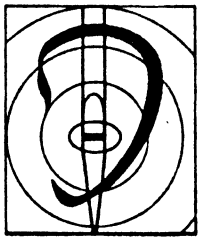


RESEARCH ASPECTS ON SINGING

AUTOPERCEPTION
COMPUTER SYNTHESIS
EMOTION
HEALTH
VOICE SOURCE

Papers given at a seminar organized
by the Committee for the Acoustics of Music

Publications issued by the Royal Swedish Academy
of Music No. 33
1981

PREFACE

This book presents the contributions to the sixth public one-day seminar jointly arranged in September 1980 by the Committee of Music Acoustics, which is sponsored by the Swedish Academy of Music, and the Department of Speech Communication and Music Acoustics, KTH. As with the preceeding seminars in this series, the purpose was to present results from acoustics research to those, who for reasons of profession or otherwise are interested, regardless of their background knowledge. This seminar focussed on the singing voice as seen from different points of scientific observation. Therefore, the contents of this book is rather varied. What unifies the different articles is, however, not only their focussing on the singing voice, but also that they deal with aspects, which are either new or frequently overlooked. Thus, Rothenberg's paper points at new theoretical possibilities regarding the voice source and particularly the reaction from the vocal tract on the vibrating glottis; Bennett's paper describes details of individual voice characteristics in singers and so stresses the rarely appreciated overwhelming complexity of even the simplest sequence of sung notes; Fonagy's research has rarely been recognized previously among voice and singing researchers even though he started to publish it a good many years ago; Fritzell's article and my own second paper are compilations of litterature on the health of the voice and on autoperception, i.e. how one perceives one's own voice during phonation. It is hoped that, in this way, the book will be found interesting and thought provoking.

The responsibility for editing and typing all articles (except Rothenberg's) has been my own, which however was a reasonable burden thanks to the availibility of a microcomputer system with a forceful text editing program. The articles have not been forced into any uniform, so the reader will find and hopefully also enjoy a variability regarding e.g. headings and figure captions. Erik Jansson and Sten Ternström have taken the responsibility for the figures and the mastertape of sound illustrations, respectively. Karin Holmgren has read and improved the manuscripts written by non-English authors from a language point of view.

KTH, October 1981

Johan Sundberg
President of the Committe for
Music Acoustics

THE VOICE AS A SOUND GENERATOR

by professor JOHAN SUNDBERG, Department of Speech Communication and Music Acoustics, K T H (RIT), Stockholm

What it is that is going on when we generate sounds with our voice? What is that gives to the voice sounds their timbral characteristics? These are the two questions which the present article will try to answer.

Fig. 1 offers a schematical view of the components involved in sound production by means of the voice organ. In terms of functions, the voice

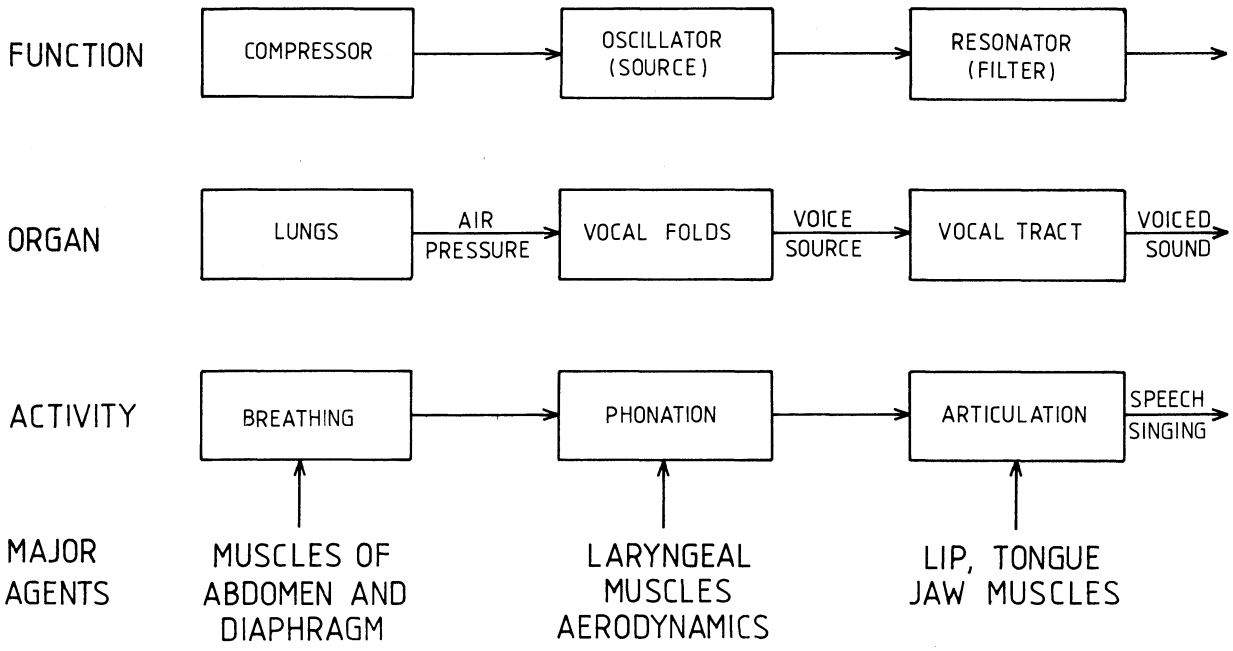


Fig. 1. Block scheme of voice production in engineering terminology (upper row), in voice terminology (middle) and in muscle terminology.

organ can be said to consist of a compressor driving an oscillator, which is connected to a resonator. This way of describing the system would be helpful to an engineer, but hardly to the layman. To him it might be more elucidating to say that what was just called an oscillator can be regarded as a sound source while the resonator is a filter, or a simple sound transmitter.

Evidently, different parts of the vocal apparaus are rsponsible for these various functions. The compressor function is handled by the lungs and the respiratory muscles. The action of this system is called breathing, as we all know. The resulting overpressure of air in the lungs drives the vocal folds, which start vibrating. The result of this is a pulsating flow of air. For each cycle of vocal fold vibration, one air pulse is generated.

This pulsating flow of air can also be described as a sound, the voice source, which is composed of several partials. The frequencies of these partials constitute a harmonic series. The partial with the lowest frequency is called the fundamental. It has a frequency equal to the frequency of the vocal fold vibration. Apart from the fundamental there is a large number of overtones in the voice source.

This entire family of simultaneously sounding tones is sent into the vocal tract with the air outside the lip opening as the next destination. The ability of the vocal tract to transfer these tones is highly variable depending on the frequency of the particular tone to be transferred. The frequencies which are most successful in travelling through the vocal tract are called resonance or formant frequencies. Those partials in the voice source which lie closest to such a formant frequency leave the lip opening with greater amplitude than other partials even if they entered the vocal tract with approximately the same amplitude. In this way, the formants enhance certain partials in the voice source and give them a greater amplitude than other partials. Therefore the vocal tract resonances can be said to give to the radiated sound its final acoustic form. This is the reason why the resonances of the vocal tract are called formants.

The formant frequencies are determined by the shape of the vocal tract, or the articulation, i. e. the adjustment of the lip and the jaw openings, the tongue shape, the soft palate, and the larynx. As soon as all these articulators have been adjusted in a given way, all formant frequencies are locked to certain values. And as soon as the position of one of the articulators is changed, the formant frequencies also change. Thus, note that the formant frequencies depend on articulation only, and the pitch of the sound produced is determined by the vibration frequency of the vocal folds, which seems to be an entirely different system. This is the reason why oscillator and resonator are placed in separate boxes in Fig. 1.

By now it should be quite clear that the acoustic characteristics of voiced sounds, including the personal voice timbre, is determined by two

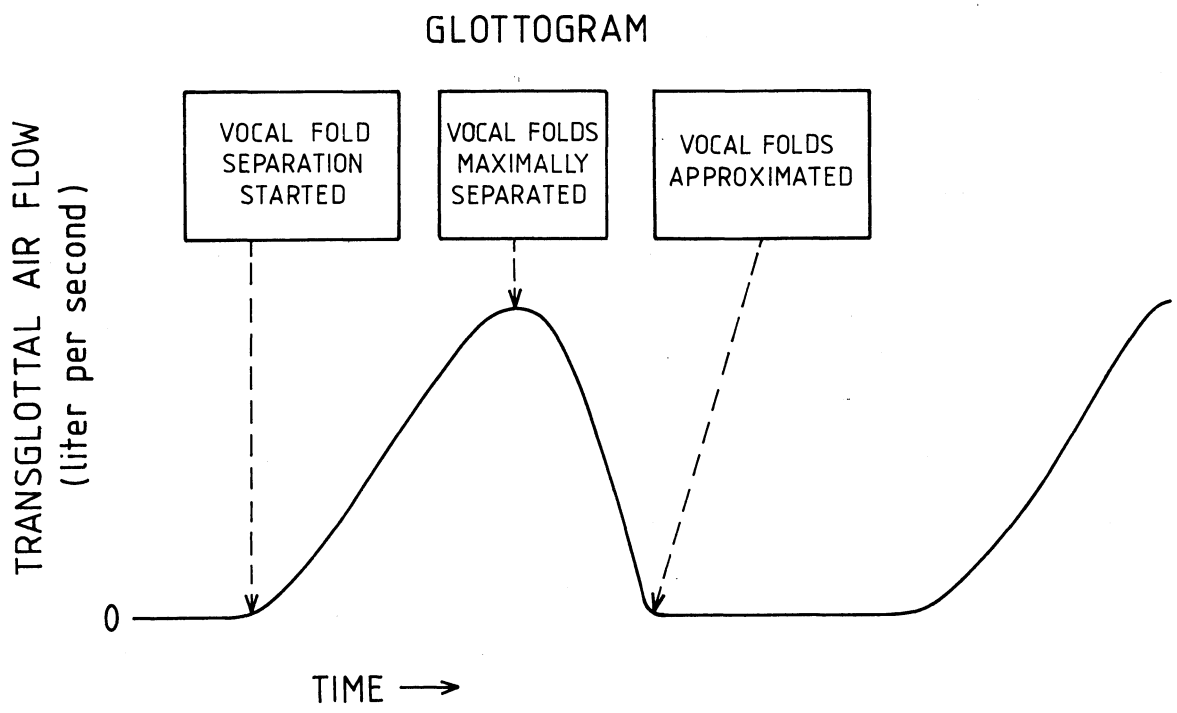


Fig. 2. Schematic illustration of voice source waveform (glottogram) showing the transglottal airflow during phonation. The glottogram also informs about certain aspects of vocal fold vibration.

separate factors: (1) the voice source which is controlled by the vibrations of the vocal folds, and (2) the constellation of formant frequencies as determined by the articulatory configuration. The articles included in this book will focus on the voice source, while the articulation and the formant frequencies will be largely disregarded. On the other hand the formant frequencies has been described in detail in previous publications in this series (Sundberg 1977 and 1978, Chowning 1980). The main question to be considered here is: what are the contributions to the personal voice timbre of the voice source, i. e. the vibrations of the vocal folds?

The air stream, which travels across the glottis during phonation, normally varies between the value of zero liters/sec, which occurs when the vocal folds close the glottis entirely, and a different maximum value, which reflects the fact that the vocal folds open the glottis maximally. Fig. 2 shows a typical although schematical record of the airflow across the glottis during phonation. The horizontal part of the curve corresponds to zero flow, reflecting the fact that the glottis is closed. The curve rises as soon as the glottis opens and falls when the glottis closes. The advantage of this sort of record of the transglottal airflow during phonation is that it is most revealing both as regards the way the vocal folds vibrate and the effect on the voice timbre. This type of record is known as a glottogram.

The informative power of glottograms is easily demonstrated if we compare a couple of glottograms selected from different types of phonation. Fig. 3 offers a typical example of an untrained voice phonating at different degrees of vocal effort. The small ripple in these glottograms should be disregarded, they are artifacts. The point is, instead, the fact that the airflow decreases from maximum to minimum value quite slowly in soft phonation and quite abruptly in loud phonation. As mentioned above this is a typical observation. It seems that this is the way the voice source regulates the amplitude of voiced sounds (Fant 1980, Gauffin & Sundberg 1980). A second observation can also be made: at low degrees of vocal effort the glottogram never reaches the value of zero liter/sec. This value is reached only in higher degrees of vocal effort. This implies that the vocal folds fail to close the glottis completely in soft phonation. This is typical of untrained voices.

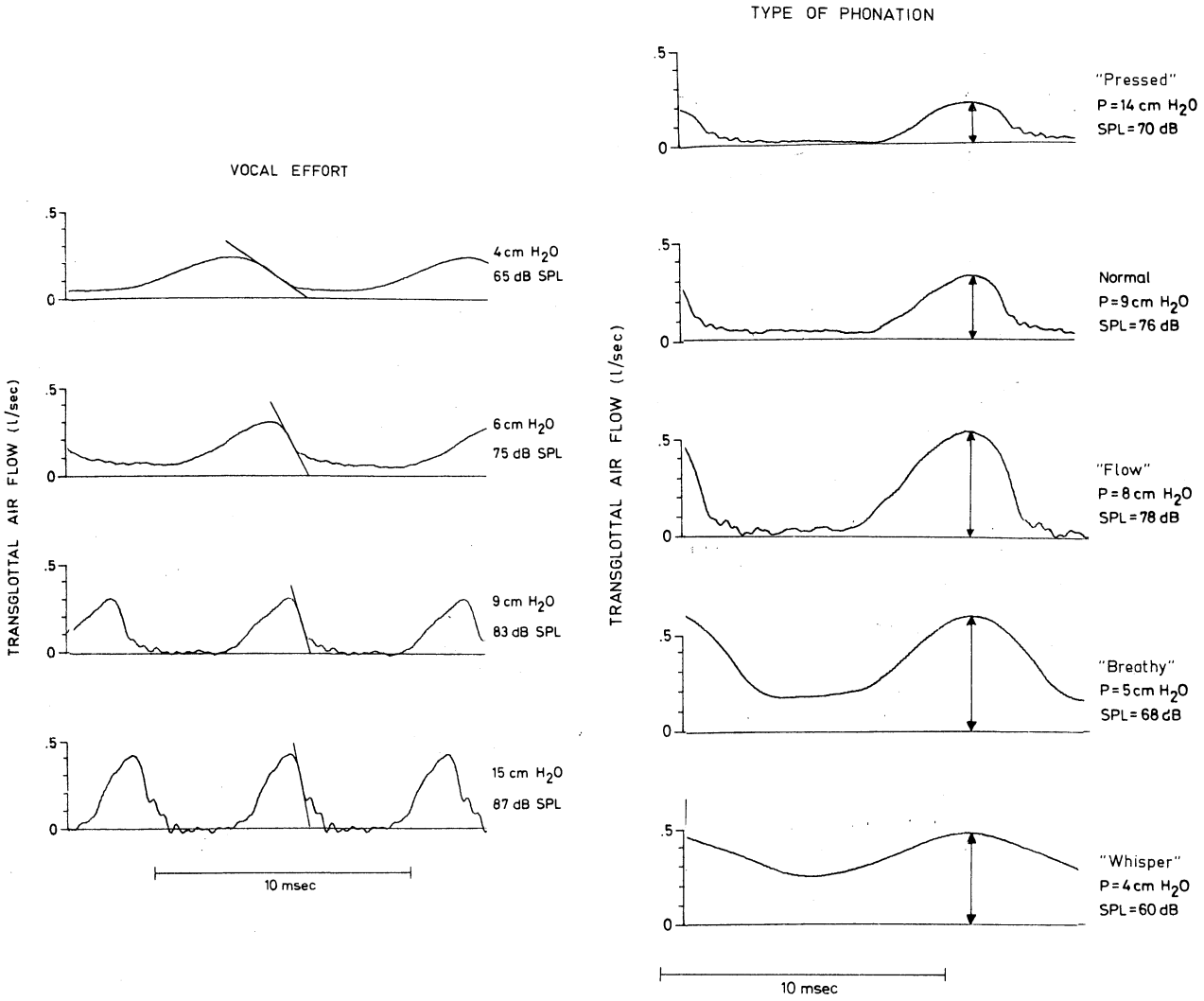


Fig. 3 (left graph). Glottograms pertaining to different degrees of vocal effort as measured in an untrained subject. Note that the subglottic pressure, which is given in cm H_2O, is decisive to the resulting sound pressure level (SPL).

Fig. 4 (right graph). Glottograms for different types of phonation in an untrained subject. Note the effect on the glottogram amplitude of changes along the phonatory parameter which ranges from "pressed" over "flow" to "breathy" phonation. Note also that, in this case, the "flow" phonation generates the loudest vowel sound.

