Uppsala University
**Dept. of Speech Communication and Music Acoustics, KTH
***Dept. of Diagnostic Radiology, University Hospital, Uppsala, Sweden

Abstract

Articulation and formant frequencies were studied from conventional radiographs of two professional female singers, an alto and a soprano. The subjects: (1) pronounced the vowels /o, i, u/ in a speech mode, and (2) sang the same vowels at a low, a middle and a high pitch. The jaw and lip openings, and, as regards the soprano, also the tongue shape and the larynx height changed systematically with fundamental frequency, in such a way that vowel differentiation mostly decreased with rising pitch. The frequency of the first formant was found to be close to that of the fundamental in cases where the fundamental would otherwise be higher than the first formant. As expected from the articulatory data, the formant frequencies showed a trend to increasing neutralization with rising pitch.

Introduction

Comparatively little is known about the characteristics of the female voice as compared to the male voice. The background is the high fundamental frequency range of the female voice, which makes formant frequency estimates uncertain and, hence, information on the voice source unsafe. The situation is, of course, much worse in the case of singers than in the case of speakers, since singers typically use fundamental frequencies that are much higher than those used by speakers.

Some years ago one of the authors (JS) attempted to estimate the formant frequencies of a professional soprano singer for vowels sung at fundamental frequencies between 250 and 700 Hz, approximately. Those estimates were based on data obtained from external vibration of the vocal tract during "silent singing" as well as from matchings of the radiated spectra on a formant synthesizer. The results suggested that the soprano tuned her first formant frequency to the vicinity of the fundamental frequency, as soon as otherwise the first formant would be lower than the fundamental. Photos of the lip opening of several female singers singing various vowels at different pitches supported the hypothesis, that the jaw opening is an important articulatory parameter behind this pitch dependent tuning of the first formant frequency. The generalizability of the results was supported by the naturalness obtained in a synthesis of female singing where this pitch dependent tuning of the first formant frequency was applied (cf., the sound illustration in Sundberg, 1977). However, the questions of: (1) the details of articulation, and (2), individual differences among singers with respect to articulation were left open. The aim of the present investigation is to supplement some articulatory data from a soprano and an alto singer, and also to attempt to make an acoustical evaluation of these data.

Experiment

The subjects were two well known professional Swedish singers, one alto who sings at the Stockholm Opera and one soprano who earns her life as a concert singer.

The entire vocal tract including the contours of the lips, the hard and soft palate, the back pharynx wall, the tongue, and the glottis was reproduced by conventional radiography. The technique and equipment thereby used were identical with those described in a parallel investigation of "kölning" (see Johnson et al., 1983).

Both singers sang the vowels /u, o, i/ at three fundamental frequencies, about 150, 300, and 600 Hz for the alto, and about 230, 470, and 950 Hz for the soprano. Also, radiographs were taken when the singers sustained the same three vowels in a normal speech mode. Finally, one picture was taken during rest with the subject holding a ruler in front of her lips, so as to obtain the scale factor.

From these radiographs area functions were derived, and the formant frequencies associated with these area functions were determined by means of an acoustical model of the vocal tract. The procedure thereby used will be accounted for later.

Results

A: Physiological data

Several physiological data are available in a lateral radiograph of the vocal tract. In the present study only such articulatory variables have been examined that are known to be acoustically relevant.

The vertical position of the larynx was measured by means of the template, shown in Fig. 1, where 0 represents the midpoint of the upper

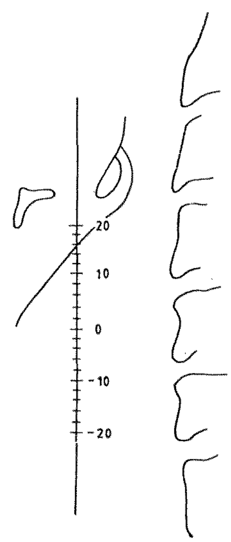


Fig. 1. Template used for measuring the vertical positioning in mm of the larynx. The value of 0 shows the rest position of the glottis contour.

glottis contour during rest. Adapting this template on the other radiographs caused no problems, as both singers kept their head in the same when they sang the various notes as during rest.

Fig. 2 shows the data obtained. There is a clear difference between the subjects. For the soprano's lowest pitches, the larynx position is lower than what is observed in the spoken vowels. However, she raises her larynx with fundamental frequency, so that she reaches a level similar to that of the spoken vowels at her top pitch. The alto uses a lower larynx in the sung vowels than in the spoken vowels in most cases but shows no clear relationship between larynx height and fundamental frequency.

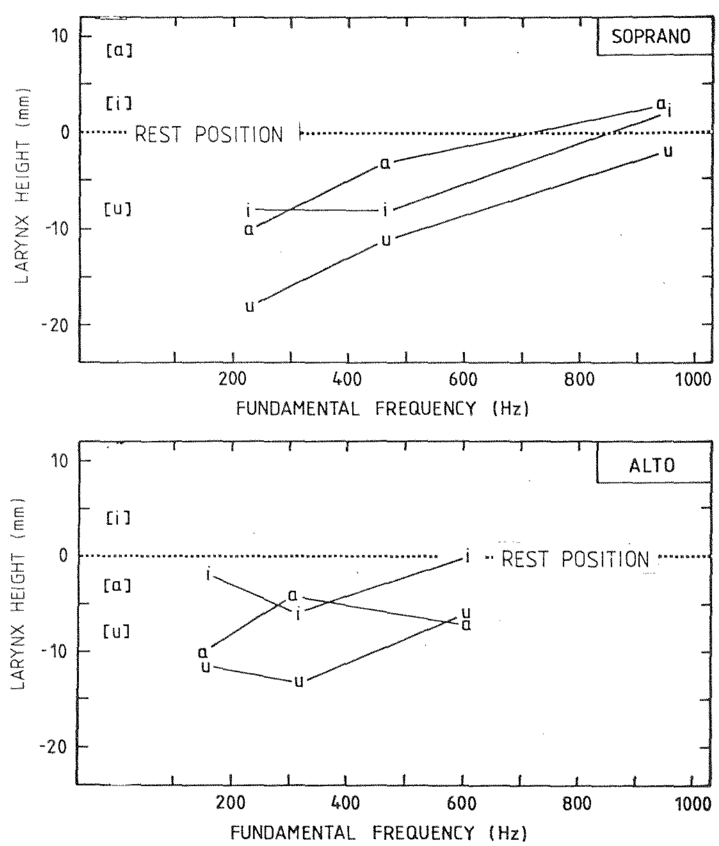


Fig. 2. The vertical position of the larynx as function of fundamental frequency in the two subjects. Vowel symbols within brackets pertain to spoken vowels, other symbols refer to sung vowels.

In a previous investigation of larynx height in singing Shipp and Izdebski (1975) found a lowering of the larynx that tended to increase with rising fundamental frequency in trained male singers. The data pertaining to our alto subject agree with this observation, while the opposite is true for the soprano. Therefore, the data were checked in the case of the soprano. Direct observation by means of fluoroscopy, when the soprano subject sang rising and falling triads, confirmed that she actually moved her larynx up and down in a harmoniously balanced synchrony with pitch.

The jaw opening was measured at the increase in distance between the

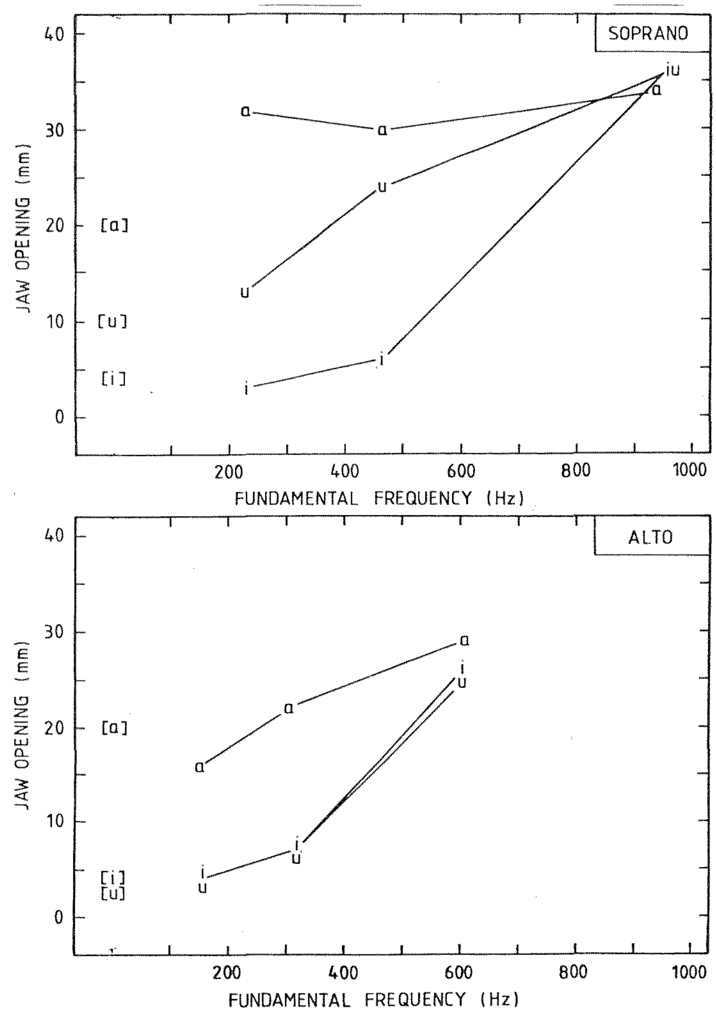


Fig. 3. The jaw opening as function of the fundamental frequency in the two subjects. Vowel symbols within brackets denote spoken vowels, other symbols refer to sung vowels.

upper and lower incisors beyond rest position, with clinched teeth. The data are shown as function of fundamental frequency in Fig. 3. In this respect both subjects behave similarly in that they increase the jaw opening with rising fundamental frequency. Thus, their jaw opening values for the top pitches are far beyond what is observed in their corresponding spoken vowels. This is in agreement with previous observations (Ondráčková, 1969; Sundberg, 1975).

The lip opening is instrumental in tuning the formant frequencies. Two measures were taken from the lip contours. The vertical distance between the upper and lower lip (h in Fig. 4) is shown as function of the jaw

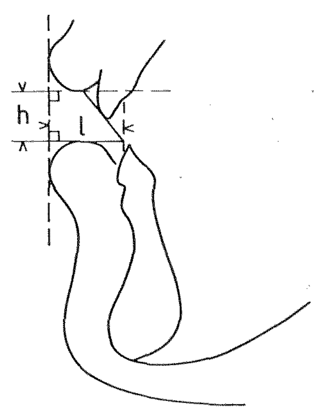


Fig. 4. Schematical illustration of the measurement of the vertical distance between the upper and lower lip, l, and of the retraction of the mouth corners, l.

opening in Fig. 5. It can be seen that this distance is highly dependent on the jaw opening, the correlation coefficient (linear regression) being 0.94 and 0.79 for the alto and the soprano, respectively. The weaker correlation in the case of the soprano reflects the fact that she reduced the vertical distance between her lips for all /u/-vowels and increased it for all /i/-vowels, except in the case of the top pitch. It seems that in this respect the alto adopts a more relaxed articulation than the soprano.

The distance between a line joining the frontmost contours of the lips and the contour of the lip corners was also determined (l in Fig. 4).

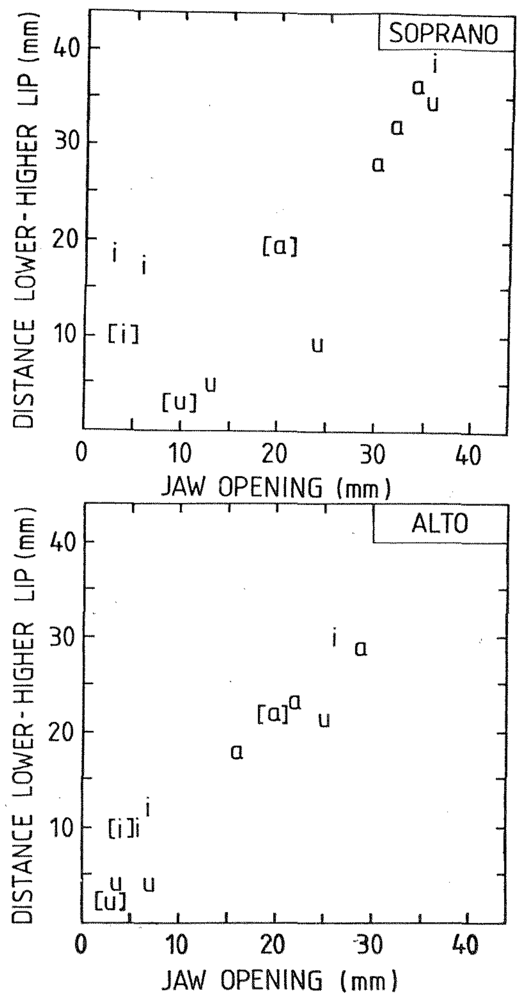


Fig. 5. The distance between the contours of the upper and lower lip as function of the jaw opening. Vowel symbols within brackets denote spoken vowels, other symbols refer to sung vowels.

When the radiograph showed two mouth corner contours because of parallax error, the average of the resulting two distances was used. The resulting measure, which we will call mouth corner retraction, is shown Fig. 6. Here again, the soprano seems to exert more articulatory activity than the alto; the alto's data points are all close to one single curve, while the mouth corners seem to be actively retracted in the soprano's /i/, again except for the highest pitch. In the case of the alto and in the case of the soprano's lower fundamental frequencies, the three vowels show different values of mouth corner retraction, while these differences almost disappear in the soprano's highest fundamental frequency.

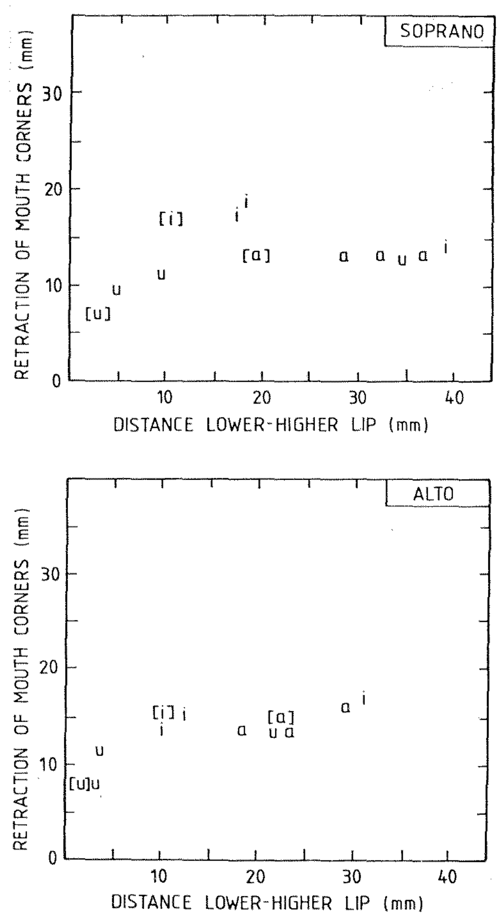


Fig. 6. Retraction of the mouth corners as function of the distance between the contours of the upper and lower lips. Vowel symbols within brackets denote spoken vowels, other symbols refer to sung vowels.

The tongue contours related to the lower jaw are shown in Fig. 7. The tongue contours are similar to those of the spoken vowels in the vowels sung at the lower fundamental frequencies. Also, both in the alto and in the soprano the trend illustrated in Fig. 7a can be observed, that the tongue shapes are slightly neutralized, as fundamental frequency is increased. At the soprano's highest pitch, the tongue contours for all three vowels are similar to the tongue contour of the spoken /o/, as can be seen in Fig. 7b. Thus, while the alto uses at least two distinct tongue shapes for the three vowels at her highest fundamental frequency, the soprano uses practically identical tongue shapes for all vowels when she sings the highest note.

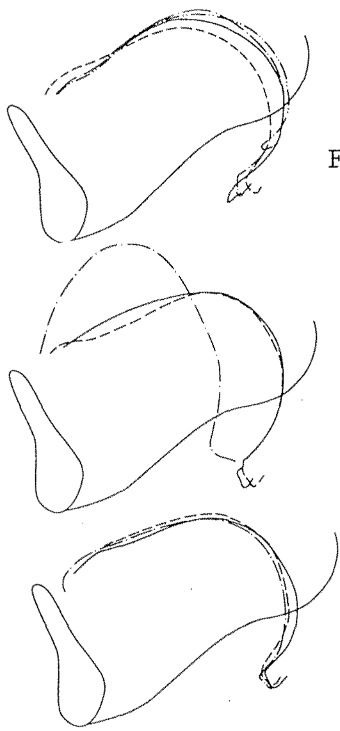


Fig. 7. Tracing of the midsagittal contours of the tongue body with the contour of the lower mandible as the reference. The upper group of .ongue shapes are from the soprano's /o/ sung at the fundamental frequencies of 230, 465, and 940 Hz (solid, dotted-dashed, and dashed contours). The double-dotted-dashed curve is from the spoken version of the same vowel. In the middle a group of tongue contours are from the alto's /o, i, u/ vowels (solid, dashed, and chain-dashed curves) sung at 610 Hz fundamental frequency. The lowest group shows the corresponding tongue shapes for the soprano's /o, u, i/ (solid, dashed, and chain-dashed curves).

B: Acoustical data

Area functions were derived from the radiographs using the procedure described previously (Lindblom and Sundberg, 1971) and schematically illustrated in Fig. 8. First, points along the midline in the midsagittal plane of the vocal tract were determined. These points were derived as mid-points between the two wall countours' intersections with the lines of grid of a semi-polar coordinate system. This system was anchored on contours of the hard palate and the cervical vertebrates, and was the

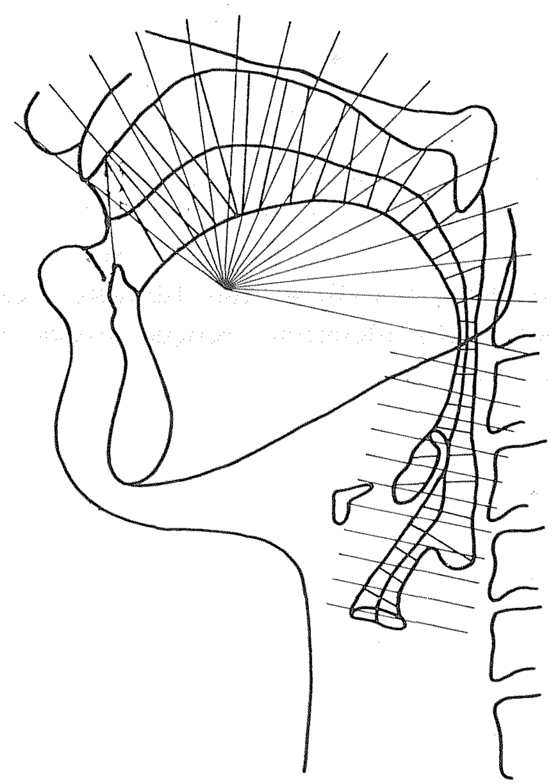


Fig. 8. Schematical illustration of the procedure used in order to estimate the dimensions of the vocal tract, see text.

same for all pictures. Then, these mid-points were smoothly joined and the resulting curve was regarded as a vocal tract midline. The distance between the vocal tract wall contours normal to this midline was determined at .5 cm intervals along the midline. The resulting mid-sagittal distances were then converted to cross-sectional area by means of two distance-to-area plots; one for the pharynx, and one for the mouth cavity. Lacking better data, the first-mentioned curve was the same as was used in a previous investigation (Sundberg, 1969; the possibilites of improving this procedure by means of recent development in radiology will be explored in a future investigation). The curve pertaining to the mouth cavity was derived from direct measurements on a plaster cast of the subjects' hard palates. Thus, for each radiograph an area function was obtained, i.e., a table containing estimates of the vocal tract cross-sectional area at each .5 cm along the mid-sagittal vocal tract midline.

The next step was to determine the formant frequencies associated with each of these area functions. This was realized by piling plexiglass discs, .5 cm thick, each with a center hole of accurately determined size. By chosing discs with center holes in accordance with a specific area function, the resulting pile of discs constituted an acoustic model of this area function. These models were then excited by a sinewave of variable frequency by means of the STL Ionophone (Fransson and Jansson, 1973), so that the formant frequencies could be determined.

As is well known from acoustic theory of voice production, an increase of the jaw opening will raise the frequency of the first formant in particular, because it narrows the open-closed vocal tract tube in the closed end and expands it in the open end. Fig. 9 shows that this is the case with the formant frequencies obtained from the procedures just described, eventhough the soprano's spoken /o/ constitutes an exception. The reason for this is that the tongue shape of the spoken /o/ in itself constricts the pharynx and expands the mouth cavity, regardless of the jaw opening. Thus, the data given in Fig. 9 support the assumption that the formant frequency data represent reasonably reliable information.

The formant frequencies derived from the radiographs are shown in Fig. 10. It is evident that the data points pertaining to the three lowest formants form an organized pattern. This again supports the assumption that the formant frequency data are reasonably reliable.

For low fundamental frequencies the first formant is kept similar to what it is in the spoken version of the vowel. However, it tends to rise as soon as the fundamental frequency is raised above the frequency value of the spoken vowel. In other words, the situation seems to have avoided that the fundamental is higher than the first formant. At the top pitch, the frequency of the first formant is slightly lower than that of the fundamental, particularly in the case of the soprano. However, it has been shown that in reality the first formant frequency is raised by

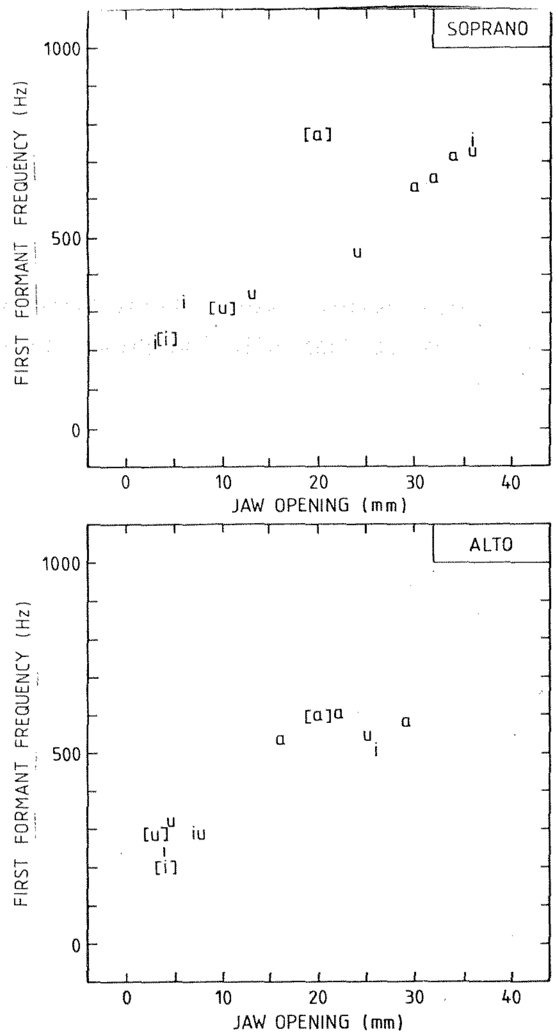


Fig. 9. Correlation between the frequency of the first formant and the jaw opening in the two singer subjects. Vowel symbols within brackets denote spoken vowels, other symbols refer to sung vowels.

yielding vocal tract walls, and this effect increases with frequency. Thus, our first formant frequency estimates must be too low at high frequencies such as the soprano's top pitch. We conclude that probably the soprano and the alto both tune their first formant frequency to the vicinity of the fundamental, as soon as the normal value of the first formant would imply that the fundamental was higher than the first formant.

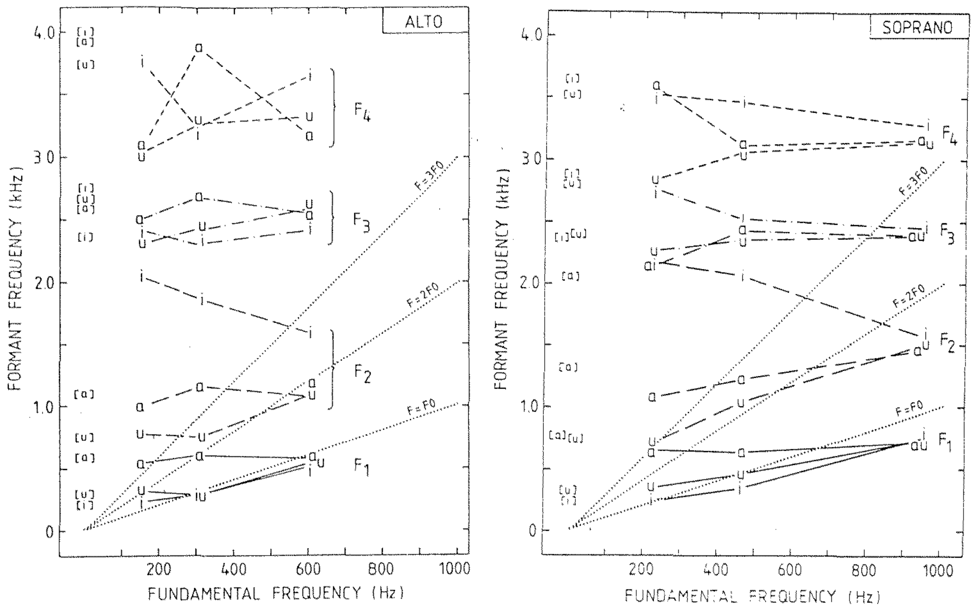


Fig. 10. Estimates of the frequencies of the first four formants as function of the fundamental frequency in the two singer subjects. Vowel symbols within brackets denote spoken vowels, other symbols refer to sung vowels. Symbols falling close to the dotted lines suggest that the formant frequency was tuned to the vicinity of the frequency of a harmonic partial.

With rising fundamental frequency, the second formant frequency rises in the back vowels /o/ and /u/ and falls in the front vowel /i/ in both singers. In the case of the soprano this leads to the situation that the second formant frequency is practically identical for all three vowels at the top pitch.

The frequency of the third formant seems rather unaffected by pitch. The fourth formant frequency data seem unsystematic, even though they look strikingly similar to the larynx height data in the case of the alto.

Discussion and conclusions

With respect to the first and second formant frequency of the various vowels sung at the different fundamental frequencies, our data are in good agreement with the results from the previous investigation of a soprano singer (Sundberg, 1975). This is interesting, given both the uncertainty of and the substantial mothod differences between the two investigations. Moreover, the subject used in the previous investigation was not identical with any of the subjects in the present study. Regarding the third and fourth formants, the results do not agree with those of the previous study. However, the data concerning these formants cannot be very accurate in the present investigation. The reason for this is that with rising frequency, the formants become increasingly sensitive to details of the area function. For this reason these data do not merit any interpretation. We conclude that our data on the articulation and on the two lowest formant frequencies are reliable and that they possess a certain degree of generality.

The articulatory maneuvres underlying the tuning of formant frequencies involve jaw opening, larynx height, and tongue shape. The most important one would be the jaw opening, as both subjects were found to change this articulatory parameter with pitch; also, similar findings have been reported in previous studies of female singers (Ondrackova, 1969; Sundberg, 1975). The point of tuning the first formant to the vicinity of the fundamental in high pitched singing is a gain in SPL, which may be quite considerable (cf., Sundberg, 1982). In high pitched singing it implies the need of raising the first formant to very high values.

The finding that the soprano raises her larynx with increasing fundamental frequency is interesting, since it is in clear opposition to the general opinion among many singing teachers, who claim that a rise of the larynx is harmful to the voice. Still, our soprano subject has been a successful professional solo singer for a number of years. We must conclude that an elevation of the larynx is not necessarily harmful to the voice (cf., Askenfelt and Sundberg, 1981).

The pitch dependent larynx height is also in clear opposition to previously published data regarding trained male singers (Shipp and Izdebski, 1975). However, from an acoustical point of view a soprano has good reasons for raising her larynx with fundamental frequency. Since she has to sing at very high fundamental frequencies, thereby preferbaly tuning her first formant to the vicinity of the fundamental, she has an extreme need of raising her first formant frequency. An elevation of the larynx shortens the vocal tract and, thus, adds to the possibilities of raising the frequency of the first formant.

As a consequence of the formant strategy observed, the acoustical vowel differentiation is reduced with rising pitch. Thus, particularly in the soprano's highest note, which, incidentally, is by no means the highest pitch that a soprano may have to sing, the formant frequencies of the three vowels become very similar. This is also in agreement with the articulatory data; with increasing pitch the articulatory similarity between the vowels increase, so that, in fact, the soprano uses almost the same articulation for all vowels in the case of the highest fundamental frequency (see Figs. 1, 2, and 3). It may be noted that, from the point of view of vowel quality perception, this strategy is harmless (Sundberg and Gauffin, 1982).

Even though there are reasons to assume that the data presented here represent reliable and representative information, it is evident that they do not allow any inferences regarding typical differences between alto and soprano voices. However, we would like to speculate that, apart from the obvious factor of vocal fold morphology, the pharynx length may be relevant. One support for this speculation is that the pharynx length is varied systematically in the soprano, viz. by means of larynx height. The typically "darker" vowel quality of an alto singer would arise from certain formant frequency combinations; the lower second formant in /i/, and the lower first formant in /o/ would lead to a timbral difference of this type. These two formants in these specific vowels are both particularly sensitive to the dimensions of the pharynx. The difference in the larynx height data between the subjects may be of relevance in this connection. Also, typical differences in the higher formants are likely, which have not been revealed in the present investigation (see Dmitriev and Kiselev, 1979).

References

Askenfelt, A. and Sundberg, J. (1981): "Larynx height and voice source. A relationship?" STLPSR 2-3/1981, 23-36.

Dmitriev, L. and Kiselev, A. (1979): "Relationship between the formant structure of different types of singing voices and the dimensions of the supraglottal cavities", Folia Phoniatrica 31, 238-241.

Fant, G. (1960): Acoustic Theory of Speech Production, Mouton, The Hague.

Fransson, F. and Jansson, E. (1973): "The STL-Ionophone: Transducer properties and construction", J.Acoust.Soc.Am. 58, 910-915.

Lindblom, B. and Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw, and larynx movement", J.Acoust.Soc.Am. 50, 1166-1179.

Ondráčková, J. (1969): "Some remarks on the analysis of sung vowels", pp. 407-418 in Study of Sounds XIV, Phon. Soc. of Japan.

Shipp T. and Izdebski, K. (1975): "Vocal frequency and vertical larynx positioning by nonsingers and singers", J.Acoust.Soc.Am. 58, 1104-1106.

Sundberg, J. (1975): "Formant technique in a professional female singer", Acustica 32, 89-96.

Sundberg, J. (1982): "Perception of singing", pp. 59-98 in (D.Deutsch, ed.) Psychology of Music, Academic Press, New York.

Sundberg, J. and Gauffin, J. (1982): "Amplitude of the voice source fundamental and the intelligibility of super pitch vowels", pp. 223-228 in (R.Carlson & B.Granström, eds.) The Representation of Speech in the Periferal Auditory System, Elsevier Biomedical Press, Amsterdam.

VOCAL TRACT GESTURES IN SOPRANO AND BASS:
a xeroradiographic-electro-laryngographic study

F. MacCurtain and G. Welch
The Middlesex Hospital, London, Great Britain

Introduction

The vocal tract appears to function as a web of interconnecting muscles. The tract extends from the lower borders of cervical vertebra seven to the base of the skull (spheno-occipital bone). The entire region contains pharyngeal musculature.

Singers demonstrate that the pharyngeal cavities are capable of continually changing shape. The present study investigates how these changes in the pharyngeal cavities contribute to contrastive registers, and whether these changes are related to, and may influence, laryngeal adjustments.

In this preliminary study, two voice types have been observed: soprano and bass.

Theory

Research evidence supporting this viewpoint is provided by: Sonninen (1970) who points out that the muscles of the external frame function in the control of pitch, and Sundberg (1974) who demonstrates the correlation in singers between formant frequencies and pharyngeal resonators. Complementary findings are also reported on the adjustments of the larynx (Gauffin, 1977) in voice, and vocal tract constrictions (Fujimura, 1977).

It is known that the intrinsic laryngeal muscles contribute to the excitation of the airstream. Vennard and Hirano (1973) suggest that the extrinsic laryngeal muscles may support the intrinsic laryngeal muscles as they change in mode of vibration (i.e., as required for differences of 'timbres'). This would indicate that the extrinsic laryngeal muscles can affect both resonator shapes and the mode of vibration of the folds (MacCurtain, 1982).

Work on voice disorders has shown that when excess musculo-skeletal tension arises, the quality of the sound becomes distorted (Aronson, 1980) and the singer may experience pain in the region of the thyroid cartilage and/or hyoid bone, even though no abnormality may be detected by otolaryngological examination. The distortion may be manifest both in the aberrant mode of vibration of the vocal folds and in the aberrant shape of the resonators (Berry et al., 1982). There is some evidence that inadequate air pressure may also be involved (Iwata et al., 1972).

The hypothesis that the whole of the vocal tract may be involved in voice production has been subjected to empirical evaluation by MacCurtain (1982). A technique has been developed whereby certain changes in the soft tissue of both parynx and larynx could be displayed. The technique involved the combination of both xeroradiography and electro-laryngography (electro-glottography).

Technique

Xeroradigraphy-electro-laryngography (XEL) (MacCurtain, 1982)

Two known techniques have been combined from Oncology and Experimental Phonetics. Xeroradiography of high quality has been recorded simultaneously with acoustic and electro-laryngographic signals.

Xeroradiography (Boag, 1973) is a radiographic imaging technique originally designed for mammography in which the conventional silver coated film is replaced by a statically charged, selenium coated plate. The photostatic image has an increased exposure latitude, i.e., good resolution in bone and soft tissues on the same image (Julian et al., 1981)[1].