Apart from the rate at which the airflow across the glottis change from its maximum to its minimum value there is another important property of phonation that can be seen in glottograms. This is demonstrated in Fig. 4, and some extreme examples can be listened to in SOUND EXAMPLE 1. In one extreme case the subglottal pressure is high and the vocal folds are forcefully pressed together, or adducted. This type of phonation will be referred to as "pressed" phonation. Phonation is also possible with lower subglottic pressure and more moderate adduction activity. If this is done phonation will turn into to what we call "flow phonation", and if phonation is changed still more in the same direction the phonation will change into "breathy". Then, the vocal folds do not make contact any longer. Thus, we have a phonatory dimension which extends from "pressed" over "flow" to "breathy".

A change of phonation along this dimension can be seen in the glottogram. When phonation moves away from "pressed", the glottogram amplitude increases (Gauffin & Sundberg, 1980). This can be observed in Fig. 4. "Pressed" phonation is characterized by a long horizontal portion (closed phase) of the glottogram and a small glottogram amplitude. "Flow" phonation, on the other hand, is characterized by a considerably greater glottogram amplitude and a shorter closed phase.

Up to now we have seen that there are two glottogram characteristics which correlate quite nicely with properties of phonation: (1) the degree of vocal effort is reflected in the rate of airflow change from maximum to minimum value; and (2) the position along the phonatory dimension "pressed/flow" is reflected in the glottogram amplitude.

There is one more point: these glottogram differences affect in a known way the voice timbre as well as the overtone content, i. e. the spectral characteristics of the sound radiated from the lip opening. The background is the following.

A glottogram is an example of a sound waveform. The waveform of a sound reflects the content of partials of that sound. The more abrupt changes a waveform exhibits, the greater the number of high overtones in the corresponding sound. Thus, a sound with very weak high overtones has a waveform describing a very smoothly changing curve. A sound with many strong

high overtones, on the other hand, has a waveform with sharp discontinuities.

Returning to the glottograms in Fig. 4, we note that the glottogram characteristic reflecting the degree of vocal effort can be described as the abruptness with which the curve moves from maximum to minimum. Hence, we conclude that this glottogram characterstics corresponds to the wealth in high overtones in the spectrum of the voice source. This agrees well with the fact that the loudness of voiced sounds is normally dependent on an overtone. As regards the glottogram amplitude it can be shown that it is closely related to the amplitude of the first partial, i. e. the fundamental of the voice source spectrum.

Thus, to summarize, there are two properties of glottograms which are relevant for voice timbre. One is the rate at which the airflow across the glottis decreases from maximum to minimum. This rate sets the overtone content of the voice source spectrum and normally also the acoustic overall amplitude of the sound radiated from the lip opening, given the formant frequencies. The other is the glottogram amplitude which sets the amplitude of the fundamental of the spectrum radiated from the lip opening, also given the formant freqencies.

From the above we conclude that these voice source characteristics can be heard in voiced sounds. This is both a trivial and a sensational statement. It is trivial in the sense that we all know very well that we can hear properties of the voice source, i. e. the way in which the vocal folds function, in the sound of the voice. Of course we are able to hear from the voice timbre when a person is using "pressed" phonation and when a person speaks loudly. But these close relationships between glottogram and voice characteristics are sensational in the sense that it is quite rare that such relationships are this simple. In most cases we have only vague ideas about the way in which sound production characteristics are manifested acoustically. We know them simply by intuition.

It has been suggested above that the formant frequencies, i. e. the articulation is also relevant as a determinant of the spectrum of voiced sounds as radiated from the lip opening. Before we infer certain voice source properties from this spectrum we must compensate for the influence

of the formant frequencies. Thus, the fundamental may be quite strong in the radiated spectrum because its frequency is close to the first formant, as in female high-pitched opera singing (Sundberg, 1975) or because phonation was close to the "flow" extreme. It seems that voice teachers are able to realize which of these two reasons applies in the practical case.

If we want to find out the same thing by technical means, though, the task is quite complicated. Then an inverse filter is needed, as described in the article Martin Rothenberg in this volume.

The singer's formant is a typical spectral characteristic of voiced sounds sung by male opera and concert singers (see e .g. Sundberg 1975). Acoustically it can be described as a peak in the spectrum envelope appearing somewhere in the neighborhood of 3 kHz. In this frequency region, then, the partials radiated from the lip opening are particularly strong. Articulatorily the singer's formant can be generated by adjusting the pharynx width so that it is considerably wider than the area of the entrance to the larynx tube. If this is done, the formants number three, four, and probably also five are clustered and the ability of the vocal tract to transport sound in this frequency range is very much improved. The result of course is that the voice source partials in this frequency range gain in amplitude. The singer's formant has the effect that the singer's voice is more easily to discern against the background of an orchestra, as is demonstrated in the sound illustrations accompanying one of my previous articles (Sundberg 1977). Thus, the singer's formant is very important to an opera singer.

However, it will be clear that the amplitude of the partials underlying the singer's formant are dependent not only of the vocal tract sound transfer characteristics, but also on the voice source characteristics, or, in other words, the initial amplitude of the partials as they enter the vocal tract. As mentioned above, this initial amplitude depends on the rate of change from maximum to minimum airflow value. An interesting question is how this rate can be manipulated. We have seen that it increases as vocal effort is increased. Vocal effort is raised primarily by increasing subglottic pressure, so this pressure seems important. The rate of decrease in the airflow is also influenced by some other fac-

tors. As Rothenberg & Zahorian (1977) has shown the values observed in glottograms of singers seem too high to be explained by movement of physical structures such as the vocal folds. In his article in this book Martin Rothenberg will explain this more in detail.

REFERENCES

CHOWNING, J. M. (1980): "Computer synthesis of the singing voice", in Sound Generation in Winds, Strings, Computers, Publications issued by the Royal Swedish Academy of Music #29, 4-13

GAUFFIN, J. & SUNDBERG, J. (1980): "Data on the glottal voice source behavior in vowel production", Speech Transmission Laboratory Quarterly Progress and Status Report 2-3/1980, 61-70

FANT, G. (1980): "Voice source dynamics", Speech Transmission Laboratory Quarterly Progress and Status Report 2-3/1980, 17-37

ROTHENBERG, M. & ZAHORIAN, S. (1977): "Nonlinear inverse filtering for estimating the glottal-area waveform", Journ. Acoust. Soc. Amer. 61, 1063-1071

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

SUNDBERG, J. (1977): "Singing and timbre", in Music, Room, Acoustics, Publications issued by the Royal Swedish Academy of Music #17, 57-81

SUNDBERG, J. (1978): "Rent och falskt i klingande praxis", in Vår Hörsel och Musiken, Publications issued by the Royal Swedish Academy of Music #23, 78-101

THE VOICE SOURCE IN SINGING

by professor MARTIN ROTHENBERG, Department of Electrical and Computer Engineering, Syracuse University, Syracuse, New York

The acoustic theory of speech production, as first proposed and as generally now implemented in formant-based speech synthesis, models the speech production mechanism during vocalic sounds with three relatively independent subsystems (Fant 1960, Flanagan 1972). These subsystems, shown diagrammatically in Fig. 1, are (1) the respiratory system, which produces a slowly-varying tracheal air pressure, (2) a time-varying glottal flow resistance whose valving action creates quasi-periodic air pulses, and (3) a supraglottal vocal tract that shapes the spectrum of the glottal flow pulses. Though each of the systems interacts with the other two systems to some degree, order-of-magnitude calculations, model studies and early measurements have indicated that for many applications it is sufficient to consider these three subsystems as operating independently, at least during voiced sounds with no strong supraglottal oral constriction.

However, as we look for more precise models of the voice source, whether this be for higher quality speech synthesis, the synthesis of the singing voice, or the study of voice pathology, it is necessary to return to an

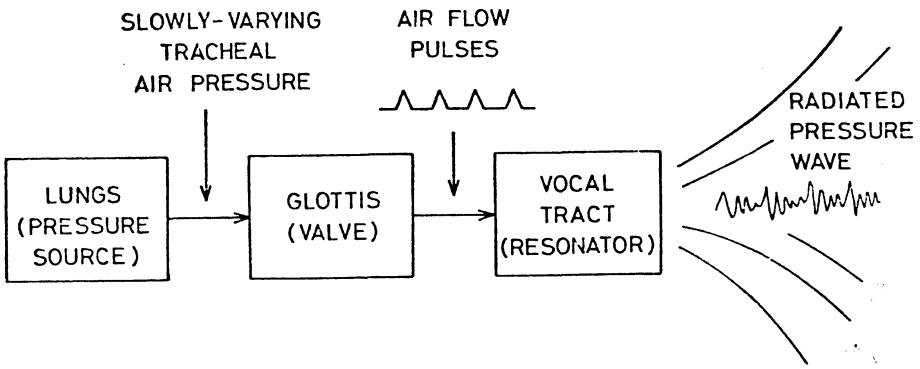


Fig. 1. Schematic representation of a non-interactive model of voiced speech production.

interactive model. This paper presents a model of the voice source that includes in a relatively simple way the acoustic interaction between the voice source and subglottal and supraglottal systems, in order to see what effect this interaction might have on the singing voice.

If we look at the minimally breathy and relatively tonal (not purposely rough or aperiodic) voice most common in singing, we find that the vibrating vocal folds open and close periodically to allow puffs or pulses of air to pass from the trachea into the pharynx. The period of these pulses (time between repetitions) is the basic determinant of the pitch of a sung note, whilethe waveshape of the airflow pulses (the shape of the plot of flow vs. time within each pulse) helps determine the quality of the note. Since the acoustic interaction we speak of here affects primarily the voice quality, and not the pitch, we will concentrate on the quality of the voice in the following discussion.

The pattern of vibration of the vocal folds during voice production is often described by the "projected glottal area," (abbreviated here as PGA), i.e., the area of the opening that would be seen from directly above or below the glottis. Measurements on high speed or stroboscopic motion pictures or from recordings of the light projected through the glottis from a source either above or below (the "photoglottograph" technique) have shown that the waveform of projected glottal area tends to consist of rather triangular pulses separated by flat portions at or near zero area. The former represent the open portions of the glottal cycle, while the latter represent the periods during which the folds are closed at some level along their vertical dimension. The apex of the triangle is often (but not always) found to be rather pointed. This pointed triangular appearance, when it occurs, is generally believed to be due to a phase difference between the movements at the upper margins and lower margins of the folds. The rising segment of the triangle would represent the area at the upper margins of the folds as they open (the lower margins having opened previously do not effect the projected glottal area during this period). The decaying segment represents the area at the lower margins of the folds as they close (while the upper margins are still open). The triangular area pulse (whether pointed or rounded) can have a small dissymmetry, either to the right (with the opening phase slower than the closing) or to the left, but tends to be rather symmetri-

cal.

The significance of projected glottal area (PGA) in simple models of vocal tract acoustics is that if during the glottal cycle the variations in air pressure just inferior and superior to the glottis were relatively small compared to the average transglottal pressure (the condition assumed in the non-interactive model for voice production), the waveform of the volume air flow (volume velocity) through the glottis would tend to have a shape rather similar to that of the projected area, and it is this glottal flow that supplies the acoustic energy to the vocal tract for voice production.

However, the glottal air flow (abbreviated here as GAF) during voice production has been generally found to have a waveform which is considerably more tilted to the right than is the PGA waveform, especially during open vowels (Miller 1959, Holmes 1963, Lindqvist 1965, Rothenberg 1973). The typical GAF waveform will have a slower and smoother increase in the glottal opening phase, and a more sharply decreasing closing phase, without the peak at the apex that one can find in many PGA waveforms. Acoustically, this difference is very meaningful, since the acoustic quality, or pattern of amplitudes in the harmonics or overtones, would be very different for these two waveforms, as explained below.

Fig. 2 shows some representative PGA and GAF waveforms to illustrate these differences. The two PGA waveforms 2a,b are approximations of PGA from a photoglottograph in which the light was introduced just below the thyroid cartilage, at the centerline of the neck, and picked up by a photocell at the back of the pharynx, just above the glottis. The glottis is not uniformly illuminated by the light source, as would be necessary for a true PGA measurement, however, since the light source is placed so that there is a clear band of light across the glottis, somewhere near the center of its length, the resulting waveform should show the general characteristics of the variation of glottal area. Waveform 2c is also an approximation of the PGA from measurements of the glottal width near its center. However, in this case the glottal width was obtained from frame-by-frame measurements of high speed motion pictures, taken from above.

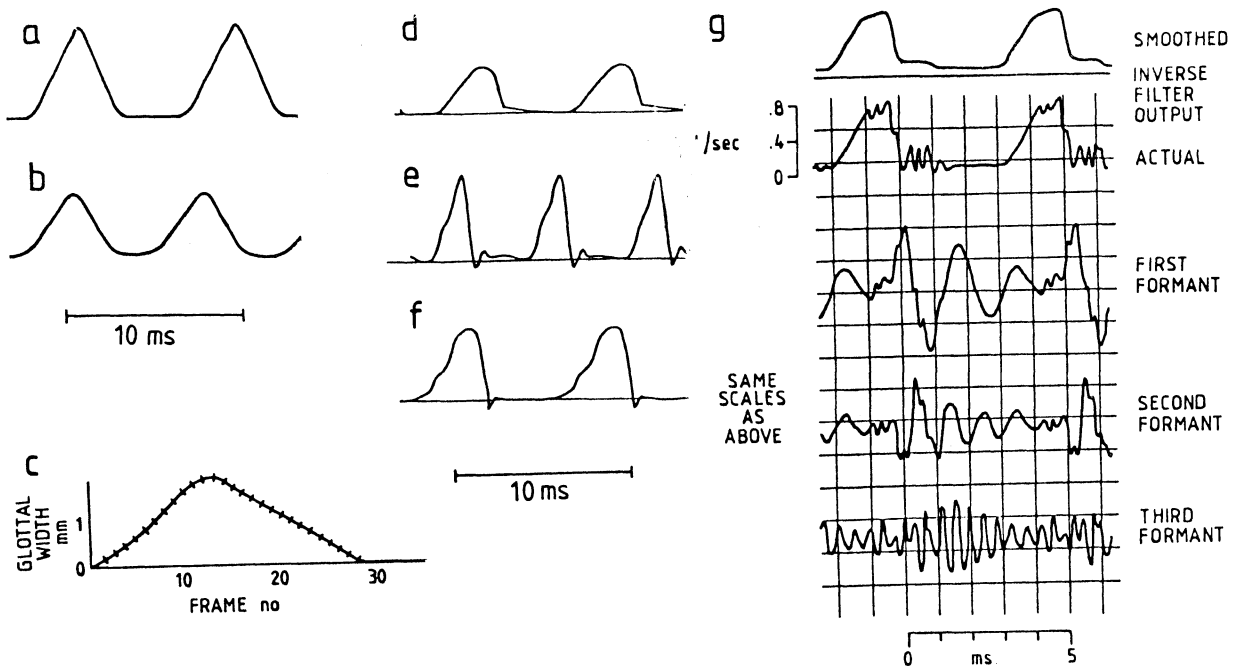


Fig. 2. Glottal waveforms during voiced speech.

a. and b. Approximation to the projected glottal area using the photoglottographic technique with an adult male subject. From Kitzing (1977). a is at 113 Hz and b at 136 Hz.

c. Approximation to the projected glottal area using frame-by-frame measurements of glottal width on high speed motion pictures. From Hirano, et al (1981). The abscissa is the frame number, each frame being about 0.2 msec. Subject was a 22-year-old male bass singer phonating in a modal register at a fundamental frequency of 128 Hz.

d, e, f, and g. Volume air flow at the glottis obtained by inverse filtering the air flow at the mouth. All subjects were male adults, with those in e, f, and g either amateur of professional singers. The lowest three formants were cancelled by the inverse filter, with higher vocal tract resonances attenuated by low pass filtering. The vowel in d, e, and f was either /a/ or /ae/. The vowel in g was /a/, sung with a strong "ring". It contains some remanent energy at a strong vocal tract resonance above the third formant. The amplitude scale shown holds only for g. The small negative "overshoot" after the glottal closure in e and f may be an artifact due to imperfect inverse filtering. Figure g also shows the energy at F1, F2 and F3 that was removed from the oral flow signal by the inverse filtering process.

The three waveforms in Fig. 2d,e,f show some typical GAF waveforms during open vowels in speech and singing. They were all obtained by the inverse filtering the air flow at the mouth, using a system with a response time of about 0.3 ms (Rothenberg 1977). This means that changes in the waveform occuring in roughly 0.3 ms may actually represent a change in flow that was significantly faster. Note that the decay in air flow in these waveforms is considerably faster than the rise in flow. In the waveforms taken during singing, the time required for the final decrease in flow is close to or possibly below the response time of the measurement system. No such rapid termination of the glottal pulse has ever been reported in area waveforms.