Xeroradiography provides good resolution of bone, soft tissue and airway on the same image (e.g., the ventricular space can be measured in relation to cervical vertebra). Another property of the final image is that contrast boundaries are edge enhanced. This increases the visualisation of low contrast areas (e.g., aryepiglottic folds are seen to change gesture from low to high notes).

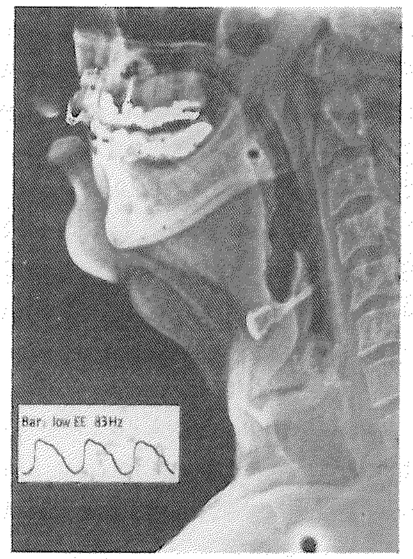


Plate 1

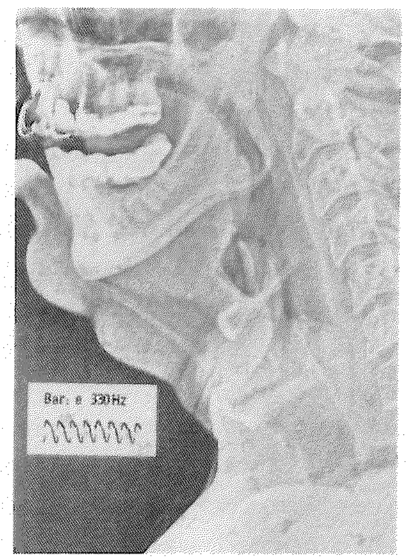


Plate 2

[1]Extensive work was undertaken to reduce the dosage (Bryant and Julian, 1978; Noscoe, 1980): with the use of a copper filter .2 mm thick backed by an aluminium sheet, radiation was reduced by 30%. The Ethical Committee of The Middlesex Hospital Medical School has sanctioned the use of the technique under the International Code of Practice (1982) for examining subjects referred by Otolaryngologists. The radiation dose of each lateral xeroradiograph is 3 mGy (.3 rad). This dosage may be associated with negligible risk: the working limit for radiographers is 2 Gy (2 rad). Dr. G.M. Adran, Consultant Radiologist, Radcliffe Infirmary, Oxford and Chairman of the National Committee of Radiological Protection, is in charge of the safety measures. The technique, however, provides quantitative information on antero-posterior dimensions of the airway, being lateral views of the vocal tract. The frame includes the spinal column, upper naso-pharynx (excluding the orbits of the eye), lips and lower cervical vertebra seven.

There are specific disadvantages which we have not succeeded in overcoming: the static nature of the imaging (it is not possible to retain the sharp outlines of soft tissue in dynamic techniques); and the long exposure time: 0.1 seconds (e.g., even in the lowest note of the Bass, seven vibratory cycles of the folds have occurred during the one exposure).

Electro-laryngography (Abberton and Fourcin, 1971) operates on the basis of a constant voltage monitoring of translaryngeal electrical conductance. Gold-plated guarding electrodes are placed on each side of the thyroid cartilage. On one wing of the thyroid cartilage a constant voltage is applied at a frequency of 2 milliamps whilst, on the opposing wing, an identical electrode arrangement is connected to a low impedance current detecting circuit. The equipment is electrically self adjusting for each subject and electrode jelly is not required, (MacCurtain and Fourcin, 1982). The electro-laryngograph, coupled to an oscilloscope, provides a waveform which relates to the pattern of soft issue contact at vocal fold level, termed L_x (electro-glottogram: E.G.G.).

The disadvantages of the technique include the problem of positioning the electrodes in the same spot each time for repeatability studies, the fact that certain necks are not suited, the wide variation that is to be found within normal wave forms (Tulp and Schutte, 1982), and the non-quantitative aspects of the findings.

Although the electro-laryngographic wave form is not a unique indicator of disorder (Wechsler, 1977) it may contribute to a more comprehensive diagnosis, e.g., the auditory impression of hypernasality may mask excessive subglottic coupling (as in breathiness). The ear picks out the nasality but is deceived as to the breathiness: the laryngographic display gives measurable information on the relationship of open to close phase in each cycle. The above two techniques: xeroradiography and electro-laryngography have been combined to display soft tissue changes throughout the vocal tract. The technique is known as XEL[2].

Method

During the routine clinical trials at The Middlesex Hospital, it was noted that 'normal' informants (who acted as control for the voice disorders) included some professional singers. The XEL technique revealed that the singers exhibited clear changes in soft tissues formation when moving the pitch of their voices from low to high. Accordingly, it was decided to investigate these changes more thorougly, utilising the XEL findings in conjunction with van Deinse's (1982) theory on vocal registers.

Over the past twelve months, a range of voice types has been investigated. For the purposes of the present study, however, it was decided to focus on register contrasts in two of these: Basses and Sopranos.

Subjects

Nine sopranos and five basses were examined. The qualification for inclusion in the study was that each subject should have a minimum of four years professional training and should currently be earning their living as a professional singer. The ages of the subjects range from 22-45 years.

Procedure

Van Deinse's classification of vocal registers (1982) indicates that there are several vocal registers for each voice type. He indicates that

[2] The XEL technique was designed, tested and put into clinical trial at The Middlesex Hospital Medical School, London. It is now in routine clinical use in the Ear, Nose and Throat Department and Speech Therapy Department (1978-83). To date, 480 subjects have been examined, including 150 normal informants and 80 larynectomees. Five Regional Centres in the UK plan to start their own XEL units in 1984.

female voices (e.g., sopranos) may have four registers, i.e., 'chest', 'head', 'little' and 'whistle'. These range from approximately 150 Hz at the lowest extremity, to over 2000 Hz at the highest. The male subjects (e.g., basses) have two registers, i.e., 'chest' and 'falsetto', ranging from approximately 64 Hz to 750 Hz.

He states that the vocal registers are a product of two factors, i.e., the interaction of the muscles with the airstream, and the change in the coupling mechanism of resonators and vocal folds. The combination of these factors produces registers of the 'first order'. In addition, a third factor may be operating as within these main registers are the registers of the 'second order'. These second order registers are a product of supplementary action by the muscles themselves, e.g., as in the transition from chest to falsetto register, or in the action of 'covering' the voice (due to the action of the cricothyroid muscle).

Frequencies were selected for vocalization in accordance with three constraints:

1) Ten pictures only were permitted for each subject (Ethical Committee requirement)

2) Each type of register should be included (wherever possible)

3) Different voice types could be contrasted in the singing of the same frequency

Accordingly, the following frequencies were selected: Fig. 1 Sopranos, Fig. 2 Bass. Each subject sang eight pitches[3] to the vowel [i] (this vowel allowed maximum space in the pharyngeal cavity.

---

[3] The use of the term 'pitches' rather than 'frequencies' is in accordance with the evidence that pitch is a psychological phenomenon (e.g., Rostron, 1976).

Fig. 1

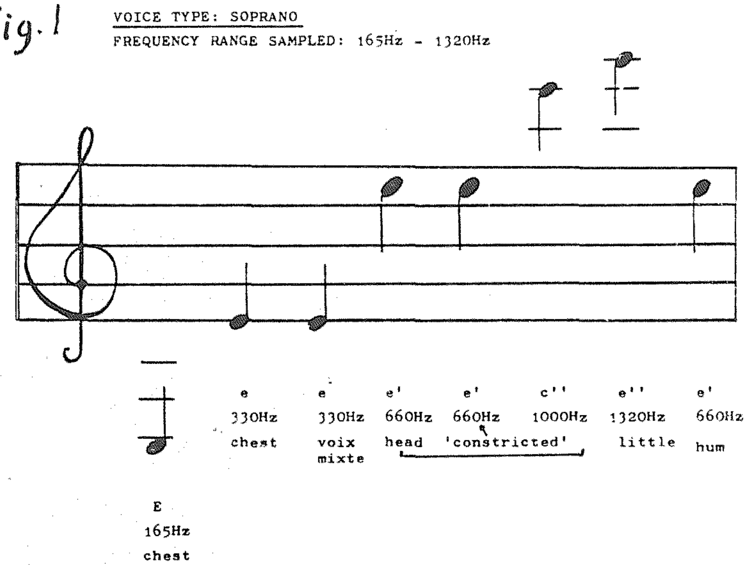
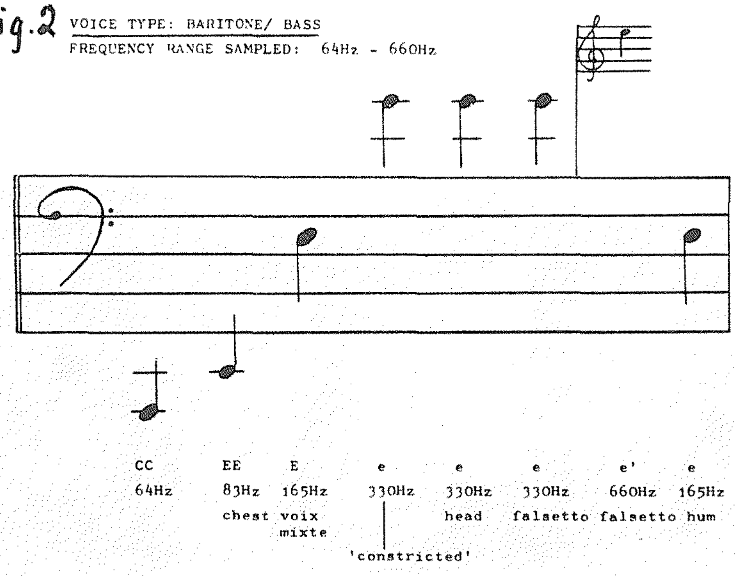


Fig. 2



The soprano pitches ranged from 165 Hz, E3, in the ´chest´ register to 1320 Hz, E6 in the ´little´ register (see Fig.I). The ´whistle´ register was not investigated as the majority of soprano subjects found this inaccessible.

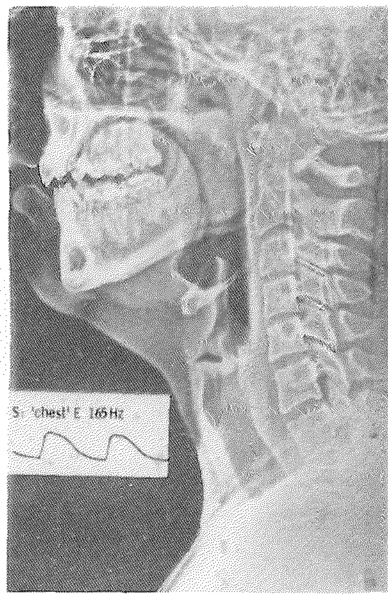


Plate 3

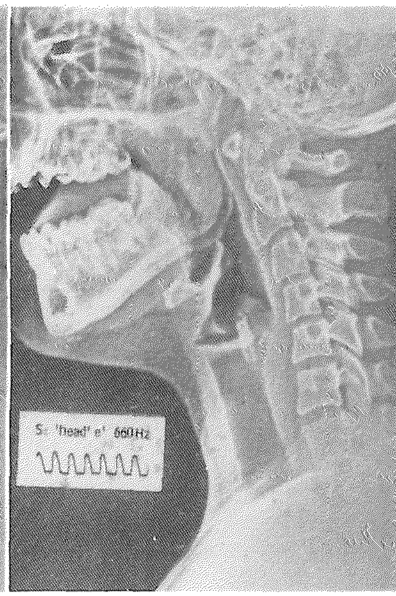


Plate 4

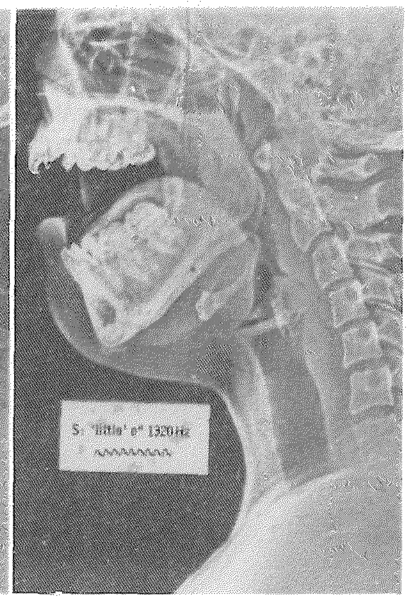


Plate 5

The bass pitches ranged from 64 Hz, C2, in the low 'chest' register to 660 Hz, E5, in the falsetto register. (The male sample included subjects who might be classified as baritones, as well as basses).

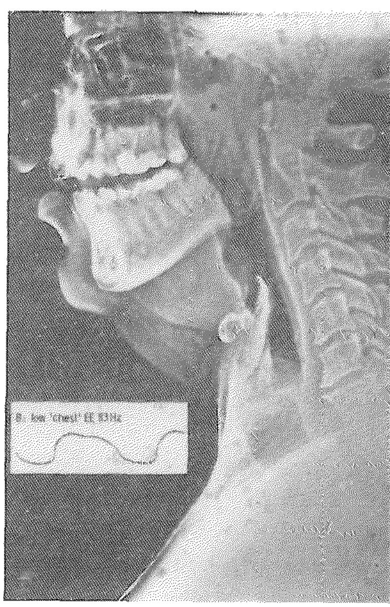


Plate 6

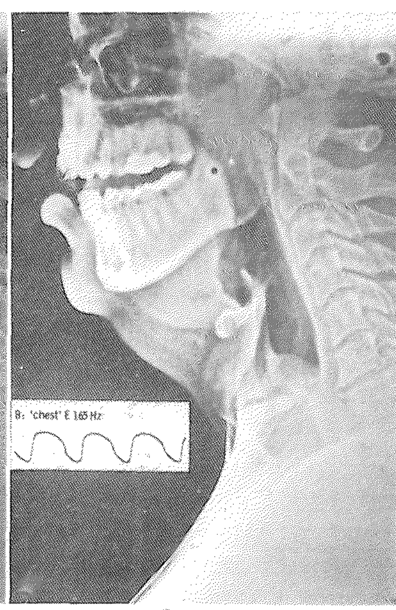


Plate 7

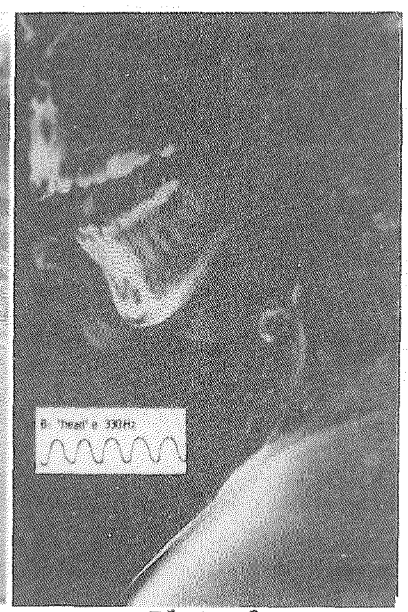


Plate 8

The pitches were sung by each respective voice type in the order given in Figs. 1 and 2.

Each pitch was sung twice; one for each part of the XEL technique. First, the subject was connected to the electro-laryngograph and a polaroid photograph was taken of the L_x wave form on the oscilloscope screen during vocalization of each of the eight pitches. A simultaneous audio recording was made onto a cassette recorder.

Secondly, xeroradiographic images were taken of each sung response. It was not possible to take xeroradiographic recordings and electrolaryngographic recordings simultaneously because the electrodes obfuscate the laryngeal area.

As well as the eight pitches, we filmed the subject at rest, to provide a base line against which soft tissue changes could be measured. The tenth picture was a repeat note. Each sound was sung mezzo forte. The pitch of the sung note was checked against the frequency measurement derived from the electro-laryngographic trace and aurally by the authors.

The xeroradiographs were mea. red at six points to obtain dimensions in millimeters. These points have been selected as being the most significant from an original descriptive framework of twelve points (MacCur-

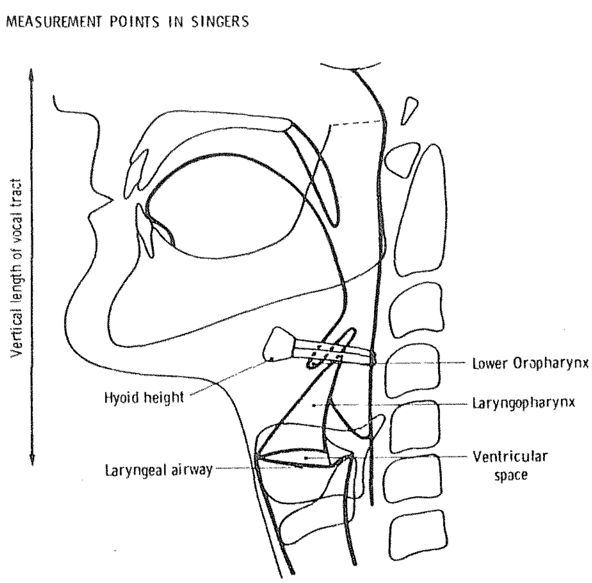


Fig. 3. These six recognisable anatomical parameters are measured at the following levels:
1. Laryngeal airway: anterior commissure to posterior of arytenoid lumen - posterior cricoid tip.
2. Laryngopharynx: epiglottic rib to pharyngeal line, at the level of mucous covering of arytenoids.
3. Lower oropharynx: anterior of vallecula to pharyngeal line.
4. Hyoid height: vertical height measured from supero-posterior border of hard palate to lower inferior border of hyoid body.
5. Vertical length of vocal tract: spheno-occipital bone to superior border of vocal folds.
6. Ventricular space: supero-inferior dimension at widest point.

tain, 1983), which derived from a study of 400 subjects (MacCurtain, op cit).

The mean dimension of the six anatomical points was computed for each group of singers. The final mean figures presented below make allowance for a magnification factor on the xeroradiograph of 1.18.

Results

Tables 1 and 2 show the mean measurements of the six significant parameters for the vocalizations in each register.

Table 1. MEAN DIMENSIONS (LENGTH IN MM) OF 6 SIGNIFICANT PARAMETERS OF THE VOCAL TRACT

VOICE TYPE BASS/BARITONE N = 6

	Hz	Ventricular space	Laryngeal airway	Laryngo-pharynx	Lower oropharynx	Hyoid height	Vertical length of tract Partial	Full
1 At rest		2.1	33.2	22.2	21.4	81.4	108.9	127.7
2 Middle E	165	4.0	37.5	23.5	29.9	88.0	113.7	129.2
3 Low EE	83	2.2	33.7	15.7	22.9	93.0	113.6	133.2
4 Low CC	64	1.4	32.6	13.4	18.8	93.0	114.4	132.8
5 Falsetto e	330	3.4	39.6	23.1	29.2	86.6	109.3	118.5
6 Constricted e	330	1.1	38.0	21.2	30.6	85.3	106.9	120.6
7 head e	330	2.5	40.0	23.1	30.0	93.2	117.4	131.7
8 Falsetto scp e'	660	1.0	40.4	18.8	27.3	86.4	107.4	123.7
9 Hum Middle E	165	2.7	38.6	23.6	26.9	90.7	115.9	135.1

Adjusted for magnification factor

Table 2. MEAN DIMENSIONS (LENGTH IN MM) OF 6 SIGNIFICANT PARAMETERS OF THE VOCAL TRACT

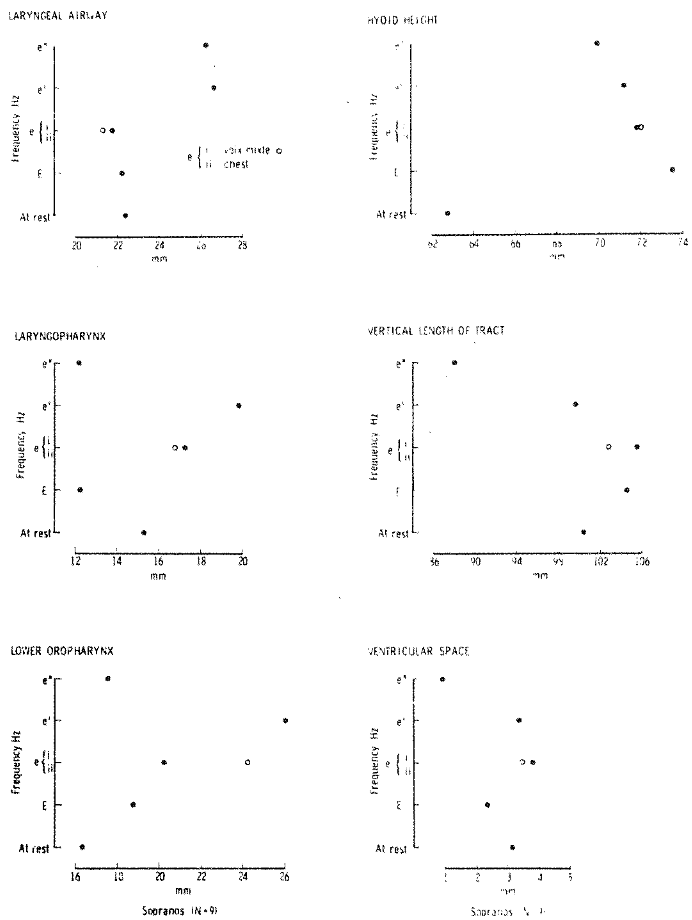
VOICE TYPE SOPRANO N = 9

	Hz	Ventricular space	Laryngeal airway	Laryngo pharynx	Lower oropharynx	Hyoid height	Vertical length of tract Partial	Full
1 At rest		3.2	22.4	15.3	16.3	62.9	83.3	100.3
2 E	165	2.5	22.2	12.3	18.9	73.5	87.2	104.4
3 Chest e	330	3.9	21.9	17.3	20.2	71.7	88.6	105.8
4 Voix mixte e	330	3.6	21.3	16.9	24.4	71.8	86.4	103.6
5 Constricted e'	660	1.6	26.4	15.0	20.3	66.1	76.8	93.7
6 Head e'	660	3.3	26.6	19.7	26.0	71.1	83.0	99.7
7 c''	1000	1.1	26.6	14.2	18.9	69.3	76.2	90.6
8 e''	1320	0.9	25.8	12.1	17.0	69.7	76.3	88.9
9 hum e	660	2.5	26.3	15.9	19.2	63.6	78.3	96.1

Adjusted for magnification factor

The data may be displayed in graphic form, and for ease of presentation only the pitch E3 (165 Hz) and its octave displacements are indicated.

Table 3.



In the Soprano group, the following characteristic trends seem to be evident:

(a) Laryngeal airway: There is an excursion of 4 mm associated with octave displacement. The airway remains relatively unchanged until 'head' voice is entered, when the airway opens out in a posterior excursion of 18%. This length is maintained in the 'little' register: one octave higher.

(b) Laryngopharynx: In moving from resting position to low 'chest' the laryngopharyngeal size decreases. In moving from low 'chest' to upper 'chest' and 'voix mixte' (330 Hz), there is a corresponding increase in the antero-posterior dimension. This direction of excursion is maintained over the next octave. In 'little' register (1320 Hz), however, the laryngopharyngeal dimension returns to the narrow opening manifest in lower 'chest'.

(c) Lower oropharynx: There is a continuous widening from the lowest sung pitch up to head, followed by a significant contraction from 'head' to 'little'. It should be noted that there is a marked contrast in dimension between two notes sung at the same frequency of 330 Hz (E4), i.e., 'voix mixte' and upper 'chest' timbres.

(d)(e) Vertical length of tract and hyoid height: As frequency increases, there is a corresponding shortening in the length of the vocal tract (allowing for some minor adjustment in 'voix mixte'). The shortening of the tract is most marked in the transition from 'head' to 'little' registers. In the soprano gestures, the vertical height of the hyoid appears to be in close correspondence with the overall length of the tract.

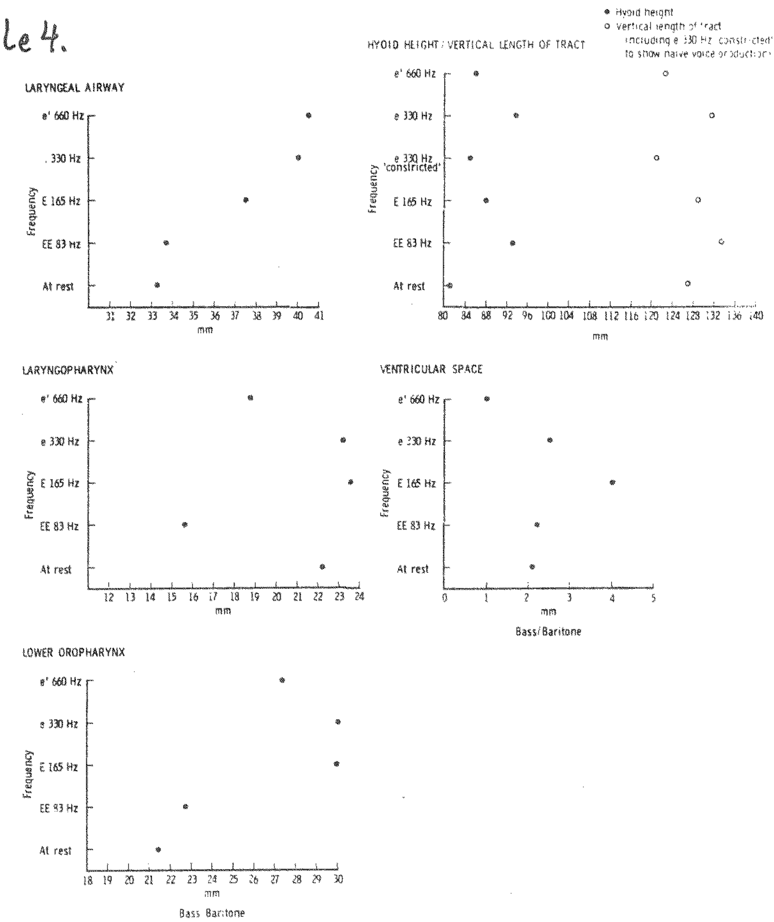
(f) Ventricular space: There would seem to be some changes in ventricular space with octave displacement.

In the Bass group, the following characteristic trends seem to be evident: See Table 4 on the following page.

(a) Laryngeal airway: Extension occurs with the increase in pitch with the largest excursion occurring in the transition between low 'chest' → 'voix mixte', and from 'voix mixte' → 'head'.

(b) Laryngopharynx: In moving from resting to the low 'chest', a contraction of 32% occurs. Transposition up an octave shows a significant opening out: an antero-posterior excursion of 53%. This wider opening

Table 4.



appears to be maintained over the next octave, i.e., 'voix mixte' —> 'head'.

(c) Lower oropharynx: This region follows a similar patterning to the laryngopharynx, but without the initial narrowing in the low 'chest'.

(d) Vertical length of tract/hyoid height: In moving from resting to the lowest note, the vocal tract shortens. In singing up an octave, the vocal tract lengthens (at 165 Hz, E3), and goes on lengthening in 'head'/ 'covered' register (330 Hz, E4). However, when the note (330 Hz) was sung in a 'naive' (i.e., untrained) manner the tract became shortened, (c.f., graph no 4d).

With professional use of the voice at 330 Hz (E4) not only does the vocal tract lengthen, but there is an associated change in the angulation of the vocal folds, i.e., a downward tilt of the anterior prominence of the thyroid cartilage occurs (associated with measurable contraction of the crico-thyroid vizor, see Ardran and Kemp, 1967). This evidence of increase in vertical length of the tract and tilting of the larynx appears to substantiate the notion of 'covering' at the top of the 'chest' register (e.g., van Deinse, 1982).

(e) Ventricular space: There would appear to be some changes in the ventricular space associated with octave displacement similar to those found in the sopranos. In both voice types the greatest space appears to be in the middle of the voice, i.e., 'voix mixte'.

Electro-laryngographic findings:

The findings are related to soft issue contact variations (L_x) The evidence is measurable in one aspect only: frequency rate of adduction/abduction (c.f., XEL prints 3-8).

At a descriptive level, one may ask whether the same note sung at equal decibel count may produce a difference in wave form if the 'timbre' is contracted: e.g., the Soprano (PD) singing at 330 Hz (E4), once in 'voix mixte' and once in 'chest'.

(See plates 9 and 10 on the following page.)

The evidence suggests that there is a difference in soft tissue contact variation at the sound source, relating to the contrasting voice quality. (The resonators measurably change shape, enhancing the harmonics required in each quality.)

It is worth noting that the open phase of the singing cycle tends to be brief. In speaking, the normal open phase = 40% of the total cycle. In singing, the open phase is kept to a minimum to reduce subglottic damping.

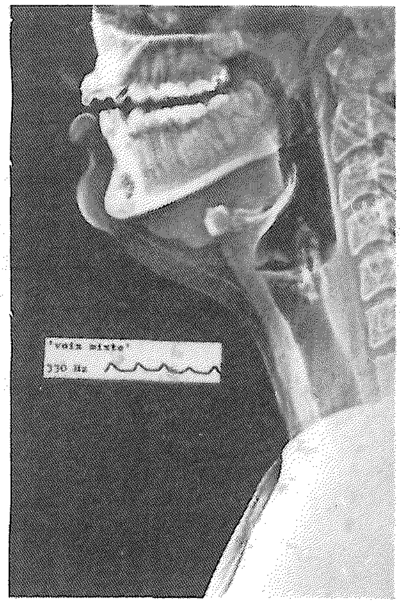


Plate 9

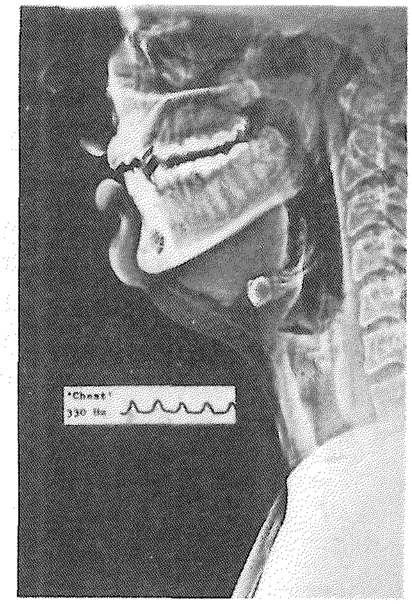


Plate 10

Repeatability

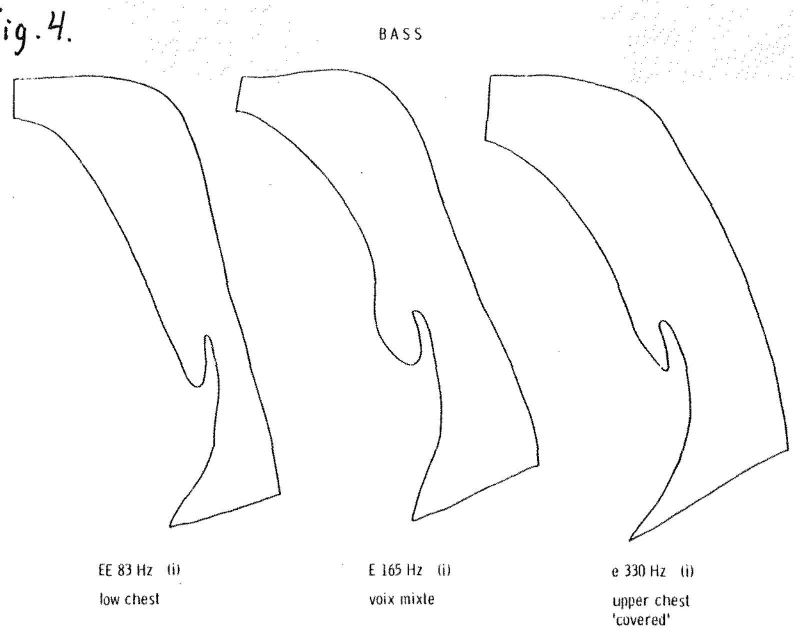
In both pilot study (five professional and five amateur singers) and present study (nine sopranos and six basses), each informant showed high repeatability in use of muscular patterning to produce a particular 'timbre' on the one note. Within the group of English professional singers, however, large intrasample variability has been found. The one auditory target may be achieved by different muscle strategies. For instance, the xeroradiographic evidence demonstrates a considerable range of muscular patterning to achieve the common auditory target of Top C (1000 Hz). In 20% of the sopranos, the hyoid bone overlays the thyroid cartilage descending lower than the level of the vocal folds. However, despite the anatomical variations and individual exploitation of the framework available (in both hard and soft tissue), there appears to be evidence of characteristic trends in the soft tissue parameters outlined above.

Discussion

Vocal tract gestures have been measured in Bass and Soprano voice types, following van Deinse's classification of vocal register. Characteristic trends have been observed, despite large intra-sample variability. Specific muscle strategies are involved, which differ from 'chest' to 'head' (to 'little').

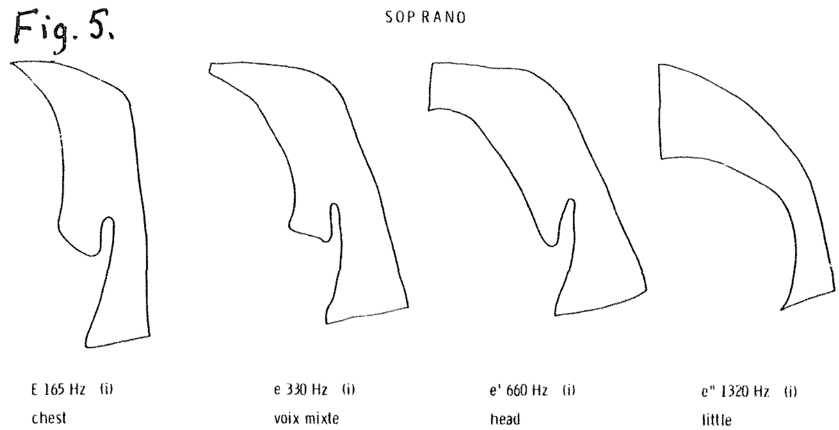
In 'chest' register, the vocal tract progressively widens with increase in pitch (found in both Bass and Soprano). Upper 'chest' in Bass (following van Deinse's classification 330 Hz (E4) with 'head/covered' quality) is produced by a change in muscle strategy. The larynx tilts by 20º of arc, the hyoid drops by 10%, the tract lengthens and the spine flexes (all measurements taken in relation to resting position, c.f., Ardran and Kemp, 1967). (See Plates 2 and 8.) This gesture is not employed by sopranos.

Fig. 4.



In moving from 'head' (660 Hz, E5) to 'little' (1320 Hz, E6) the Soprano produces yet another muscle patterning. There is a narrowing and shortening of the tract, a laryngeal tilt, and extreme flexion of spine (see Plates 4 and 5). This gesture is not used by the Bass singer, even in 'falsetto' range. (See Fig. 5 on the following page.)

Fig. 5.



'Covering': Upper 'chest' (van Deinse) (330 Hz, E4) is produced by a gesture distinct from other 'chest' notes, and might therefore be considered as belonging to a separate category. Whatever the classification used, at 330 Hz the Bass adopts that physiological gesture associated with 'covering'. The tract lengthens and there is a measurable contraction of the crico-thyroid vizor. Both hyoid and larynx tilt (see Plates 7 and 8). There appears to be no corresponding change of gesture at the equivalent part of the Soprano voice. If the Soprano does 'cover' at the top of the 'head' register, she does not appear to use the same strategy.

It remains to be seen if these findings are applicable to a larger sample, and whether the trends outlined above are indeed true indicators of vocal register for these voice types.

Conclusions

Bass and Soprano demonstrated vocal tract gesture in octave displacement. Characteristic trends have been observed, despite large intra-sample variability. For example, both Bass and Soprano progressively widen the vocal tract in moving from 'chest' to 'head' register.

However, in 'head' register the muscle strategies differ: the Bass (at 330 Hz, E4) tilts the larynx with 20° angulation of arc and lengthens the tract. The Soprano does neither. As the Soprano moves from 'chest' (at 660 Hz, E5) to 'little' (at 1320 Hz, E6) yet another muscle strategy is involved: extreme spinal flexion with narrowing and shortening of vocal tract. The Bass has no equivalent gesture.

Each register requires a different muscle strategy which involves the entire vocal tract.

Acknowledgements

This work is funded by the Cancer Research Campaign. The authors gratefully acknowledge the members of the interdisciplinary team: together we designed and developed the technique at The Middlesex Hospital Medical School and The Middlesex Hospital: Professor R.J. Berry and Mr. N.J. Noscoe (Oncology), Mr. M. Bridger (Otolaryngology), Mrs. Ruth Epstein (Speech Therapy), Dr. N. Brown (Nuclear Medicine), and from University College, London: Professor A.J. Fourcin (Experimental Phonetics). We are indebted to the singers who gave their time and patience as we worked together over the past years.

References

Abberton, E. and Fourcin, A. (1971): "A new laryngograph", Med.Biol.Illus., 21, 172-182.

Ardran, G.M. and Kemp, F.H. (1967): "The mechanism of the larynx in the epiglottis and closure of the larynx", Brit.J. Radiology, 40, 372-389.

Aronson, A. (1980): Clinical Voice Disorders, Thieme Stratton Inc.

Berry. R.J., Epstein, R., Fourcin, A.J., Freeman, M. MacCurtain, F., and Noscoe, N. (1982): "An objective analysis of voice disorders", Parts I and II, Brit. J. Disorders of Communication, 67-85.

Boag, J.W. (1973): "Xeroradiography: a review article", Physics of Medicine and Biology, 18, no. 1.

Bryant, T. and Julian, W. (1978): "Reduction of radiation dose to patients in xeroradiography", Brit.J. Radiology, 51, 974-980.

Code of Practice for Protection of Persons Against Ionizing Radiation Arising from Medical and Dental Use (1982), HMSO, London.

van Deinse, J. (1982): "On vocal registers", J.Res. in Singing, 5, 33-49.

Fujimura, O. (1977): "Control of the Larynx in speech", in The Larynx and Language (eds. G.Fant & C.Scully), Phonetica, 34, 300-305.

Gauffin, J. (1977): "Mechanisms of larynx tube constriction", in The Larynx and Language (eds. G.Fant & C.Scully), Phonetica, 34, 307-309.

Greene, M. (1981): The Voice and its Disorders (3rd ed.), Pitman Medical Books.

Hardcastle, W.J. (1976): Physiology of Speech Production, Academic Press.

Iwata, S., van Leden, J.H., and Williams, D. (1972): "Airflow measurement during phonation", Brit.J. Disorders Communications,. 5, 67-79.

Julian, W., MacCurtain, F., and Noscoe, N. (1981): "Anatomical gestures in contrastive voice quality", J. Physiol.

MacCurtain, F. (1982): "Pharyngeal factors influencing voice quality", Unpubl. Ph.D. thesis, University of London.

MacCurtain, F. (1983): "Pharyngeal gestures in contrastive voice qualities", pp. 199-213 in Abstracts of the Tenth Int.Congr. of Phonetic Sciences, Vol. IIA (eds. A.Cohen & M.P.R.v.d.Broecke), Foris Publications, Dordrecht.

MacCurtain, F. and Fourcin, A.J. (1982): "Applications of electroglottography", in Transcripts of 10th Symposium, Care of the Professional Voice, Part II, The Voice Foundation, New York.

MacCurtain, F., Noscoe, N., Julian, W., Berry, R.J., and Fourcin, A.J. (1982): "Xeroradiography-electro-laryngography as an aid to diagnosis", in Transcripts of 10th Symposium, Care of the Professional Voice, The Voice Foundation, New York.

Noscoe, N. (1982a): "High definition of radiographic analysis of voice and its disorders", Radiography, 148, 147-150.

Noscoe, N. (1092b); "An objective analysis of voice disorders. Parts I and II", Brit.J. Disoreders of Communication, 67-85.

Rostron, A.B. (1976): "Pitch control in the human voice", Quart.J. Experimental Psychology, 28, 305-310.

Sonninen, A. (1970): "Phoniatric viewpoints on hoarseness", Acta Otolaryng., 236, 68-81.

Stevens, S.S. (1935): "The relation of pitch to intensity", J.Acoust. Soc.Am., 6, 150-154.

Sundberg, J. (1974): "An articulatory interpretation of the 'singing' formant", J.Acoust.Soc.Am., 55, 838-844.

Tulp, A. and Schutte, H.K. (1982): "Interpretation of electroglottograms, based on zonal representation within the phonetogram", paper presented at the Electroglottography Workshop, Rotterdam 1982.

Vennard, W. and Hirano, M. (1973): "Mechanisms of the physiological basis for vocal registers", in Vocal Registers (ed. L. Large), Mouton.

Wechsler, E. (1977): "A laryngographic study of voice disorders", Brit.J. Disorders Communication, 12, 9-22.

INVESTIGATIONS OF THE VIBRATION MODES OF THE VOCAL CORDS DURING DIFFERENT REGISTERS

C. Ocker
Hochschule für Musik, Hamburg, BRD
and
W. Pascher and M. Röhrs
Universitäts-Krankenhaus, Hamburg, BRD

Abstract

In cooperation with the Musikhochschule in Hamburg and the Phoniatric Department of the University-ENT-Clinic in Hamburg, we have made fiberoptical stroboscopic investigations of different registers and register dysfunctions in professional singers, and the results are compared to conventional indirect microstroboscopic findings.

The register problem, one of the themes of this conference, has presented a challenge for our team.

The close coupling between the three functional areas in phonation: the wind cavity, the tone source, and the vocal tract, suggests that these systems influence each other. In register research the influence of the vocal tract upon the voice source (that is, upon the vibration patterns of the vocal folds) has proved to be probable but not of great importance, although this has not been convincingly demonstrated, as far as we know (Fant, 1970; Flanagan, 1965; Habermann, 1978; Kafka-Bützow, 1982).

In our opinion, a subliminal tendency towards uncertainty can be observed in investigations of the register problem as a result of this situation. Each author attempts to display his results as "in accordance" with the findings of other researchers.

These facts have an understandable influence on stroboscopic examinations, which are usually performed with the indirect laryngoscope technique. Schönhärl (1960), the "modern old-master" of the stroboscopic technique, writes in his monograph:

"Da der Vorgang der stroboskopischen Untersuchung immer bei geöffnetem Mund und herausgestreckter Zunge erfolgt, - wie auch bei der gewöhnlichen Spiegeluntersuchung - wird bei einer reinen Stimmuntersuchung die natürliche Tongebung wohl beeinträchtigt, was aber bei der allgemeinen Beurteilung eines schon erkrankten Kehlkopfes nicht sehr von Bedeutung ist."

However, how does this stand in the case of the healthy voice, as in the singer? Can we rely on the findings that we, as phoniatricians, have made with the stroboscope? If so, is it possible to provide a subsequent validation for the numerous visual investigations of vocal vibrations? If not, where do the differences lie? The object of our team was to examine this question.

In order to judge the influence of the vocal tract, we chose two examination methods, which were performed in one session: indirect micro-laryngoscopy (McKelvie et al, 1970; Pascher et al, 1971; Seidner et al, 1972), and the so-called fiber-stroboscopy, a method which we developed in 1976 (Pascher and Neumann, 1976). This examination procedure does not alter the vocal tract (Kitzing, 1982). Both stroboscopic investigations were performed under full local anesthesia of the nasal, naso-pharyngeal, pharyngeal, and upper laryngeal areas.

The fiber-stroboscopic procedure proved difficult in all of its details. It demanded from the investigator:

- extensive practice in the technique,
- careful psychological guidance of the sensitive singer-subjects,
- exact reporting of the findings.

Although the fiber-stroboscopic examination of ENT patients has been described as "easy" (Gould et al., 1979; Saito et al., 1978) in the case of singers, it is not. The subtle technique, which demands great practice, and the required homogeneity in the observations made it necessary that one member of our team performed the actual examinations. We are grateful to Dr. Röhrs for meeting this challenge with great personal commitment.

The singer in our team was responsible for the acoustical judgments within the investigation of the vocal registers.

Our team has discussed the register problem extensively. When we employ the usual terms, such as chest register, middle register, operatic head register, and falsetto, this does not mean that we agree with them. But it is easier for all of us, and we have fewer problems with each other, when we follow the definitions according to John Large (1982).

Registers are the results of certain functions and can be simulated in other register areas by well-trained singers. Therefore, acoustical results can often only be reported definitively as the result of a longer acquaintance with a particular voice, or as a spontaneous judgment, with the danger of inaccuracy within certain parameters. For this reason, we have confined our investigation, essentially, to students who have studied several years with the singer in our team.

Among the 16 subjects who participated in the investigation, there were five cases where we could not perform the fiber-optical examination because their nasal passages were not wide enough for the instrument in spite of the anesthetic and vaso-constriction. In eleven cases we were able to compare the results of the micro-stroboscopy with those of the fiber-stroboscopy.

The results of our investigation are summarized in Figs. 1 and 2. The upper half of the scale shows the transition areas between the middle register and the operatic head register, as they were indicated by the acoustical impression while singing upward. In both examinations, the subjects started at their natural speaking frequency and sang upward on the vowel /e/.

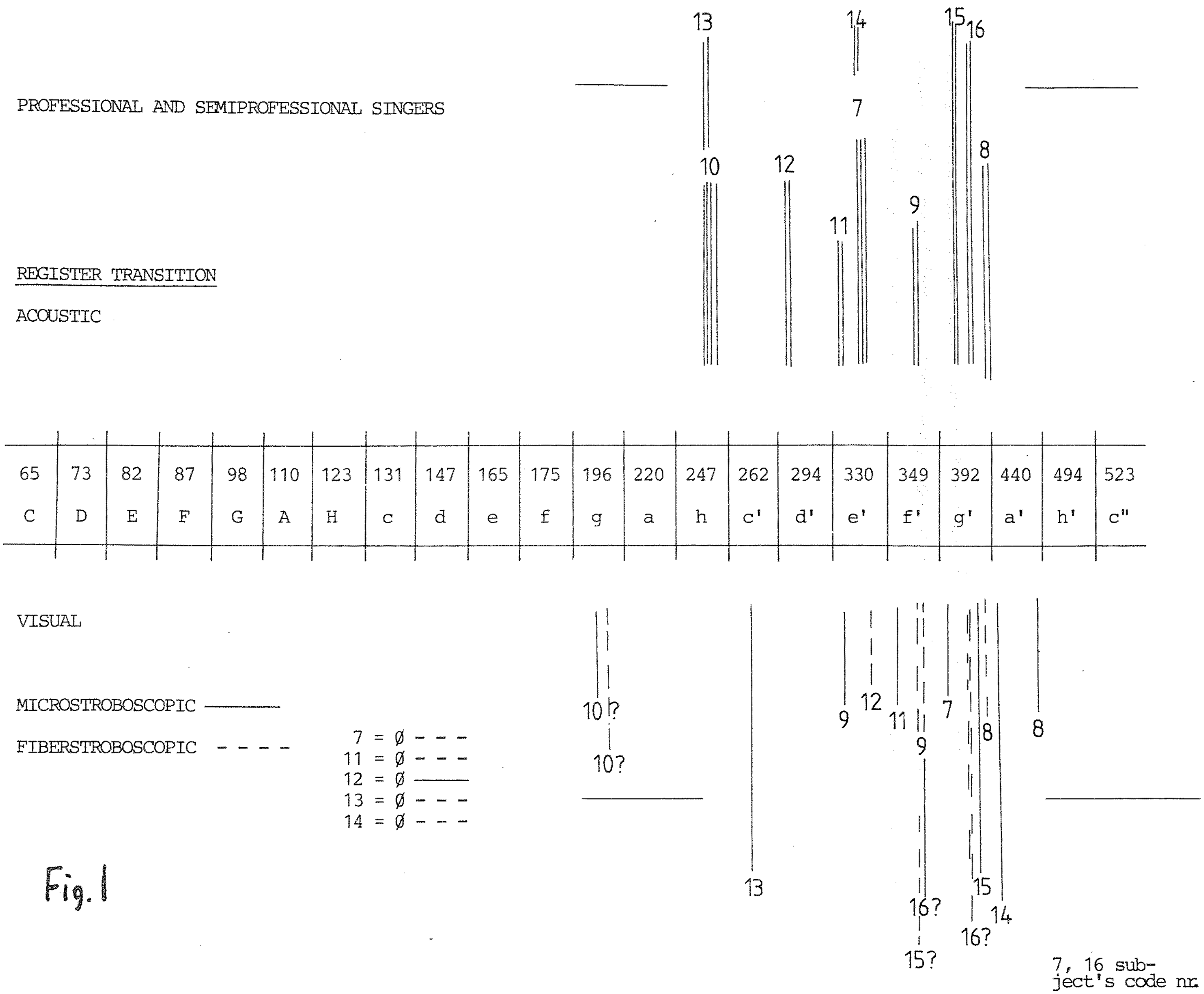

PROFESSIONAL AND SEMIPROFESSIONAL SINGERS
REGISTER TRANSITION
ACOUSTIC
13
14
15
16
7
10
12
8
9
11
65 73 82 87 98 110 123 131 147 165 175 196 220 247 262 294 330 349 392 440 494 523
C D E F G A H c d e f g a h c' d' e' f' g' a' h' c"
VISUAL
MICROSTROBOSCOPIC
FIBERSTROBOSCOPIC
7 = Ø
11 = Ø
12 = Ø
13 = Ø
14 = Ø
10?
10?
13
9
12
11
9
7
8
8
15
16?
14
16?
15?
7, 16 subject's code nr.

Fig. 1

NON PROFESSIONALS

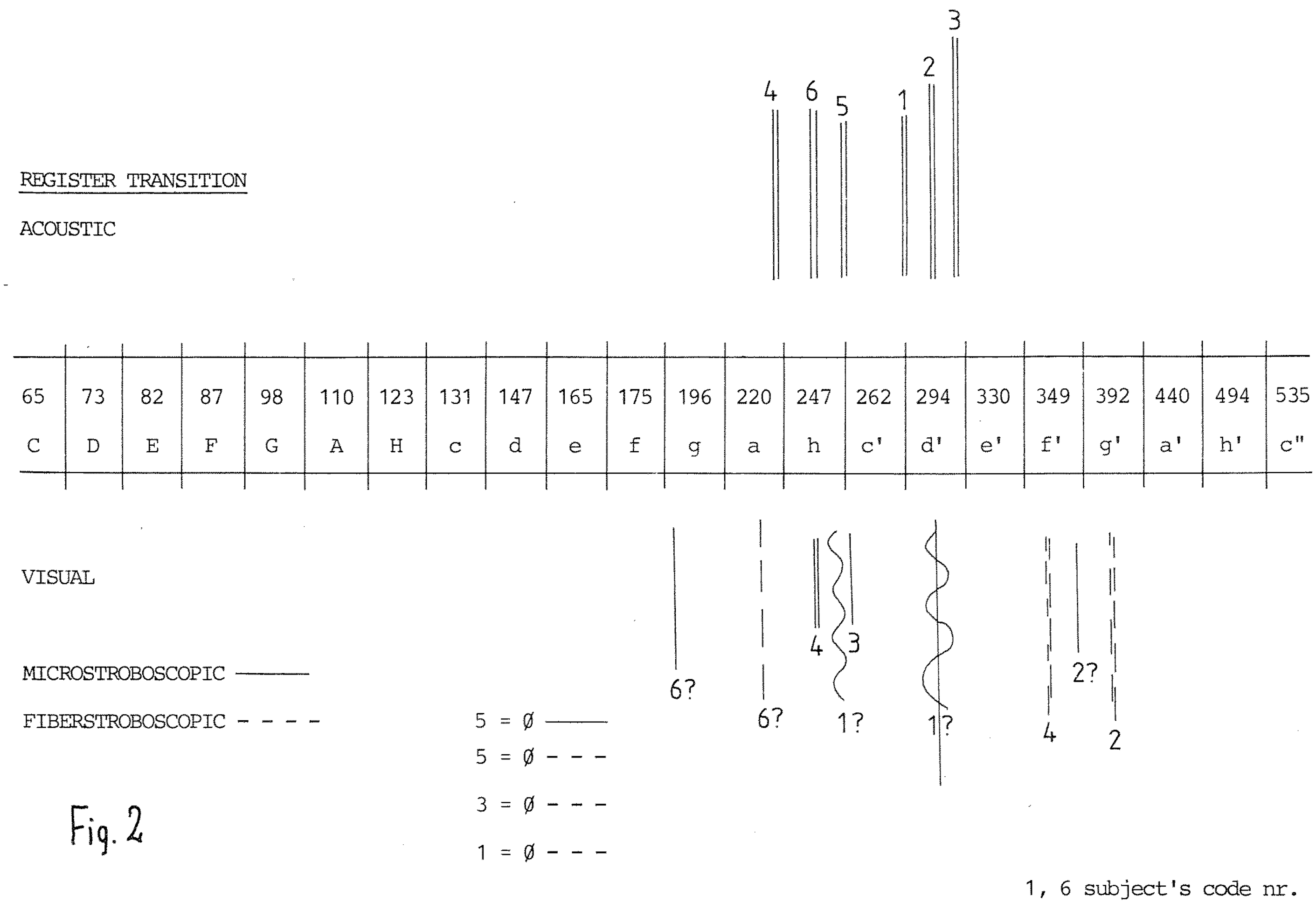


Fig. 2

In both techniques, at first we observed a gradual elongation of the vocal folds accompanied at the same time by a reduction in the amplitude and a shift in the vibrating portion of the vocal folds, with a constant, complete closure of the glottis. These visual vibration forms are in full agreement with the vibrating patterns with increasing frequency, described in the literature (Hirano, 1981; Hollien et al., 1971; Rubin and Hirt, 1960; Sonninen, 1956; Vennard, 1967).

Starting at a certain pitch, we could observe a point of maximum elongation of the vocal folds; in some cases we also observed vibration patterns that became irregular, glottal closure that was weak or becoming insufficient, and, occasionally, remedial mechanisms such as a curving in of the false vocal cords or a lowering of the epiglottis.

In addition, the fiber-stroboscopic examination revealed in three cases a complete shift in the level of the larynx, which rendered an overall view of the vocal folds impossible, even with the fiber-optical technique. For this reason, our presentation of the visually observed alterations in the vibration patterns, as shown in the figures, employs various symbols, including wavy lines and question marks.

With respect to the particular alterations as we observed them, there was no difference between "professionals" and "non-professionals", or between the micro-stroboscopic and the fiber-stroboscopic techniques. In both groups, the professionals as well as the non-professionals, the observable variations in the vibration patterns and the acoustical register transitions were concentrated in the same areas. We could observe, in some cases, distinct deviations from this trend - for the most part among the non-professionals - and, in other cases, a relatively close correspondence between these two areas - mostly among the professionals. However, the statistical analysis of these observations revealed this trend to be not significant. Also, a tendency of the visual alterations to come before or after the acoustical transition could not be definitely confirmed.

The figures do not show the observation that the total frequency areas examined with the micro-stroboscope and the fiber-stroboscope were, in all cases, practically identical, starting with the spontaneous speaking range up to approximately a third above the acoustical register transition. Thus, from a practical point of view, we conclude that both examination techniques - the micro-stroboscope in the conventional mirror examination which alters the vocal tract and the fiber-stroboscope which does not - give the same results with respect to the vibration patterns of the vocal folds in different registers.

We found that the influence of the vocal tract upon the vibration patterns of the vocal folds, if it exists, is too small to be observed with our methods. For the phoniatrician visual examination methods can be used with a clear conscience.

References

Fant, G. (1970): Acoustic Theory of Speech Production (2nd edition), Mouton, The Hague.

Flanagan, J.L. (1965): Speech Analysis, Synthesis and Perception, Springer-Verlag, Berlin.

Gould, W.J., Kojima, H., and Lambiase, A. (1979): "A technique for stroboscopic examination of the vocal folds using fiberoptics", Arch.Otolaryng., 105, 285.

Habermann, G. (1978): Stimme und Sprache, Thieme Verlag, Stuttgart.

Hirano, M. (1981): Clinical Examination of Voice Disorders of Human Communication 5, Springer-Verlag, Wien.

Hollien, H., Brown Jr, W.S., and Hollien, K. (1971): "Vocal fold length associated with modal, falsetto, and varying intensity phonations", Folia Phoniat., 23, 66.

Kafka-Bützow, A. (1982): "Physiologie der Stimme und Sprache", in Phoniatrie und Pädaudiologie (eds. P. Biesalski & F. Frank), Thieme Verlag, Stuttgart.

Kitzing, P. (1982): "Photo- and electroglottographical recording of the laryngeal vibratory pattern during different registers", in Proc. 2nd Conf. Int. Decade of Research in Singing (eds. M. Waar-Garratt & Ch. Warr), Rotterdam.

Large, J. (1982): "Toward an integrated physiologic-acoustic-theory of vocal registers", in Proc. 2nd Conf. Int. Decade of Research in Singing (eds. W. Waar-Garratt & Ch. Waar), Rotterdam.

McKelvie, P., Grey, P., and North, C. (1970): "Laryngeal strobomicroscope", Lancet, II, 503.

Pascher, W. and Neumann, G. (1976): "Fiber-Stroboskopie, Technik und Anwendungsmöglichkeiten", Arch. Oto-Rhino-Laryng., 213, 464.

Pascher, W., Homoth, R., and Kruse, G. (1971): "Verbesserte visuelle Diagnostik in der Laryngologie und Phoniatrie", HNO (Berlin), 19, 373.

Rubin, H.J. and Hirt, C. (1970): "The falsetto, a high speed cinematographic study", The Laryngoscope, 70, 1305.

Saito S., Fukuda, H., Kitahara, S., and Kokawa, N. (1978): "Stroboscopic observation of vocal fold vibration with fiberoptics", Folia Phoniat., 30, 241.

Schönhärl, E. (1960): Die Stroboskopie, Thieme Verlag, Stuttgart.

Seidner, W., Wendler, J., and Halbedl, G. (1972): "Mikrostroboskopie", Folia Phoniat., 24, 81.

Sonninen, A. (1956): "The role of the external laryngeal muscles in length-adjustment of the vocal cords in singing", Acta Oto-Laryng., Suppl., 130.

Vennard, W. (1967): Singing, the Mechanism and the Technic, Carl Fischer Inc., New York.

SIMULATION OF SINGING WITH A COMPOSITE MODEL OF SPEECH PRODUCTION *

C. Scully and E. Allwood
Department of Linguistics and Phonetics,
University of Leeds, Leeds, England

Abstract

A model of speech production has been developed from which intelligible, speech-like sounds have been generated (Allwood and Scully, 1982; Scully and Allwood, 1982). From the kinematics of the articulators, aerodynamic conditions at the glottis, the velopharyngeal port and one supraglottal articulator are computed. Acoustic noise sources, aspiration at the glottis and frication at the supraglottal constriction are derived from cross-section area and local aerodynamic conditions. A transient source from rate of change of oral pressure is included.

The voice source is derived from a functional model of the larynx. A parametric representation of the waveform of glottal flow is used (Fant, 1980). Parameter values are controlled by three variables: pressure drop across the glottis, the d.c. component of glottal area and the effective stiffness and mass of the vocal folds. The range of values and dependencies are preset by means of a table of parameter values at selected combinations of physiologcial conditions. Thus different larynx types or modes of action may be simulated. Specifically, the larynx simulation can be of a trained singer or an ordinary speaker, based on published data for inverse filtering. Noise sources and glottal acoustic losses are made to vary during the voice cycle.

Attempts to model singing will be described.

Introduction: the model

A model of speech production processes has been developed from which intelligible speech-like sounds have been generated (Allwood and Scully, 1982). Inputs to the model consist of commands for about 10 quasi-independent articulators to move to a succession of target states by means of specified articulatory transitions. Articulators for normal

---

* This paper is illustrated by sound examples.

speech include lung walls, vocal folds, walls of the pharynx, soft palate, tongue body, tip or tip-blade portion of the tongue, jaw, and lips. Lung wall-rib cage articulation is represented in the model by rate of decrease of lung volume or by air pressure within the alveoli of the lungs. The latter option was selected for the modelling described here for its simplicity of specification, although it is recognised that volume decrement is probably a better model of subglottal control in natural speech and singing. Two components of vocal folds actions are represented: first, the control of glottal area AG, which, for our purpose, is equated with abduction and adduction of the vocal folds; secondly, a global factor, called Q, to represent the effective stiffness and mass of the vocal folds. Vertical movements of the larynx have not yet been included. Articulatory coupling of the nasal cavities is represented by a constriction of cross-section area AV, the area of the velopharyngeal port. For the remaining articulators, the emphasis is on their shaping of the acoustic tube of the vocal tract. Articulatory commands define the cross-section area of the tube at each of a few parameter points. By non-linear interpolation, the area function of the whole of the vocal tract is obtained. For vowel-like vocal tract shapes, their shaping of the acoustic tube of the vocal tract. Articulatory commands define the cross-section area of the tube at each of a few parameter points. By non-linear interpolation, the area function of the whole of the vocal tract is obtained. For vowel-like vocal tract shapes, only the points shown in Fig. 1 are needed. This is the static vocal tract shape used here for the simulation of a male singer producing an vowel, with a vocal tract length, from S to L, of 17.5 cm.

The aerodynamic stage of speech production, which links sections to sounds, defines air pressures and air flows at three points: the glottis, a major supraglottal constriction, and the velopharyngeal port. Acoustic sources of voice and aspiration noise are generated at the glottis; frication noise and transient, or pulse, sources are generated at the supraglottal constriction; frication noise may be generated at the velopharyngeal port also. One simulation with the complete aerodynamic model showed that the voice source dominated for singing, while the vocal tract constriction at the P point, shown in Fig. 1, was not severe enough to affect aerodynamic conditions at the glottis. For these reasons, and to reduce computation time, subglottal pressure was defined directly by an input command instead of being computed. The properties of the voice acoustic source were defined by a functional model, to be described in more detail below.

In the next stage of the model, the acoustic sources are inserted, at the appropriate places, into the branched acoustic tube defined by the

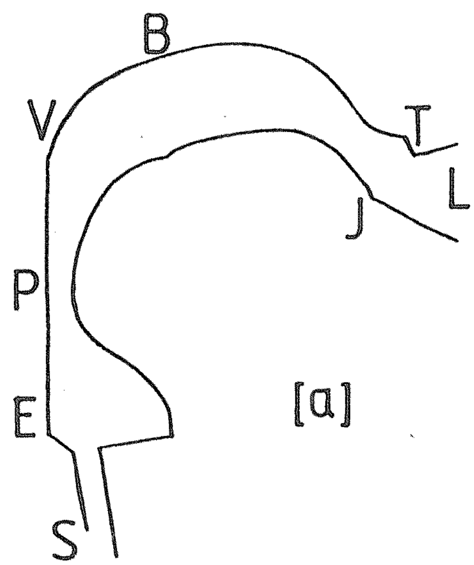


Fig. 1. Vocal tract shape, derived from the area function, for simulation of a male singer producing [ɑ] with a wide pharynx. Vocal tract length = 17.5 cm; cross-section in lower pharynx, AE = 15 cm^2; cross-section of larynx tube = 1.5 cm^2.

supraglottal articulators and a specified nasal area function, the two being linked by an orifice 0.5 or 1 cm in length of cross-section area AV. The method of Kelly and Lochbaum (1962) is used to compute the acoustic component of volume flow rate of air out of the mouth and nose.

Low-pass filtering is included in each section of the tubes, the nasal portion having more smoothing than the oral and pharyngeal portions. Losses at the open end of the tubes are represented by a fixed frequency reflection coefficient. This approximates the tube termination, but does so without regard to frequency characteristics and mouth size. Trans-glottal losses are represented by a reflection coefficient which varies with the size of the glottis AG. Three formulae have been devised for mapping from the articulatory (d.c.) component of glottal area to the rapid acoustic changes of glottal area during voicing. In one of the simulations for arpeggio singing this a.c. component of AG was used for the glottal losses. Time varying losses implemented in this way did not appear to be spectrographically or auditorily very significant. However the output waveforms are visually closer to those of real speech (Scully and Allwood, 1982); in addition, the research of Fant and Liljencrants (1979) shows that the amount of decay of formants during a voice cycle is of perceptual significance. The transformation from acoustic flow at the lips or nostrils to acoustic pressure at a microphone is represented by high pass filtering at 6 dB/octave. The model is implemented on a VAX 11/780 computer. Speech signals are down-line loaded to a PDP 11/03 microprocessor and output in real time under the control of a hardware clock at a sampling frequency of 11.9 kHz.

Functional model of the voice source

A parametric representation of Ug(t), the acoustic component of volume flow rate of air through the glottis, is used (Fant, 1982). The waveform and its parameters are shown in Fig. 2. Besides T0 to define F0, three out of four additional parameters must be defined. Problems of independence of variables arise with some of the possible combinations; for VOIA, TCR and K (an asymmetry factor) the calculation is straightforward. Here TD was included instead of K since the literature indicates strongly that limits on this "closing time" may be a characteristic of different speak-

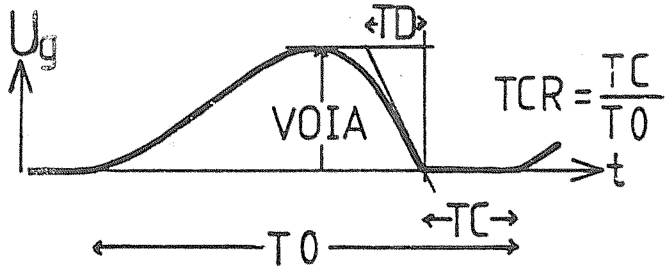


Fig. 2. Parametric representation of the waveform for the voice acoustic source. TCR = 1 - open quotient. The magnitude of the (negative) slope of Ug at the point where flow comes to zero is called |USLOPE|. |USLOPE| = VOIA/TD. F0 = Q + 4.PDIFF. Units are: PDIFF cmH20, Q Hz, AG cm^2, VOIA cm^3/s, TCR dimensionless, TD ms.

er and singer types. An iterative solution for the other wave parameters had to be used in this case. This waveform represents a "short circuit current" source. True glottal flow is not derived in the mode, but the form of the wave is made to vary with three independent controlling variables. These are PDIFF the pressure drop across the glottis, Q, and AG, both defined above. A table is set up, as shown in Fig. 3, to define the values of the dependent variables VOIA, TCR and TD at specified combinations of the independent variables PDIFF, Q and AG. By a linear mapping, values of VOIA, TCR and TD can be obtained for any combination of values of PDIFF, Q, AG. Thus, the wave shape changes continuously during a simulation, as the geometry of the glottis, the aerodynamic forces and the mechanical properties of the vocal folds change in the ways defined by their input commands.

The parameters in the table must be given realistic values. The table is intended to conform to basic physical constraints while being capable of modelling different speaker types or options selected by a given speaker. Here, we are interested in the use of a larynx for professional singing; therefore TD is allowed to become as small as 0.3 ms, VOIA is allowed to go as high as 900 cm^3/s and TCR is allowed to go as high as 0.7. Other variations are based on published data, especially those of

WAVE SHAPE PARAMETERS

WAVSW = 1 REVISED WAVE SHAPE METHOD
TKSW = 1 (VOIA, TCR, TD)

PDIFF	Q	AG	VOIA	TCR	TD
10.00	Q1 200.00	0.05	VOIA1 400.00	TCR1 0.05	TD1 3.00
5.00	100.00	0.05	VOIA2 300.00	TCR2 0.20	TD2 1.80
10.00	100.00	0.05	VOIA3 750.00	TCR3 0.60	TD3 0.50
10.00	100.00	0.15	VOIA4 850.00	TCR4 0.05	TD4 2.50
MINIMUM			0.01	0.00	TDMIN 0.30
MAXIMUM			VOIAMAX 900.00	TCRMAX 0.70	1000000.00

Fig. 3. Table relating three dependent variables VOIA, TCR and TD to three independent variables PDIFF, Q and AG, for the definition of the voice source waveform shown in Fig. 2.