The acoustic significance of the highly unsymmetrical flow waveforms has two principal aspects:*

(1) A rapid termination of the glottal pulse followed by a period of zero or a small constant flow causes a strong excitation of the higher order vocal tract resonances (formants). Because the amplitude of the high frequency energy generated at closure varies directly as (though not strictly in proportion to) the amplitude of the final slope of the decay in air flow, those GAF waveforms with a more rapid decay of flow would contain the greatest amount of energy at frequencies higher than those generated at reasonably high amplitudes by most musical instruments, and so would carry better above a musical accompaniment. Since a rapid decay of flow would cause the third and higher formants to be stronger and, to a lesser extent, also the second formant, the resulting voice quality voice would tend to be clearer and more intelligible.

It should be added that the difference between a voice source that is weak at high frequencies and one that is strong is not just in the relative amplitude of the harmonics, but usually also in their periodicity. If we use a suitable band pass filter to isolate the third or fourth formant energy in the radiated acoustic wave, and observe the

---

* Statements in this paper referring to accepted principles in the frequency analysis of waveforms or the properties of linear acoustic systems will not be documented by specific references.

waveform on an oscilloscope, it can be seen that the strong voice will contain a relatively strong oscillation at the third, and usually also the fourth formant frequency (see fig. 2g, for example). The pattern in this oscialltion will repeat itself very closely in every glottal cycle, i.e., the waveform of the formant energy will be very periodic or tonal. On the other hand, the waveform from the weak voice will often have energy at the third and higher formants that is not only weaker, but more variable in its strength and pattern in each glottal cycle, i.e., that is less tonal. This difference in tonality is significant in that it means that the weak voice cannot be made into the strong one simply by electronic or acoustic amplification of the higher formants. Such high frequency emphasis could partially correct the relative formant amplitudes, but could not improve the tonality.

(2) The smoother onset of the glottal flow pulse would mean that less high frequency acoustic energy was generated at that point. This means that with a highly unsymmetrical flow pulse, most of the energy at the second and higher formants will be generated at the instant of glottal closure. From the principles of frequency (or Fourier series) analysis it can be shown that a periodic flow waveform with only one discontinuity in slope (at closure) will have a distribution of energy in which the amplitude at the harmonics of the fundamental frequency tend to decay uniformly with increasing frequency, with the rate of decay depending on the sharpness of the cessation of air flow. On the other hand, a symmetrical flow pulse will have dips in the glottal spectrum near those frequencies at which the energy generated at the opening of the glottis is partially cancelled by the energy generated at the glottal closing, i.e., at which the ~~with~~ interval between the opening and closing is an even number of cycles. If a formant of a vowel held at a constant pitch and articulation (as in singing) were to fall near such a "low-energy" frequency, it would not be as strongly transmitted, and the resulting vowel might not be expected to be as clear as in the single-excitation (unsymmetrical) case.

Thus, if the glottal flow waveform were to follow the area waveform, as predicted by early, non-interactive models of the voice source, there would be little carrying power to the voice, with vowels less distinguishable over the sound of a loud accompaniment than is the case with a

"good" singer. Resonance effects at or above 3 kHz, such as the "singing formant" reported by Sundberg (1974), could improve the carrying power of the voice, by strengthening the higher frequencies, but, needless to say, if there were not already strong high frequency components in the spectrum of the voice source, it is doubtful that such effects could produce the richness of voice quality that can be heard with some singers.

It seems, therefore, that the key to the understanding of why the voice source in some people can have a spectrum rich in higher harmonics, and which can, if properly modulated by the supraglottal vocal tract resonance, carry clearly over the sound of most musical instruments, lies in an explanation of how the glottal air flow waveform can differ so markedly from projected glottal area and, more specifically, have a much more rapid termination of the glottal pulse.

In table 1, I have listed four factors which could possibly contribute to this type of flow pattern. Of the four, I believe that the fourth is generally the most important; however, according to the present state of our knowledge of the voice source, each could be significant under some circumstances.

Table 1.
Possible causes of the asymmetry in the glottal air flow waveform

1. Asymmetry In Projected Glottal Area

2. Different Relationship Between Area and Flow Resistance During Opening and Closing Phases

3. Air Displaced By Vocal Fold Movements

4. Acoustic Energy Storage or Reactance Forces in The Vocal Tract
   a) Potential Energy - Acoustic Compliance
      (Primarily due to the compressibility of air)
   b) Kinetic Energy - Acoustic Inertance
      (Primarily due to the inertia of air flow)

The first entry in Table 1 emphasizes that the projected glottal area itself can be significantly unsymmetrical, as in the example in Fig. 2c. Any dissymmetry in the PGA would add to, or subtract from, the dissymmetry in GAF.

Proceeding to the second entry in Table 1, it should be kept in mind that the model for the glottal aerodynamics that puts glottal flow admittance (the inverse of flow resistance) largely proportional to projected glottal area is only a simple first approximation. Consider two instants, one during the opening phase and one during the closing phase, at which the PGA is the same. Though the projected areas are equal, the configuration of the vocal folds can be quite different (Baer, 1981), and therefore, the glottal admittance values could be quite different. The relationship between the shape of vocal folds and the resistance to air flow is still not well understood; however, a better understanding of this relationship may show that for some modes of vocal fold vibration, the admittance function may be quite unsymmetrical even when the PGA waveform is symmetrical.

The third factor in Table 1 is the air volume displaced by the movements of the vocal fold masses. One may think of this as the "hand clap" effect. As the folds separate, the displaced air tends to reduce the net outward glottal flow. Conversely, when coming together, the vocal folds displace an air volume that increases the net outward flow. When these displaced air components are added to the GAF waveform, the effect is to tilt the glottal pulse to the right (Rothenberg 1973). However, rough calculations of the order-of-magnitude of this effect indicate that it could cause only a small part of the dissymmetry found in the more harmonic-rich of the naturally occurring glottal waveforms (Rothenberg 1973, Rothenberg and Zahorian 1977, Flanagan and Ishizaka 1978).

At this time, it appears that the primary factor causing the GAF waveform to differ from the PGA waveform is the fourth listed in Table 1, i.e., the influence of the acoustic reactance in the vocal tract on the pressures and flows within the glottis. The term "acoustic reactance" refers to the acoustic energy stored in the vocal tract at any instant. When the energy stored is potential energy, as in the compression of a volume of air, then the acoustic reactance is referred to as a compliance. When the

energy is kinetic, as in the inertial energy stored in the velocity of the air flow at a constriction in the vocal tract, the reactance is referred to as an inertance. When the voice fundamental frequency F_0 is below the first formant frequency, as is usually the case, the reactive part of the supraglottal vocal tract impedance, as seen from the glottis, is inertive at F_0. The subglottal acoustic impedance, as seen by the glottis, also tends to be inertive for frequencies between the highest respiratory tissue resonance (of the order of magnitude of 10 Hz in adults) and the lowest acoustic resonance (of the order of magnitude of 400 Hz in adults, as can be seen in Fig. 4 below). From a simple model of the vocal tract acoustics, it will also be shown that this type of inertive loading of the glottal source at F_0 and its lower harmonics can and most likely does cause a dissymmetry in the flow waveform of the type we have discussed above.

The subglottal and supraglottal inertive loading of the glottis is shown

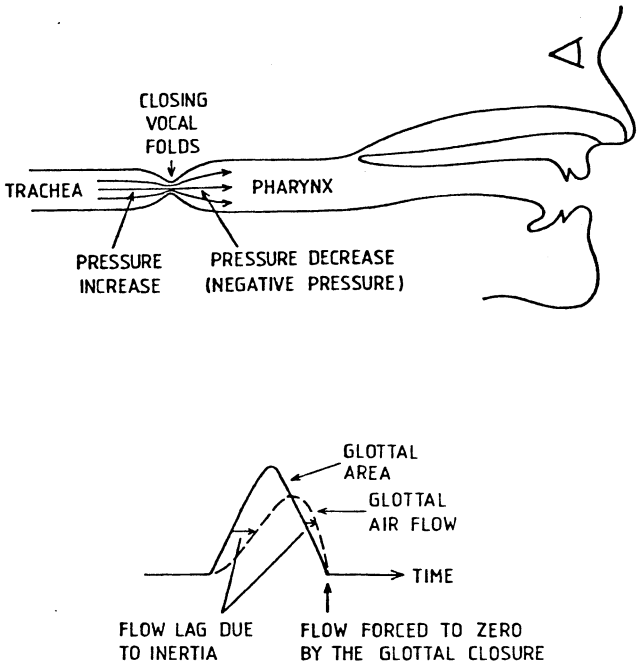


Fig. 3. Top: Diagrammatic view of the vocal tract showing the effect of flow inertance on the tracheal and pharyngeal air pressures during the closing of the vocal folds. Bottom: Diagram showing how flow inertance

in the vocal tract creates an unsymmetrical glottal air flow pulse. diagrammatically in the sketch in Fig. 3. For simplification, the vocal tract is shown as a horizontal tube, with a simple constriction representing the glottal "valve". If we picture the air masses just before and after the glottis as more inertive (mass-like) than compliant (compressible), and the vocal folds opening after being closed a long time, it can be seen that there will be a "delay" or lag in the build-up in air flow after the vocal folds open, as the lung pressure acts to overcome the inertia of the air mass. This lag is shown by the left-most horizontal arrow in the sketch of the glottal area and flow waveforms. Conversely, as the vocal folds close to reduce the flow (the condition shown in the vocal tract sketch), there is an inertive force that resists the decrease in flow. This inertive force is actualized by a momentary, inertia-induced increase in pressure in the trachea, and a decrease in pressure in the pharynx caused by the inertia of the supraglottal air pulling it away from the closing glottis.

But the inertia of the air flow does more than just delay the build-up and decay of air flow. The important added feature is that although the decay of air flow is momentarily delayed, it must finally be forced to zero at the instant of complete glottal closure (assuming that there is a complete or almost complete closure). As shown by the sketch of air flow, the requirement that the flow be zero at closure causes a sharp drop in flow just before the instant of closure that is so important in generating a strong high frequency spectrum in the voice.

Just how the vocal tract acoustic inertance affects the glottal air flow can be seen in simultaneous measurements of pressure below and above the glottis made by Koike (1980), and shown in Fig. 4. Koike used two miniature pressure transducers, one of which was suspended just above the glottis, and the other just below the glottis, with the connecting wire to the lower one passing through the glottis at its posterior end. On the figure, I have drawn in an approximation to the projected glottal area, assuming a typical waveshape, and an instant of complete glottal closure that occurs at the (simultaneously occurring) peak tracheal pressure and negative peak oral pressure. (In later measurements, Koike has verified

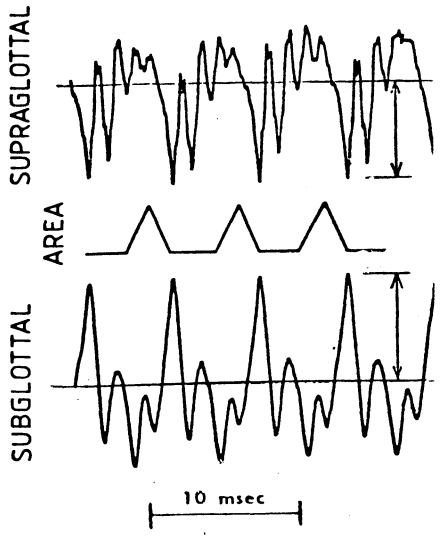


Fig. 4. Pharyngeal (supraglottal) and tracheal (subglottal) air pressures during a vowel /a/. From Koike (1981). To bring out the points emphasized in this paper, I have sketched in a simple approximation to the projected glottal area (PGA), assuming a 50% duty cycle and symmetrical opening and closing phases. The vertical arrows indicate the peak change of the pressures from their approximate average values during the glottal closing phase.

photoglottographically that the instant of closure does indeed occur as shown in Fig. 4.) Also, for similar waveforms shown on a much smaller scale, but with numerical amplitude scale given, the reader is referred to Kitzing and Löfqvist (1975).

It can be seen in Fig. 4 that as the glottal area decreases, a strong negative pressure develops above the glottis, and an increased positive pressure develops below the glottis. The net effect is to increase the transglottal pressure which is forcing air through the glottis, so as to delay the decrease in air flow that would otherwise be caused by the closing vocal folds, as was shown in the sketch of Fig. 3. When the closing vocal folds finally do cause a cessation of flow, the inertia forces are suddenly terminated, and the supraglottal and subglottal

pressures move rapidly toward their average values, in the oscillatory

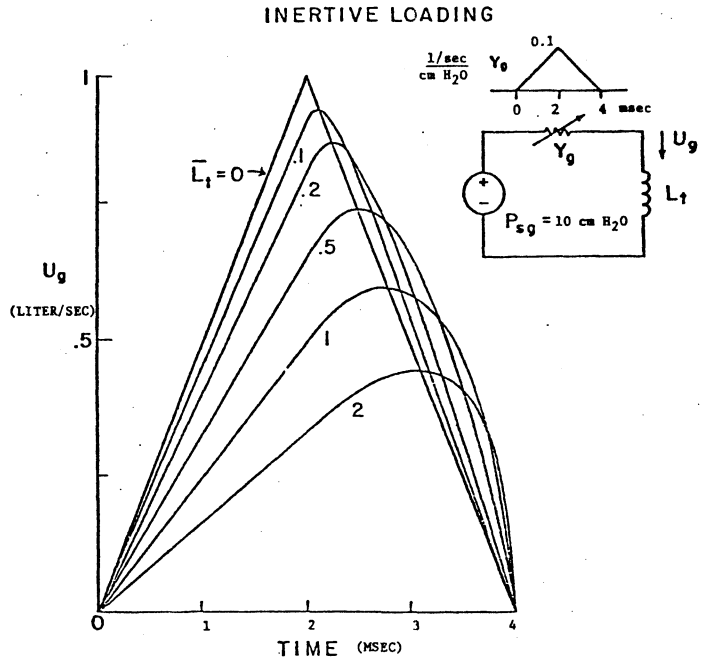


Fig. 5. The effect of inertive loading of the glottis on the glottal air flow waveform, as derived from the simple model of the vocal tract shown in the figure.

pattern caused by the respective resonances.

To see more quantitatively how the vocal tract acoustic inertance affects the glottal flow U_g, Fig. 5 shows the solution to the nonlinear differential equation formed when the glottal admittance Y_g is represented by the simple, symmetrical triangular variation of conductance (equals $1/R_g$) shown at the upper right, and the acoustic impedances of the subglottal and supraglottal systems are combined into a single inertance L_t (Rothenberg 1981). P_{sg} is the average tracheal pressure, and would be equal to the average alveolar or lung pressure if the dissipative flow resistance in the ~~average alveolar~~ bronchi and bronchioles can be neglected.

The amplitude and time scales show values that might be expected for a

male voice at moderately high vocal effort. In a given situation, the effect of a specific value of inertance depends on the amplitude of the glottal admittance and the duration of the glottal pulse. To obtain a measure of the inertance with a more invariant significance we have defined a normalized inertance $\bar{L}_t$, with actual inertance L_t and normalized inertance $\bar{L}_t$ related by the equation

$$\bar{L}_t = L_t \, 2Y_{gmax}/t_p \qquad \text{Eq. 1}$$

where t_p is the actual duration of the glottal pulse, and Y_{gmax} is the actual maximum glottal admittance. Thus, for this example, a normalized value of $L_t = 1$ is equivalent to an actual L_t of .02, in cm H_2O, l/sec units.

For the symmetrical admittance pattern in the figure, it can be shown that the terminal slope of the air flow pulse becomes infinite for $\bar{L}_t \geq 1$, and the higher harmonics therefore become quite strong as the normalized inertance approaches unity. For values of $\bar{L}_t$ less than about 0.2, the inertance has little effect on the flow pattern; however, it can be shown that for non-breathy open vowels the normalized inertance can always be expected to be at least that value, with values of at least 0.5 attainable for most speakers during normal speech.

Rough calculations indicate that it is at least possible that some individual voices could attain values of $\bar{L}_t$ of 1.0 and conceivably even as high as 2.0 with some pitch and vowel combinations. However, our knowledge to date of vocal tract acoustics does not permit the value of $\bar{L}_t$ to be tied down much more closely in the range 0.2 to 2.0. It appears though, that this is a crucial question in the understanding of voice quality. By simulating the voice source shown in Fig. 5 , using values of $\bar{L}_t$ varying in the range of 0 to 2.0, and connecting the voice to a formant-type vocal tract analog, it can be shown that differences in the value of normalized vocal tract inertance of a magnitude that could conceivably be ascribed to physiological variations between individuals can make a very significant change in the quality and carrying power of the voice.

In future research we must therefore determine what the factors are that

most strongly influence $\bar{L}_t$. Are they vocal tract features such as the degree of pharyngeal constriction? (A constriction in the flow path always tends to increase the inertance, since it increases the particle velocity at the point of constriction.) Or are there features of the vocal fold shape that influence $\bar{L}_t$ by changing the glottal admittance? (By equation 1 above, $\bar{L}_t$ is proportional to the peak glottal admittance.) There is also the shape of the entrance of the larynx to consider, and the complex pattern of air flow within the laryngeal vestibule, as a jet of air emerges from between the vocal folds during the glottal pulse. The time-varying reactive components of the flow path between the vibrating vocal folds may also be a factor; however, our experiments with an analog simulation of the time-varying inertance at the glottal constriction indicate that because of its variation in time this inertance does not add significantly to $\bar{L}_t$. (The relatively constant inertance of an emerging jet of air might be significant, however.)