Lindqvist-Gauffin (1970), Rothenberg (1973), Sundberg and Gauffin (1979), and Gauffin and Sundberg (1980). More data are needed to set up a realistic table, especially for women's voices and for children but, in decreases with Q; TCR increases to a maximum of 0.5 or 0.6 as PDIFF increases while it decreases when AG or Q increases; F0 is mainly controlled directly by Q, but to this is added a linear dependence on PDIFF, so that F0 = Q + 4.PDIFF, in keeping with much published data. The relationship of F0 with AG is not clear at present.

Acoustically, the most important wave parameters are T0, for F0 and pitch perception, the area under the curve for energy at low frequencies and USLOPE for energy at high frequencies (Fant, 1980). The general variation of |USLOPE| with PDIFF can be predicted in some cases and not in others. |USLOPE| = VOIA/TD. Since VOIA increases with PDIFF while TD decreases, |USLOPE| increases with PDIFF. Similarly it is to be expected that |USLOPE| decreases as Q increases. The relationship of |USLOPE| with AG is not obvious since both VOIA and TD increase with AG. The complexity of some results from real speech arises in part from the sometimes unavoidable use of F0 as an independent variable. Increases of Q and PDIFF have opposite effects on TCR as they do on TD. Therefore TCR and TD may arise or fall as F0 rises. These effects have been investigated, using the Ishizaka-Flanagan two-mass model of vocal fold vibrations, by Monsen et al. (1978).

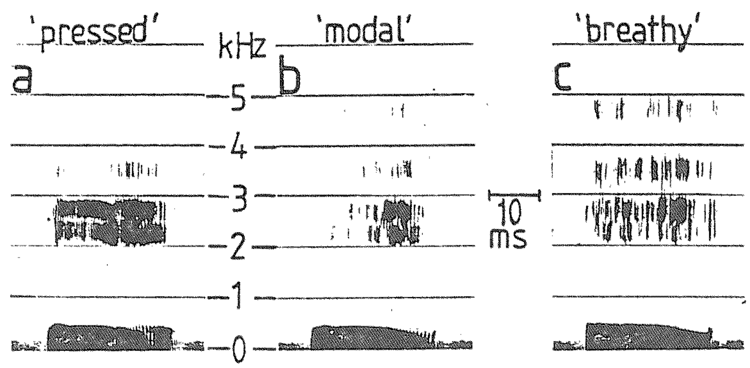


Fig. 4. Spectrograms for three phonation types with a simulated male [i] vowel. Glottal area AG is 0.0125 cm^2 for 'pressed', 0.05 cm^2 for 'modal' and 0.15 cm^2 for 'breathy'. An alternative term for 'modal' is needed since this is the name given to the middle register, opposed to 'vocal fry' and 'falsetto'.

Preliminary checks suggest that, when speech is simulated, the growth of radiated acoustic pressure as a function of air pressure drop across the glottis follows a power law, approximately $p_r \alpha \Delta P^{1.6}$ (for example, Ladefoged and McKinney, 1963; Isshiki, 1964). Different spectral emphasis is obtained by varying AG, for the control of phonation type, as shown in Fig. 4. The model is probably too simplified to represent the full range of vibration modes of the vocal folds and associated flow waveshapes described by, for example, Titze (1979).

Simulation of singing

Vibrato

In the modelling, the auditory target of a particular singer type and quality has to be attained by trial and error with monotoring. First attempts at vibrato singing were auditorily reasonably acceptable, especially where nasal coupling was introduced. Spectrograms of the acoustic outputs from the model and some of the variations in voice source parameters are shown in Figs. 5 and 6. The importance of the ratio of pharyngeal area to larynx tube area for the production of the singing formant (Sundberg, 1974) was confirmed, by varying the cross-section area AG from 15 cm^2 to 2 cm^2. It was assumed that vibrato is the result of formant (Sundberg, 1974) was confirmed, by varying the cross-section area AG from 15 cm^2 to 2 cm^2. It was assumed that vibrato is the result of fast transitions of muscle forces, as shown by Hirano (1981). Vibrato rates may be related to neurological control loops for the muscles and to movement tremor (Shipp et al., 1980). It may be a mechanism for reducing the fatigue of the fast- twitch intrinsic laryngeal muscles. Fast muscles tire more quickly than slow, postural, muscles; their maximum muscle tension can be maintained voluntarily for a short time only, perhaps for as little as a few hundred milliseconds (Stålberg and Young, 1981). The problem probably does not arise in normal speech, where abductor and adductor muscles alternate and pitch is changing most of the time, but it could be highly relevant for professional singing. Modelling of vibrato was done in two ways: first by varying subglottal pressure cyclically: secondly by varying Q cyclically, which seems likely to be a more realistic method. An articulatory transition time of 100 to 72 ms results in a vibrato rate of 5 to 7 Hz. As laryngeal articulatory transitions these are only a little faster than those found in speech, so the view of vibrato as voluntary activity seems tenable.

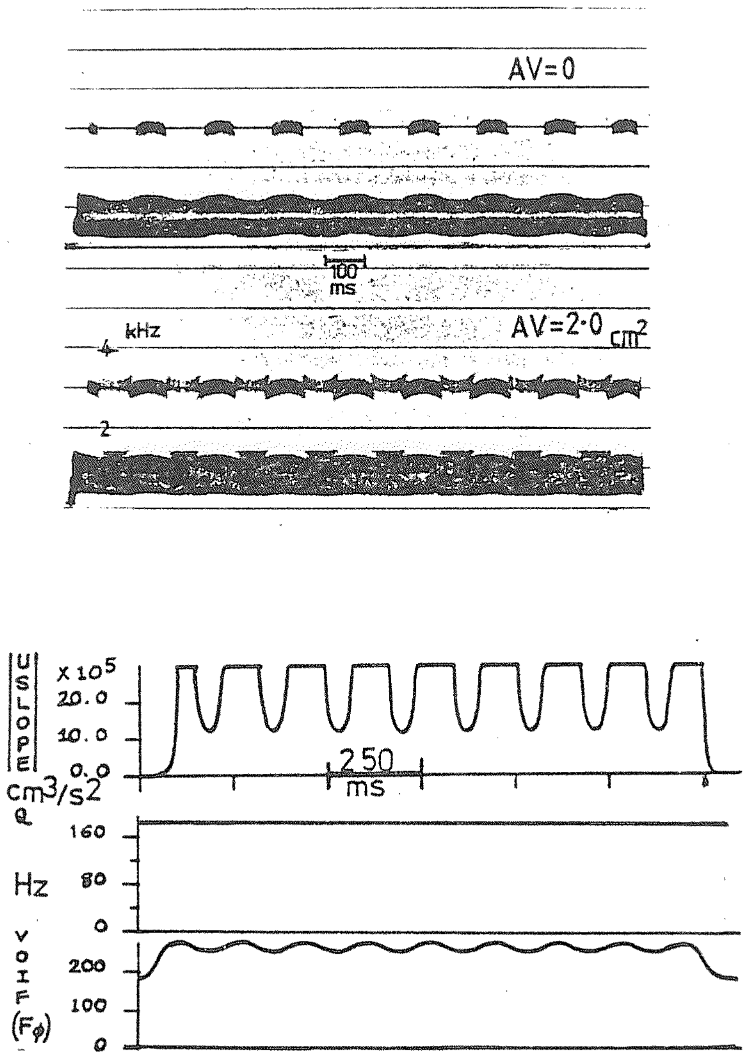


Fig. 5. Simulation of vibrato by oscillations in subglottal pressure. Top, the spectrogram for sung [ɑ] (male), without nasal coupling; middle, the spectrogram with nasal coupling; below, some voice wave parameter variations arising as a result of varying pressure drop across the glottis.

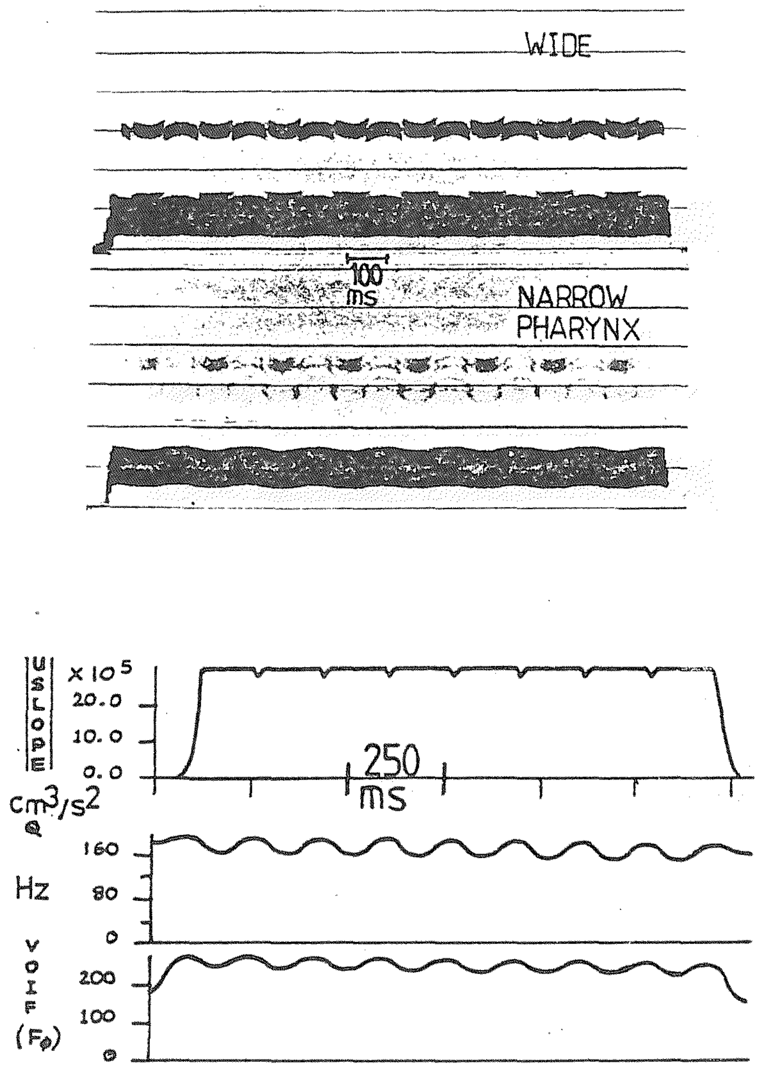


Fig. 6. Similar to Fig. 5, but here Q oscillations are amployed for the vibrato. Top, with wide pharynx (AE = 15 cm^2); middle, with narrow pharynx (AE = 2 cm^2); below, some waveshape parameter variations and the time function for a controlling variable Q.

Changing pitch

Changing notes were introduced using Q transition times consistent with data from real singing (Fujisaki, 1981). 'Non-legato' note changes were modelled with a transition time of 120 ms. Pitch changes which give a speech-like effect can be generated with the model by giving transitions to Q, with or without additional transitions for PDIFF and AG.

Changing loudness

Changes of loudness were modelled in five different ways. First, PDIFF was increased, keeping AG constant at 0.05 cm^2. The aim here was to increase loudness at constant pitch, as Q had to be reduced to compensate for the increase in PDIFF. To simplify the calculation, increases in PDIFF were made linear. No doubt a professional singer learns to perform much more complex computations for achievement of the acoustic and auditory goals. Loudness was increased, secondly, by increasing PDIFF without Q compensation. The result was like neither singing nor speech. Thirdly, AG was reduced from 0.12 to 0.015 cm^2 to simulate increasingly 'pressed' phonation, with PDIFF maintained at 10 cmH20. After an initial loudness increase, the sound faded away, when the glottal area became very small. Fourthly, in addition to the AG reduction from 0.12 to 0.015 cm^2, PDIFF was increased from 10 to 30 cmH20. By this expected. Finally, loudness was increased by gradually increasing the swing of PDIFF for vibrato. Acoustic intensity level traces for one mezzo-soprano singer showed increased loudness associated with an increasing magnitude of swing in intensity level about a steady mean value associated with a good, clear and regular vibrato (Lennox, 1982). This was the effect to be simulated. When PDIFF increased by 1 cmH20 on each swing the effect was that of an explosive [hɑ-hɑ]-like sound: with the rate of pressure increase reduced, the effect was more natural for singing.

Female singing

First attempts at the simulation of female singing demonstrated, as expected, that merely shortening the vocal tract (to 14 cm) was not sufficient to create the right auditory effect. A reduction of the volumes of the lower pharynx and the oral cavity did not improve the simulation significantly. Almost certainly a different table of voice wave parameter values will be needed. More emphasis on data collection for the speech and singing of women should be given a high proirity for researchers in the future.

Acknowledgements

The research is supported by the Science and Engineering Research Council, project GR/B/34874. The VAX 11/780 system is part of the Interactive Computing Facility of the SERC at Leeds University. Colleagues in the Department of Mechanical Engineering offered useful ideas in the development of the linear mapping used in the functional model of the voice source.

References

Allwood, E. and Scully, C. (1982): "A composite model of speech production", ICASSP82 Conf.Rec., Vol. 2, 932-935.

Fant, G. (1980): "Voice source dynamics", STL-QPSR 2-3/1980, 17-37.

Fant, G. and Liljencrants, J. (1979): "Perception of vowels with truncated interperiod decay envelopes", STL-QPSR 1/1979, 79-84.

Fujisaki, H. (1981): "Dynamic characteristics of voice fundamental frequency in speech and singing; acoustical analysis and physiological interpretation, Proc. 4th FASE Symposium, Venice.

Gauffin, J. and Sundberg, J. (1980): "Data on the glottal voice source behaviour in vowel production", STL-QPSR 2-3/1980; also given at the 10th ICA, Sydney, 1980.

Gauffin, J. and Sundberg, J. (1980): "Data on the glottal voice source behaviour in vowel production", STL-QPSR 2-3/1980, 61-70; also given at the 10th ICA, Sydney, 1980.

Hirano, M. (1981): "The function of the intrinsic laryngeal muscles in singing", pp. 155-170 in (K.N.Stevens & M.Hirano, eds.) Vocal Fold Physiology, Univ. of Tokyo Press.

Isshiki, N. (1964): "Regulatory mechanism of voice intensity variations", J. Speech and Hearing Res. 7, 17-29,

Kelly, J.L. and Lochbaum, C.C. (1962): "Speech synthesis", pp. 1-4 in Proc. 4th ICA, Copenhagen, paper G42.

Ladefoged, P. and McKinney, N.P. (1963): "Loudness, sound pressure and subglottal pressure in speech", J.Acoust.Soc.Am. 35, 454-460.

Lennox, D. (1982): "The use and control of pitch in speech and singing with particular reference to vibrato", unpubl. M.A.Diss., Dept. of Linguistics and Phonetics, University of Leeds.

Lindqvist-Gauffin, J. (1970): "The voice source studied by means of inverse filtering", STL-QPSR 1/1970, 3-9.

Monsen, R.B., Engebretson , A.M. and Vemula, N.R. (1978): "Indirect assessment of the contribution of subglottal air pressure and vocal-fold tension to changes of fundamental frequency in English", J.Acoust.Soc.Am. 64, 65-80.

Rothenberg, M. (1973): "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing", J.Acoust.Soc.Am. 53, 1632-1645.

Scully, C. and Allwood, E. (1982): "Glottal losses in a composite model of speech production", Proc. Institute of Acoustics, Nov, 1982, A1.1-1.5,

Shipp, T., Leanderson, R. and Sundberg, J. (1980): "Vocal vibrato", pp. 46-49 in (V.L.Lawrence & B.Weinberg, eds.), Transcripts of the Eighth Symposium Care of the Professional Voice, Part I.

Stålberg, E. and Young, R.R. (eds.) (1981): Clinical Neurophysiology, Butterworth Int., Medical Rev.

Sundberg, J. (1974): "Articulatory interpretation of the 'singing formant'", J.Acoust.Soc.Am. 55, 838-844.

Sundberg, J. and Gauffin, J. (1979): "Waveform and spectrum of the glottal voice source", pp. 301-320 in (B.Lindblom & S.Öhman, eds.), Frontiers of Speech Communication Research. Festschrift for Gunnar Fant, Aademic Press.

Titze, I.R. (1979): "Physical and physiological dimensions of intrinsic voice quality", in (H.Hollien & P.Hollien, eds.), Current Issues in the Phonetic Sciences, Vol. 9, John Benjamins P.V., Amsterdam.

DEPENDENCE OF THE HIGH SINGING FORMANT ON PITCH AND VOWEL IN DIFFERENT VOICE TYPES

W. Seidner*, H.K. Schutte**, J. Wendler*, and A. Rauhut*
*HNO-Klinik der Humboldt-Universität zu Berlin, DDR
**University Hospital, Groningen, The Netherlands

Abstract

Five outstanding professional singers were checked by narrow band spectral analysis in loud phonation of the vowels /a/, /u/, and /i/. The following parameters were investigated: Main frequency of the high singing formant between 2000 to 5000 Hz, relative intensity of this formant and its 15 dB bandwidth. Additionally, averaged spectrograms of sustained vowels and long-term-average-spectra (LTAS) of continuous singing were carried out.

The high singing formant is located at a higher frequency in high voices than in low voices. There is no relation of main frequencies of the singing formant. There is a dependence not only on pitch but also on vowel. As for female singers we observed that the high singing formant appears as a double peak in the range of 2500 to 3000 Hz and 3000 to 4000 Hz, respectively. The relative intensity of the high singing formant was higher and the bandwidth was narrower in male singers as compared to female singers.

Introduction

The acoustics experts in this audience will probably be disappointed by our report; they may miss a theoretically based discussion of the results. However, we would like to point out that we performed these measurements starting from a clinical-practical standpoint.

We started our investigations with the aim of improving the registration of the dynamic range of the voice depending on pitch. Simultaneously, we included the high singing formant in our investigations (Seidner et al., 1981; Stürzebecher et al., 1982; Wendler et al., 1981; Seidner and Schutte, 1982). We tried to find out whether a filter of 2000 to 4000 Hz bandwidth may be used for measuring the high singing formant.

Subjects and methods

There have been few systematical experiments as to the problem of marking the high singing formant distinctly under conditions of differing pitch, vowel and voice types (e.g., Stout, 1938; Winckel, 1971; Morosow, 1977; Sundberg, 1970, 1974; Dmitriew and Kiselew, 1979). This is why we investigated five outstanding stage and concert singers. We carried out a narrow band spectral analysis of different vowels which had been recorded on tape during loud phonation when measuring the dynamic voice range. We analyzed the following parameters: 1) center frequency of the formant in the range of 2000 to 5000 Hz; 2) relative intensity of the singing formant at various fundamental frequencies and at different frequencies of the first formant; and 3) the -15 dB bandwidth of the singing formant.

Results

The results of the total of 286 spectral analyses are illustrated in the following figures. In each figure we find the center frequency of the high singing formant in the top graph, its relative intensity in the middle graph, and the -15 dB bandwidth at the bottom. The values pertaining to the single vowels are plotted individually as a function of pitch (vowel /a/ ■, vowel /u/ ●, vowel /i/ ▲).

Fig. 1 (bass): Modest formant peaks occur at 2500 to 3000 Hz (/a/ and /u/ at 2500 Hz, /i/ at 3000 Hz), the relative intensity of which is about -10 dB. Only the vowel /i/ is marked more distinctly in the upper half of the artistically usable pitch range. The formant areas for the three

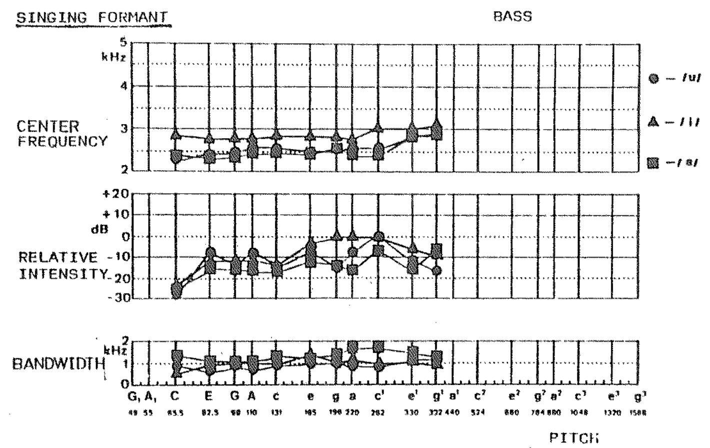


Fig. 1. Center frequency of the high singing formant (above), relative intensity (middle), and -15 dB bandwidth (below) for a bass singer singing the vowels /a/, /u/, and /i/ at various pitches.

vowels are nearly equally wide (1000 Hz). There is only a widening in the case of the /a/ at higher pitches.

Fig. 2 (baritone): The center frequency of the vowel-dependent formant peaks is similar to those of the bass (this singer's voice sounds similar

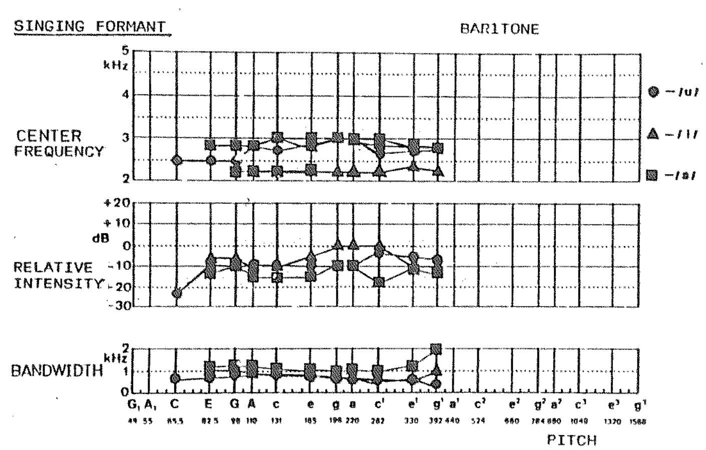


Fig. 2. Center frequency of the high singing formant (above), relative intensity (middle), and -15 dB bandwidth (below) for a baritone singer singing the vowels /a/, /u/, and /i/ at various pitches.

to the voice of the bass). The only striking thing is that the peak is double over the entire pitch range in the case of /i/ (2300 and 3000 Hz) and at lower pitches in the case of /a/. The relative intensity and the width are similar to the bass.

Fig. 3 (tenor): The main frequency of the vowel-dependent formant peak is above 3000 Hz. From c^1 (C4) and onwards additional spectral compo-

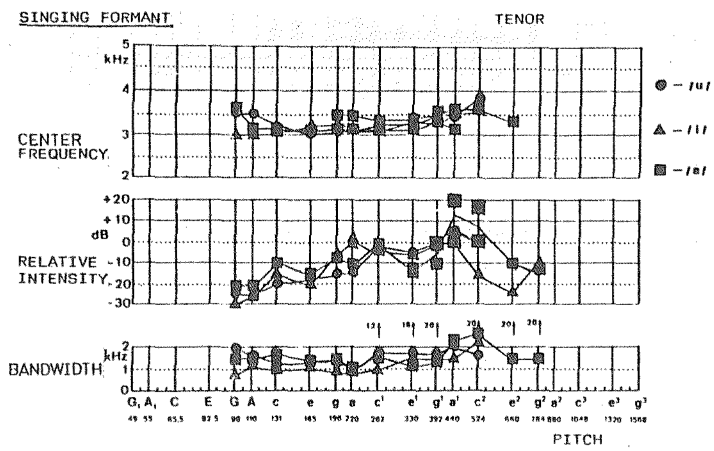


Fig. 3. Center frequency of the high singing formant (above), relative intensity (middle), and -15 dB bandwidth (below) for a tenor singer singing the vowels /a/, /u/, and /i/ at various pitches.

nents were observed in all vowels, mostly far beyond 4 kHz with decreasing intensity. This energy may result partly from an increase in sharpness with increasing pitch and partly due to distortion effects from the recording system, too. The relative intensity of the singing formant changes depending on the pitch and the vowel to a greater extent than in low voices. In the pitch range a^1 (A4) to c^2 (C5) the intensity is greater than that of the first and second formant in the case of /a/. As compared with the low voices, the width of the investigated formant peaks is altogether greater (an average of 1000 to 2000 Hz) and varies to a larger extent.

The formant peaks of female singers are striking by their doubling of all vowels being sung. This applies throughout the whole pitch range.

Fig. 4 (alto): The lower peak in the formant area is at 2500 Hz, the higher one at 3000 to 3500 Hz, additionally there is from e^2 (E5) and onwards a strong widening of the spectrum. This is similar to the voice

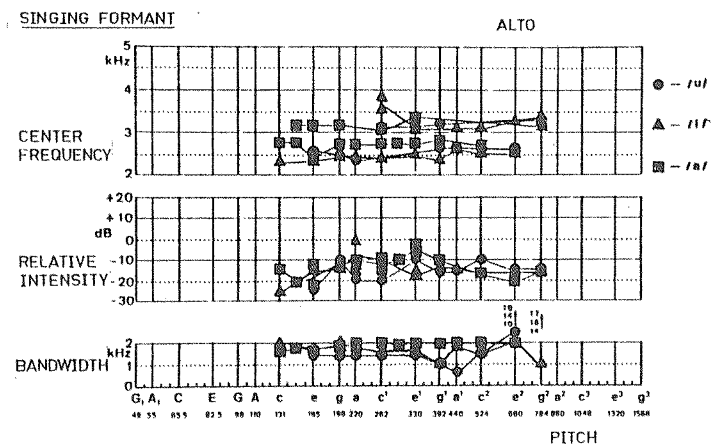


Fig. 4. Center frequency of the high singing formant (above), relative intensity (middle), and -15 dB bandwidth (below) for an alto singer singing the vowels /a/, /u/, and /i/ at various pitches.

of the tenor. The relative intensity of the singing formant is somewhat lower as compared to that of the male singers; the width is, in comparison with the tenor, somewhat smaller and shows from e^1 (E4) and onwards larger changes with pitch than that of the tenor.

Fig. 5 (soprano): The two peaks of the singing formant area are higher and wider in comparison with the alto. The low peak can be seen at 2500 to 3200 Hz, the higher peak at 3300 to 4300 Hz. Furthermore, the two peaks of /a/ are displaced in the direction of the higher frequencies. The relative intensity of the singing formant shows the lowest values in comparison with the other voices. While the curves of the vowels /a/ and /i/ are almost lying horizontally, the curve of /u/ falls at lower pitch-

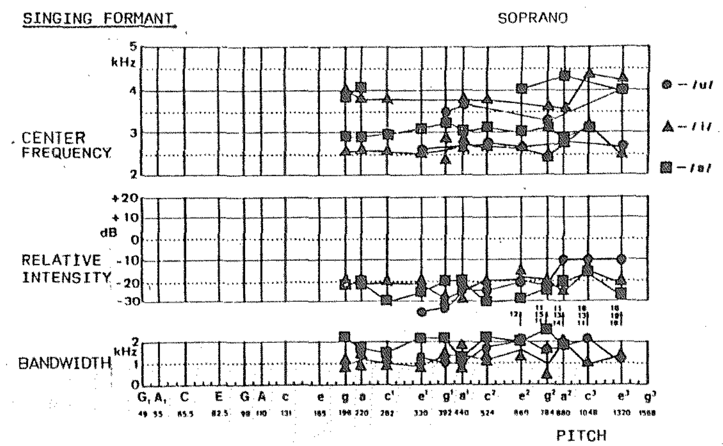


Fig. 5. Center frequency of the high singing formant (above), relative intensity (middle), and -15 dB bandwidth (below) for a soprano singer singing the vowels /a/, /u/, and /i/ at various pitches.

es and increases at higher pitches. The width of the singing formant is, first of all, different from that of the alto, as there are differences between /a/ and /i/ (wide in /a/, narrow in /i/).

Summarizing, we can say:

1. The high singing formant has its maximum at higher frequencies in high voices and at lower frequencies in lower voices, but, there are not strict correlations between the main frequencies of the singing formants and certain voice types. The main frequencies of the formants vary depending on vowels as well as on pitches with no regular tendencies detectable from our investigations.

2. As to female singers two peaks can occur which are located at 2500 to 3000 Hz and at 3000 to 4000 Hz.

3. The relative intensity of the high singing formants is altogether higher in male singers, than in female singers. In addition, there is

also a vowel and pitch dependence. In low male voices the vowel dependence is more marked.

4. The high singing formant of male voices is narrower as compared with that of female voices. In all voice types the vowel dependence of this width is a little more distinct of higher pitches than at lower pitches.

The acoustical investigations were completed by vowel-dependent averaged spectrograms of sustained vowels at loud phonation over the whole pitch range as well as by long-term-average-spectra (LTAS) of continuous singing. This showed that the high singing formant was more distinctly marked in men than in women. Furthermore, it revealed that the separation of the singing formant area from the vowels' formant area occurs in a more striking way in men than in women.

Finally, we can state that regarding a filter of 2000 to 4000 Hz, it is possible to observe with sufficient exactness the high singing formant in different voice types.

References

Dmitriew, L. and Kiselew, A. (1979): "Relationship between the formant structure of different types of singing voice and the dimensions of supraglottic cavities", Folia Phoniat. 31, 238-241.

Morosow, W.P. (1977): Biofisitscheskie osnowy vokalnoi retschi, Nauka, Leningrad.

Seidner, W. and Schutte, H.K. (1982): "Empfehlung der UEP: Standardisierung Stimmfeldmessung/Phonetographie", HNO-Praxis 7, 305-307.

Seidner, W., Wendler, J., Wagner, H., and Rauhut, A. (1981): "Spektrales Stimmfeld", HNO-Praxis 6, 187-191.

Stout, B.(1938): "The harmonic structure of vowels in singing in relation to pitch and intensity", J.Acoust.Soc.Am. 10, 137-146.

Stürzebecher, E., Wagner, H., Becker, R., Rauhut, A. and Seidner, W. (1982): "Einrichtung zur simultanen Registrierung von Stimmfeld und hohem Sängerformant", HNO-Praxis 7, 223-226.

Sundberg, J. (1970): "Formant structure and articulation of spoken and sung vowels", Folia Phoniat. 22, 28-48.

Sundberg, J. (1974): "Articulatory interpretation of the 'singing formant'", J.Acoust.Soc.Am. 55, 838-844.

Wendler, J., Seidner, W., Rauhut, A., Jerzinski, P., and Anders, L. (1981): "Vocal registers in spectral analyses", Paper presented at the 8th International Symposium of CoMeT, Amsterdam.

Winckel, F. (1973): "How to measure the effectiveness of stage singers' voices", Folia Phoniat. 23, 228-233.

SESSION IX: BREATHING AND PHONATION

LARYNGEAL MUSCLE ADJUSTMENTS FOR SUSTAINED PHONATION AT LUNG VOLUME EXTREMES

Thomas Shipp*, Philip Morrissey*, and Stig Haglund**
*VA Medical Center, San Francisco, CA, USA
**Karolinska Hospital, Stockholm, Sweden

Abstract
Three young adult male subjects underwent an experimental procedure that was an addendum to a primary study. They were required to sustain phonation for 3 secs at a frequency high in their vocal range and then, without inhaling, expire as much air as they could until they were barely able to produce phonation and then phonate at the previous high frequency. EMG signals were obtained from the posterior cricoarytenoid, interarytenoid, thyroarytenoid, and cricothyroid muscles. When compared with measures obtained in the high lung volume condition, all three subjects demonstrated a marked increase of CT activity in the low lung volume condition. Two of the three subjects had dramatic decreases in a measure of medial compression for low lung volume, while the third showed a small increase in this measure. The difference in cricothyroid muscle activity as a function of lung volume is best explained by the inherent change in vertical forces acting on the vocal tract. The differences in medial compression will be examined in terms of optional physiological strategies available to subjects for controlling this factor.

Introduction

The literature in vocal pedagogy has historically stressed the importance of "proper" breathing techniques for the production of good singing quality and for the prevention of vocal fold damage. In this assertion there is a clear recognition that some type of interaction exists between respiratory and laryngeal structures to produce voice of acceptable quality in a non-abusive manner.

Unfortunately, there is little information in the research literature that bears directly on this subject. One exception is the identification (Adzaku and Wyke, 1980) of subglottal pressure receptors affecting reflexogenic motor signals to the laryngeal musculature. These sensors respond to air pressure changes and airflow variability and adjust the laryngeal structures accordingly. Our interest is in another aspect of pulmonary-laryngeal interaction involving possible laryngeal adjustments that accompany changes in lung volume level. We wanted to examine whether intrinsic laryngeal muscle adjustments occurred to maintain phonation frequency and intensity over the full range of lung volume levels. Specifically, we wanted to assess whether muscles controlling both the vocal fold adduction-abduction movement and tissue stiffening would adjust activity during phonation at lung volume extremes. It seemed apparent that the radical difference in mechanical forces and aerodynamic potential at lung volume extremes could be accompanied by some type of laryngeal compensatory mechanism to maintain voice frequency and intensity values.

Two groups of subjects participated in the experiment. The first comprised three young adult males with no training in voice production, although they could match their vocal pitch to a target stimulus. These three men were participants in a larger study involving many physiologic measures obtained during a vocal reaction time experiment. Their participation in the current lung volume task was an addendum to the major study. This fact should be kept in mind to understand the lack of rigid voice frequency and intensity controls exercised with these three subjects. Simultaneous EMG activity was sampled from three intrinsic laryngeal muscles: cricothyroid (CT), posterior cricoarytenoid (PCA), and interarytenoid (IA) with a measure of subglottal air pressure (P_s) obtained directly via an intratracheal catheter. The task for these untrained subjects was to sustain the vowel /a/ in the upper part of their frequency range beginning at or near full lung volume. After three seconds, phonation was terminated by vocal fold abduction allowing a free flow of expired air to pass through an open glottis. When he was "low" in his tidal volume level, he resumed phonation at the same frequency and intensity as before and continued until his air supply was exhausted.

Subsequently, each subject repeated this same task in the low part of his voice frequency range.

The second subject group comprised three highly-trained singers, two of whom had extensive experience singing professionally in concert and opera. Hooked wire electrodes were inserted unilaterally to each subject's cricothyroid muscle. Inadvertently, one additional muscle was sampled from one subject. During initial electrode placement the wires were determined to be in the sternothyroid (ST), a strap muscle responsible for lowering the larynx. Since the next insertion attempt accurately placed another electrode pair in the CT, both muscles were sampled during his experimental tasks.

Following verification of accurate electrode placement, the subjects sustained the vowel /a/ at the two lung volume extremes. For these trained singers, however, assessment was made of their full concert frequency range, which was then converted to semitones and divided into quarters so that each subject produced phonation at his 25%, 50% and 75% frequency levels. At the start of the experimental procedure, a physiologic calibration maneuver was performed by all subjects to put the sample muscles through their full contractile range. These minimum and maximum muscle signals were averaged and integrated over a 500 msec window and assigned a value of "zero" and "100" respectively. Any subsequent muscle activity obtained during the study was compared to that subject's metric and assigned a value along the muscle activity scale. For the experimental tasks the first 500 msec of muscle activity at the onset of phonation at high lung volume and the penultimate 500 msec window in the low lung volume phonation were extracted and the resulting values used to represent a muscle's participation at each lung volume extreme.

Results

Figs. 1 and 2 display the individual subject data from singers and non-singers of their CT activity at lung volume extremes. The three non-

Fig. 1.

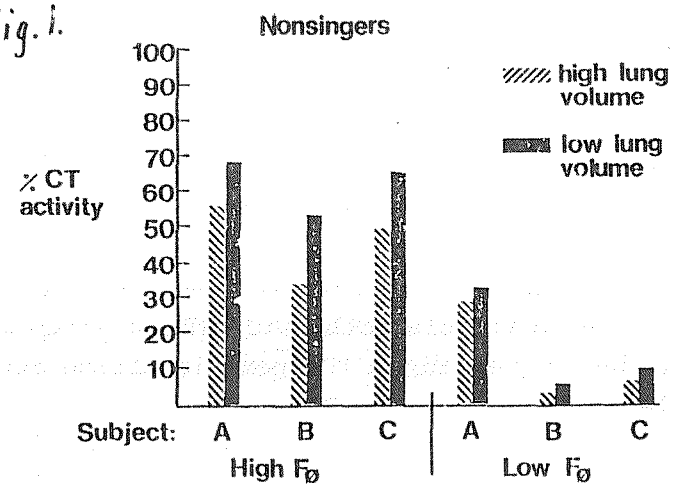
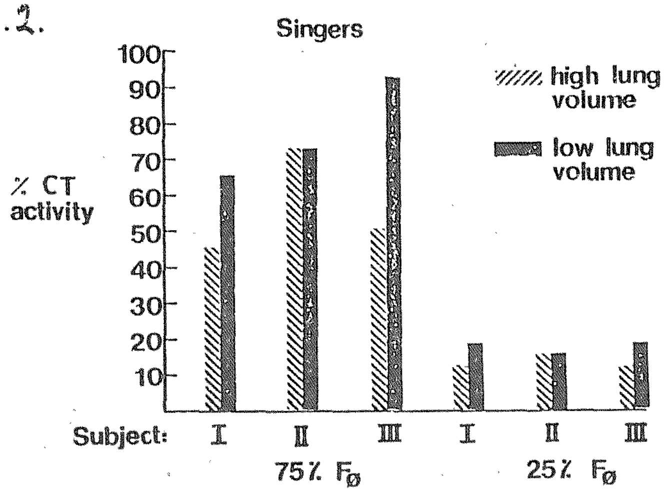


Fig. 2.



singers show a similar pattern of CT activity consistently being greater at low lung volume compared to high regardless of voice frequency. The magnitude of muscle increase between lung volume levels varied between subjects, but the direction of change was the same. A similar pattern is observed for the singers except Subject 2 who maintained the same CT level at lung volume extremes at both voice frequency levels.

Fig. 3.

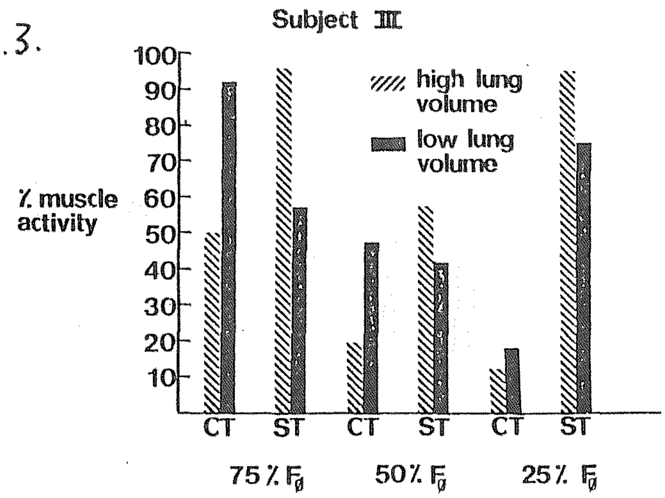
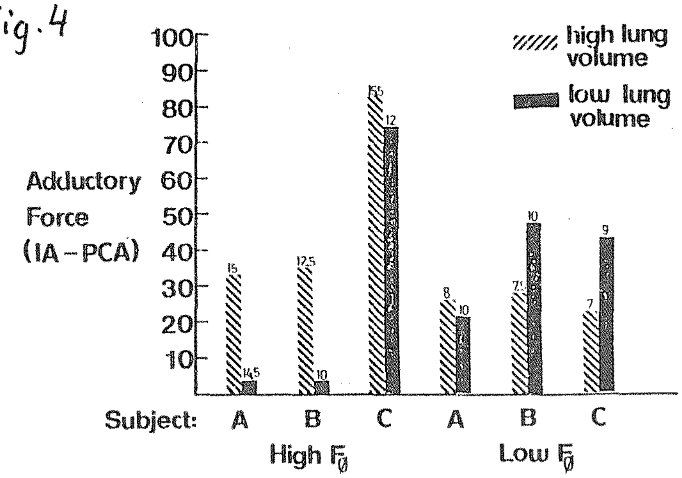


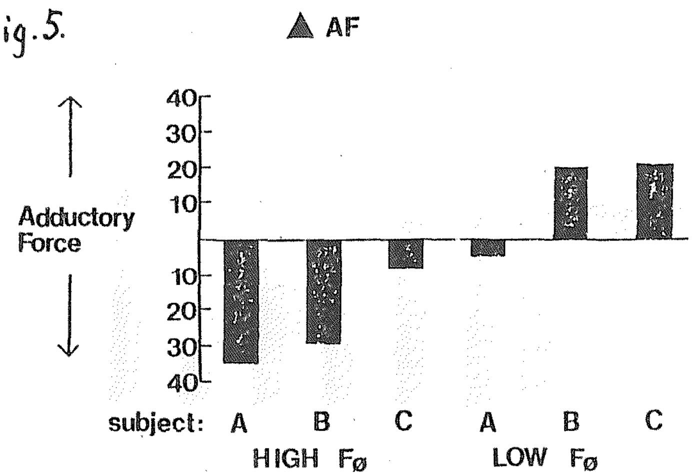
Fig. 3 shows subject 3's CT and ST activity at the three voice frequency levels sampled. It can be seen that the ST displayed a reciprocal pattern from that of the CT; that is, as CT activity increased going from high to low lung volume at each frequency, ST activity decreased.

Fig. 4



Only with the three untrained subjects were data obtained on adductory-abductory gestures as a function of lung volume. It should be remembered that vocal effort was not controlled. Since the measure we

Fig. 5.



were trying to extract was one indicating the magnitude of vocal fold adduction in the various conditions, we utilized a measure termed "adductory force" (AF) which was computed by taking the percentage of IA activity and subtracting the PCA percentage for that segment. These values are shown in both Figs. 4 and 5. In the high F_0 condition, AF declined in all three subjects when going from high to low lung volume. It can be noted that these large differences in AF were accompanied by relatively small adjustments in P_s as represented by the number atop each histogram bar. In the low F_0 condition the opposite trend appears. With the exception of Subject A, the other two increased AF when going from high to low lung volume.

Discussion

These results answer the experimental question affirmatively: muscles involved in frequency control and those controlling adductory force appear to make adjustments as a function of lung volume level. There is even an indication from one subject that at least one extrinsic laryngeal muscle responds differentially to lung volume extremes during phonation. Why these changes occur is, of course, open to many interpretations. My own is based principally on previous research by me and my colleagues conducted on both trained and untrained voice users.

Our studies of vertical laryngeal position in singers showed that they maintained their larynx low in the neck while producing phonation throughout their voice frequency range, as contrasted with untrained subjects who positioned their larynx upward as voice frequency was elevated. Further, singers' laryngeal position was unaffected by experimentally manipulated jaw positions, while non-singers were. From these data and others we concluded that singers use a low laryngeal position principally to maintain a desirable voice quality throughout their voice frequency range; whereas non-singers position their larynx upward to assist in voice frequency control. The mechanics of this action seem to be related to the vertical stiffening forces generated by any mechanical action that elongates the structures comprising the tracheobronchial tree. These forces can be generated by elevating the larynx in the neck and/or contracting the diaphragm fully to achieve full lung volume. This stretching thins out and stiffens all tissues lining the vocal tract including the vocal fold, the result of which is increased vibration frequency and, to some extent, a less complex vibratory pattern. By this we mean that the thinning and stiffening process inhibits the complex rolling, undulating, phase-shifting type of vibratory cycle that high speed motion pictures of the larynx show to distinguish singers from non-singers. We presume that singers employ any strategy facilitating the generation of such pattern despite the state of the respiratory system or any other physical posture that might interfere. It stands to reason that the singer must employ increased ST activity to keep the larynx low and, perhaps, it may be that singers always utilize some degree of CT contraction for this purpose.

Now, specifically with respect to the findings of cricothryroid activity, it seems that all subjects supplemented CT activity with some degree of vertical stiffening provided by the descended diaphragm during the full lung volume condition. When the diaphragm is subsequently elevated in low lung volume, this reduction in stiffening must be counteracted by increased longitudinal stiffening, i.e., augmented CT activity. Thus, singers and non-singers usually increase CT activity in low lung volume phonation to maintain voice frequency.

For singers to offset tissue stiffening by diaphragm descent, they increase ST activity to put more slackness in vocal fold tissue to encourage the efficient phase shifting vibration (Titze, 1980) and they increase CT activity to maintain the target frequency. When the diaphragm elevates in low lung volume, the folds are in a much more slackened condition and less ST activity is needed to produce the desired vibratory motion.

Examining the results of adjustments found in the adduction-abduction gesture indicate that the pattern of decreased AF in the high F_0 condition when going from high to low lung volume might well be a laryngeal gesture calculated to facilitate continued vocal fold vibration in the face of greater vocal fold impedance caused by the previously noted increased CT activity. The opposite effect of increased AF noted in low F_0 phonation may be to conserve air and control vocal intensity (Fant, 1983). In low F_0 phonation the folds are thick and slack having small impedance; therefore, increased AF would conserve transglottal airflow and augment P_s without interrupting vibration. AF acts in a 3-fold way to control aspects of phonation:

1. Maintains a relatively constant subglottal air pressure for controlling vocal intensity in the face of radically different mechanical conditions.

2. Conserves air when the supply is limited.

3. Sets up a vocal fold configuration facilitating vibration under markedly different respiratory and laryngeal conditions.

Summary

The present study provides evidence that there exists a motor control mechanism that adjusts laryngeal behavior to accomodate for altered forces that result from changes in lung volume level. For the most part these adjustments are automatic and some may manifest themselves differently in trained and untrained voice users.

References

Adzaku, F.K. and Wyke, B.D. (1980): "Laryngeal subglottic mucosal mechanoreceptor reflexogenic influences on laryngeal muscle activity", Ann. Royal Coll.Surg.Engl. 62, 426-431.

Fant, G. (1983): "Preliminaries to analysis of the human voice source", STL-QPSR 1/1983, 1-27.

Izdebski, K. and Shipp, T. (1979): "The effect of vertical laryngeal position on singers' vowel formants", J.Res.Sing. 2, 1-9.

Ohala, J. (1972): "How is pitch lowered?" J.Acoust.Soc.Am. 52, 124.

Shipp, T. (1975): "Vertical laryngeal position during continuous and discrete frequency change", J.S.H.R. 18, 707-718.

Shipp, T. and Izdebski, I. (1975): "Vocal frequency and vertical laryngeal positioning by non-singers and singers", J.Acoust.Soc.Am. 83, 1104-1106.

Titze, I.R. (1980): "Comments on the myoelasthc-aerodynamic theory of phonation", J.S.H.R. 23, 495-510.

ACTIVATION OF THE DIAPHRAGM DURING SINGING.
A study of transdiaphragmatic pressures.

J. Sundberg*, R. Leanderson**, C. von Euler**, and H. Lagercrantz**
*Dept. of Speech Communication and Music Acoustics, KTH
**Karolinska Institutet, Stockholm, Sweden

Abstract

The oesophageal and gastric pressures during singing are measured in four professional singers performing singing tasks requiring rapid changes of subglottal pressure. Evidence for a consistent use of the diaphragm is found in all subjects. Two subjects activate the diaphragm in the first part of the phrase, presumably as long as the passive expiratory recoil forces of the breathing system is higher than the subglottal phonatory pressure. One subject exhibits a diaphragmatic activity through the entire phrase.

Introduction

There has been much discussion on the role of the diaphragm in regulating subglottal pressure during singing (see e.g., Hixon, 1979; Proctor, 1980). Previous studies have concluded that diaphragmatic activity only occurs during a fairly short period in the beginning of phonation (Bouhuys et al., 1966). Fig. 1 by Draper et al. (1959) illustrates this case. Here, the diaphragm is seen to relax already at the onset of phonation and to remain passive throughout the utterance, while the inspiratory intercostals are left alone to counteract the gradually decreasing passive expiratory recoil forces during the first part of the phrase. Later in the phrase the expiratory muscles are seen to be recruited increasingly in order to maintain a stable subglottal pressure. This is shown also in Fig. 2 from an EMG study of the inspiratory external intercostals and the expiratory internal intercostals and tidal volume during sustained phonation (Sears and Newsom Davis, 1968).

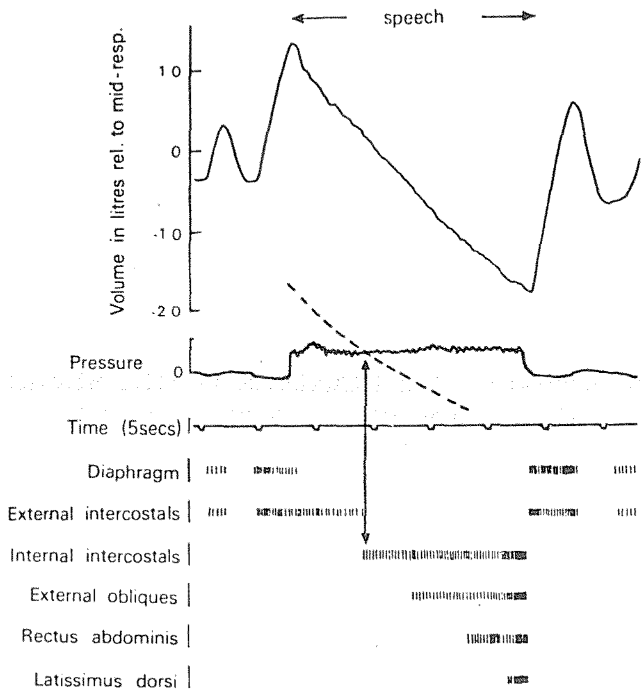


Fig. 1. Timing relationships between lung volume and the EMG signals from various respiratory muscles during sustained phonation. From Draper et al. (1959).

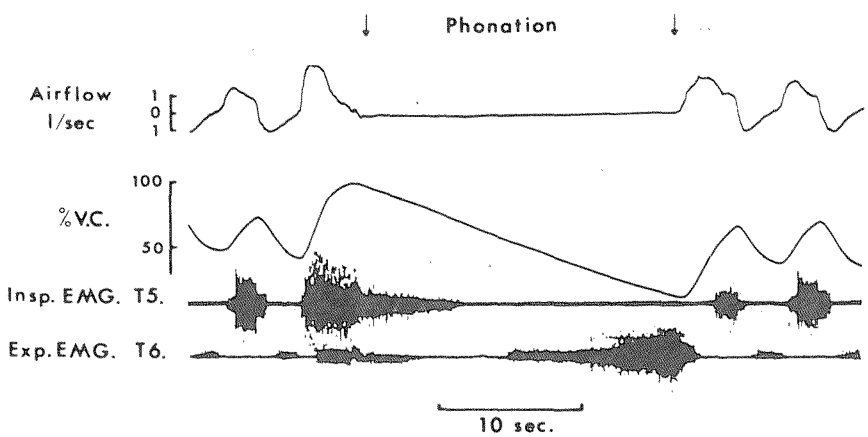


Fig. 2. Timing relationships between airflow and lung volume expressed in % of vital capacity (two top curves), and EMG signals from insipratory and expiratory intercostal muscles (two bottom curves) during sustained phonation. From Sears and Newsom Davis (1968).

On the other hand, many singers and singing teachers feel convinced that the diaphragm is active throughout the entire phrase for the purpose of regulating the subglottal pressure. So far, however, the case that the diaphragm participates in this counteracting of the passive expiratory forces has not been documented.

There have been arguments against the assumption of a diaphragmatic activity throughout the phrase. Electromyographic studies have suggested that the intercostals and the oblique abdominal muscles are comparable to limb muscles regarding a highly developed proprioceptive feedback control system (Lansing and Meyerink, 1981). The diaphragm, on the other hand, even though it is the main inspiratory muscle, has been found to possess a poor supply of proprioceptive receptors (Corda et al., 1965; von Euler, 1979). This has caused some to believe that the diaphragm might be a more clumsy tool than the intercostal and abdominal muscles for accurate and rapid control of subglottal pressure. However, the diaphragm can be guided by the important intercostal-to-phrenic reflexes, which, by way of both proprio-spinal and spiro-bulbar-spinal pathways join the intercostal proprioceptors to the phrenic motoneuron pool (Decima et al., 1969; Decima and von Euler, 1969; Remmers, 1973).

Previous studies of the role on the diaphragm in singers have focused on sustained phonation mainly. The purpose of the present investigation was to reexamine the role of the diaphragm under conditions of rapid changes of subglottal pressure in singing. In particular, we wanted to test the possibility that the diaphragm might be used for balancing the expiratory forces (passive as well as active) during a longer portion of the phrase (von Euler, 1982).

Experiment

Our investigation was carried out on four highly trained singers. In a natural upright position they performed different singing tasks. The respiratory volume events were recorded by inductive spirometry ("Respitrace"), which takes into account the contributions to the total volume from both the rib cage and the abdomen.

The oesophageal and gastric pressures were measured by means of a catheter furnished with two pressure transducers, one at the tip and one about 10 cm from the tip ("Gaeltec"). The subjects swallowed the catheter so that the tip was located in the gastric ventricle and the other transducer in the oesophagus. The pressure difference, i.e., the transdiaphragmatic pressure was recorded directly from a differential amplifier. The transdiaphragmatic pressure reflects in a quantitative manner the tension of the diaphragm caused by active contraction (see, e.g., Newsome Davis et al., 1976), or by passive stretching in its uppermost position during expiration. The oral pressure during /p/-occlusion was recorded from a pressure transducer with a thin plastic catheter, that the subject held in the mouth corner. Fundamental frequency was measured from a contact microphone fastened to the subject's throat, and the relative sound level was recorded by a dynamic microphone about 30 cm from the lips. During the experiments, these signals, except for the sound level, were simultaneously recorded on a multichannel rectilinearly writing pen recorder.

In a previous paper we presented results showing that singers sometimes use different subglottal pressures during a /p/-occlusion and a preceding/following vowel (Leanderson et al., 1983). Here we will present and discuss data recorded when the subjects' task required rapid changes of subglottal pressure.

Measurements

Fig. 3 gives a survey of the various parameters studied. It was recorded during repeated pronunciation of the syllable /pa/. The top curve shows the oral pressure and the second curve from top shows the changes in lung volume. This curve sometimes showed artifacts, and, hence, the lung volume curves could not be appropriately calibrated. The next two curves in the same figure represent oesophageal and gastric

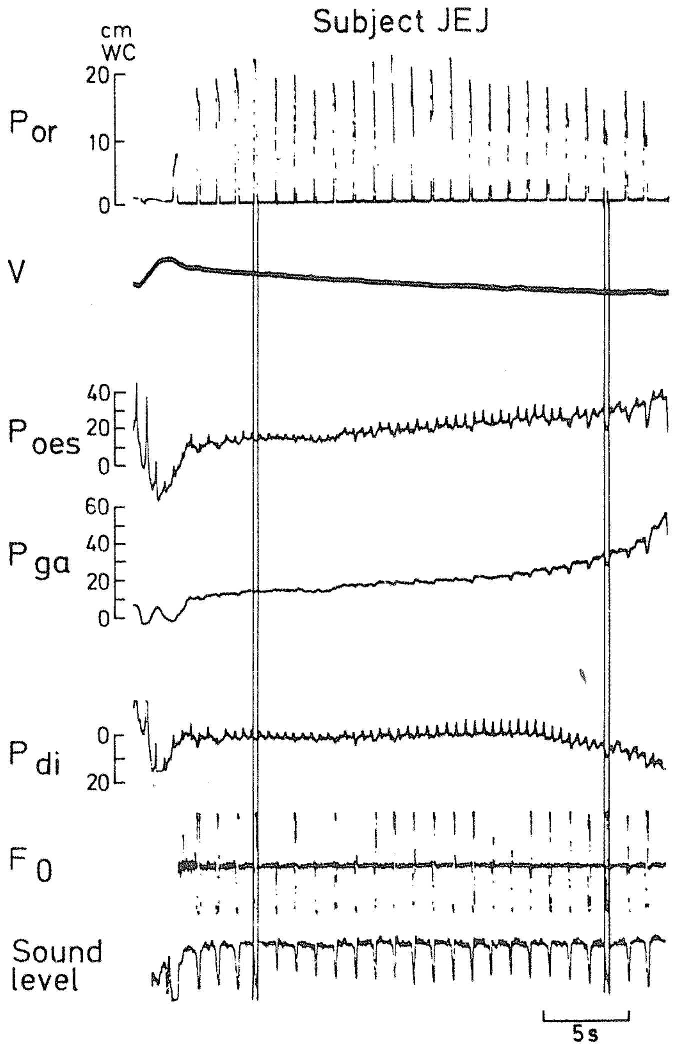


Fig. 3. Survey of the various parameters studied: P_{or} is the oral pressure during /p/-occlusion; V is the lung volume; P_{oes} and P_{ga} are the pressures in the oesophagus and in the gastric ventricle, and the P_{di} is the transdiaphragmatic pressure; F_0 is the fundamental frequency. The bottom curve shows the sound level. The recording was obtained when the subject repeatedly pronounced the syllable /pa/.

pressures. The transdiaphragmatic pressure is shown in the next graph. The two bottom curves show fundamental frequency and sound level.

The phrase is started with a strong negative pressure reaching about -20 cm water column (WC). Probably, this low pressure reflects the resistance associated with nasal inhalation, which is typical for most trained singers. In this example, oesophageal and gastric pressures are similar through about three quarters of the phrase. Hence, the transdiaphragmatic pressure stays close to zero (only disturbed by the cardiogram). In the last quarter of the phrase the transdiaphragmatic pressure rises to about 15 cm WC. Presumably, this mirrors a passive stretching of the diaphragm.

Results

Fig. 4 illustrates the intersubject variability. In this experiment the subjects were asked to sing a sustained tone alternating between

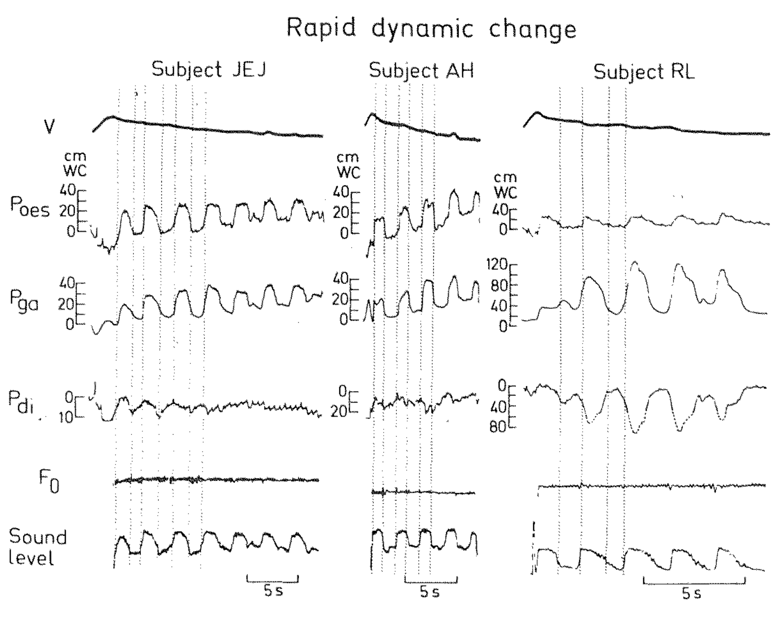


Fig. 4. Recordings obtained when the three subjects sang a sustained tone, the intensity of which was suddenly and rhythmically changed between <u>subito forte</u> and <u>subito piano</u>. Symbols as in Fig. 3.

subito forte and subito piano (i.e., with rhytmically and suddenly changing intensity). Different subjects applied different strategies with respect to the diaphragmatic activation during phonation, as can be seen in the figure. In all three subjects the oesophageal pressure changed rhythmically, approximately in synchrony with the sound level, even though subject RL apparently preferred gradual rather than sudden decreases in subglottic pressure and sound level.

With respect to the average transdiaphragmatic pressure, the subjects behaved differently. Subject JEJ showed essentially parallel curves for oesophageal and gastric pressures. In the beginning of the phrase the transdiaphragmatic pressure is varied in synchrony with the sound level changes. It is close to 0 cm WC during the loud segments of the tone and increases somewhat during the soft segments. This implies the absence of constant transdiaphragmatic pressure during the loud portions of the tone.

In subject AH the gastric pressure was also fairly parallel to that of the oesophageal pressure, so that the transdiaphragmatic pressure showed only small variations in synchrony with the changes in sound level. However, this pressure varied around a rather constant value of 10 cm WC through two-thirds of the phrase, and then approached zero. This indicates a rather moderate but constant diaphragmatic activity in the first part of the phrase.

Examination of the details in the registrations reveals that in both subject JEJ and subject AH the transdiaphragmatic pressure increased at the moment when a reduction of subglottic pressure was required for the decrease in loudness. These momentary increases in diaphragmatic activity ceased during the later part of the phrase in subject JEJ. This implies that these subjects reduced the pressure generated by the abdominal wall contractions by activations of the diaphragm.

In subject RL the transdiaphragmatic pressure variations in synchrony with the loudness shifts were much greater than in the two other subjects. Moreover, his strategy seems to be quite opposite; in decreases of loudness, there seems to be no diaphragmatic activity at all, while during the loud parts of the tone, the transdiaphragmatic pressure reached very high values, over 80 cm WC. This implies that the diaphragm is forcefully contracting against the pressure generated by the abdominal wall in loud singing.

Fig. 5 illustrates the behavior of the same subjects when they were repeatedly singing a sequence of octave intervals. All subjects were

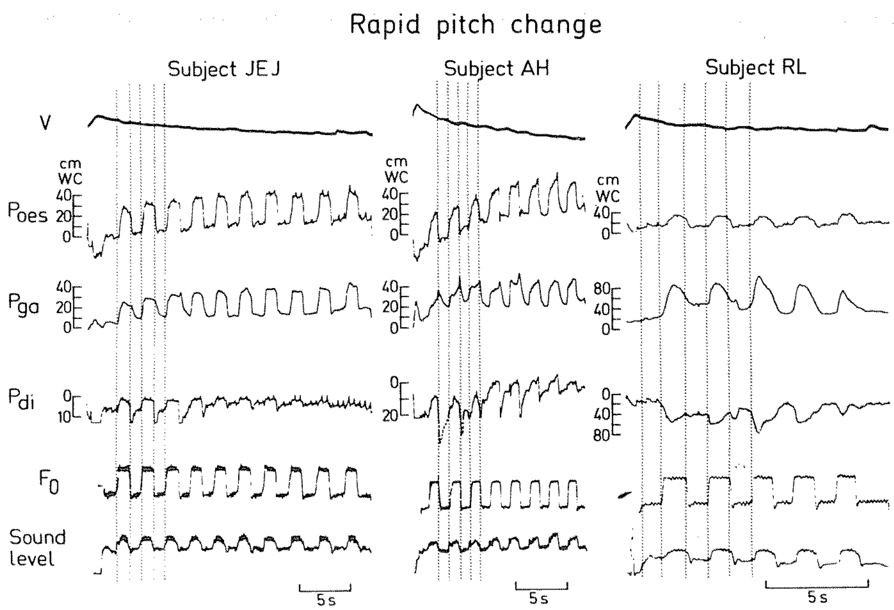


Fig. 5. Recordings obtained when three subjects repeatedly sang a sequence of octave intervals. Symbols as in Fig. 3.

seen to use a higher oesophageal pressure for the higher than for the lower pitch. The increases in the oesophageal pressure are achieved by a rhythmical increase of the gastric pressure. The transdiaphragmatic pressure indicated rhythmical contractions of the diaphragm in all subjects, at least during the first part of the phrase.

Regarding the individual subjects, JEJ showed a regular pattern of diaphragm activation through the first half of the phrase. He contracted his diaphragm strongly and suddenly at the moment when he left the high pitch. Then, he retained a moderate diaphragmatic activity, and relaxed it during the high pitch, so that the transdiaphragmatic pressure returned to zero. Finally, during the late part of the phrase, he relaxed his diaphragm completely. Subject AH showed a similar diaphragmatic behavior, but he retained a considerable transdiaphragmatic pressure also when singing the high pitches during the first third of the phrase. Thus, these two subjects seemed to use their diaphragm in order to reduce the subglottal pressure for the lower pitch.

Subject RL, again, exhibited the opposite behavior. He showed maximum transdiaphragmatic pressure difference during the high pitch. His diaphragmatic activity resulted in pressure differences as high as 50 cm WC or more. As when singing the tone with varied loudness, this subject showed a more gradual change in gastric pressure than the other subjects. Another difference between subject RL and subjects JEJ and AH was that subject RL maintained his diaphragmatic activity throughout the phrase, keeping a transdiaphragmatic pressure of about 20 cm WC for the last tone in the phrase.

Discussion

In contrast to what has been concluded in previous investigations (Ladefoged, 1967; Proctor, 1980), we have found variations of the transdiaphragmatic pressure during singing, apparently reflecting a diaphragmatic activity. In rapid phonatory changes, such as producing a voiceless consonant in a VCV sequence, or a sudden change of pitch and intensity, as in the present investigation, the diaphragm probably acts synergistically with the external intercostals both in counteracting the strong elastic expiratory recoil forces and in establishing rapid and precise changes in subglottal pressure. In one subject, RL, there were indications of a strong diaphragmatic activity throughout the phrase, and the diaphragm was apparently used in order to reduce a fairly strong 'overpressure' generated by the abdominal muscles.

A contraction of the diaphragm contributes somewhat to an expansion of the rib cage (see, e.g., Goldman and Mead, 1973; Mortola and Sant'Ambrogio, 1978). In this way a diaphragmatic contraction extending over the entire phrase seems to be in accordance with the practise of many pedagogues and singers who strive to hold the rib cage expanded as long as possible through the phrase. This behavior probably requires a motor synergy of both the external intercostals and the diaphragm. The purpose of this strategy may be associated, habitually or reflexively, with a laryngeal abduction activity. It is probably typical for untrained voices to use a high subglottal pressure combined with a high adduction activity ("medial compression") during phonation at high pitch and/or high intensity (Gauffin and Sundberg, 1980). The result then is "pressed phonation" (Sundberg and Gauffin, 1979; Rothenberg, 1972). If a diaphragmatic contraction tends to be associated with laryngeal abduction during phonation, such contraction may help the singer to avoid pressed phonation. The testing of this hypothesis is left to future investigations.

Conclusions

From the data presented above we conclude that:

(1) there are considerable intersubject variations with respect to the diaphragmatic activity in singing phrases which require rapid changes in subglottal pressure;

(2) in some subjects the diaphragm is active in the first part of the phrase only, probably for lung volumes associated with expiratory elastic recoil forces greater than the subglottal pressure; this diaphragmatic activity is used for accomplishing rapid drops in subglottal pressure; in other subjects a diaphragmatic activity is used during the entire phrase in high pitched and/or loud tones in order to reduce the 'overpressure' generated by the abdominal wall muscles.

References

Bouhuys, A., Proctor, D., and Mead, J. (1966): "Kinetic aspects of singing", J.Appl.Physiol. 21(2), 483-496.

Corda, M., von Euler, C., and Lennerstrand, G. (1965): "Proprioceptive innervation of the diaphragm", J. of Physiol. (Lond.) 178, 161-169.

Decima, E.E., von Euler, C., and Thoden, U. (1969): "Intercostal-to-phrenic reflexes in the spinal cat", Acta Physiol.Scand. 75, 568-579

Decima, E.E, and von Euler, C. (1969): "Excitability of phrenic motoneurones to afferent input from lower intercostal nerves in the spinal cat", Acta Physiol.Scand. 75, 580-591.