From the above discussion it should not be concluded that acoustic interaction between the glottal source and the vocal tract is simply a matter of an inertive effect that is either more or less in a given voice. There are some other implications of this interaction which could be highly significant in our understanding of the singing voice. I will discuss here only two factors which seem to me to be the most significant.

First, I would like to mention only briefly that the degree of acoustic interaction is reduced when there is not a fairly complete glottal closure following the glottal flow pulse (Rothenberg 1981). Thus a voice that is breathy in the sense that there is never a complete glottal closure attained during voicing cannot develop the added carrying power that inertive vocal tract loading can bring.

Second, it is important to point out that source-tract acoustic interaction varies in type and intensity depending on the vowel (Rothenberg 1981). Vowels with a high first formant, such as /a/ or /ae/ (as in "father" and "hat," respectively) appear to have the strongest interaction of the type described above. When the first formant is low, or when the voice fundamental frequency F_0 is high, so that the ratio F_1/F_0 is low, the interaction can be much more complex and does not necessarily improve or strengthen the voice. Let us use F_R to refer to the frequency

of the lowest acoustic resonance of either the subglottal or supraglottal system. As the ratio F_R/F_0 is reduced, the interaction with the resonance energy becomes more important. As when F_R/F_0 is high, the significant factor determining the high frequency energy generated by the closing of the vocal folds is the pressure increase or decrease in the vocal tract near the glottis when the vocal folds are closing. As can be seen in the pressure recordings of Fig. 4, there are strong oscillations in pressure below and above the glottis due to the subglottal and supraglottal resonances, respectively. When F_R/F_0 is high (as in Fig. 4), these oscillations have time to decay between glottal closings, and have little effect. However, if the ratio F_R/F_0 is equal to or less than about three, then there may be a significant resonance-related peak in pressure occurring as the vocal folds are closing.

Considering the effect on supraglottal pressure first, if the timing is such that the first formant causes a positive pressure peak during the closing of the glottis, reducing the transglottal pressure, then the glottal air flow will be reduced prematurely, before the instant of glottal closure. This can greatly reduce the high frequency energy generated at the closure. Likewise, a negative pressure peak occurring while the vocal folds are closing will increase the transglottal pressure and thus increase the high frequency energy.

Though less is known about subglottal resonances than is known about supraglottal resonances, it is clear that a resonance-induced increase or decrease in tracheal pressure occurring while the vocal folds are closing can also affect the high frequency energy generated at the closure. An increase of tracheal pressure would increase the energy generated and vice-versa. However, the higher damping of a subglottal resonance (see Fig. 4, for example) means that it is less significant than a supraglottal resonance of comparable frequency in effecting voice quality.

At this time it is not clear to what degree a trained singer makes use of formant or subglottal resonance to enrich voice quality. Sundberg has shown that there sometimes appears to be a shifting of the articulation of a sung vowel so as to "tune" a lower order formant to the vicinity of a multiple of the voice fundamental frequency in female singing. However, is the singer's goal to increase the energy at the formant being tuned,

as would be predicted by linear, non-interactive theory, or is the goal to enrich the voice spectrum at higher frequencies also, as would be predicted by the nonlinear interactive model? (It is interesting to note that the tuning of a formant for maximum high frequency energy is slightly different from the tuning for maximum formant energy, since the latter is always at an exact multiple of the fundamental, while the former may not be.) Also, is there a possibility of "tuning" the subglottal resonances? (This seems doubtful.)

Putting these questions another way, one can ask whether in some styles of singing the trained singer is not only trying to select vocal tract tunings that enrich voice quality but is also striving to avoid improper tunings which would dilute voice quality. Such tuning adjustments could be conceivably accomplished through small changes of vowel articulation which would vary with the note sung; however, there may be other, less obvious methods. For example, a slight nasalization can increase the damping of the vocal tract formants, and thus reduce the formant energy from a given glotttal pulse that is still present in the supraglottal system during the next closing of the vocal folds. Can nasalization therefore be used as a device to produce a more uniform voice quality as pitch is varied?

In conclusion, I would like to add some rather speculative remarks on the possible relationship between a nonlinear, interactive model of the voice source, the computer-based synthesis of the singing voice and the philosophy of singing pedagogy. It appears from the model presented above that a part of the task of the trained singer is to develop a style of singing in which the acoustic interaction of the voice source with the vocal tract tends to increase the acoustic energy at the second and higher formants produced at the instant of glottal closure. However, this must be done in such a way as to avoid large differences in voice quality as the voice pitch and the vowel are varied. In other words, the singer must learn to make use of the source/vocal tract coupling in improving voice quality, while avoiding or compensating for its undesireable effects. Though these may be awesome tasks for the human speech production mechanism, however, they are simple ones for the electronic voice. Electronically, one may set the richness of the "voice" spectrum at any level desired. Likewise, the voice source of the electronic voice can be easily

isolated from the vocal tract it drives. If we add to these "advantages" of the electronic voice the ability to select the proper voice fundamental frequency to any desired accuracy, including the characteristics of pitch change and vibrato, we might conclude that the computer can be a better singer in some sense than a human being can be, at least during vowels. (Many consonants cannot be accurately synthesized as yet.)

Said differently, it may be possible that some aspects of classical singing pedagogy have been inadvertently aimed at making the voice a good "computer", just as some aspects of classical painting pedagogy were aimed at making the painter a good "camera". But I wonder if, in the future, we will find that the computer will easily outdo the human singer in producing "technically perfect" singing (accurate in pitch control, and rich and uniform in voice quality). And as that develops, may we not see a shift to more expressive singing styles that leave to the computer that which it does best?

ACKNOWLEDGEMENTS

Most of the concepts presented in this paper were developed during my tenure during 1980 as a guest researcher in the Department of Speech Communication and Music Acoustics at the Royal Institute of Technology in Stockholm. While there, I had the benefit of many lengthy and fruitful discussions with staff members, especially Professor Gunnar Fant and Dr. Jan Gauffin. I am indebted to them for their help. Also, the remarks in this paper on the influence of the first formant on source-tract acoustic interaction stem primarily from my experience with an analog simulation, designed and constructed during my stay there, of an interactive voice source model similar to that in Fig. 5, but including elements simulating the action of the first formant. This simulation was connected to the laboratory's singing synthesizer MUSSE for perceptual experiments. I very much appreciate the assistance of Professor Johan Sundberg and Mr. Björn Larsson, its designer, in connecting the interactive voice source to MUSSE and in interpreting the results.

REFERENCES

BAER, T. (1981): "Observations of vocal fold vibration: measurement of excised larynges", Vocal Fold Physiology, (K.N. Stevens & M. Hirano, eds.) University of Tokyo Press, 181-189.

FANT, G. (1960): Acoustic Theory of Speech Production. s-Gravenhage: Mouton.

FLANAGAN, J.L. (1972): Analysis Synthesis and Perception of Speech. Springer-Verlag, Berlin, second edition.

FLANAGAN, J.L. & ISHIZAKA, K. (1978): "Computer model to characterize the air volume displaced by the vibrating vocal cords", J. Acoust. Soc. Amer. 63, 1559-1565.

HIRANO, M., KAKITA, Y., KAWASAKI, H., GOULD, W.J., and LAMBIASE, A. (1981): "Data on high speed motion picture studies", Vocal Fold Physiology, (K.N. Stevens and M. Hirano, eds.) University of Tokyo Press, 85-91.

HOLMES, J.N. (1963): "An investigation of the volume velocity waveform at the larynx during speech by means of an inverse filter", Proc. Speech Communication Seminar, Stockholm, 1962, Vol. 1, paper B-4 (Royal Institute of Technology, Stockholm).

KITZING, P. (1977): "Methode zur kombinierten photo- und elektroglottographischen Registrierung von Stimmlippenschwingungen", Folia phoniat. 29, 249-260.

KITZING, P. & LÖFQVIST, A. (1975): "Subglottal and oral air pressures during phonation - preliminary investigation using a miniature transducer system," Medical and Biological Engineering, Sept. 1975b, 644-648.

KOIKE, Y. (1981): "Sub- and supraglottal pressure variation during phonation", Vocal Fold Physiology, (K.N. Stevens & M. Hirano, eds.) University of Tokyo Press, 181-189.

LINDQVIST, J. (1965): "Studies of the voice source by means of inverse filtering technique", Congr. Rep. 5th Int. Congr. Acoust., Liege, Vol. 1, paper A35.

MILLER, R.L. (1959): "Nature of the vocal cord wave", J. Acoust. Soc. Amer. 31, 667-679.

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

ROTHENBERG, M. (1977): "Measurement of air flow in speech", J. Speech Hear. Res. 20, 155-176.

ROTHENBERG, M. (1981): "Acoustic interaction between the glottal source and the vocal tract", Vocal Fold Physiology, (K.N. Stevens & M. Hirano, eds.) University of Tokyo Press, 305-323.

ROTHENBERG, M. & ZAHORIAN, S. (1977): "A nonlinear inverse filtering technique for estimating the glottal area waveform", J. Acoust. Soc. Amer. 61, 1063-1071.

SUNDBERG, J. (1974): "Articulatory interpretation of the 'singing formant'", J. Acoust. Soc. Amer. 55, 838-844.

SINGING SYNTHESIS IN ELECTRONIC MUSIC

By GERALD BENNETT, Professor of composition, Musikhochschule Zurich, Schweiz

It is a special honor for me as a musician to have been invited to a scientific symposium on the voice as a musical instrument. My work on the singing voice has not shared the objective goals of that of my fellow speakers today. While they have primarily been concerned with the understanding of complex physiological, acoustical, and linguistic phenomena for knowledge's sake, I have tried to answer certain acoustical questions in order to make music using the computer.

I will present the results of a study of four soprano voices. The purpose of the study was to improve a program for sound synthesis by computer which had as model the human vocal tract. I will speak principally about the results of this research; at the end of the talk, however, I shall try to indicate the importance of this and similar research for computer sound synthesis in particul r and electronic music in general.

Before I begin, I want to acknowledge two debts. The first is to Gunnar Fant, whose theory of speech production is of central importance for this project. The second is to Johan Sundberg, without whose wise and precious assistance this project could never have begun. Neither of these debts can be repaid; I am proud to be able to speak of the fruits of their inspiration and assistance in the presence of both men.

In December, 1978, at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris, Xavier Rodet, Johan Sundberg and I began to synthesize singing voices using a computer, not with any of the available languages for computer sound synthesis, but rather by means of a very powerful and efficient program for speech synthesis developed by Rodet some years earlier at the laboratories of the Commission à l'Energie

Atomique (CEA) at Saclay, France. Although the study I will speak of today is my own, the vocal synthesis project at IRCAM has been carried out by Rodet and me jointly; without his sound synthesis program, none of this research would have been done.

Details of the techniques of synthesis Rodet uses can be found in Rodet & Bennett (1980). Briefly, his program models the resonant response of an imaginary vocal tract to a virtual source, not as is usually done, by filtering the source to obtain the resonant peaks, or formants, but rather by synthesizing directly the sound of each formant and then combining the formants to a single voice. This method allows the user to define and control the resulting sound with great precision. Other seminars in music acoustics at Stockholm have dwelt at length on techniques of sound production and reproduction by computer. I shall not recapitulate any of these discussions here, except to direct the curious reader to the standard text Mathews (1970).

To begin with, listen to two examples of singing voices synthesized on the computer using Rodet's method. The first (SOUND EXAMPLE 1) is a phrase from the madrigal "Moro lasso" by Carlo Gesualdo (ca. 1560-1613). The second (SOUND EXAMPLE 2) is a composite of computer voices I used in a piece for baritone, five instruments and tape: Aber die Namen der seltnen Orte und alles Schöne hatt' er behalten

Rodet and I were happy that we had been able to synthesize sounds bearing at least some resemblance to natural singing voices, but we found our efforts disappointing from many points of view. The most glaring fault seemed to be the lack of individuality of the voices, and after I had finished the piece for baritone, instruments and tape in November, 1979, I set out to remedy this failing.

By this time, after a year's work, we had learned the following things: We knew how to synthesize various sung vowels by defining formant frequencies, amplitudes and bandwidths; we could distinguish male and female voices in synthesis by differences in formant position for the same vowel, by differences in spectral richness, and by differences in the behavior of the formants with changes of fundamental frequency; we were able to link perceived amplitude and spectral richness in a convincing

way (so in SOUND EXAMPLE 2 every crescendo and decrescendo brings about a change of timbre in the voice); we understod something of the nature of both vibrato and the apparently random fluctuations of fundamental frequency we observed in every singing voice; we modified an algorithm of Fant's defining formant amplitude and bandwidth in terms of formant frequency so as to simplify greatly the use of the synthesis program. But we always synthesized the same soprano or the same baritone; we did not know how to define several voices of the same register, each having its own "personality". The literature offered only average values for acoustical parameters - those we had used at the beginning of our synthesis - but gave no clue as to how these values vary among similar voices. I was looking for specific values for the parameters which are given to the synthesis program: formant frequencies, vibrato rate, etc; since most of our work hitherto had been with male voices, I decided to find these values by analysing carefully four different soprano voices, three of professional women singers and one of a boy soprano, aged 11 years.

I made extensive recordings (two to three hours each) of the three female voices, less extensive ones for the boy's voice. I asked the singers to sing single notes on specific vowels, at varying subjective dynamic levels throughout their entire ranges. In addition I asked special studies of each of the women (for example many different attacks on one pitch, different types of vibrato on the same note, etc). The recordings provided the material for the present study.

My first and certainly most important discovery was that specific single values for the parameters I sought to define were inadequate to characterize a specific voice. No one vibrato rate would distinguish voice A from voice B; much more important turned out to be the way in which the vibrato rate of one voice varied under specific conditions, compared to that of another voice. I very soon realized that I was looking not for unique values but rather for rules of behavior of the most important acoustical features for each voice. A first analysis suggested that the following five aspects of each voice should be studied carefully:

1. timbre (spectrum)
2. overall sound pressure level
3. vibrato rate and amplitude
4. random variation of fundamental frequency
5. attack patterns

I shall speak about each of these five aspects separately.

1. Timbre

For each of the four voices I tried to model the acoustical spectrum as precisely as possible. In our synthesis program this meant simply defining center frequencies for each formant, for the program calculated the appropriate formant amplitudes and bandwidths itself. In fact, this automatic calculation of formant bandwidth and amplitude worked well for the women's voices; it did not work at all for the boy's voice, presumably because our algorithm was derived from measurements of adult voices. Hence, SOUND EXAMPLE 3 gives only examples of synthesis of the women's voices. In the sound example one can compare the timbres of the natural voices with those of the synthesized voices; in the example, a short section of a sung note is followed by the synthesized timbre for each of the three singers.

Fig. 1a and 1b show the spectrum of one of these sung notes and that of my imitation respectively. Frequency is on the horizontal axis and ampli-

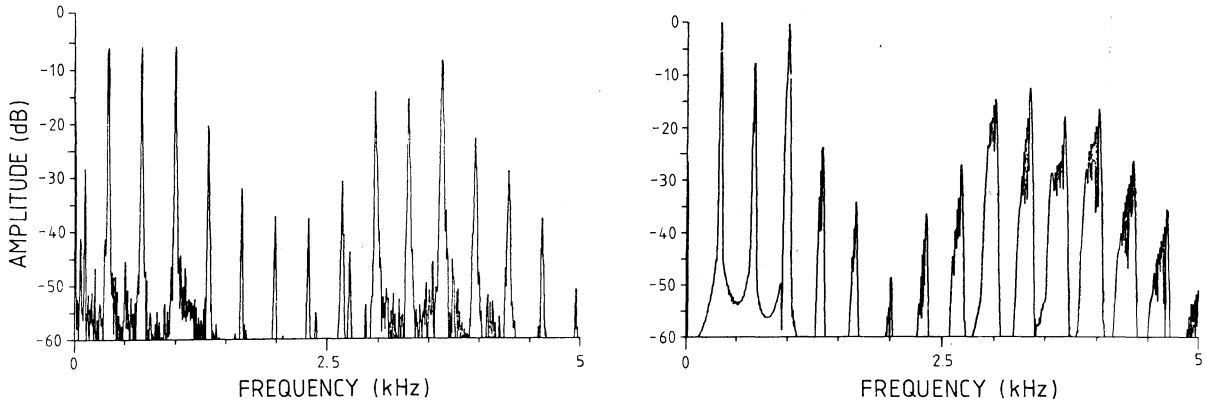


Fig. 1.

tude of the vertical, measured in dB less than the maximal amplitude found in the spectrum. The regularly spaced rays represent the sound's partials.