Draper, M.H., Ladefoged, P., and Whitteridge (1959): "Respiratory muscles in speech", J. Speech and Hearing Res. 2, 16-27

von Euler, C. (1982): "Some aspects of speech breathing physiology", in (S. Grillner, ed.): Speech Motor Control, Pergamon Press, Oxford, 93-103

Gauffin, J. and Sundberg, J. (1980): "Data on the glottal voice source behavior in vowel production", STL-QPSR 2-3/1980, 61-70

Goldman, M.D. and Mead, J. (1973): "Mechanical interaction between the diaphragm and the rib cage", J. of Appl.Physiol. 22, 407-422

Hixon, T.J. (1979): "Chest wall shape in singing" Transcripts of the VII Symposium Care of the Professional Voice, 1978, The Voice Foundation, New York

Ladefoged, P. (1967): Three Areas of Experimental Phonetics, Oxford University Press, London

Lansing, R.W. and Meyerink, L. (1981) "Load compensating responses of human abdominal muscles", J. of Physiol. 320, 253-268

Leanderson, R., Sundberg, J., von Euler, C., and Lagercrantz, H. (1983): "Diaphragmatic control of the subglottic pressure during singing", to be published in Transcripts of the XII Symposium Care of the Professional Voice

Mortola, J.P. and Sant'Ambrogio, G. (1978): "Motion of the rib cage and the abdomen in tetraplegic patients", Clinical Science and Molecular Med. 54, 25-32

Newsom Davis, J., Goldman, M., Loh, L., and Casson, M. (1970): "Diaphragm function and alveolar hypoventilation", Quart.J. of Med. New Series XLV, No. 177, 87-100

Proctor, D. (1980): Breathing, Speech, and Song, Springer-Verlag, Wien

Remmers, J.E. (1973): "Extra-segmental reflexes derived from intercostal afferents: phrenic and laryngeal responses", J. of Physiol. 233, 45-62

Rothenberg, M. (1972): "The glottal volume velocity waveform during loose and tight voiced glottal adjustment", Proc.Seventh Int.Congr. of Phonetic Sciences, Montreal, Mouton, The Hague, 380-388

Sears, T.A. and Newsom Davis J. (1968): "The control of respiratory muscles during voluntary breathing", Ann. New York Sci. 155, Art 1, 183-190

Sundberg, J. and Gauffin, J. (1979): "Waveform and spectrum of the glottal voice source", in (B.Lindblom & S.Öhman, eds.) Frontiers of Speech Communication Research, Academic Press, London, 301-320

ARTICULATORY PERTURBATION OF PITCH IN SINGERS DEPRIVED OF AUDITORY FEEDBACK

Sten Ternström*, Johan Sundberg*, and Anders Collden**
*Dept. of Speech Communication and Music Acoustics, KTH, Stockholm
**State Conservatory of Music, Stockholm, Sweden

Abstract

Ideal intonation in choir singing requires a high accuracy regarding fundamental frequency. Experience of choir conductors suggests that manipulation of the vowel timbre sometimes affects intonation. Articulatory changes, such as shifting the shape of the toungue body, might affect pitch. In the ideal case a singer would detect such a pitch effect by auditory feedback and compensate for it accordingly; however, poor auditory feedback is known to occur in choral practise.

Singer subjects wearing headphones with noise masking the auditory feedback were asked to change the sung vowel while maintaining a constant pitch. The associated changes of phonation frequency were analyzed. The vowels /i/ and /y/ were found to raise the fundamental frequency. The effects were found to be larger in the absence of auditory feedback.

Introduction

Accurate control of voice fundamental frequency (FØ) is a primary technical concern for choir singers. In a previous investigation (Ternström & Sundberg, 1982), it was found that the intonation of the individual singer may be affected by certain properties of the sound from the rest of the choir. Absence of vibrato, strong high partials, and prominence of the common partials in a consonant dyad (fifth and major third) were found to facilitate beat-free intonation.

Given the mechanical complexity of the voice organ, a change in articulation is likely to affect also the fundamental frequency. The "intrinsic pitch of vowels" is in fact a well established concept among speech researchers. A singer, on the other hand, is likely to avoid such effects by means of auditory and/or proprioceptive feedback. In practice, however, the auditory feedback may be poor or even absent.

The purpose of the present study was to investigate the influence of articulatory movements on fundamental frequency in experienced choir singers. Subjects singing with and without auditory feedback were asked to change from one vowel to another while attempting to maintain a constant pitch.

Experiment

The six subjects (two female, four male) were students of Choir Leader Pedagogy at the State Academy of Music in Stockholm; they all had formal vocal training. The experiment was made a part of their course in Voice Acoustics.

The subjects sang both with and without masking of the auditory feedback. The feedback was effectively eliminated by means of noise presented over headphones (cf. Elliott & Niemoeller, 1970). An appropriate masking was achieved at a comfortable masker SPL by band-pass filtering the noise.

A short written description of the experiment was given in advance. The subjects were instructed to sing the two consecutive vowels in a given vowel pair at a comfortable loudness and pitch, and, in particular, to attempt to maintain the same pitch regardless of the vowel change. The subjects were recommended to imagine the pitch in advance, and to make fairly rapid vowel shifts. They were instructed to start singing the vowel pair as soon as they heard the masker noise, and to stop before the noise was terminated after 5 seconds. The vowel pair was then to be repeated without masker noise. There was no prior practice.

The vowel pairs were chosen so as to involve changes in only a few articulatory parameters. Thus, pairs 1-3 (see Table I) are back vowels and involve a lowering of the jaw. In pairs 4-6 the jaw is lowered and the shape of the tongue body is changed. Pairs 7-9 involve lip spreading mainly. Pairs 10-12 require a raising of the tongue body. The subjects were given a list with the vowels in CVC key words with the vowel characters underlined.

Table I Vowel pairs.

	Swedish CVC key words Vowel 1	Vowel 2	IPA symbols Vowel 1	Vowel 2
1	bor	borr	/u/	/ɔ/
2	borr	barr	/ɔ/	/a/
3	bor	barr	/u/	/a/
4	bil	väl	/i/	/ɛ/
5	väl	barr	/ɛ/	/a/
6	bil	barr	/i/	/a/
7	byt	bit	/y/	/i/
8	bit	bet	/i/	/e/
9	byt	bet	/y/	/e/
10	bål	böl	/o/	/ø/
11	böl	by	/ø/	/y/
12	bål	by	/o/	/y/

Each subject made two runs separated by a pause of about one hour. Each run lasted for two and a half minutes. In the second run, the word pairs were reversed to give transitions between the same vowels as in the first run, but in the opposite direction.

The equipment used was arranged as shown in Fig. 1. The subject was seated alone at a table in a small and quiet room. On the table there was a microphone, the list of CVC word pairs, and a paper with the instructions, which also were explained to each subject by the experimenter. The subjects wore headphones and a contact microphone on the throat near the larynx. The rest of the equipment and the experimenter were in a separate room.

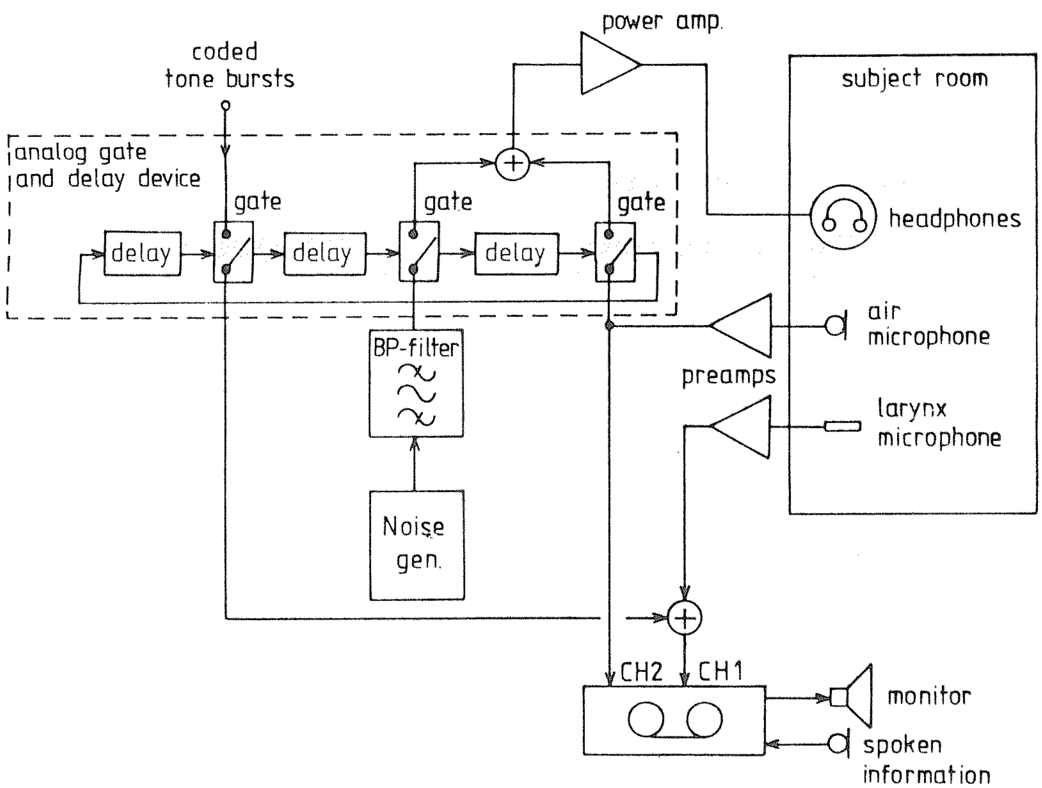


Fig. 1. Connection diagram for experimental equipment. This setup allows the experiment to run for one subject at a time without operator intervention.

The headphones with the masker noise were of the "closed speaker"-type (Clark model 100). A band-pass filtered white noise and the sound from the air microphone on the table were automatically gated in alternation to the headphones. The gating sequence is illustrated in Fig. 2. The SPL and roll-off frequencies of this masking noise were adjusted for each subject before his or her run of the experiment. Typical values were 110 dB SPL, and 50 Hz high-pass, 1250 Hz low-pass, with a constant filter slope of 24 dB/oct. Complete masking without discomfort was always ob-

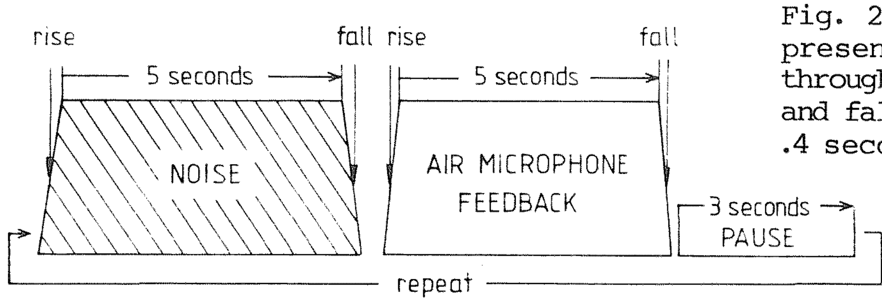


Fig. 2: Sound sequence presented to subjects through headphones. Rise and fall times are about .4 seconds.

tained except occasionally for the vowel /i/ in male subjects. In those cases the subjects reported that they could hear that they were singing something, but that they were unsure of the exact pitch.

The contact microphone and air microphone signals were recorded on separate tracks of a tape recorder. The signals from the contact microphone were intersparsed with pulsed tone bursts (inaudible to the subjects) for easier identification during the measurements.

First, the subject was correctly and comfortably installed. The gate was then started, and the cyclic sequence in Fig. 2 was allowed to run two or three times to let the subject acquaint himself with the time intervals involved. The run was then started: the sequence was allowed to cycle 12 times, one for each vowel pair on the list.

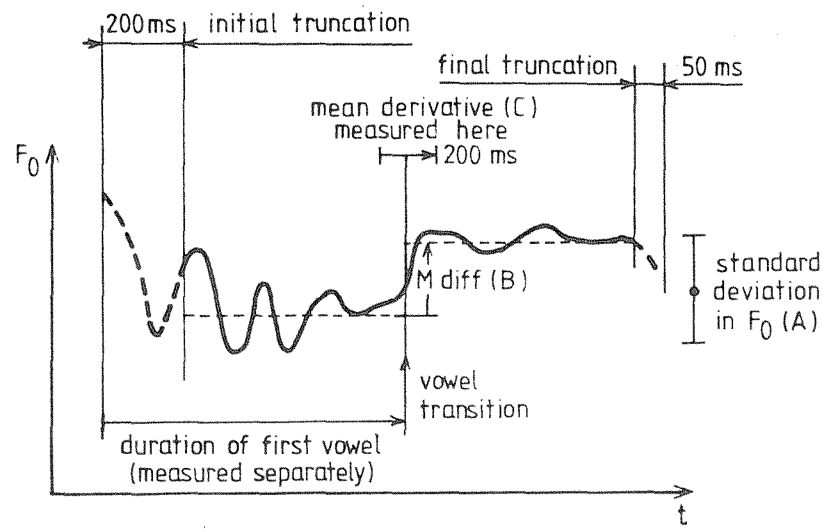
Analysis

The moment of vowel transition in each vowel pair (i.e., the duration of the first vowel) was determined in the following way. The level of the air microphone signal passed through a low-pass filter (-24 dB/oct at 315 Hz) was recorded on a Brüel & Kjaer 2307 level recorder. The filter was then switched to high-pass, and the procedure was repeated. The formant frequency changes associated with the vowel transitions appeared clearly in either or both of these signals. This gave a localization of the transition to an accuracy of 50 ms or less.

A computer program for the analysis of the fundamental frequency behavior of audio signals has recently been written by Thomas Murray and Sten Ternström at the Dept. of Speech Communication and Music Acoustics (Ternström, 1983). The program samples a hardware period-time counter and an amplitude meter, 100 times per second, and records the obtained sample values in a data file. The fundamental frequency and its derivative as functions of time are calculated and plotted together with their mean and standard deviation values. The time window and/or resolution in frequency may be chosen, and a number of numerical filters may be included as desired. The program also produces FØ and derivative histograms.

The following fundamental frequency contour characteristics were measured for each recorded vowel pair (cf. Fig. 3):

Fig. 3: Example of FØ curve with the analyzed quantities indicated (see text). Typical duration of entire vowel pair is 2.5-4 seconds; typical variation in FØ is ± 20 cent. Not drawn to scale.



A) standard deviation of FØ over the entire vowel pair

B) difference in mean FØ before and after the vowel transition (Mdiff)

C) average FØ derivative in a 200 ms interval surrounding the vowel transition.

A median filter over five samples (50 ms) was always used, to remove sporadic errant values from the period-time counter. The standard deviation included initial and final FØ transients. For the mean-value measurements, 200 ms at the beginning and 50 ms at the end of each vowel pair were truncated, so as to lessen the influence of onset and offset transients. For the derivative measurement, a 170 ms mean-value filter was added, which reduced the influence of vibrato.

Results

A typical example of the graphic output of the FØ program is given in Fig. 4. It shows the fundamental frequency curve and histogram of subject MK singing /u/ -> /ɔ/, with and without masker noise. The vowel transitions (as measured separately) are marked with arrows. The subject's vibrato (± 30 cent) was eliminated by means of a 190 ms mean-value filter. This token also serves well to illustrate several aspects of typical subject behavior.

There is a large initial transient in FØ in the case with masker noise. When the masker is removed, the subject centers on a different, higher pitch than before (although he would believe that he is singing the same pitch!). The pitch difference is large enough to produce two quite separate peaks in the histogram. The peak pertaining to the case without masking is much taller and narrower than the other, reflecting a higher pitch precision in the absence of masking. We noted that for some subjects the pitch difference was systematic, and smaller during the second run; however, in the present study we are not concerned with pitch-matching.

A) The Standard Deviation, SD, was usually larger with masking. This implies a less stable FØ in the absence of auditory feedback. The effect was small but statistically highly significant (see Table II). Standard deviation caused by tape flutter was typically less than 5 cents, i.e., well below that of the subjects.

B) Let <u>Mdiff</u> denote the difference between mean FØ before and after the vowel transition. Consider first each subject's absolute value of Mdiff as averaged over all 12 vowel pairs (Table III). This value reflects the size of the average pitch change. In most subjects this

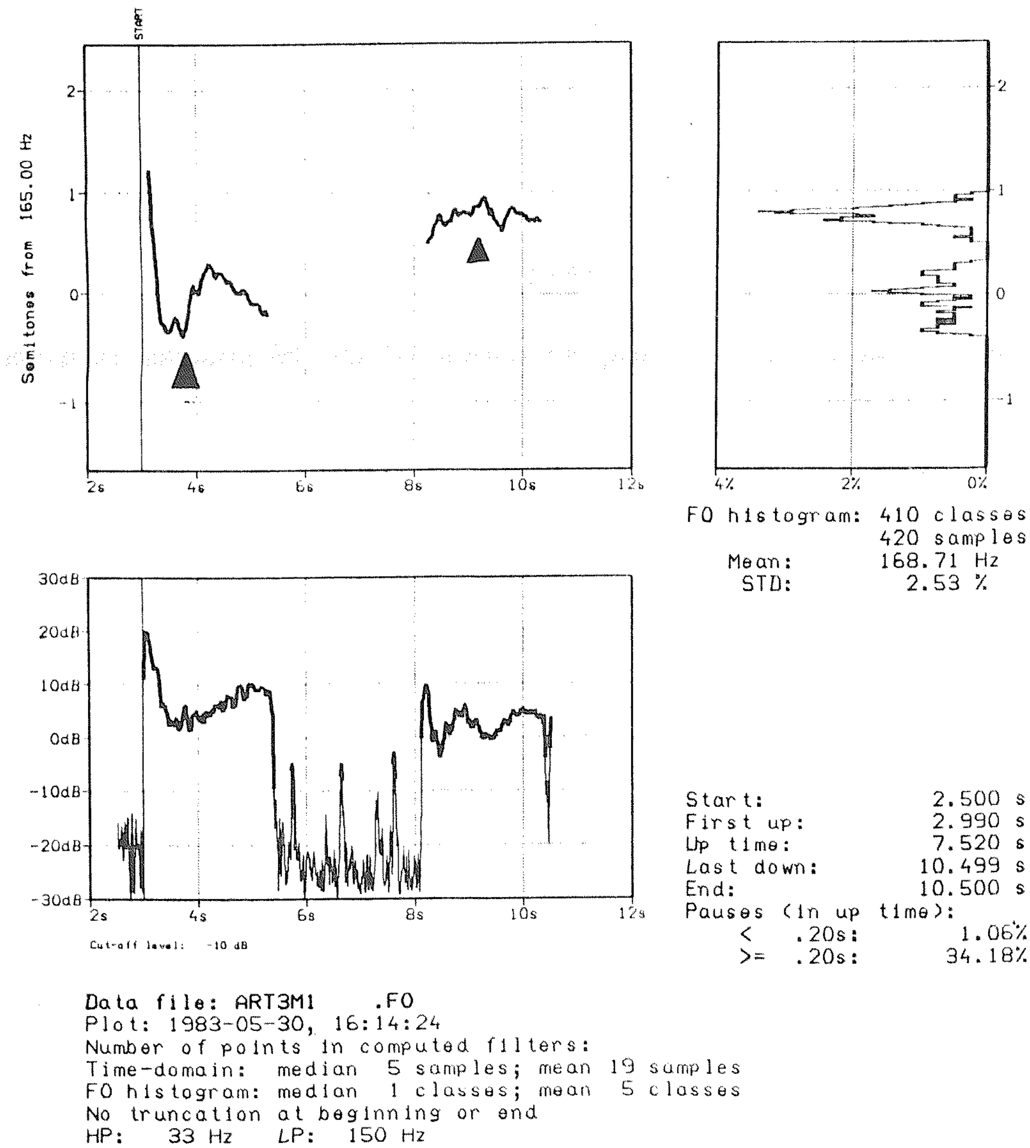


Fig. 4. Example of graphic printout from the "FØ" program. Clockwise from top left are: FØ vs. time ("pitch contour"), FØ histogram with statistics for the present time window, timing information, data identification and current filtering, and the relative amplitude vs. time. The frequency scale is logarithmic in semitone increments; one semitone is about 6%. See text for comments on the data shown.

Table II. Sample statistics for the standard deviation SD in FØ. Based on about 70 observations of each. Values in cent.

	average SD	SD of SD	Student's t
1st direction			
- masking	29.0	11.0	$\frac{29.0-21.5}{\sqrt{(11.0^2+8.8^2)/70}} = 4.4$
- no masking	21.5	8.8	
2nd direction			
- masking	22.7	8.5	$\frac{22.7-18.4}{\sqrt{(8.5^2+5.8^2)/70}} = 3.5$
- no masking	18.4	5.8	

Table III Average over 12 vowel pairs of the absolute value of Mdiff. The values are given in cent.

Subject:	MK	HW	RH	EW	LN	MN
1st direction						
- masking	34.1	24.3	17.6	14.3	23.2	11.0
- no masking	25.2	7.3	9.8	14.2	14.4	26.5
2nd direction						
- masking	9.0	19.6	16.3	14.8	23.8	12.9
- no masking	6.8	8.1	10.2	17.3	12.3	13.1

quantity was clearly reduced when the masker noise was removed, thus lending more support to the assumption that pitch control is facilitated by auditory feedback. One of the subjects, MN, behaved in the opposite way. Her Mdiff is larger when auditory feedback is eliminated; examination of the FØ plots revealed that with auditory feedback her pitch became more stable but also that she made larger pitch changes at the vowel transition. This presumably reflects a dependence of her perceived pitch on spectral signal properties.

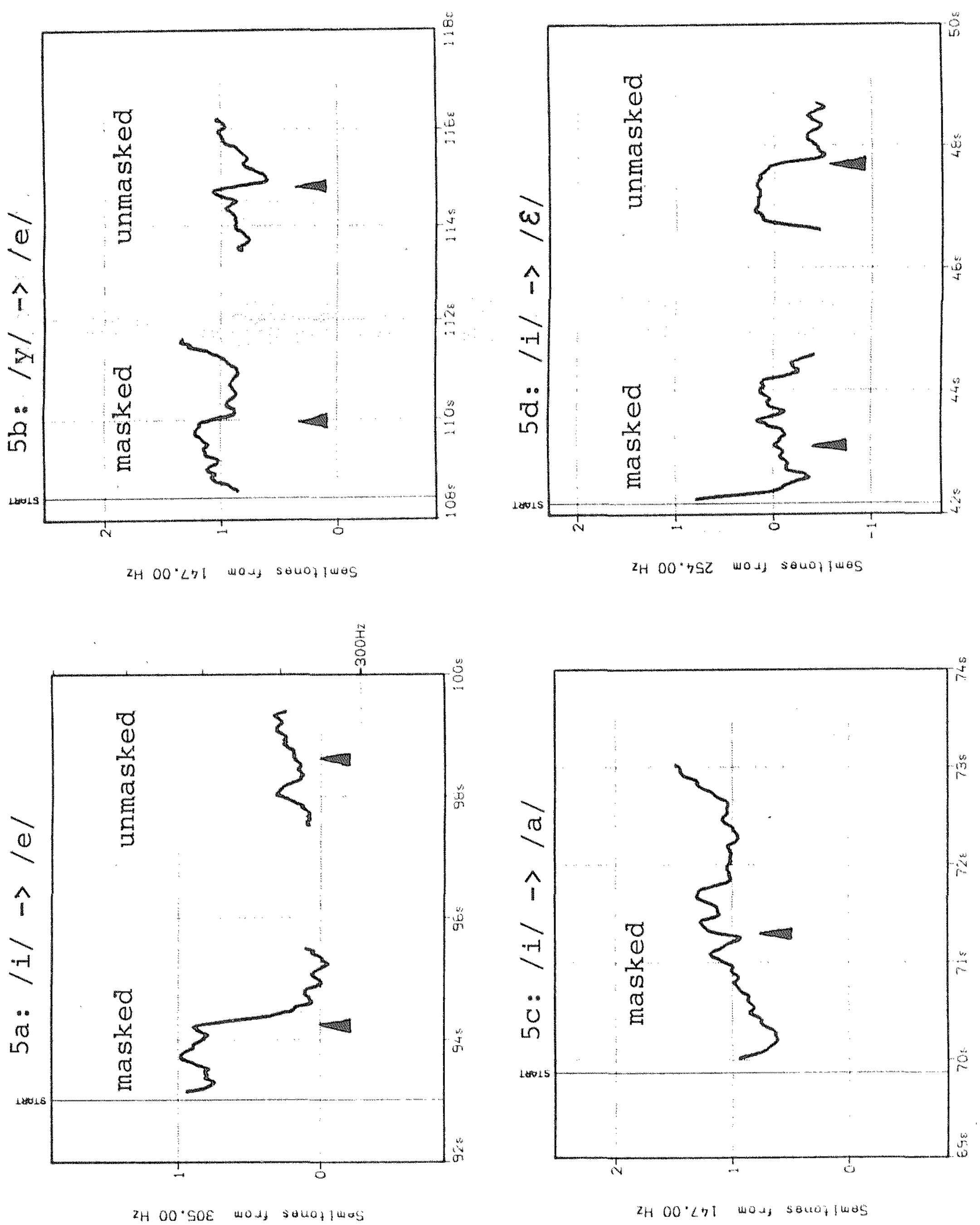


Fig. 5. Fundamental frequency vs. time for four interesting tokens (see text). The frequency scale is logarithmic and in semitones; the time scale is in seconds. A 50 ms median-value and 170 ms mean-value filter were used throughout to remove pitch extraction errors and vibrato.

Consider next the Mdiff of each vowel pair, including the sign, when averaged over the six subjects. In terms of this quantity, the different vowel pairs were found to be fairly similar. Even for the vowel pairs which differed the most, the differences were significant only at the 90% level. The intersubject scatter in Mdiff was large, that is, the subjects all behaved rather differently for a given vowel pair. For example, none of the vowel pairs gave an Mdiff which was of the same sign for all subjects; nor would the Mdiff of a given vowel pair as sung by a given subject tend to change sign when the direction of vowel change was reversed. Our overall impression was that irrelevant FØ motions were corrupting the Mdiff measure, thus obscuring any existing effect.

Mdiff was a often poor measure of FØ perturbation, namely when the fundamental frequency varied not only during, but also before and after the vowel change. To circumvent this shortcoming, the FØ derivative was measured in a 200 ms interval surrounding the vowel transition, but this also did not yield any significant results. Figs. 5a-d shed some light on subject behavior and on the problems we encountered with the Mdiff and derivative measurements.

Fig. 5a gives a clear-cut example of articulatory pitch perturbation: in noise, the singer makes a large pitch change exactly coincident with the vowel change, but when her feedback is restored, the pitch is stable at the transition.

Fig. 5b shows "the Mdiff problem": though there are clear steps in FØ, there is no big difference in average FØ before and after the transition. This applies both for the masked and unmasked cases.

Fig. 5c reveals a limitation of the derivative measure: FØ is erratic mostly during the vowel transition; i.e., something is happening, but measuring the derivative would be meaningless.

Fig. 5d is typical of subject 6 (MN). The clear pitch change arises only in the presence of auditory feedback.

The computer plots showing fundamental frequency versus time exhibited a clear pitch change at the vowel transition in about one in every four cases of individual vowel pairs. For this reason only these "relevant" tokens were selected for a closer examination, while the remainder was disregarded. This selection relies on the assumption that there is an effect of articulatory changes on pitch, for which most subjects managed to compensate. From the point of view of the present investigation, however, precisely those cases are of interest in which the subjects failed in this respect. The "relevant" tokens were selected using the following criteria:

- the FØ plot should show a clear pitch change within 50 ms of the vowel transition, and
- the duration of the pitch shift should be similar to the duration of the vowel transition.

Of the 284 instances of phonation, 74 were found to meet the above criteria. The direction (disregarding magnitude) of the pitch change in these cases is given in Fig. 6. We see that the vowel /i/ shows a marked

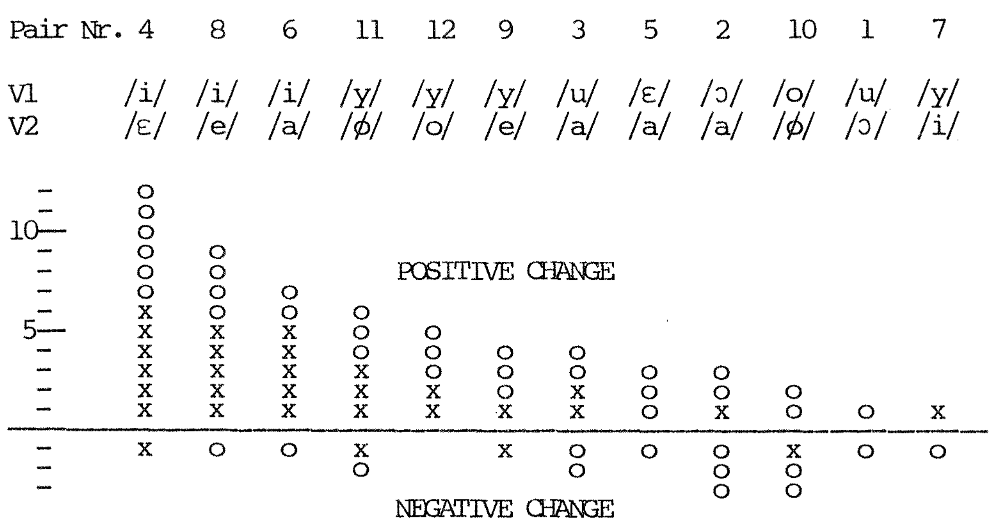


Fig. 6. Sign of pitch changes resulting from each pair V1-V2 of vowels. The figure includes only those cases where there was a clear pitch effect of the vowel transition. The signs x and o refer to the conditions with and without the auditory feedback, respectively. Marks above the dashed line indicate that the pitch was higher during V1. No distinction is made between the two opposite directions of vowel change.

tendency to raise the pitch. For /y/ the effect is less marked but still quite clear. The vowel /a/ exhibits a slight bias toward lower pitch. Tokens produced with and without feedback are about equally frequent.

Discussion

Decreased stability and precision in FØ control in the absence of auditory feedback has been observed before by Elliott & Niemoeller (1970), Ward & Burns (1978) and others; our computer program has enabled us to quantify these effect in a quick and precise manner.

The observed effect that FØ tends to rise for the "high" vowels /i/ and /y/ agrees with results obtained in speech research (e.g., Bush, 1981; Petersen, 1976). Several hypotheses have been advanced to explain the effect; these are reviewed by Ohala (1978) and others. The widely accepted so called "tongue-pull" hypothesis claims that the raised tongue position used in high vowels exerts an upward pull on the larynx, indirectly causing a longitudinal stretching of the vocal folds and thereby increasing FØ.

Bush (1981) studied, among other things, the vowel articulation of deaf speakers as compared to a hearing control group. She found intrinsic vowel pitch effects which were larger in the deaf group, and concluded that intrinsic pitch is an involuntary effect for which hearing people can compensate to some extent. This is in agreement with our findings on singers.

The English vowel /u/ is also classified as "high" and pitch-raising. The Swedish /u/, on the other hand, is classified as a back vowel. This could explain why it did not give any noticeable effect in the present study.

Summary of conclusions

- Sung vowels exhibit intrinsic pitch effects which are rather weak but similar to those observed in speech. In the limited selection of vowel pairs made for the present study, the front non-rounded vowels /i/ and /y/ were found to raise FØ.

- Both the accuracy and the precision of the pitch control of singers deteriorate markedly in the absence of auditory feedback. The existence of this effect implies that adequate feedback is an important requirement in choral performance and practice. Further studies are required to determine what is meant by "adequate".

Acknowledgments

Geoff Plant of the National Acoustics Laboratories in Sydney is gratefully acknowledged for suggesting useful references. This project is supported by the Swedish Council for Research in the Humanities and Social Sciences.

References

Bush, M.A. (1981): "Vowel articulation and laryngeal control in the speech of the deaf", part of doctoral thesis, MIT.

Elliott, L. and Niemoeller, A. (1970): "The role of hearing in controlling voice fundamental frequency", Int. Audiology 9, pp. 47-52.

Ewan, W.G. (1979): "Can intrinsic vowel F_o be explained by source/tract coupling?", J.Acoust.Soc.Am., 66(2), pp. 358-362.

Ohala, J.J. (1978): "Explaining the intrinsic pitch of vowels", Rpt. of the Phonology Lab., No. 2, Univ. of Calif. at Berkeley, May, pp. 118-125.

Petersen, N.R. (1976): "Intrinsic fundamental frequency of Danish", Rpt. of the Inst. of Phonetics, Univ. of Copenhagen, 10, pp. 1-27.

Ternström, S. (1983): User's Manual for the "FØ" computer program for analysis of fundamental frequency in time. 30 pages. Dept. of Speech Comm. & Music Acoustics (unpublished).

Ternström, S. and Sundberg, J. (1982): "Acoustical factors related to pitch precision in choir singing", STL-QPSR 2-3/1982, pp. 76-90.

Ward, D. and Burns, E. (1978): "Singing without auditory feedback", J. Res. in Singing, 1, 2, pp. 25-44.

SOME RADIOLOGICAL OBSERVATIONS OF VOCAL SOURCE-VOCAL TRACT

G.J. Troup* and H. Luke**
*Physics Department, Monash University
Clayton, Vic., Australia
**Radiological Department, Alfred Hospital
Prahran, Vic., Australia

Abstract

A radiological study of the vocal tracts and larynges for eight singers, in various states of training, has been made. Among the singers was a pupil of Lucy Manén's, and a pupil of Hussler and Rodd-Marling. All singers had the 'singing formant'. Similarities observed between the singers were: (a) greatly expanded piriform sinuses; (b) the lowering of the larynx below resting or speaking position during singing, leading to an approximation of the hyoid bone to the thyroid cartilage; and, (c) a large laryngeal ventricle. The larynx was fully closed before the onset of the note only for Manén's pupil. For some singers, the velum was not fully closed on certain vowels during a sung portion of an aria. The form of the epiglottis and of the laryngeal ventricle changed markedly for certain singers (notably the pupil of Manén and of Hussler and Rodd-Marling) for each vowel (Italian [a], [e], [i], [o], [u]) while in perhaps the most experienced singer, little change in either took place, but more pharyngeal adjustments were observed. All these various changes are physiological evidence of voice-source, vocal-tract interaction. A discussion of the results will be given.

Introduction

The research described in this paper did not set out originally to study vocal-tract, vocal source interaction. It was initiated by a paper of Manén's (1980) in which statements of Francesco Bennati's (1830) were quoted, categorising the various elements of 'voix orotunde' in which he was trained:

(1) there is a contraction below the insertion of the tongue,

(2) a particular direction is given to the expiratory flow of air,
(3) the tone starts within the ventricles of the larynx,
(4) there is an approximation of the hyoid bone to the thyroid cartilage,
(5) without this approximation the voice loses all quality.

It is obvious that (1), (4), and (5) are very close indeed to the mechanism for the 'singing formant' as suggested by Sundberg (1975). Further, they are verifiable by X-radiology. Mânen links (3) with her own statements (1974, 1980) that the laryngeal ventricles must meet to 'seal up' the larynx before the tone is begun. This last again is readily verifiable by X-radiology.

For Manén's paper (1980), there is also a reference to a work by Gemelli (1945), in which the different conformation of the epiglottis and laryngeal ventricle for the different vowels [a], [i] and [u] (Italian) were reported. Here is a translation (G.J.T.):

From the (x)*-radiography carried out, it results that: a) in each of the three vowels a, i, u (which in my researches on the vowels I have shown to be the fundamental ones, to which all the other vowels can be reduced), the vocal cords have different behaviour. While in a the vocal cords are mainly pressed against each other and are as if thickened at their edges; in i the approach of the two cords is minimal, as is also their thickness; in u we have instead an intermediate approach of the two vocal cords and the vocal cords have an intermediate thickness of their edges;

b) in parallel one has a different behaviour of the order parts of the larynx; in a the ventricle has almost disappeared, and the ventricular folds are also close together and thickened; the epiglottis is widened; in u the ventricles (sic)* are separated to the maximum; they have a rounded and almost re-inflated bottom, while the epiglottis is moderately lengthened; in i the epiglottis is lengthened to its maximum and the ventricles are less separated than in a and in u.

---

* Emphasis added by G.J.T.

In conclusion; one cannot consider the vocal cords as standing alone, and one cannot discuss their mechanism considering it alone; if one must make an account of their mechanism, it is necessary to consider the vocal cords as one of the elements of the soft parts of the phonatory tube, all of which take up a characteristic behaviour for each vowel; hence the comparison of the larynx to a musical instrument is founded on the erroneous concept of the isolated function of the vocal cords; the different vowels are given by the concurrence, together with the tension and thickness of the vocal cords, of the lengthening or thickening of the epiglottis, of the dilation of the ventricles (and)* of the thickening of the (ventricular)* folds.

Gemelli made use of a technique in which the x-ray film is introduced into the pharynx. Since this work was carried out in 1945, it seemed also worthwhile to carry out a radiological study for verification, with more modern techniques.

However, because Hussler and Rodd-Marling (1976) were brought to our attention, we were able to obtain a copy of Raoul Husson's Thesis (1950).

This alludes to the work of Greiner (1938) in the 1930's, using x-ray tomographic methods. Gemelli (1945) seems unaware of this material. Husson's Thesis (1950) concerns itself considerably with vocal tract-vocal source interaction, and regards the different behaviour of the larynx for each vowel, as evidence of vocal tract-vocal source interaction. It is in the spirit of Husson that we wish to present our results. Perhaps another translation is in order (G.J.T. and M.J.B.D.).

"In a word, the reaction of the resonator on the make-up of the laryngeal sound can be expressed in the following way.

The timbre of the larynx sound is drawn by the timbre of the "proper sounds" (formants)* of the pharyngeo-buccal resonator."

There is here the "law of affinity" (pulling)* which, according to us, governs the reaction of the pharyngeo-buccal resonator on the make-up of the laryngeal sound.

---

*Clarification added by G.J.T.

Husson then goes on to discuss the 'pulling' of an oscillator by a coupled resonator, and invokes "le principe des petits mouvements" - i.e., the principle of linearity. Subsequently:

> "...seeing that the partials of the pharyngeo-buccal resonator vary with the vowel emitted, the application of the law of affinity leads us to admit that the larygneal make-up itself varies with each vowel, another way of saying that each vowel will be already prefabricated by the larynx. Now the fact is exact, because the tomo-graphics of the larynx taken along with the emission of different vowels on the same note reveal different laryngeal attitudes (differences in the thickness of the cords, the width of the glottis, the contraction of the ventricles of MORGAGNI) (see the thesis of GREINER...).

Protocol, subjects, apparatus

The protocol chosen to examine the behaviour of the whole vocal tract and larynx, was as follows:

1. speak |a|, |e|, |i|, |o|, |u|
2. sing 'doh' to 'soh' legato and staccato and back on each vowel (pitch chosen by singer),
3. sing legato and staccato the first musical phrase of 'Caro mio ben' (to..."Languische il cor") in Italian at a comfortable pitch.

The sequence to be completed for each of sideways viewing by the x-ray apparatus (lateral view) and full fronted viewing (frontal view). The apparatus consisted of a Siemens tridoros 5S 3-phase generator in conjunction with an image intensifier and Pantascope table. The x-radiograph could be viewed directly in real time by means of the image, intensifier, and the image and accompanying audio where stored on videotape for later replay and processing - e.g., the extraction of single frames for analysis. The environment was that of the hospital x-ray centre; the acoustics were not very good, and all subjects criticized the vocal quality produced as certainly not their best. If any dissatisfaction was expressed either by subject or researcher, the appropriate section was repeated. There was no danger of radiation damage at any time to any of the subjects, as radiation dose was closely monitored. In fact, when it was discovered that subject C had undergone a large number of x-rays for a neck injury, she was only allowed to participate in protocol 3, and that wearing considerable antiradiation bodily protection. Those subjects not allergic to heavy-metal compounds were asked to 'gargle' with one prior to x-ray examination. This did not noticeably affect vocal performance, but was a considerable aid to outlining the borders of the vocal tract, epiglottis, piriform sinuses, etc.

The subject details are as follows:

Subject A: Male, 50, voice professionally trained beginning at age 39, sings regularly in choirs and does solo work.

Subject B: Male, 24, vocally trained for 2 years, has subsequently become 'runner up' in an important aria contest in Australia (pupil of Subject E).

Subject C: Female, 24, vocally trained for 2 years, unable to participate fully by reason of previous exposure to x-ray dose.

Subject D: Female, 53, a pupil Manén's. Has sung in Europe on contract for many years, and now teaches singing.

Subject E: Male, about 37. A pupil of Hussler and Rodd-Marling. Has sung in Europe, has sung in opera in Australia, teaches singing, runs a very fine (performing) choir.

Subject F: 57, a very experienced Oratorio singer, has studied 'vocal production' all her life, now teaches singing and directs a School of Music.

Subject G: 45, sings regularly in "early music" groups. Has 'spinous processes' on his vertebrea, which makes x-ray viewing of the larynx from the front extremely difficult.

Subject H: A contralto in her late 70's, well known in Australia. Unfortunately had laryngitis at the recording session.

Subject J: A fine young soprano, aged 33, has sung in Europe, and in opera in Australia.

Only extracts from recordings of subjects A, B, D, E, F, H and J were shown. In all cases, the different behaviour of the 'sinus morgagni' on the different vowels as sung is clear. Subjects E and F show clearly a 'pulling down' of the larynx during singing. Subject D has actually pulled the larynx up during singing and is the only one to have shown any attempt at total closure of both 'false' and 'real' vocal cords at the start of voice production.

Subject J shows a surprising rise of the larynx with the rising scale. Such a phenomenon for a good singer is not unknown (Ruth, 1963). Subject D shows a clearly different technique from the others: the hyoid bone goes forward, not down, and the larynx is actually pulled up from the rest position. It may be that she sang in the "ee" (I am excited!) mode rather than the "ah!" (I am pleased) mode (Manén, 1980)

The tentative conclusions we form from the study are:

(1) The total closure of both 'false' and 'real' vocal cords, prior to addressing a note is not necessary prerequisite for a well-trained voice.

(2) There is no doubt that different vowels demand a different shape and position of the sinus morgagni. Hence verification of the work of Husson and Gemelli, and clear evidence of vocal-source, vocal tract interaction.

(3) Singing is like a sport. Once the basics are learned (correct stance, correct breathing etc.), the singer must perhaps make his/her own way with the personal apparatus.

Acknowledgements

We have many people to thank, especially the most willing and helpful subjects. Special thanks are due to Miss Joan Arnold, Director of the

Melba Meororial Conservatorium of Music, 16 Hoddle Street, Abbotsford, Victoria, Australia, without whose encouragement and help this work would not have been possible. We would also like to specially thank Ms. Junewyn Jones and Mr. David Ashton-Smith for help with this project. Mr. M. Ryan helped with making up the conference video tape from the raw data. Dr. M.A.B. Deakin checked the translation from Husson.

References

Bennati, F. (1830): "Du mechanisme de la voix humaine dans le chants", Memoires de l'Academic Royale des Sciences, Paris.

Gemelli, A. (1945): "Nota preventiva sul meccanismo d'azione delle corde vocali nella fonazione", Pont.Acad, Scientiarum IX, 7, 69.

Greiner, G-F. (1938): "La tomographie du larynx", Thesis, Faculty of Medicine, Strasbourg.

Hussler, F. and Rodd-Marling, Y. (1976): Singing the Physical Nature of the Vocal Organ, Hutchinson, London.

Husson, R. (1950): "Étude des phénomènes physiologiques et acoustiques fondamenteux de la voice chantee", Thesis, Faculty of Sciences, Paris.

Manén, L. (1974): The Art of Singing, Faber Music, London.

Manén, L. (1980): "Vocal timbres, the essence of the classical school of singing", Studies in Music 13, 34.

Ruth, W. (1963): NATS Bulletin, May, 2.

Sundberg, J. (1975): Acustica 32, 89.

SINGING VOICE: BRIGHT TIMBRE, SINGER'S FORMANTS AND LARYNX POSITIONS

Shiqian WANG
RLE, Massachusetts Institute of Technology
Cambridge, MA, USA

Abstract

This study shows that bright timbre, especially in high voices, can be produced by trained singers in different singing styles (Chinese, Western early music and Western Operatic) with different laryngeal position but similar acoustic features - two to four high energy peaks in the spectrum, including an extra formant (the so-called "singing formant") near F_3 or F_4 in the bright timbre energy range.

The supporting data come from perceptual judgment of timbre, measurement of larynx position, and spectral analysis of various trained tenors' voices in video and audio demonstrations.

My findings, which differ from those of Shipp and Izdebski (1975) and Sundberg (1972, 1977) are:

1. Laryngeal height is not a reliable way to distinguish a trained singer from an untrained one. Not all trained singers maintain the low laryngeal position.

2. In both the high and the low laryngeal styles of singing, trained singers are capable of producing bright timbre and its associated acoustic features. Lowering of the larynx therefore does not explain the production of brightness. There is no reason to conclude that singing with a high larynx necessarily produces a voice of poor quality (at least in respect to brightness) or poor vocal health.

Introduction

The relationship between bright timbre (or brilliance), the singer's formants, and larynx location during phonation in the Western Operatic singing voice ("covered voice") has been studied extensively.

Frommhold and Hoppe (1965) concluded that vocal quality is inversely related to vertical movement of the larynx, a high larynx being associated with poor vocal quality. Shipp and Izdebski (1975) found that non-singers position the larynx upward in increased voice frequency and tend to maintain a level at or above the physiologic resting position, whereas trained singers maintain a laryngeal position below the physiologic resting position throughout their vocal range. Some other researchers had similar assumptions or findings (Faaborg-Andersen and Vennard, 1964; Sundberg, 1969).

The formants associated with bright timbre in singers have been studied by Bartholomew (1934), Rzhevkin (1956), Vennard (1967), and others. The singing formant, as defined by Sundberg, is an extra high spectrum envelope peak near 2.8 kHz located between the third and fourth formants of vowel sounds produced in male Western opera and concert singing. (Throughout my article a distinction is made in the use of the term "singer's formants" (of high range), as described in the Results (C.1.), and the term "singing formant", as defined by Sundberg.)

Sundberg (1974, 1977) concluded that the singing formant is responsible for the perception of bright timbre, and that larynx lowering explains the singing formant, that is, lowering the larynx may cause the laryngeal "tube", or ventricle, to act as an independent resonator, thereby creating the singing formant. He and others (Shipp and Izdebski; Vennard) found that the larynx lowers progressively as the sung tone rises in pitch.

This explanation is of more than academic concern, since it seems to be consistent with the widely held opinion that singing with the larynx in high position may lead to damage of the vocal cords (Luchsinger and Arnold, 1959, as cited in Sungberg, 1969; and Sundberg, 1976**). In other words, singing with a high larynx seems to be seen as "singing in the pain" for both the singer (voice damage) and the audience (perception of poor timbre).

Questions

In apparent conflict with this opinion is the impression gained from subjective observation of some trained Chinese and Western early music singers, who apparently sing with a high larynx and whose voices are bright and appear to be healthy. (Although these singers may sing in more than one style, their voices are here labeled as exhibiting either Chinese or early music style.)

Three main questions arise from the observation that some traditional Chinese and early music singing voices are produced with the larynx in high position:

1. Do some singers of Chinese and of some early music styles produce a timbre rated as bright by professionals?

2. Do all trained singers, regardless of style, maintain a laryngeal position below the physiologic rest position?

3. If there is a bright timbre in some Chinese and early music singing voices, what are the acoustic features, and how are they related to bright timbre and larynx position in the singing voice?

Method

To answer the questions, ten trained tenor voices representative of the three styles of singing were chosen for this study. Each subject had at least fifteen years of training and experience and was apparently in good vocal health. The study consisted of:

1. Perceptual judgment of timbre.

2. Measurement, during phonation, of the vertical distance of the larynx from its physiologic position.

3. Measurement of acoustic features, including the first five formants, the singing formant, and the fundamental frequency.

Perceptual judgments were made by asking each subject to sing three vowels: [ɑ], [i],[u] in full voice. The sounds produced were rated as to brightness on a scale of 1 to 7 by 19 voice teachers and singers at a meeting of the Boston Chapter of the National Association of Teachers of Singing.

Vertical movement of the thyroid cartilage away from its resting position was recorded and measured by the use of photographic slides and videotapes, while the voice was simultaneously recorded on audiotape during phonation. The measurement was facilitated by placing a narrow strip of colored adhesive tape on the skin of the neck, at a right angle to the airway, overlying the tip of the thyroid prominence in the physiologic resting position. The extent of vertical movement as measured on the slides and videotapes was then multiplied by a coefficient obtained by measuring a suitable facial landmark and comparing it to the measurement on the slide or videotape.

Acoustic features were determined by the spectral analysis program on a PDP-11 computer with a UNIX operating system. The program was designed to digitize a series of waveform files and perform several kinds of spectral analysis. The formants were determined by using both the Discrete Fourier Transform (DFT) and Linear Prediction Coding (LPC), and the fundamental frequency was obtained from the corresponding Cepstrum. A 25.6 msec segment of voice preemphasized at 6 dB/oct was selected visually, multiplied by a Hamming window, and converted into both a magnitude spectrum using the DFT and an idealized spectrum using the all-pole model LPC.

Results

A. Results of the perceptual judgments of timbre were as follows:

1. The voices of all subjects were rated as being in the bright range (each average rating number was greater than 4) although, as might be

expected, they had different degrees of brightness.

2. The vowels [i] and [ɑ] were judged as being significantly brighter than [u] ($P<0.001$).

3. The voices of the Chinese singers in this study were judged as having the greatest degree of brightness for all vowels.

B. The results of measurements of larynx position for the three styles of singing were as follows: (see Figs. 1, 2)

1. In the Chinese and early music singers, the larynx generally remained above the physiologic resting position, especially in the high range of the voice.

2. The larynx of each subject in the Western Operatic style was usually below the resting position, as has been determined by other researchers. This finding was very clearly demonstrated by two subjects who could sing in the Western Operatic, and either the Chinese or the early music style.

3. While the height of the larynx varied from singer to singer, the position of the larynx was highest for the vowel [i], nearly as high for [ɑ], and lowest for [u] in all styles of singing.

4. In the Chinese and early music singers the height of the larynx increased as the pitch went up. In contrast, in the Western Opera singers, the larynx moved progressively lower. These findings held roughly true for pitches up to about G_4, or 392 Hz.

C. Measurement of acoustic features yielded the following results:

1. Two to four high energy peaks (singer's formants), in some cases including an extra peak (the so-called** singing formant) near F_3 or F_4 were found in the spectra of not only the Western Operatic voices, but also the Chinese and early music singers in this study. The peaks were

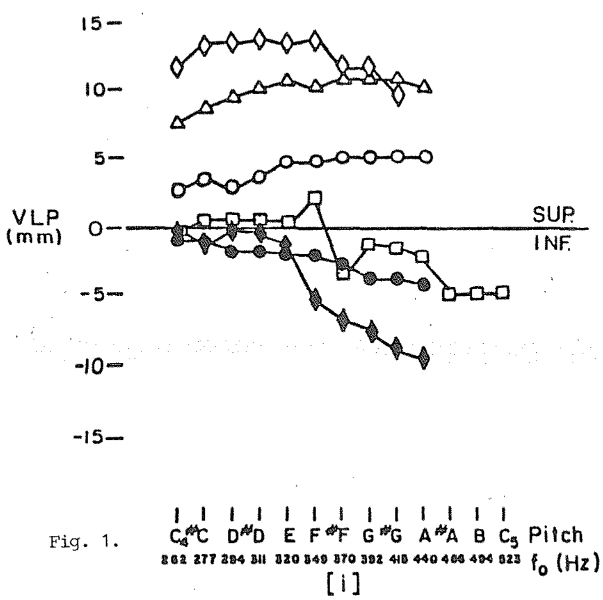


Fig. 1.

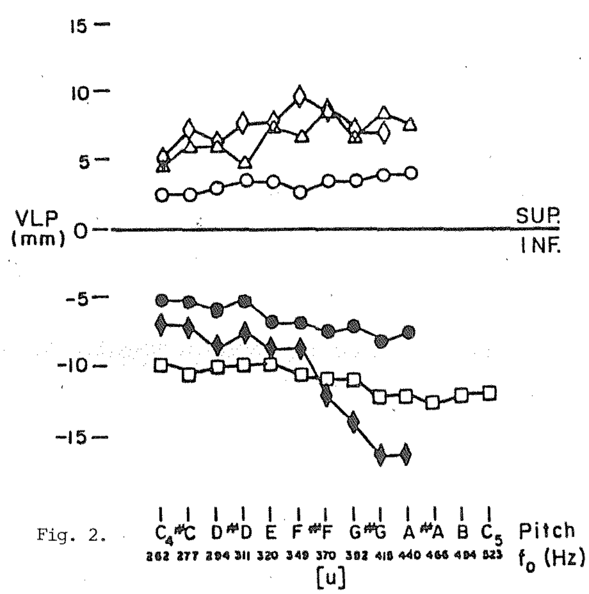


Fig. 2.

△ Subject A (early music singing voice)
○ Subject B (Chinese singing voice)
● Subject B (covered voice)
◊ Subject C (early music singing voice)
♦ Subject C (covered singing voice)
□ Subject D (covered voice)

Figs. 1 and 2. Measures of superior or inferior vertical larynx position (VLP) with respect to the physiologic rest position (zero line) of four singers producing sustained phonation at selected fundamental frequencies (f_0).

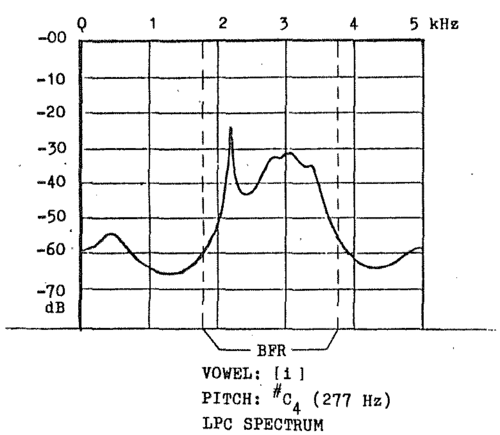


Fig. 3. Chinese singing voice

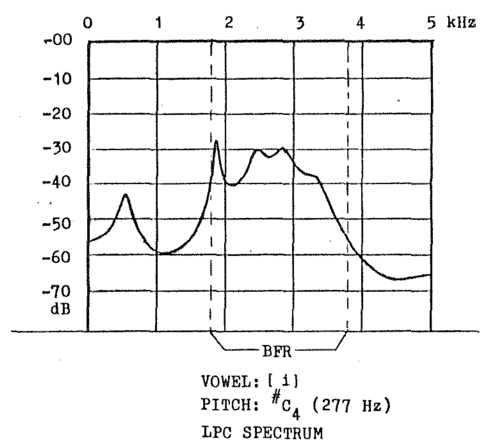


Fig. 4. Western operatic voice

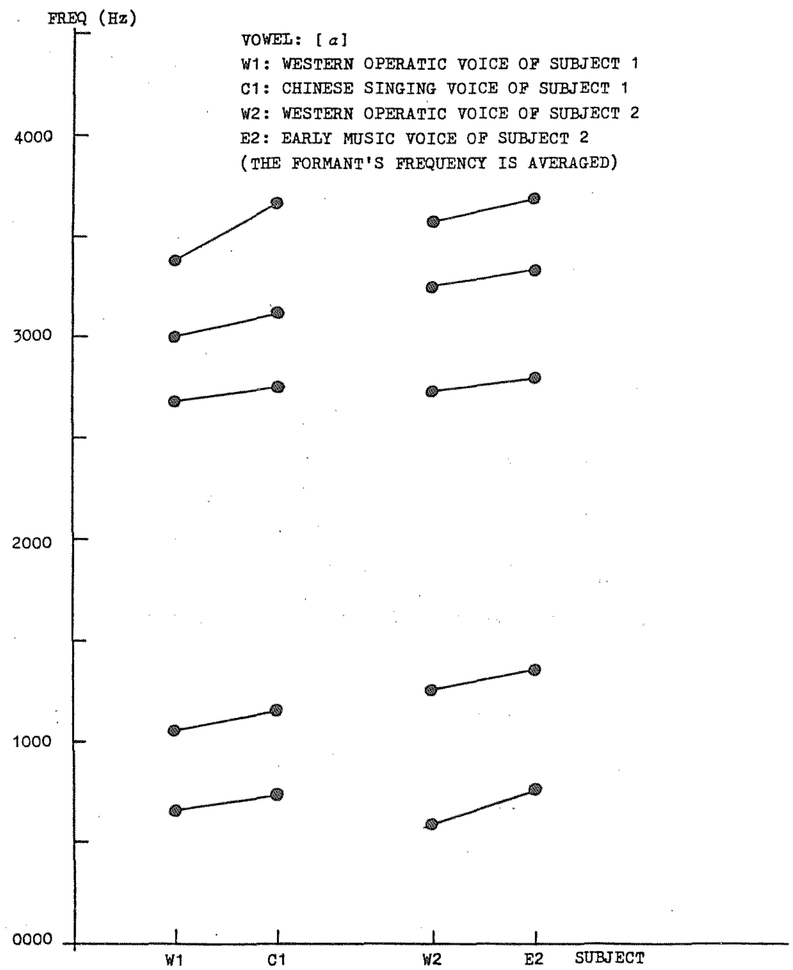


Fig. 5. Comparisons of singer's formants in different styles

found roughly between 1.8 and 3.8 kHz, which we shall call the "Bright Timbre Frequency Range", or "BFR" (see Figs. 3, 4).

2. In all three styles of singing, the relative amplitude of the formants in the BFR increased as the fundamental frequency increased.

3. All of the formants were higher for the Chinese and early music voices when compared to the Western Opera voices (see Fig. 5). These measurements were taken in two subjects, each singing in two styles, and generally held true except for F_1 in the vowel [i], in which a very slight inverse change occurred.

Discussion and summary

A. Although there are differences in timbre perceptible to trained ears, both the high and low laryngeal styles of singing are capable of producing bright timbre and concentrations of high energy in the BFR.

B. Although an extra peak ("singing formant") near F_3 or F_4 may be responsible for bright timbre, bright timbre can also be associated with concentration of high energy in the BFR in the absence of this extra peak.

C. This study does not confirm previous assumptions or findings that singing with a high larynx, in and of itself, is a distinguishing characteristic of the untrained voice. Not all experienced professional singers maintain the low laryngeal position; however, the technique for singing with a high larynx used by some well-trained singers must be different from that used by untrained or poorly trained singers.

D. Sundberg's explanation, based on studies of the Western Opera voice, that larynx lowering is responsible for the "singing formant" and bright timbre, does not account for their presence in this study. Therefore, lowering the larynx is not necessary to produce bright timbre and its-associated acoustic features including the so-called "singing formant".

Finally, on the basis of this study, there is no reason to conclude that singing with a high larynx necessarily produces a voice of poor quality or causes poor vocal health, although there are many untrained or poorly trained singers who use a high larynx position.

Determination of the physiologic and the acoustic differences between good and bad singing with a high larynx would be an interesting topic for future research. The results could have important implications for the training and care of the professional voice.

Acknowledgments

The author expresses his thanks to his colleagues at the Speech Communication Group of the Research Laboratory of Electronics at the Massachusetts Institute of Technology and the Department of Speech Pathology and Audiology, Teachers College, Columbia University. Warm thanks are due to Kenneth Stevens, Gunnar Fant, Ronald Baken, Joseph Huand, and Ingo Titze. Special thanks go to George Geyer for his help in preparation of this paper. Gratitude is expressed to Weide Wang and Karen Komar, to singers in China, Norway, and the USA who were the subjects of this study, and to the nineteen teachers and singers who served as jury in the perceived brightness study.

This research was supported in part by a grant from The Voice Foundation.

Footnotes and references

*A videotape, "Singing Voice: Bright Timbre, Acoustic Features and Larynx Position", based upon this paper has been produced and is available upon request to the author.

** In Sundberg, personal letter of February 26, 1984. He had notified this author that he also realized recently "that singing with a high larynx occurs in professional singing and does not need to be harmful to the voice function" which coincides with the finding of the papers of this author presented at the 105th Meeting of ASA, May, 1983 and SMAC, July, 1983.

Bartholomew, W. (1934): "A physical definition of good voice quality in the male voice", J.Acoust.Soc.Am. 6, 25-33.

Faaborg-Andersen, K. and Vennard, W. (1964): "Electromyography of extrinsic laryngeal muscles during phonation of different vowels", Ann.Oto-laryng.Rhinolaryng. and Laryng, 73, 248 ff.

Frommhold, W. and Hoppe, G. (1965): "Tomographische Studien zur Funktion des menschlichen Kehlkopfes", Folia Phon. 17, 83-91; 18, 81-90.

Luchsinger, R. and Arnold, G.E. (1959): Lehrbuch der Stimm- und Sprachheilkunde, Springer, Wien.

Rzhevkin, S. N. (1956): "Certain results of the analysis of a singer's voice", Soviet Physics-Acoustics 2.

Shipp, T. and Izdebski, K. (1975): "Vocal frequency and vertical larynx position by singers and non-singers", J.Acoust.Soc.Am. 58, 1104-1106.

Sundberg, J. (1969): "Articulatory differences between spoken and sung vowels in singers", STL-QPSR 1/1969, 33-42.

Sundberg, J. (1974): "Articulatory interpretation of the singing formant", J.Acoust.Soc.Am. 55, 838-841.

Sundberg, J. (1977): "The acoustics of the singing voice", Scientific American 3, 82-91.

Vennard, W. (1967): Singing: The Mechanism and the Technic, Carl Ficher, New York.

A SCHEMA THEORY OF HOW CHILDREN LEARN TO SING IN-TUNE: AN EMPIRICAL INVESTIGATION

Graham F. Welch
Southfields, London, England

Abstract

The literature on children's singing reveals that, for each age group there are some children who cannot sing "in-tune". These children are labelled "poor pitch singers" (p.p.s.). The literature suggests that singing "in-tune" is not simply a question of "can" or "cannot" but rather a hierarchy of skill with several levels of competence. In general the number of p.p.s. within sample populations declines with age, with boys outnumbering girls by 2 or 3:1.

The traditional method of treating this disability, although implying that improvement is possible, has only shown limited success.

Reference to the psychological literature on feedback, however, suggests that, (1) learning can only take place when Knowledge of Results (KR) is present, and (2) variety of experience rather than repeated measures of the same kind may be more conductive to producing "novel" patterns of behaviour.

Applying these findings to the mechanism of singing, a schema theory of how children learn to sing in-tune is proposed, and its hypotheses subjected to empirical evaluation.

A group of sixty-six p.p.s., aged seven years, were randomly divided between six treatment groups, two of which were controls. Visual feedback and KR were provided by the use of an electrolaryngograph coupled to an oscilloscope. Both these variables were manipulated against high and low variability of practice.

Results indicate that Knowledge of Results (KR) may be crucial for p.p.s. to become pitch accurate, especially where the KR is combined with a high variability of practice.

Introduction

A paper by Welch (1982) proposed a schema theory for how children learn to sing in-tune. This present paper reports an empirical evaluation of that schema theory.

The schema theory applied to the problem of how children learn to sing in-tune suggests that singing is a specialised form of motor behaviour. Under the theory it can be hypothesised that control of vocal pitch is achieved via a motor response schema which is made up of three component parts. These are (1) a recall schema, (2) a recognition schema, and (3) an error labelling schema. The term 'schema' is used here to denote an abstraction of commonalities from previous like responses. Schmidt (1975) incorporated the theories previously postulated by Adams (1971) and Pew (1974) into a new schema theory which sought to explain how 'novel' motor responses are made.

The recall schema is hypothesised as being the psychological mechanism responsible for co-ordinating the various parts of the vocal apparatus in order to produce the sung response. It is based on the relationship between four inputs, i.e. (i) the desired pitch (= the desired outcome), (ii) the previous planned co-ordinations of the vocal mechanism (= past response specifications), (iii) an assessment as to the success of previous sung responses (= past outcomes), and (iv) the pre-response state of the vocal mechanism (= the initial conditions).

The recognition schema is hypothesised as being the psychological mechanism responsible for assessing the accuracy of the sung response. This assessment is done by comparing the actual sensory feedback of the vocal response, both physiological and auditory, with an "expectation" or "forecast" of this sensory feedback. This "expected" sensory feedback is the product of the relationship between previous expected:actual feedback comparisons, i.e. "Last time I did this, I expected it to feel/sound like this; it actually felt/sounded like that, so next time I shall expect it to feel/sound like....."

Finally, the error labelling schema is hypothesised as being responsible for labelling any perceived mismatch between the "expected" and actual consequences of the vocal response. An example of this "label" might be "too high", "too low", or just "wrong". Such labelled errors are then fed back to the recall and recognition schemata as "outcomes" so that the schema rules accomodate to new experiental data prior to the next vocal response.

Reference to the psychological literature on feedback suggests that efficient error labelling, and thus "learning", could only take place if the sung response were rated in some way. This rating, termed Knowledge of Results (KR) would be made by a source external to the child, and would provide an artificial measured outcome of the response (i.e. artificial in the sense that the units of measurement would be man-made, not absolute). It is hypothesised that the child or adult would only be able to label errors accurately without KR (a psychological process termed "subjective reinforcement") if sufficient previous responses with KR had been made.

Implicit in such a schema theory is the notion that, because the schemata are generated from the relationships between a variety of inputs, any increase in the range of experience within this variety will result in an increase in the strength of the schemata for producing "novel" (= in-tune) response. (Researches by Williams and Rodney (1978), Moxley (1979), Carson and Wiegand (1979) into the advantages of variability of practice in other areas of motor behaviour support this hypothesis, - for a fuller account, see Welch, 1982, pp. 177-178.)

The customary approach to the improvement of out-of-tune singing is a drill condition where only a single pitch is practised (e.g. Wolner and Pyle, 1933; Joyner, 1969, 1971; Roberts, 1972).* Once this pitch is secure, other adjacent pitches are added by the same method of continuous repetition. The schema theory, however, predicts that learning can only take place when Knowledge of Results (KR) is present, and in a form that can be comprehended by the child, otherwise errors cannot be labelled accurately. To inform a young child, for example, that they are "sharp" or "flat", or "above" or "below" the note presupposes a grasp of related musical concepts. The schema theory additionally predicts that a variety of experience may facilitate abstraction of certain facets of that experience, and this abstraction could then be incorporated into a schematic framework, within which future experience could be processed.

*see Welch (1979a), pp. 52-54.

To test these hypotheses, an experiment was run in which a group of young children took part. Each child had been observed on a screening procedure to sing out-of-tune (termed "poor pitch singers" or p.p.s.). S's were given training in six learning conditions to evaluate the relative effects of Knowledge of Results and Variability of Practice for learning to sing in-tune.

Method

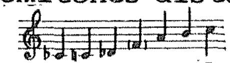
Subjects

Subjects were 66 boys and girls aged between 7 and 8 years drawn from five London Primary schools. Each subject was selected from a large parent sample on the criterion that they were observed to make gross pitch errors in a screening procedure which required them to sing two test melodies. Prior to the screening the melodies had been taught to them in a class situation. The subjects were then randomly assigned among the six learning conditions, with the restriction that groups were matched for sex and degree of disability. (The population of out-of-tune singers was found to fall into two broad categories. Children in the upper half showed a degree of pitch variation when singing, which occasionally coincided approximately with the melody. The children in the lower half appeared to be vocally fixed within a narrower pitch band, and exhibited little or no variation in sung pitch.)