You will remark that in the lower part of the spectrum the imitation is quite good, each partial having nearly the amplitude of the corresponding partial in the original. In the higher end of the spectrum the imitation is less exact: this synthesized sound shows somewhat more energy in the upper formants. This imprecision may be due to the inexactness of the automatic calculation of amplitude and bandwidth of each formant, or it may be due to a difference in the source spectrum slopes of the real and the synthetic voices (Sundberg 1981). In the following discussion we shall see how to correct this difference.

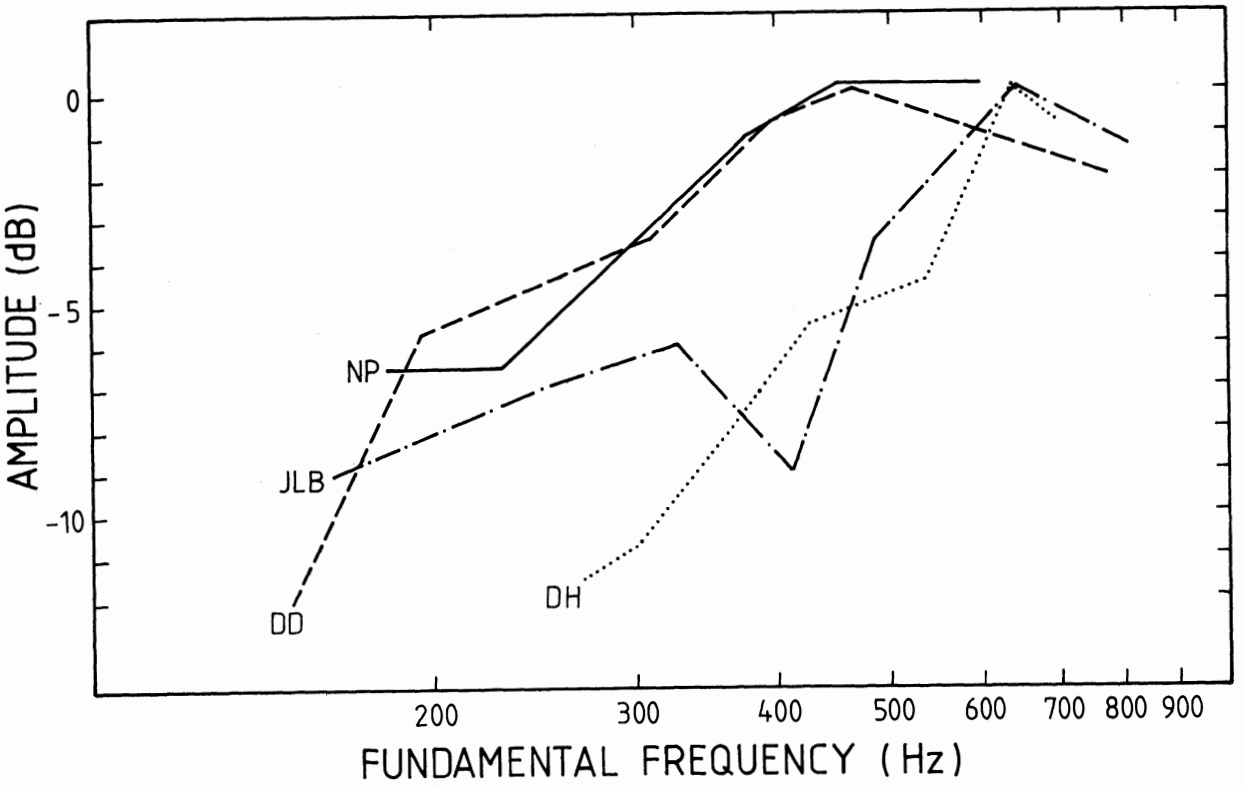


Fig. 2

2. Overall sound pressure level

I asked each singer to perform arpeggios throughout her or his range, keeping the sound pressure level (SPL) as nearly equal as possible. Each of the four singers sang more loudly in the upper part of her or his range than in the lower, but for each of the singers the SPL increased amplitude at a different rate and reached its maximum at a different point in the range. Not the specific values for any one pitch, but rather the behavior of the SPL over the entire range was characteristic and individual for each singer. Fig. 2 illustrates the results of the measurements of overall SPL as a function of pitch.

Here the change in amplitude clearly seems to be a function of fundamental frequency. The vertical scale shows how much softer than the loudest note each note of the arpeggio was. You see for instance that singer DD had the greatest dynamic change (but also the greatest pitch range), singer NP the least. You also see that JLB has a "break" in her voice around 415 Hz (G#4). Singer DH, the boy soprano, has a less dramatic break around 550 Hz (C#5). In my imitations, I took account of the slope of the increasing SPL, the frequency at which it was at a maximum and the slope of its decrease.*

When a singer makes a crescendo, he changes not only the loudnesse of the note sung, but also its timbre. In particular, the amplitude of the upper formants increases more rapidly than that of the fundamental as the note gets louder. This change in spectrum reflects change in the voice source due to the increased subglottic pressure. We had already modelled this timbral change in our original synthesis program, thanks to data given us by Johan Sundberg. We had originally assumed that formants two and higher

* These measurements of SPL are very approximate. For more precise measurements, microphone position and distance would need to be controlled more exactly. Since the goal of these studies was to trace patterns and not to define unique values, this degree of precision seemed unnecessary, especially in view of the much more important amplitude changes singers constantly make for expressive reasons.

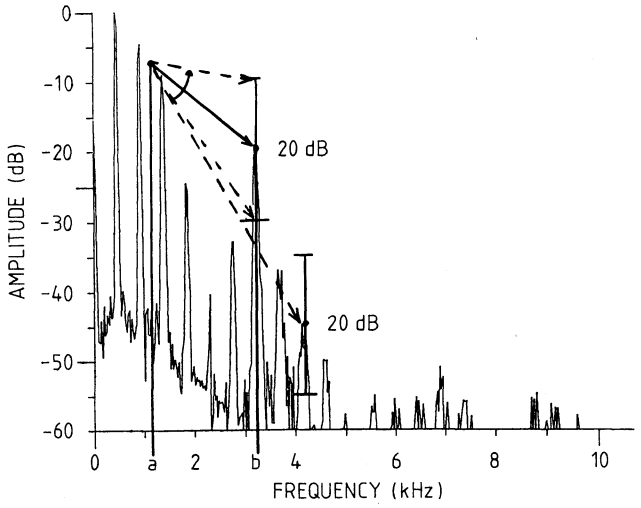


Fig. 3.

all changed amplitude by the same amount. Later we felt it more appropriate to increase the amplitude of the higher formants faster than that of the lower formants. Fig. 3 shows the principle of the present algorithm.

In analysing my recordings, I found the amplitude of the formant "hill" around the third and fourth formants of the loudest notes to be about 20 dB greater than in the softest notes (in the illustrations of spectra all amplitudes are scaled to the loudest partial, usually the fundamental, but sometimes - see Fig. 1a - the second partial). In our synthesis we now model this 20 dB change, linking it approximately monotonically to the overall SPL of the note. In addition, we can differentiate individual voices by defining a) the highest frequency whose amplitude does not increase faster than the overall SPL, and b) the frequency to which the observed 20 dB change applies. These points are marked a and b respectively in Fig. 3. Intermediate amplitudes are scaled proportionally. Our synthesis allows no direct control over the spectrum of the source assumed to be exciting the formants. Therefore we must mimic the effects of spectral change in the source in the resulting sound itself. The algorithm which does this is designed so that the amplitudes of all formants below frequency b vary with overall SPL less than 20 dB, while the amplitudes of the formants above frequency b vary more than 20 dB. The

lower the frequency of b, the greater the increase in amplitude of the higher formants. The result is roughly equivalent to varying the slope of the spectrum of the voice source with the overall SPL. In SOUND EXAMPLE 4 you hear scales sung by synthetic voices using the timbral models of the three woman singers and imitating the patterns of SPL evolution shown in Fig. 3 (without the break in JLB's voice).

3. Vibrato rate and amplitude

Vibrato is a more complex phenomenon than either spectral definition or SPL behavior for the characterization of an individual voice, because changes in vibrato are more intimately linked to expressive intentions. My remarks here apply only to a neutral vibrato of the kind most singers employ unconsciously.

We must consider two aspects of vibrato: vibrato rate and vibrato amplitude. Vibrato is a more or less regular change of pitch around a nominal center frequency; by vibrato rate I mean the frequency of this regular change, by amplitude its depth or excursion about this center pitch. Fig. 4a illustrates the vibrato of a sung Bb4.

You see that the note is attacked from below; the vibrato begins, increases to a maximum of about plus and minus 16 Hz, or more than a

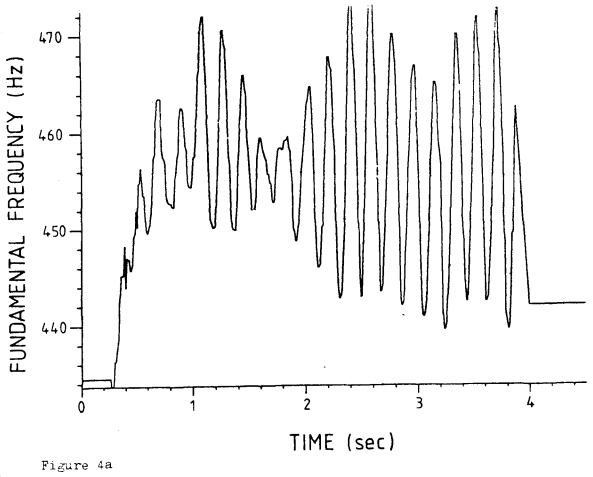


Figure 4a

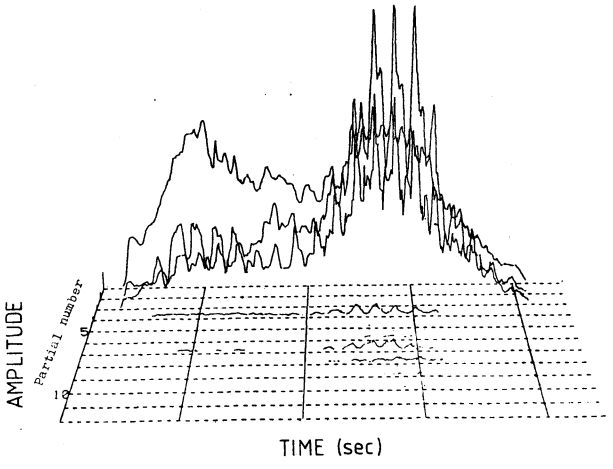


Figure 4b

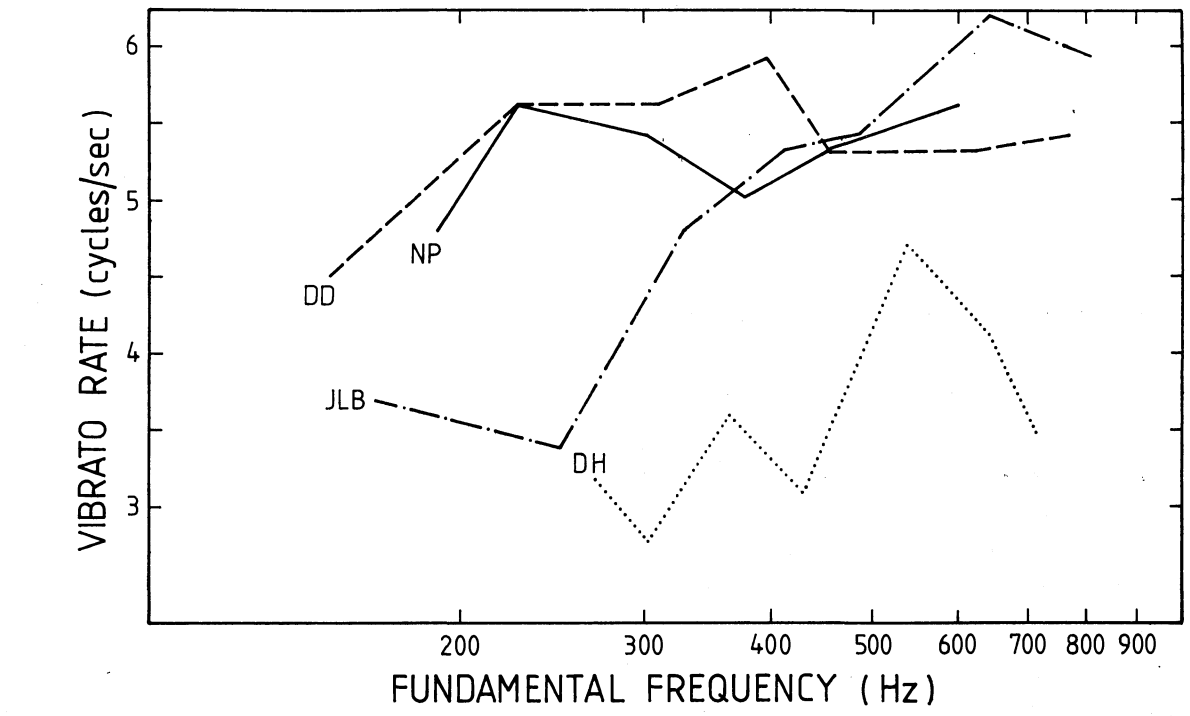
Fig. 4a and b.

quarter tone in each direction, then decreases and increases once more at the end of the note. The vibrato rate is rather slower at the beginning of the note, reaching an apparent target value (about 6 Hz) after approximately two seconds. Fig 4b shows the spectral evolution of the same note (low partials are at the back, higher ones in front, amplitude is in the vertical plane).

Here you see a strong initial attack in the first partial (maximum at about 0.9 seconds), then a decrease of amplitude followed by an increase with maximum at 2.5 seconds, and a final decrease of amplitude. You also see the entrance of the higher partials as the note gets louder, a very clear illustration of the way timbre changes with amplitude. (Note that the amplitude scale in this figure is linear and so gives a quite different picture of the relationship between lower and upper formant areas than do the other figures.)

By comparing Fig. 4a with Fig. 4b we can see to what extent vibrato amplitude is dependent on a note's fundamental amplitude. Only once do we see a discrepany between the two: at the end of the note the amplitude of the fundamental decreases rapidly while the vibrato amplitude increases. I have noted this increase in vibrato amplitude at the end of a note in most of the singers both male and female whom I have recorded. I imagine that the increased vibrato amplitude helps keep the soft sound interesting and relatively full in timbre.

In Fig. 4b we also see large fluctuations of amplitude in each of the partials having the same period as the vibrato variations in fundamental frequency. These fluctuations do not reflect changes in amplitude of the source signal synchronous with the changing fundamental frequency (there are none). Instead the changes are produced by the signal's passing through the narrow bandwidth resonances of the vocal tract: as a voice source partial changes frequency because of vibrato, it may move closer in frequency to the center of a resonance. As it does so, its amplitude increases, and as it moves away, its amplitude decreases again. Since the vocal tract's resonances are unlikely to be tuned to the partials of a note (although it seems that certain singers, particularly sopranos, make some effort to tune their formants to the nearest partials of long notes), different partials will reach their maxima at different times. In


VIBRATO RATE (cycles/sec)
6
5
4
3
NP
DD
JLB
DH
200
300
400
500
600
700
800
900
FUNDAMENTAL FREQUENCY (Hz)

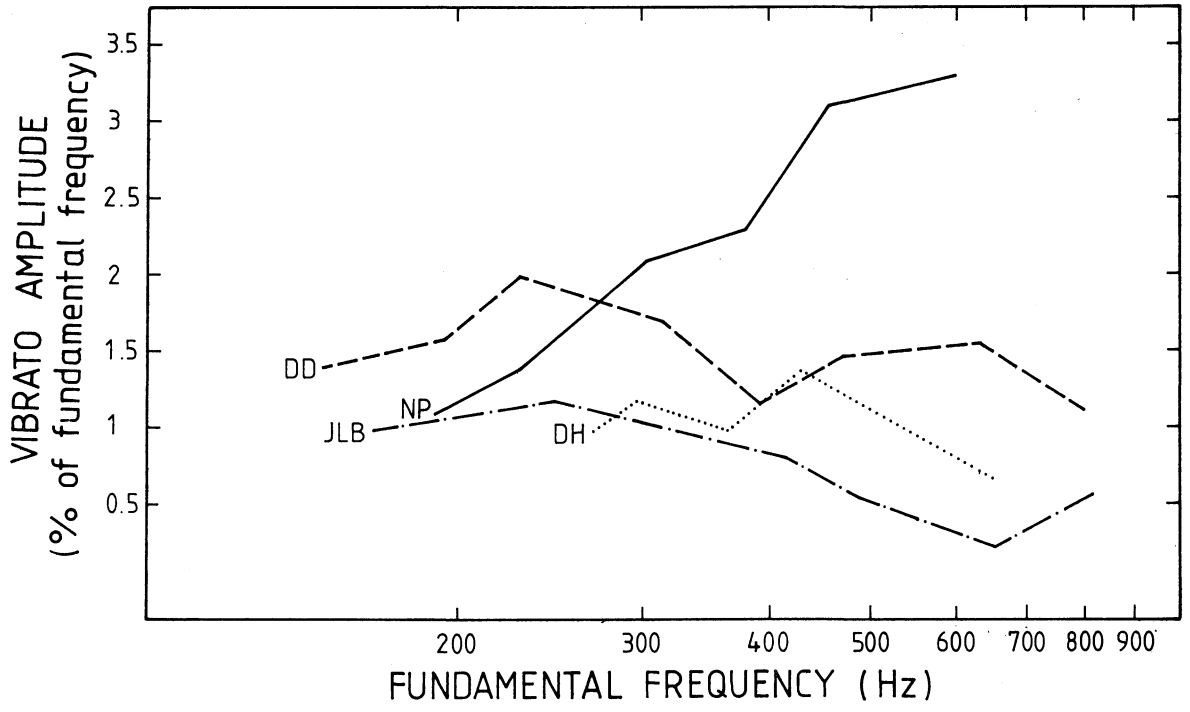

VIBRATO AMPLITUDE
(% of fundamental frequency)
3.5
3
2.5
2
1.5
1
0.5
DD
NP
JLB
DH
200
300
400
500
600
700
800
900
FUNDAMENTAL FREQUENCY (Hz)

Fig. 5a and b.

fact, this can be seen in Fig. 4b: compare partials 4 and 7 moving through two different resonances). The importance of the vibrato for a singer seems to me to lie in vibrato's ability to amplify the voice and to enrichen its timbre by ensuring that the partials near a formant actually bring that formant to greatest resonance.