Apparatus

Each subject was taken individually through the treatment drill procedure for the group to which he/she had been assigned. Two electrodes were placed against the skin of the subject's neck, on either side of the laryngeal assembly. These were able to pick up the electrical potentials from the vocal folds of the larynx and transmit them to an electrolaryngograph. This device analysed the incoming signals, and extracted the frequency information from them; the output was then fed to an oscilloscope capable of transforming it in such a way that it was represented as a linear trace that moved upwards and downwards on the oscilloscope screen, as a visual correlate of the vocalised pitch. Subjects sat facing the oscilloscope with their line of vision at an angle of 90^{o} to the screen, and at a distance of 25-30 cm from it. Subjects listened to pitch stimuli provided via earphones from a cassette taperecorder. Each tone lasted for four seconds, and was followed by a six second silent interval. Subjects were instructed that they were to attempt to vocally reproduce the pitch of each stimulus tone during the silent interval which followed it.

Design

Two sets of learning stimuli were prepared, one set each of two levels: Level I = low variability, presented tones at a single pitch target, with F4  as the target. Level II = high variability, presented six pitches, randomly ordered, spaced not less than ±4 semitones distance from the target, i.e. B♭3, C4, D♭4, A4, B4, C5  but did not include the target. The range of these 6 pitches was determined by reference to a) the method adopted in a previous experiment conducted by Williams and Rodney (op.cit.) and b) to the literature on children's "comfortable" vocal range (Welch, 1979b).

Against these two main levels of variability of learning experience (i.e. target-only and non-target) three treatments were employed.

Treatment 1: (No visual feedback, no Knowledge or Results.) Subjects in groups 1 and 2 sat in front of the oscilloscope, as for the two other treatment conditions, but the instrument was not turned on, and the subjects in this condition did not see the visual trace representing their attempted approximations to the stimulus tones they heard.

Treatment : (Visual feedback.) Subjects in groups 3 and 4 saw the visual trace on the oscilloscope, which provided virtually simultaneous feedback. The relation of their attempted approximation to the target pitch (i.e. the error between target and vocalisation), however, was not indicated to them.

Treatment 3: (Visual feedback and Knowledge of Results.) Subjects in groups 5 and 6 saw the visual trace on the oscilloscope representing their approximation to the stimulus tone. In this condition, lines were affixed to the screen in a position calibrated to represent the stimulus tone being heard. The margin of error between the visual representation of their attempted approximations and the stimulus tones was therefore apparent to the subjects, who thus had both visual feedback and Knowledge of Results. The two levels and three treatments used in the design are shown schematically in Fig. 1.

	Treatment 1 No visual feedback No Knowledge of Results	Treatment 2 Visual feedback	Treatment 3 Knowledge of Results
Level I Low Variability	Group 1 No-VF No-KR single target pitch (F4)	Group 2 VF single target pitch (F4)	Group 5 KR single target pitch (F4)
Level II High Variability	Group 2 No-VF No-KR six target pitches (B♭3,C3,D♭4,A4,B4,C5)	Group 4 VF six target pitches (B♭3,C4,D♭4,A4,B4,C5)	Group 6 KR six target pitches (B♭3,C4,D♭4,A4,B4,C5)

Fig. 1. The experimental design.

Learning trials were arranged in blocks of thirty-three groups of ten trials with a one minute interval between each block. This arrangement was devised in order to minimise vocal and attentional fatigue for the subjects. At the completion of the 30 learning trials, each subject was given a further ten trials. The stimulus for these was the target pitch F4. Neither visual feedback, nor Knowledge of Results was provided.

The relative effects of the six learning conditions were evaluated from the vocal pitch errors (measured in cents) observed on the ten No-VF, No-KR test trials by means of a levels by treatments (2x3) analysis of variance - ANOVAR. The responses on the ten No-KR test trials were recorded for all subjects onto a second cassette taperecorder via the electrolaryngograph. The pitch errors on these test trials were analysed by computer program which calculated mean deviations of the trial response in cents from the target pitch. First a speech display program was utilised to determine the frequency range in Hz of the sung response. Then the mean error in Hz over the ten test trials was computed before being transformed by a secondary computer program into the error in cents (100 cents = 1 semitone).

Results

Obtained data comprised six populations of error scores, one for each of the six experimental groups. The error scores represented the mean pitch deviation in cents of the ten test trials for each subject within that group. Since the direction of the error, above or below the pitch of the target, was not a concern of the experiment, the error scores of a group could be regarded as a distribution of scores about the group mean, and comparison of the performances on the test trials of the six groups could then be undertaken by normal parametric means.

Accordingly, the data was first subjected to an Analysis of Variance (levels by treatments). The results of this analysis are shown in Table 1.

Source	SS	df	ms	F	p
Total	3023091.7	65	-	-	-
levels (LV/HV)	27368.6	1	27368.6	0.72	n.s.
treatments	459824.4	2	229912.2	6.04	p<.005
treatments x levels	251165	2	125582.5	3.3	p<.1
Error	2284733.7	60	38078.9	-	-

Table 1. Analysis of Variance: treatments x levels

The above analysis reveals a significant difference for treatments (i.e. between groups 1/2, 3/4, 5/6) of p<.005, but no significant difference for levels (i.e. between groups 1/3/5, and 2/4/6) or for the interaction between treatments and levels (p<.1). Follow-up t-tests revealed that:

(i) the Kr groups (5/6) were significantly better than the No-KR groups (1/2, 3/4) at the p<.05 level;

(ii) there was a significant difference between the two control groups who had neither visual feedback nor KR but differed in level of variability (groups 1 and 2) at the $p<.02$ level; and

(iii) there was also a significant difference between group 1 (No-VF, No-KR/LV) and the two KR groups 5 and 6 ($p<.01$).

In addition, those children who had shown the greater degree of disability (i.e. were the most out-of-tune) at the onset of the learning procedures were significantly more accurate on the test trials if they had received KR compared with those who had not (sig. diff. $p<.001$).

Discussion

Those groups who were given extrinsic feedback in the form of Knowledge of Results were the most accurate at matching the F4 target (350 Hz) on the ten post-training No-KR trials. The effect of KR on the degree of vocal pitch accuracy is reflected in the results regardless of sex or level of disability.

The failure of levels (LV/HV) to reach significance is largely due to group 4 (who received six stimulus pitches and visual feedback). It would seem that visual feedback without KR is ineffective in generating an error labelling schema because of the lack of an objective rating of vocal pitch accuracy. High variability of practice combined with visual feedback actually appeared to confuse the subjects. The levels, however, did reach significance where no extra feedback or KR was available. The group (2) who had no visual feedback nor KR but did have variability of practice were significantly more accurate ($p<.02$) over test trials than group 1 (No-VF, No-KR, and low variability) who had repeatedly practised the target pitch.

It was noted that despite the gross pitch errors made when singing the two melodies of the screening procedure, the final mean errors in cents for the experimental subjects showed that many children were extremely close to the target pitch over test trials with almost half being within a quartertone. One may hypothesise that it was the simplicity of the experimental task, i.e. singing in a single pitch, which allowed the children to be more successful than when they attempted a simple melody.

The data is generally supportive of the hypotheses contained within the schema theory. Without KR, schema generation is seen as being deficient, because the subject is not able to acquire accurate outcome information about the sensory consequences of his vocal action. As a result, the error labelling schema and the recognition schema are both affected. When KR was present on the learning trials, subjects were able to generate sufficiently strong schemata so that they were able to maintain performance if the KR was then withdrawn. Where no KR is present on learning trials, a variety of inputs (variability of practice) is seen to produce stronger, more efficient schemata, providing that there is no other incoming information which might bring confusion to the error labelling. (In the present experiment adding visual feedback alone was not seen as being beneficial to the poor pitch singer.)

When KR is present, however, it is seen as outweighing variability of practice in learning to sing in-tune.

It is concluded that children who sing out-of-tune can become pitch accurate if the learning condition contains both qualitative information about the pitch error and sufficient practice for this information to be applied, preferably utilising a variety of pitch stimuli.

References

Adams, J.A. (1971): "A closed-loop theory of motor learning", J. of Motor Behavior 3, pp. 111-150.

Carson, L.M. and Wiegand, R.L. (1979): "Motor schema formation and retention in young children: a test of Schmidt's Schema theory", J. of Motor Behavior 11, pp. 247-251.

Joyner, D.R. (1969): "The monotone problem", J. of Research in Music Education 17, 1, pp. 115-124.

Joyner, D.R. (1971): "Pitch discrimination and tonal memory and their association with singing and the larynx", M.Phil.Thesis, University of Reading.

Moxley, S.E. (1979): "Schema: the variability of practice hypothesis", J. of Motor Behavior 11, 1, pp. 65-70.

Pew, R.W. (1974): "Human perceptual-motor performance", in B.H. Kantowitz (Ed.): Human Information Processing: Tutorials in Performance and Cognition, New York: Erlbaum.

Roberts, E. (1972): "Poor pitch singing. A survey of its incidence in school children and its response to remedial training", M.Ed. Thesis, University of Liverpool.

Schmidt, R.A. (1975): "A schema theory of discrete motor skill learning", Psychological Review 82, 4, pp. 225-260.

Welch, G.F. (1979a): "Poor pitch singing: a review of the literature", Psychology of Music 7, 1, pp. 50-58.

Welch, G.F. (1979b): "Vocal range and poor pitch singing", Psychology of Music 7, 2, pp. 13-31.

Welch, G.F. (1982): "Poor pitch singing in young children - the significance of feedback systems in learning to sing in-tune", I.D.R.S. Proceedings, Rotterdam, pp. 174-182.

Williams, I.D. and Rodney, M. (1978): "Intrinsic feedback, interpolation and the closed-loop theory", J, of Motor Behavior 10, 1, pp. 25-36.

Wolner, M. and Pyle, W.H. (1933): "An experiment in individual training of pitch-deficient children", J.Educational Psychology 24, pp. 602-608.

STROBOGLOTTOMETRIC AND ACOUSTIC MEASURES OF NATURAL VOCAL REGISTERS

J. Wendler, S. Fischer, W. Seidner, A. Rauhut, and U. Wendler
Humboldt-Universität, Berlin, DDR

Abstract

Simultaneous stroboglottometric measures and spectral analyses of sustained phonations confirm the thesis, that there are two main natural registers which are, for males, physiological and they are laryngeally based with distinctive vibratory patterns of the vocal folds resulting in typical spectral structures of the voice sound. LTAS investigations during ongoing singing of untrained and professional singers, in particular with a professional using his voice as a tenor as well as a countertenor, give good reason to believe that artistic registers are the result of vocal tract influences and/or interactions of laryngeal and pharyngeal adjustments.

Introduction

According to Garcia (1855) after "a series of succeeding sounds of equal quality on a scale from low to high produced by the same mechanical principle", the nature of the voice changes basically to "another series of succeeding sounds of equal quality, produced by another mechanical principle". The two qualities can clearly be identified by auditive perception and are generally labeled as different registers. As Seidner, et al. (1981) and Seidner and Wendler (1981) have demonstrated, the two qualities represent two physiological functions. In males, both can be realized over a rather wide pitch range, independently from one another, and not only in succeeding parts of the vocal range with a small overlapping area of amphoteric sounds. Since these functions exist naturally and not as a result of a special voice training, we can call them natural registers, the "basic" voice "chest" register, and the "other" voice "fistula" register, following, among others, J. Muller (1840) who compared the softer and weaker sound of that voice to the sound of flute pipes (fistula = tube, pipe). But, this terminology is a preliminary suggestion and we are open to further discussion (Wendler and Seidner, 1982). This paper deals with the pecularities of the vibratory pattern and the corresponding voice sound in these two natural registers.

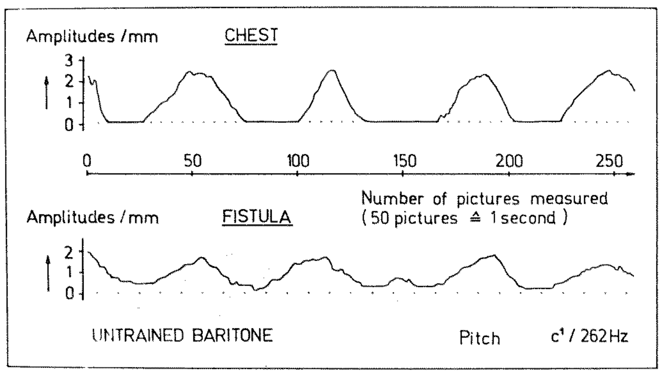
Subjects and Methods

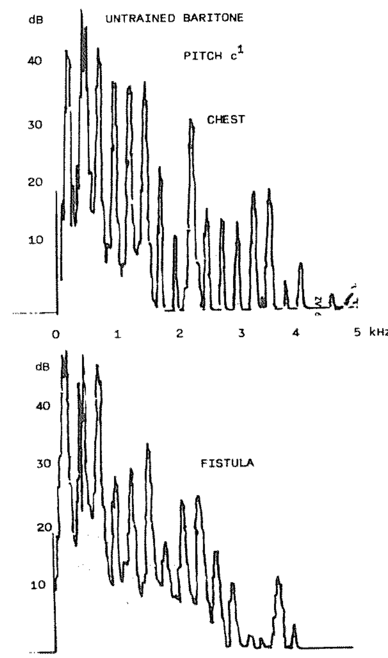
Three male persons (one untrained baritone, one singing teacher/baritone, and one professional singer/tenor, counternor) and two female persons (one non-singer/soprano and one singing actress/alto) served as subjects. They were investigated by means of stroboglottometry (Wendler and Tostman, 1981) which is based on indirect microstroboscopy of the larynx with simultaneous video-recording of the pictures and the voice sound on a Philips Series 87 recorder by means of TuR equipment for microstroboscopy and stroboglottometry as well as a Bruel & Kjaer measuring microphone 4132.

The subjects phonated at the same pitch in the chest and the fistula registers. Then the amplitudes were measured in the middle of the vocal folds using the picture-by-picture operation of the recorder. 200-250 measures per phonation were taken (50 pictures = 1 second), so, depending on the pitch, ca 500-3000 cycles formed the basis for a schematic outlining of the vibratory pattern.

We know, of course, the problems connected with the use of stroboscopy and the reservations in comparison to high-speed motion pictures. But both of them have their advantages and their disadvantages. A high-speed film shows each cycle in precise detail, but the section to be measured must necessarily be rather short. In addition, the noisy camera makes simultaneous voice recording for acoustic analysis considerably difficult. In stroboscopy, the low noise of the flash lamp can practically be neglected, and the vibrations of the vocal folds can be recorded and measured over a great number of cycles. The vibratory patterns obtained do not show the real vibrations. They only allow one to visualize the vibrational behavior in a kind of long-term average procedure, and thus they provide us with an overall impression of the glottic activity during phonation in terms of amplitudes, slope of opening and closing movements, and closure. The great number of irregularities, which obviously characterize the human voice as a human output, thereby do not allow certain statements in the time-domain.

The sustained sounds delivered during stroboscopic mirror-investigation were analyzed by means of a Bruel & Kjaer narrow band analyzer Type 2031 (40-Hz bandwidth, averaging of 128 spectra). Additionally, LTAS analysis was carried out with the same equipment (based on 1064 spectra) for each of the subjects singing a folksong ("Sah ein Knab ein Röslein stehn") in chest and fistula register, both of the versions in the same tune and pitch.



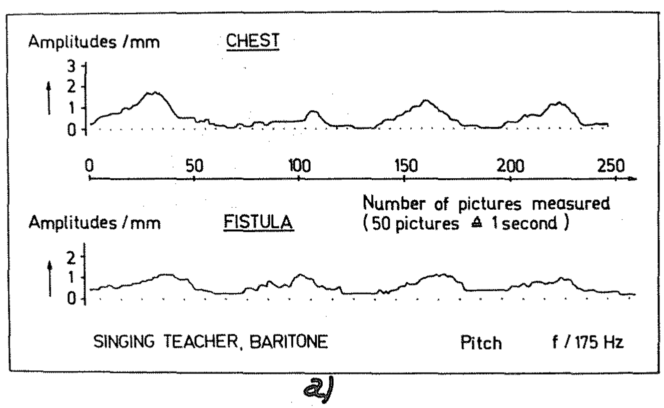


b)

Fig. 1. a): Stroboglottometry
b): Spectra during mirror phonation

Results

Figs. 1a and 1b selected for demonstration confirm, at first glance, the thesis of Sonninen (1962) that "the key to the problem of registers seems to be primarily laryngeal and not pharyngeal". The vibratory patterns differ distinctly in chest and in fistula register. In chest voice there is always a clear closure and the course of the curve is, simplified, triangular. In fistula voice, generally no closure can be observed, the curve is smoother, more sinusoidal. The corresponding acoustic outputs reveal a poorer spectrum in fistula phonation compared to chest phonation, a difference which - as is well known - mainly depends on the closure of the folds during the vibratory cycle. One of our subjects, the singing teacher, was capable to vary the amplitudes of his vocal folds arbitrarily from very narrow to very wide while loudness, timbre, and pitch were being kept fairly constant. As can be seen from Figs. 2 and 3, wide amplitudes were associated with a remarkable decrease of energy in the area of the high singing formant.



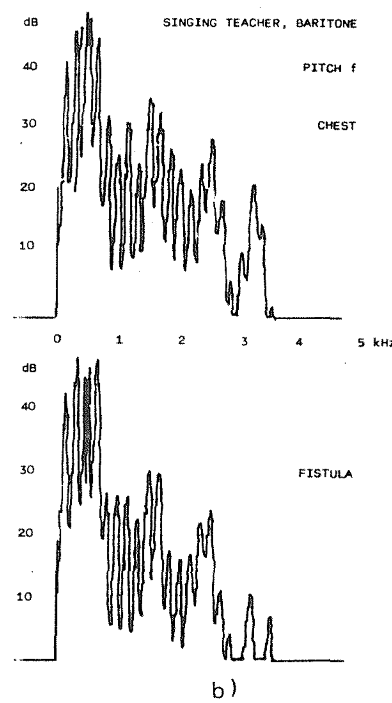


Fig. 2. a): Stroboglottometry
b): Spectra during mirror phonation

During the phonations with the laryngeal mirror positioned in the subjects' throats, the influence of the vocal tract on the spectral structure of the sounds was certainly reduced and probably equalized to some degree. That means that the acoustic differences observed concerning registers are more related to the source than to the filter, and vocal education or special training of the singing voice had little or no effect.

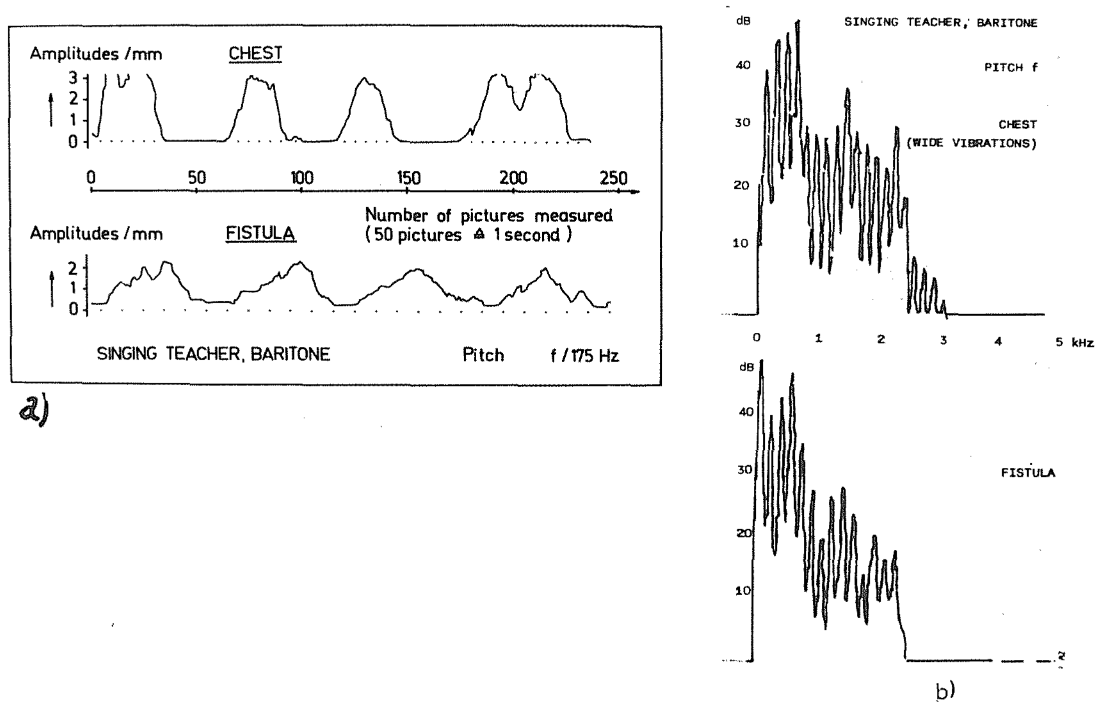


Fig. 3. a): Stroboglottometry (same pitch, intensity, timbre as in Fig. 2, but wide amplitudes)
b): Spectra during mirror phonation

The case is quite the reverse when looking at the LTAS-analyses in on-going singing under normal conditions. Here we can see in Figs. 4-6 that the singing teacher forms a noticeably high singing formant in his chest register, but not so high in his fistula register. The professional singer shows a considerably high singing formant when using his tenor voice as well as when using his countertenor voice. A comparison with the corresponding vibratory patterns makes it evident that the tenor voice is based on the natural chest register and the countertenor voice on the natural fistula register (Figs. 7a and 7b). The untrained baritone, despite having a good command of a healthy and powerful voice exercised in choir singing, developed no high formant at all, neither in chest nor in fistula functions.

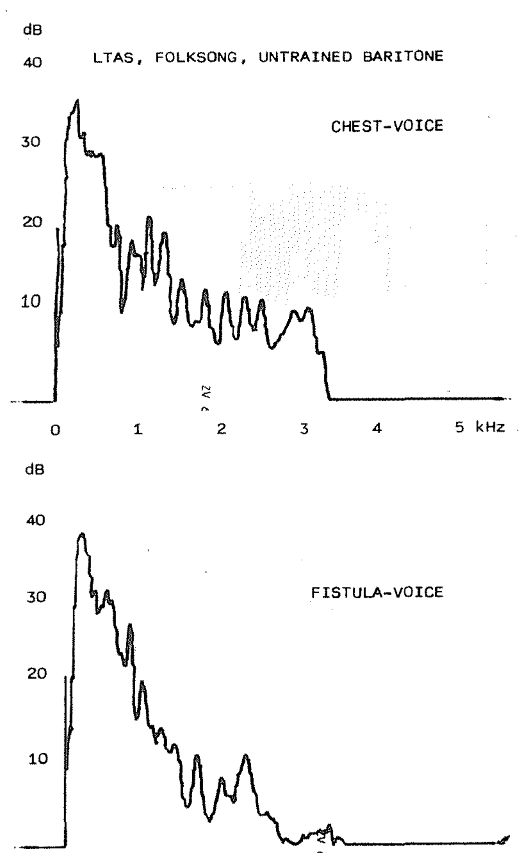


Fig. 4.

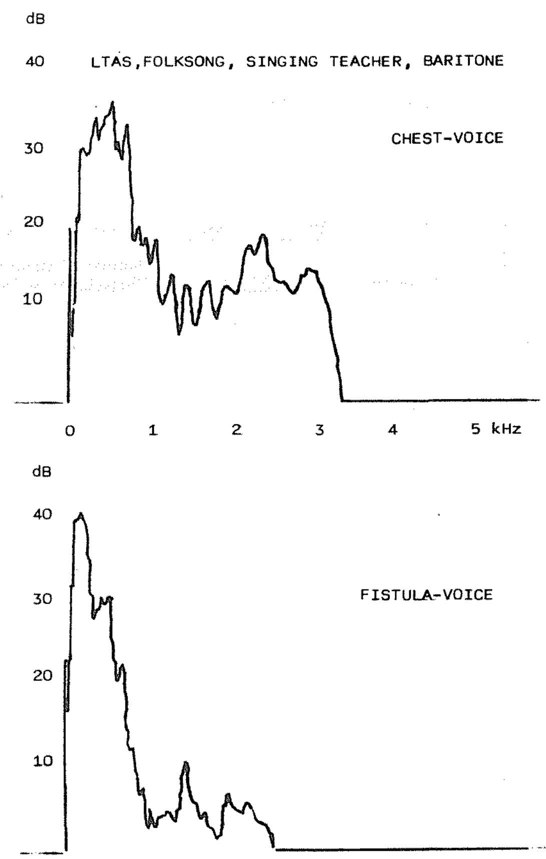


Fig. 5.

Figs. 4-6. LTAS of the folksong "Sah ein Knab ein Röslein stehn", untrained singer (4), singing teacher (5), professional countertenor (6)

After evaluation of our results as a whole, we would like to modify Sonninen's thesis in the following way: There are two physiological or natural main registers caused by two distinctively different ways of glottal operation; these registers are primarily laryngeal. They form the basis for the development of artistic registers which seem to be the result of vocal tract influences and/or interactions between laryngeal and pharyngeal adjustments.

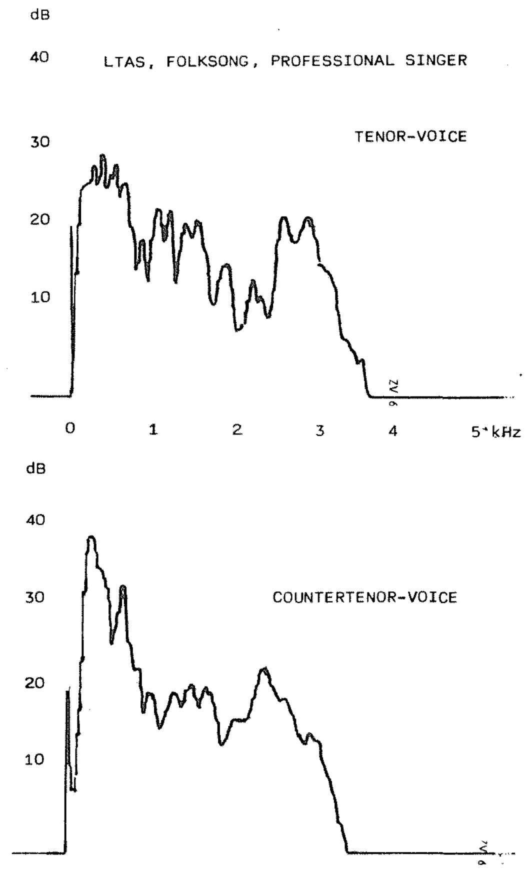


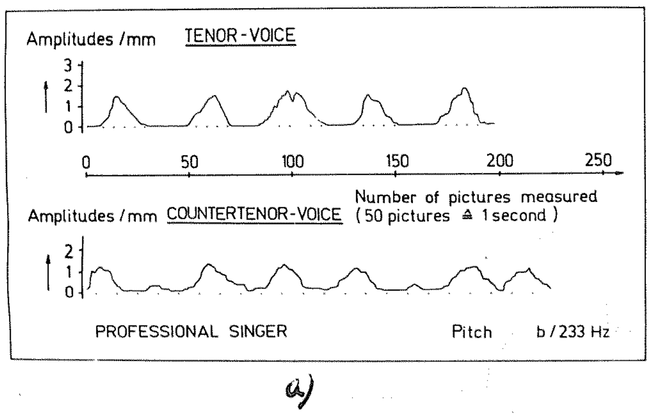
Fig. 6.

The above statement is true only for male voices. For females, it is highly exceptional that they are able to phonate in different natural registers. We will return to that point in connection with a more comprehensive study on both natural and artistic registers.

<u>References</u>

Garcia, M. (1855): "Observations on the human voice", Proc. Royal Society, <u>3</u>, pp. 399-408.

Muller, J. (1840): <u>Handbuch der Physiologie des Menschen, Bd. II</u>, Hölscher, Coblenz.



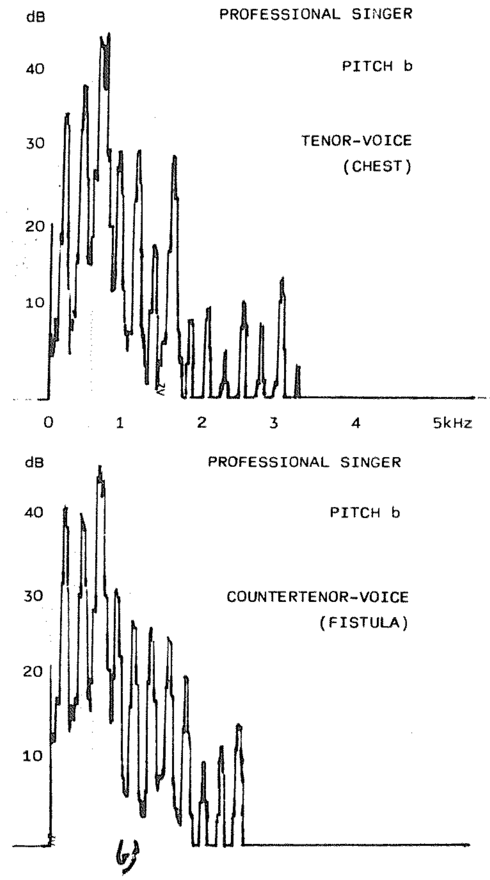


Fig. 7. a): Stroboglottometry
b): Spectra during mirror phonation

Seidner, W. and Wendler, J. (1981): "Vocal registers in voice-area measurements", paper presented at the 8th Int.Symp. of CoMeT, Amsterdam.

Seidner, W., Wendler, J., Wagner, H., and Rauhut, A. (1981): "Spektrales Stimmfeld", HNO-Praxis, 6, pp. 187-191.

Sonninen, A. (1962): "Paratasisgram of the vocal folds and the dimensions of the voice", Proc. 4th Int.Congr. of Phonetic Sciences, Helsinki 1961, pp. 250-258.

Wendler, J. and Seidner, W. (1982): "Considerations on vocal registers", Vox-Humana (Sonninen-Festschrift), Papers in Speech Research, 5, pp. 153-164. The Institute of Finnish Language and Communication, Univ. of Jyväskylä.

Wendler, J. and Tostmann, R. (1981): "Stroboglottometric measures in phoniatric voice diagnostics", Proc. 4th F.A.S.E. Symp. on Acoustics and Speech, Venice, p. 113.

** SOUND EXAMPLES **

SIDE A

G. BROCK-NANNESTAD: The Masking of Characteristic Spectra by the Acoustico-Mechanical Recording Process

Track 1

a) Adelina Patti as found, including both high note, change of register and piano interlude
b) Adelina Patti corrected, identical section as in Ex. 1.

Track 2

a) Jussi Björling: last notes of "La donna è mobile", filtered, played twice.
b) Enrico Caruso: last notes of "La donna è mobile", filtered, played twice.

M. CASTELLENGO, B. ROUBEAU & C. VALETTE: Study of The Acoustical Phenomena Characteristic of the Transition between Chest Voice and Falsetto

Track 3 Ascending and descending scale sung by a countertenor, Fig. 6 (12").

Track 4 Isoparametric sounds of Fig. 10 (19").

Track 5 Short example of yodel, Fig. 11, played twice (11").

J. ESTILL, T. BAER, K. HONDA & K.S. HARRIS: Supralaryngeal Activity of Three Male Voices of Different Quality

Examples of sung voice qualities, each example contains an ascending scale (repeated) and a short melody.

Track 6 Speech quality (41")

Track 7 Falsetto (43")

Track 8 Low larynx (46")

Track 9 Twang (39")

Track 10 Opera (43")

Track 11 Belting (32")

SIDE B

A. JOHNSSON, J. SUNDBERG & H. WILBRAND: "Kölning". Study of Phonation and Articulation in a Type of Swedish Herding Song

Examples of Swedish hearding song in field recordings.

Track 1 Karin Edvardsson-Johansson (b. 1909), Transtrand, Dalecarlia (1´25")

Track 2 Kristina Nordin (b. 1906), Klövsjö, Jämtland (42")

Track 3 Matilda Nord (b. 1898), Ytterhogdal, Jämtland (56")

Track 4 Dansar Edvard Jonsson (b. 1893), Malung, Dalecarlia (33")

C. SCULLY 6 E. ALLWOOD: Simulation of singing with a composite model of speech production

Track 5 (11")
- a) Early attempt of singing
- b) Male singing (ɑ), Fig. 1 with vibrato by cyclical variations in subglottal pressure, with a wide pharynx, without nasal coupling.
- c) The same example as in b) but with nasal coupling, see Fig. 5.
- d) (ɑ) with vibrato by cyclical variations in vocal fold ´tension´ Q, with a wide pharynx and with nasal coupling included, see Fig. 6.
- e) The same example as in d) but with a narrow pharynx.

Track 6 (7")
- a) Changing notes (non-vibrato) with ´non-legato´ transitions.
- b) The same example as in a) but a slightly different larynx type.
- c) The same example as in a) but with varying glottal losses during the voicing cycle.

Track 7 (5")
- a)-d) Four examples of ´speech-like´ pitch changes on an (ɑ) vowel.

Track 8 (5")
- a)-d) Four examples on "toffee" using the complete model, including turbulence noise and transient acoustic sources.

Track 9 Controlling loudness five ways: (15")
- a) Subglottal pressure raised while Q is reduced so as to maintain fundamental frequency constant at 262 Hz.

b) Subglottal pressure raised with Q kept constant so that fundamental frequency changes.
c) Glottal area decreased with subglottal pressure kept constant.
d) Glottal area decreased as in c) but subglottal pressure raised.
e)-g) Increasing magnitudes of swings in subglottal pressure.

Track 10 First attempts of female singing: (ɑ) with a vocal tract 14 cm long (3").
a) C4 (middle C)
b) C5 (an octave higher)