In my arpeggio recordings I measured both vibrato rate and vibrato amplitude as a function of fundamental frequency. Fig. 5a and 5b summarize the results of these measurements.

The vibrato rate increases with higher pitch in all four voices up to a maximum, after which it decreases again. The pattern is much like that of amplitude evolution, although much more detailed studies than mine would be necessary to explore the relationship here. On the other hand, vibrato amplitude, which in Fig. 4a and 4b seemed clearly related to fundamental amplitude, acts in Fig. 5b much more like an independent factor, or rather like a factor more dependent on aesthetic choice than on the uncontrolled physical behavior of the voice. Three of the sopranos narrow the amplitude of their vibratos as they move up their ranges. One increases the amplitude of her vibrato with higher pitch, presumably because she prefers a deep vibrato on high notes and not because she cannot sing otherwise. The SOUND EXAMPLE 5 illustrates a synthesized voice singing the same phrase twice, with a different vibrato evolution each time.

4. Random variation of fundamental frequency

If we study Fig. 4a again carefully, we find that the fundamental frequency is by no means stable, but rather fluctuates in an irregular movement.

As far as we know, these fluctuations are truly random and reflect minute, uncontrolled variations in the tension of the muscles of the vocal folds. These small variations are essential to any convincing synthesis of the singing voice: without them even the most carefully placed formants and the most elegant vibrato sound machine-like; in their presence, the ear excuses small faults of timbre and vibrato. SOUND EXAMPLE 6 shows

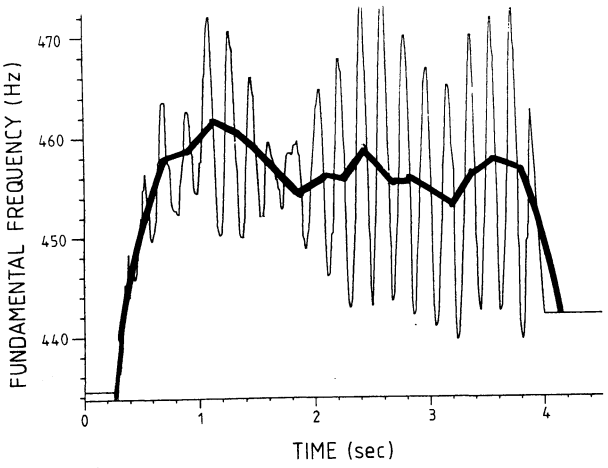


Fig. 6.

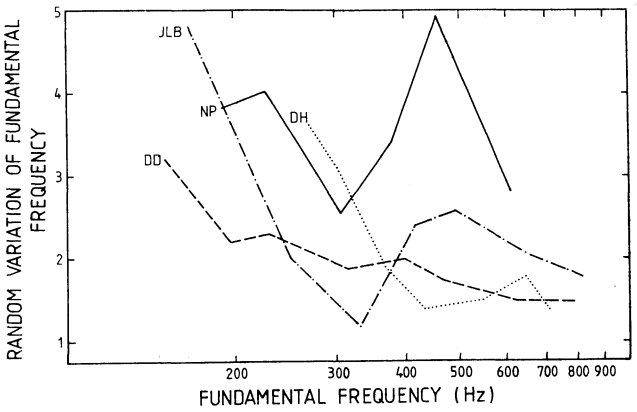


Fig. 7.

the importance of random variation of fundamental frequency: the two synthesized notes are identical except for the presence of random variation of fundamental frequency in the second note.

Fig. 7 shows the results of the measurement of random fundamental frequency variation as a function of pitch for our four sopranos.

In general the random variation decreases with ascending pitch. The largest variations are to be found at the lower end of the arpeggios, outside the actual soprano range, where control over pitch and timbre is naturally less precise because of the great relaxation of the vocal folds. As the voice moves up through the range, the tension of the vocal folds is increased, and the random variations of pitch become smaller.

Of the four aspects of the voice thus far discussed, random fundamental frequency variation seems to show the least difference in behavior between individual voices. Nevertheless, I think it important to model the general decrease in the excursion of the random variation with increasing pitch, and one can easily imagine designing a synthetic voice where the pattern of change of the random fundamental frequency variation would be an important part of the "personality" of the voice.

5. Attack patterns

In a general way, it seem that each singer has a repertoire of attack patterns, more or less extensive, depending on the attention he or she has given the matter. In my recordings, each of the singers had characteristic attacks. JLB, for example, usually overshoots the target pitch of a held note by between 8% at E4 and 4% at G#5 (the overshoot reaches 15% - 20% at the lower extreme of the range below C4). The length of the overshoot (that is the time for the beginning of the note until the target pitch is stable) is about 0.5 second at the lower end of the range and about 0.1 second at the upper end. Fig. 8a - 8f show the fundamental frequency patterns of the six notes in an E-major arpeggio from B3 to G#5. The vertical axis represent frequenccy in Hz, the horizontal time in seconds. Note particularly the overshoot at the beginning of each note and the irregular vibrato, compared with the vibrato in Fig. 4a.

Singer NB also attacks the notes in the lower part of her range from above (Fig. 9a-9d), but she takes much longer than soprano JLD to reach a stable pitch (from one to two seconds). She attack higher notes from below and reaches target pitch somewhat faster than for lower notes; her vibrato is much more regular than JLB's (cf Fig. 9e and 9f).

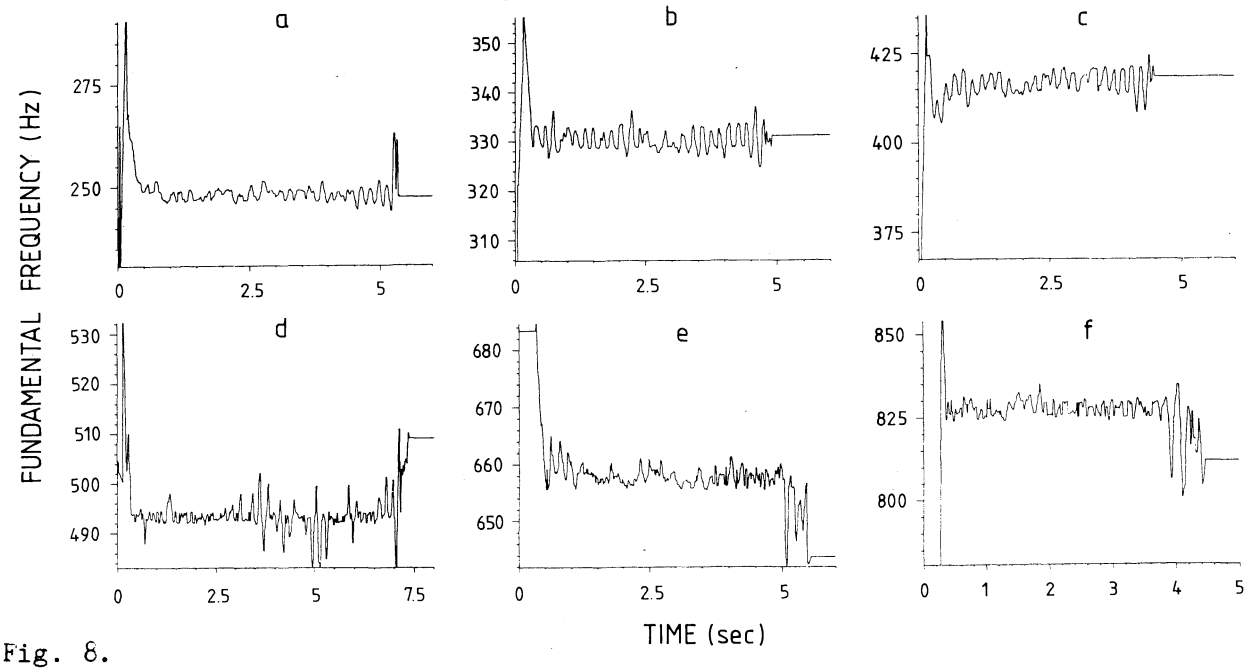


Fig. 8.

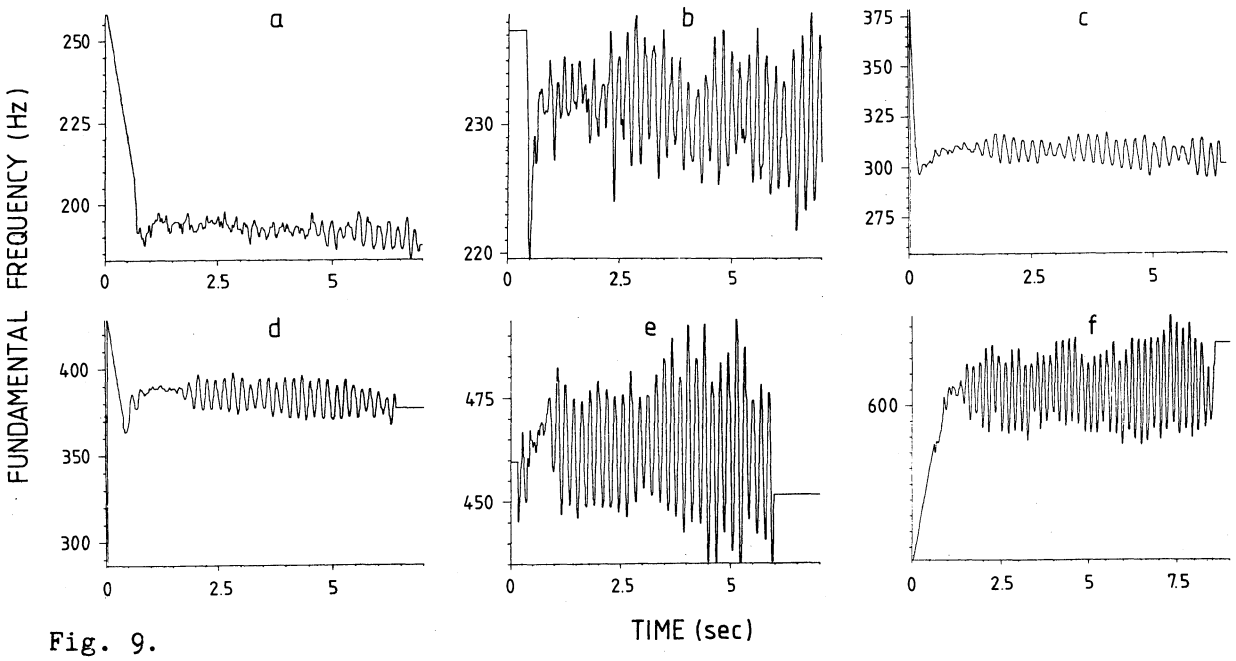


Fig. 9.

Clearly, a computer program offering dynamic control over pitch and amplitude can realize attack patterns like these without difficulty. But besides studying attack patterns, we must also consider the behavior of the global fundamental frequency of individual notes. Singer DD, for instance, always flattens fundamental frequency by a few Hz when she makes a crescendo.

6. Discussion

What is the interest of such analysis for electronic music? Of course, for me the most immediate interest has been in improving the synthesis of singing voices by computer. This has been the goal of all my work, and my research has been formulated to provide specific answers to specific problems of synthesis. But why did I choose to imitate singing voices at all?

My original interest in having the computer synthesize singing voices was compositional. Of course, straight imitation is not very exciting: there is no sense spending such time and effort on something singers will always do better than the computer. But I could imagine many compositional situations where I wanted an extension of the human singing voice, either in a direction no singer could go, or without conveying the sense of physical effort a human singer would expend to realize the effect. Here only synthetic voices would do; at the same time, the intrinsic quality, the complexity and the subtlety of the synthesized voices should be comparable to those of the natural voice; hence the interest in achieving a good imitation. But to imitate well necessitates understanding well what singers do when they sing; therefore studies like the present ones must be made.

This research has a more complicated relation to electronic music, less directly applicable than measurable results, to be understood rather in an illustrative sense. These studies give the composer a glimpse of how complex and how delicate the inner structure of sound is and show him interrelationships between physical dimensions of sound which may give guidance for constructing new and rich imaginary sonorous worlds.

Electronic music has always seemed to offer composers unlimited possibilities for making new sounds. The technical difficulty has always been to avoid using this richness in a merely anecdotal or decorative way, but rather to choose from the infinity of possibilities sounds varied and satisfying to the ear and yet unified at some deeper level. Many compser have experienced how synthetic sounds empirically chosen for a piece, each in itself interesting and complex, but without structural relation to the others, seem to loose much of their interest and complexity when combined. Here the ear seems very demanding, as if it were capable of performing analysis to determine the derivation and the inner constituency of sound. Johan Sundberg has spoken of how sensitive the ear is to the quality of "naturalness" in synthetic voices: the fact that most physical characteristics of the voice can vary only within strictly fixed boundaries; even the slightest transgression of the boundaries makes the voice sound "unnatural", while within these boundaries great variation is possible. Sundberg believes (if I understand him correctly) that this judgement is based on an analysis of how the sound is produced; if the

analysis shows that the sound could not have been produced by natural means, we immediately loose a great deal of our discriminatory power for the sound.

I am certain that the same phenomenon occurs not only in imitations of nature but in electronic music in general: the perception's attempts to define the origin of a sound give insight into the sound's structure. A succession of sounds unrelated at this level may in many contexts be aesthetically less satisfying than a succession whose sounds have some kind of inner structural relationship. These studies of singing voices show some of the types os complexity the ear is apparently good at dealing with; thus they cast light on the organization of sound at a level deeper than mere surface interest, offering composers a conceptual model of the subtle nuances of sound's inner structure.

However, the research has had consequences more direct than simply to illustrate the complexity of the organization of sound. The discovery that most of the acoustical features important for defining an idividual voice are not unique values but dynamic factors dependent on the momentary context meant that we had to change our synthesis program so that it expected, in place of a single value for, say, vibrato rate, a rule allowing it to derive vibrato rate from the pitch and the intensity at the moment. This change in the program had an important consequence we had not originally foreseen: the ability to vary synthesis parameters dynamically and as a result of decisions taken by the program meant that the computer was no longer merely an instrument for sound production, but could be incorporated into the compositional process itself. This is certainly not the first time composers have used the computer to help them write music, but whereas the computer generally does little more than produce a score to be played either by traditional instruments or by a separate synthesis program, in our program one can pass on as much control to the computer as one wishes, from the creation of a score to the adjustment of the smallest detail of the sound synthesis itself.

This research into singing has, we believe, led to a significant advance in both the theory and the practice of sound synthesis. Our model, in contrast to most sound synthesis programs, is a physical model, that of the vocal tract. Certain acoustical features of the voice have been found

to be more closely interrelated than others. These interrelationships have defined the stucture of the program. We believe strongly that these relationships will keep their validity even when the program is used to produce sounds bearing no apparent similarity to vocal sounds.

Let us conclude by listening to SOUND EXAMPLE 7, a short section of a piece for tape alone, Winter(1980). Here the synthesis program produced both singing voices and sounds which do not sound vocal. The musical organization of such heterogeneous material was satisfyingly strict; since both vocal and non-vocal sounds had the same general structure, I could subject both, despite their superficial differences, to the same operations. Thus it was the synthesis program which suggested many of the aspects on which to base the piece's musical development, namely those acoustical features lining apparently very different materials at a deep structural level. But it was also the synthesis program which organized much of this diverse material. Here musical imagination and acoustical analysis moved forward together, each inspiring the other, each enriched by the other's experience.

REFERENCES

FANT, G. (1970): The Acoustic Theory of Speech Production, Mouton, The Hague

FANT, G. (1973): Speech Sounds and Features, The Massachusetts Institute of Technology Press, Cambridge

MATHEWS, M. V. (1970): The Technology of Computer Music, The Massachusetts Institute of Technology Press, Cambridge

RODET, X. & BENNETT, G. (1980): "Synthèse de la voix chantée par ordinateur", Conférences des Journées d'Etudes 1980, Festival International du Son, 73-91

SUNDBERG, J. (1978): "Synthesis of singing", Swedish Journ. Musicology 60:1, 107-112

SUNDBERG, J. (1979): "Perception of singing", Speech Transmission Laboratory Quarterly Progress and Status Report 1/1979, 1-48

SUNDBERG. J. (1981) Personal communication

EMOTIONS, VOICE AND MUSIC

by professor IVAN FÓNAGY, Centre National de Recherche Scientifique, France

1. DYNAMIC DISTINCTIVE FEATURES

The distinctive prosodic features of emotive speech were studied in different languages by a variety of methods: (a) acoustic analysis of prosodic parameters such as fundamental frequency, intensity and speaking rate in stage speech; (b) the method of constant content; (c) the elimination of so-called endo-semantic (i.e. lexical, grammatical) information by means of filtering or laryngographic recordings; or simply by using pseudo-English or pseudo-Hungarian sentences, i.e. nonsense sentences retaining the transition probabilities in English or Hungarian phoneme sequences; (d) modification of the input signal; or (f) synthesis of prosodic versions of a sentence.

The results obtained for basic emotions were very similar for related and non-related languages (English, Huichol, Hungarian, Japanese, Swedish etc.), and are generally stated in terms of mean values of intensity, rate of speech, frequency level, frequency range. They can be, and, indeed, sometimes are presented in the form of a distinctive feature matrix, see Table I.

However, results of this kind seem to me both trivial and unsatisfactory in most cases. In examining matrices of distinctive features we are moderately surprised by the unbroken series of plus signs in all prosodic dimensions in the case of excitement. Also, we are embarrassed by the overlap between anger, joy and ecitement (provided that we accept the plus signs obtained for frequency level and frequency range by Lynch 1934, Fairbanks & Pronovost 1938, Hadding-Koch 1961, and Gårding & Abramson 1965).

Disagreements in the dimensions of frequency range and frequency level in the case of anger, and in the dimension of frequency range in the case of fear (high level according to Fairbanks & Pronovost 1938 and Huttar 1967, and low level according to Höffe 1960 and Ostwald 1963) may reflect interindividual differences or might be due to terminological divergencies. Binary feature analysis has been arbitrarily transferred from the domain of segmental phonetics to that of prosodic analysis. The distribution of formant frequencies may be revealing in the phonetic analysis of vowels; the distribution of fundamental frequency averages adds little, if anything, to the characterisation of the prosodic expression of anger, hatred, tenderness, and joy. In terms of such distributions the emotions anger and joy are almost overlapping, and hatred comes closer to tenderness than to anger (Fónagy 1978).

Table I. Matrix of prosodic features "characterizing" some basic emotions. The data were adapted from tests and measurements published by Lynch (1934), Skinner (1935), Cowan (1936), Fairbanks & Pronovost (1938), Fairbanks & Hoaglin (1941), Abe (1955), Chang (1958), Eldred & Price (1958), Grime (1959), Höffe (1960), Uldall (1960, 1964), Hadding-Koch (1961), Kaiser (1962), Fónagy & Magdics (1963, 1967), Ostwald (1963, 1973), Davitz (1964), Mahl-Schultze (1964), Gårding & Abramson (1965), Huttar (1967).

	Intensity	Frequency	Melodic interval	Speed of utterance
Excitement	+	+	+	+
Joy	+	+	+	+
Sadness	-	-	-	-
Anger	+	+-	+-	+
Fear	-	+-	?	+

Intonation is a time function. A static grid such as a matrix of distinctive features obviously cannot represent a process. A dynamic theory of intonation is implicitly contained in the Hungarian term "hanglejte's", 'intonation', literally 'vocal dance'. Similar concepts of fundamental frequency events underlie other metaphorical terms, such as musical scale or even rise and fall of the pitch (cf Aristoxenos, first century A.D, see Harmonische Fragmente, ed P. Marquard 1883, p 10 f.)

Over the last 20 years or so we have tried to describe the prosodic expression of some emotions and attitudes in terms of dynamic features. The material has been emotive speech, mostly produced in plays and experiments by actors and actresses along with recordings of spontaneous manifestation of anger, complaint, persuasion, and some other attitudes (Fónagy & Magdics 1963, 1967, Fónagy 1971a, 1972, 1978). I will just give a brief account of the principal dynamic features of some emotions and attitudes.

Anger, as manifested in a violent quarrel seems to be characterized in English, French, German, and Hungarian emotive speech by features such as:

(a) forceful expirations, very intense activity of the expiratory

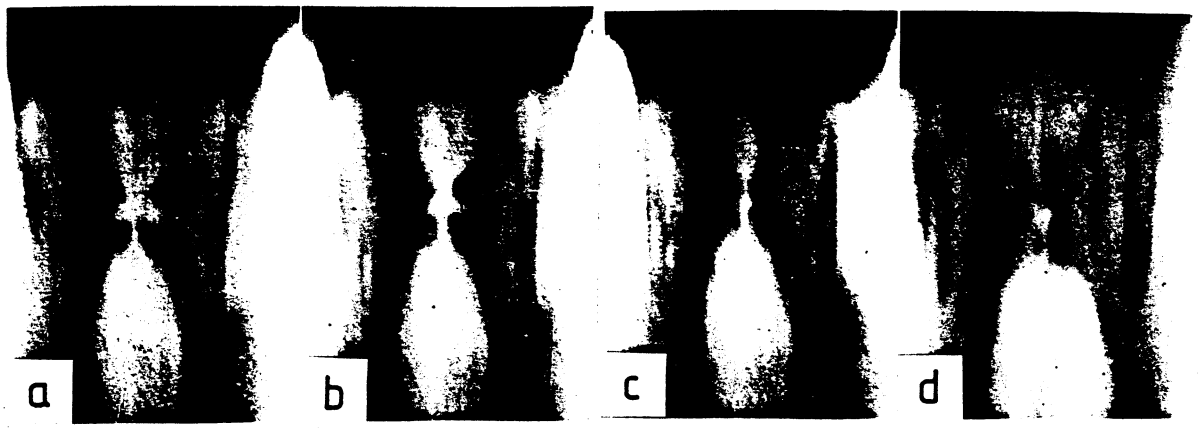


Fig. 1. Tomographic recordings of the laryngeal configuration in (a) tender voicing; (b) tender whispering, (c) angry whispering, (d) voicing during simulated anger. From Fónagy 1962.

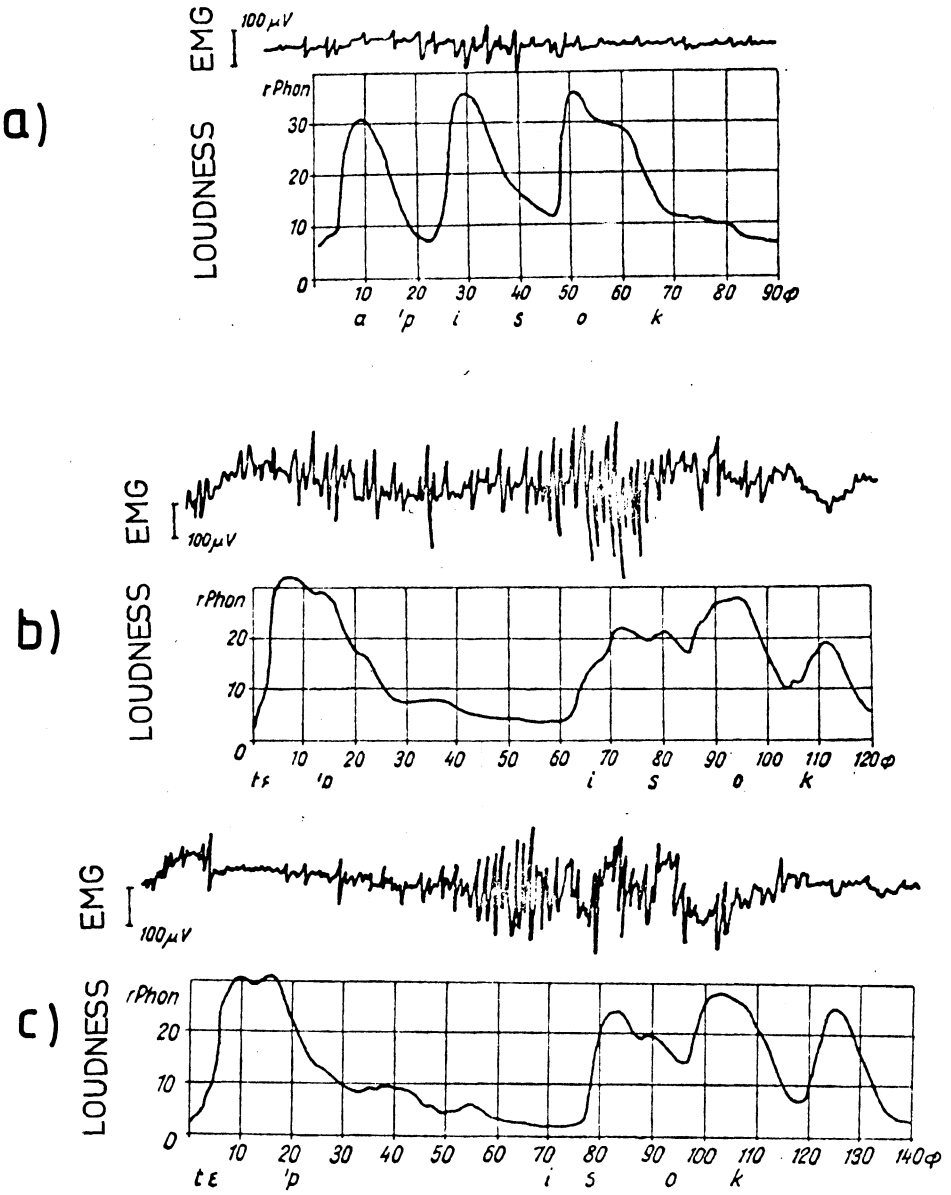


Fig. 2. Action potentials of the internal intercostal muscles (upper curves) and loudness level (bottom curves) in different versions of the Hungarian word piszok (=dirt) pronounced (a) as a neutral statement (A piszok, The dirt); (b) in simulated anger (Te piszok!, You dirty fellow!); and (c) in simulated hatred. From Fónagy 1958.

muscles, which incidentally is a dynamic feature noticed already by Hippocratus;

(b) imperfect phonation, such as breathy voice;

(c) a rigid metrical pattern with equally distributed heavy stresses (--+ --+ or +-- +--) which occasionally fall on linguistically unstressed syllables;

(d) a rigid melodic base-line on mid or mid-low pitch level with sudden rises corresponding to the interval of a fourth, fifth or sixth; the resulting peaks form a virtual even line, which, however, may also be slightly rising or falling;

(e) phonation in chest register with dark vowel colour;

(f) high speaking rate.

In order to avoid the pitfall of metaphor, we have to distinguish thoroughly between direct or concrete laryngeal gesturing and indirect or projective tonal gesturing by means of imaginary spatial pitch movements. Tomographic recordings as well as direct laryngeal X-ray recordings clearly show that the vocal strategy is entirely different for agressive vs. tender emotive attitudes. Fig. 1 shows the highly divergent laryngeal configurations for tender voice, tender whispering, angry whispering and phonation in simulation of hatred. Tender voice is characterized by a complete but smooth contact of the vocal folds, widely separated false vocal folds and wide laryngeal ventricles. In hatred, on the contrary, the laryngeal ventricles are compressed, probably because of a spasmodic contraction of the adductor muscles. As a consequence of this, the false vocal folds are approximated so that they touch each other, and, hence, the vibrations of the true vocal folds are highly disturbed. (For further details see Fónagy 1962.) The final consequence of all these muscle contractions is that, in spite of the considerable effort by the expiratory muscles and the resulting high subglottic pressure, the intensity of the sound produced might be lower than that produced with much less effort in tender speech, see Fig. 2a-c.

The ratio between the input physiological energy and the resulting acoustic energy, which is greater than or equal to unity in hatred, could be considered as a symbolic measure of the involvement of aggressive emotions.

Let us now turn to the prosodic projection of laryngeal mimicry. SOUND

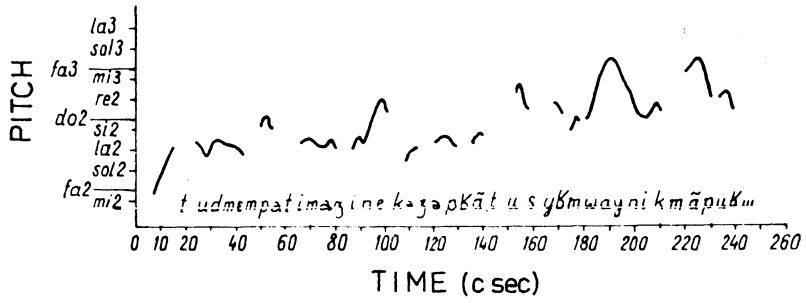


Fig. 3. Fundamental frequency curve of a fragment of an angry discussion: "Tu vas tout de meme pas t'imaginer que je prends tout sur moi, uniquement pour ....(te faire plaisir)" as pronounced by the French actress Marie-Claude Mestral.

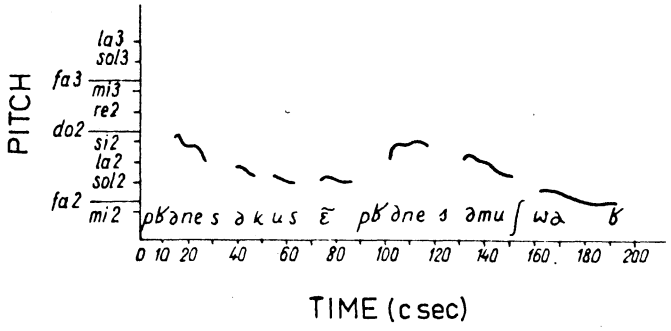


Fig. 4. Fundamental frequency curve fo a French sentence: "Prenez ce coussin, prenez ce mouchoir" as pronounced in a tender attitude by Marie-Claude Mestral.

EXAMPLE 1 offers some typical examples of anger and tenderness in English, French and Hungarian speech. Figs. 3 and 4 compares the rigid and peaky melodic contour of anger with the slowly undulating pitch curve in tender speech. The relevance of smooth vs. violent, sudden changes of direction was tested by means of synthesis of French and Hungarian utterances. The sentence was emotionally ambiguous, "C'est autant de gagné tu ne risque rien!". When synthesized with only one slow rise and fall of fundamental frequency, as shown in Fig. 5, the majority of the subjects, who had to chose one out of four labels (statement, persuasion, consolation, and reproach) interpreted the utterance as a consolation (see Fónagy & al 1979). These synthesized versions can be listened to in SOUND

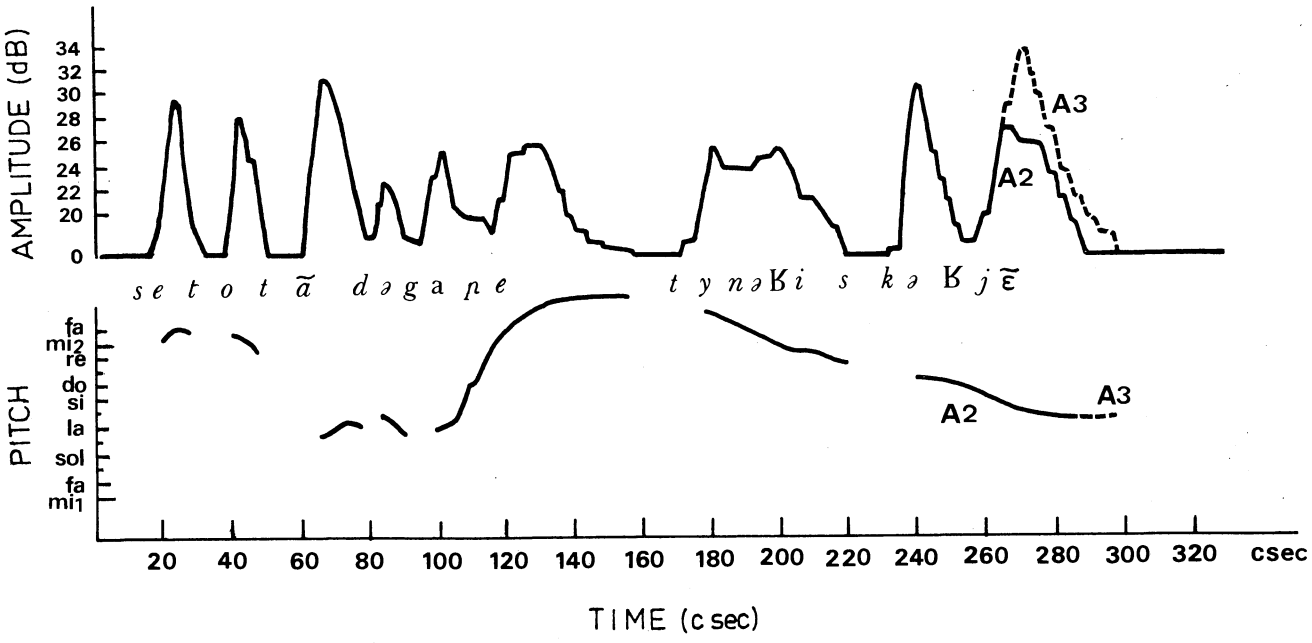


Fig. 5. Intensity level (upper graph) and fundamental frequency (lower graph) of two versions (A2 and A3) of the synthezized French sentence "C'est autant de gagnè, tu ne risque rien!". A majority of listeners found both these versions expressing comfort.

EXAMPLE 2.

When this same sentence was presented with four pitch rises, it was interpreted either as a persuasion for the contour marked C1 in Fig. 6, or as a reproach for the contour marked C2 in the same figure. When the listeners were asked to place each version on a 8-point semantic scale of "aggressiveness", this latter version produced a significantly higher score in the dimension of aggressiveness (5.8 vs. 4). The main difference between the pitch contours C1 and C2 in the lower part of the figure concerns the rate of pitch change, which is higher in C2 and the abruptness of the pitch changes, which is greater in C2. The prosodic opposition of a rigid and peaky melodic line vs. a smoothly undulating melodic line serves as a distinctive feature of quarrel vs. tenderness (aggression vs. affection) seems to hold even for the language of European music. Typical examples are given in SOUND EXAMPLE 3.

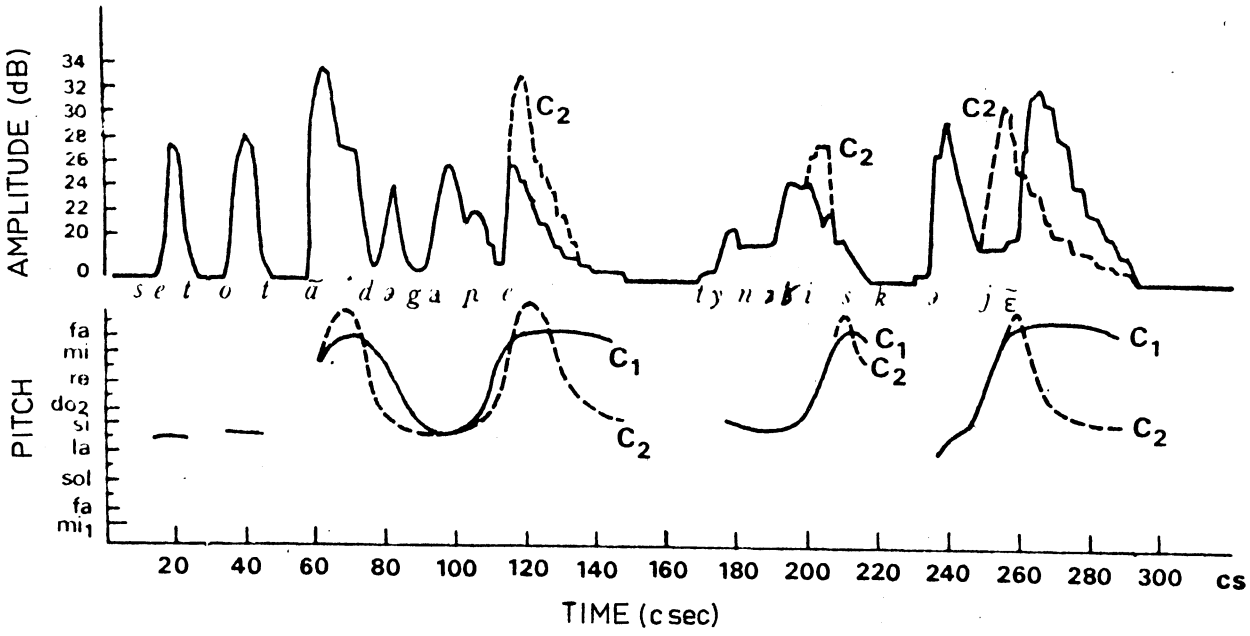


Fig. 6. Intensity level (upper graph) and fundamental frequency (lower graph) of two synthesized versions of the same sentence as in Fig. 5. The C2 version containing the more abrupt changes obtained a significantly higher score in the dimension of aggressiveness.

How interpret the apparently para-linguistic tendencies inherent in the direct and indirect (laryngeal and melodic) performance induced by anger or tenderness? In fact, a mere description of laryngeal and prosodic gesturing usually contains some elements of interpretation of the underlying psychological process, probably because of an isomorphism of psychic and phonetic events. I think of terms such as spasmodic contraction, strangled voice, or images such as smooth undulation vs. rigid melodic base line and recurrent peaks reflecting retention followed by repeated sudden outbursts, explosions, smooth vs. violent changes of direction.

In postulating a certain isomorphism between prosodic gesturing and emotive content, we imply that the different emotions may be interpreted in terms of mental gesturing, a kind of symbolic acts. Indeed, the term emotion which derives from e-movere (to move away from, viz. the neutral state) reflects such a dynamic concept of emotive attitudes. According to the classical theory of emotions, as formulated by Hippocratus, Democritus, Seneca (De ira, Opera v. 1, 1654) emotions are merely disturbances,

pathologic states of mind, which lead to abnormal and irrational behavior. Kant (Anthropologie 1798, sect. 71) considered emotions as manifestations of mental illness. The theories proposed by Young (1943, pp 28-48) or Dumas (1948, pp 278 sq) do not differ susbstantially from the views of stoic philosophers.

This theory has a very restricted explanatory power for the interpretation of the prosodic expression of emotions. An alternative theory has been proposed by Charles Darwin (1894). Darwin consideres the symptoms of emotive behavior as vestiges of a once purposive activity. Thus, e.g. the acceleration of heart beat and respiration in both anger and fear increases the blood-circulation and hence the potential of muscular activity. Such an increased muscular potential is required for fight as well as for flight. i.e. for the two ancestral situations to which Darwin retraced anger and fear. As formulated by G. W. Crile: "Anger is a phylogenetic product of fight, fear reproduces flight, and love recapitulates copulation" (1915, p. 76). In this framework vocal and prosodic features of anger, fear, and other emotions can be interpreted in terms of gestures, i.e. in terms of symptomatic or symbolic bodily movements.

Let us consider a few examples of this for the case of anger:
(a) hyperventilation c uld be considered an attempt to produce forceful exhalations which can eliminate an object causing irritation or tension. The word to eliminate as well as the Latin word perdere denote at the same time to drop and to kill.The metaphor inherent in the English and the Latin word probably reflects such an unconscious identification of the two actions of dropping and killing.
(b) increased loudness as well as phonation in chest register may serve the purpose of frightening the rival, the enemy or the victim;
(c) muscular tension on all levels may be interpreted as a preparation for a fight;
(d) the "strangled" voice corresponding to a laryngeal constriction as observed in hatred could be regarded as a symbolic attempt to strangle the enemy (cf Freud 1940-46, vol I:238; Ferenczi 1939, vol 3:448, Fónagy 1971 a, Fónagy 1981).
(e) the rigid melodic base-line could be a projection of a tense, rigid body posture preceding and preparing the attack;

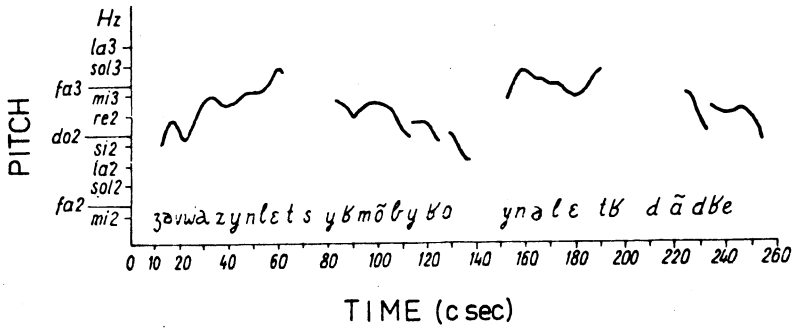


Fig. 7. Fundamental frequency curve of the French sentence "Je vois une lettre sur mon bureau, une lettre d'André" as pronounced in an outburst of joy by the French actress Marie-Claude Mestral.

(f) the sudden rises in pitch frequency correspond to successive outbursts of aggression and may substitute for sudden blows;
(g) the high degree of regularity of the prosodic structure may seem to match the high degree of organisation necessary for fighting successfully.

Also if we turn to articulation, the gestures characterizing emotive speech can be interpreted in an analogous way: the stiffening and the staccato movements of the tongue in speech simulating anger or hatred; the increased maxillary angles; the upper incisors "bite" on the lower lip in articulation of /w/.* The nearly metrical organization characterizing anger is generally absent in jubilation, i.e. in the expression of an emotion corresponding to a triumphant situation as is illustrated in Fig. 7. Incidentally, this situation does not require self command.

* Radiocinematographic analysis of Hungarian and French emotive speech as recorded from one female and one male untrained speaker subject as well as from three actresses and one actor clearly demonstrates that anger, hatred, tenderness, joy, and fear are expressed by analoguous articulatory gestures in unrelated languages (Fónagy 1976, Fónagy & al., fortcoming).

Semantic tests were carried out on 23 synthesized versions of the pseudo-Hungarian sentence "/kisera me:ra ba:vatag/". The results showed that those versions, which exhibited an equal distribution of violent stresses, a straightening of the base-line and a straightening of the level reached by the pitch maxima, are more likely to be interpreted as expression of anger than versions having irregular pitch rises reflecting violent but less well coordinated body movements (Fónagy & Fónagy, unpublished).

The slow speed and the slow gradual pitch changes in affectionate, tender speech seem to reflect smooth, caressing movements. They would express the desire not to hurt the partner by unexpected, inconsiderate movements.

This kind of encoding is archaic both at the level of content and at the level of expression. The contents communicated are pre-conceptual, and the way of communicating these contents is itself based on a pre-rational type of mental processing; it seems to rely heavily on magical presuppositions:
(a) the behavior of the part may be used to represent the behavior of the whole; the posture and the movement of the body may be projected on the larynx;
(b) the apparent movements of the pitch in the acoustic domain are representing the movements of the body;
(c) the speaker may identify with the listener; thus, the "strangled" voice may be equivalent with the throttling of a present or absent adversary;
(d) the verbal product can be equated with any object, inanimate or animate; thus, tearing the sentence into pieces by means of violent and irregular stresses may be a substitute of an action of violence directed against the listener or against a third person.

I believe that most of the paralingustic features of emotive prosody may be interpreted on the basis of such magical assumptions, in terms of direct or indirect (projective) vocal gesturing.

A prolonged state of fear or anguish is characterized by an extremely narrow pitch range in all languages, which I have investigated, i.e.

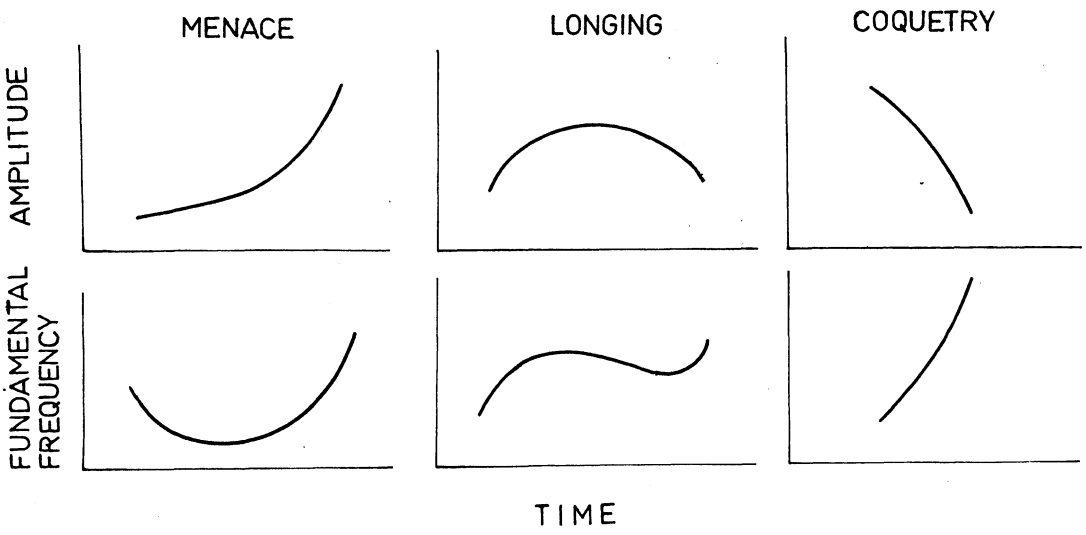


Fig. 8. Schematized intensity and frequency curves in synthetic versions of a Hungarian (nonsense) sentence which most successfully elicited the attitudes specified.

English, French, German, and Hungarian, as well as in vocal music. The pitch rises about a semitone in the stressed syllables, and then returns to a mid level, where it becomes so to speak paralyzed. Papageno's frightened "O wär' ich eine Maus" from the first act of Mozart's Magic Flute offers a striking example. The extreme narrowing of the pitch might reflect a maximal contraction of the body. If we hunch up and do not move, we are likely to stay out of sight/remain unrevealed. Low vocal intensity and a breathy voice could serve the same purpose.

The three configurations shown in Fig. 8 are stylized reproductions of synthesized versions of the pseudo-Hungarian sentence mentioned above /kisera me:ra ba:vatag/. Fortysix synthesized versions of this phrase were presented to two groups of Hungarian subjects in a forced-choice test. One of the groups consisted of 42 high-school students, and the other of 25 university students in a forced-choice test (see Fónagy & Fónagy, forthcoming for further details). From the results the following emotional profiles emerged:

(a) menace: a parallel increase in intensity and pitch frequency;

(b) longing: a gradual increase and decrease of intensity, a rise and a fall in pitch frequency followed by a slight final rise which is synchro-

nized with a drop in intensity;
(c) a quick but smooth pitch rise in the last syllable combined with a simultaneous rapid drop of intensity appeared to be best for evoking the impression of a luring, coquettish attitude.

Similar vocal strategies characterize menace, longing and coquetry in English, German, and French, as can be seen in Fig. 9 as well as in Western European music, cf Fig. 10-11. According to the principle of isomorphism we have to consider the simultaneous rise of sound pressure and pitch frequency as an expression of an increase of physical tension. Such a gradual increase in tension can be interpreted as a growing threat. As a projection of bodily display, the slow, gradual pitch rise, supported by a parallel rise in intensity seems to reflect an uprise and a slow, threatening approach. Similarly, we might consider the slow dynamic and tonal rise as observed in longing as a reflection of increasing tension. However, the rise in intensity is always preceded by an offbeat -- by two or more unstressed syllables on mid or mid-high level -- and followed by a slow descent.

The contradiction between the changes in intensity and in pitch could be

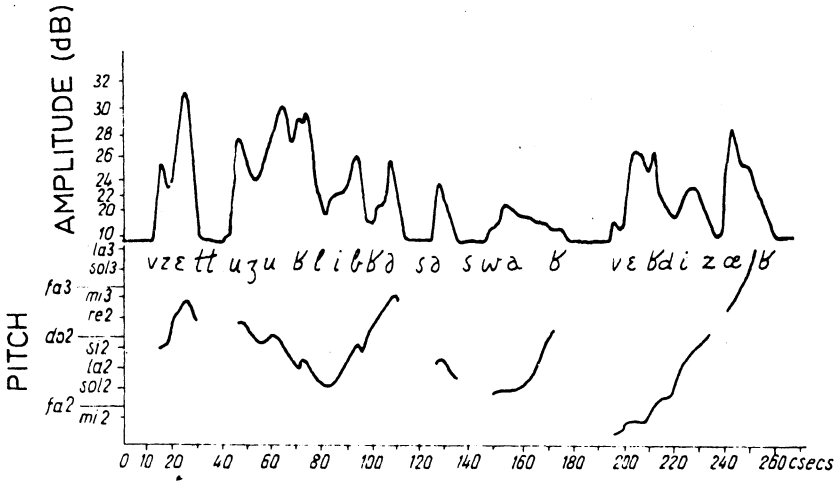


Fig. 9. Intensity (upper graph) and fundamental frequency (lower graph) of the French sentence "Vous etes toujours libre ce soir?" as pronounced in an attitude of coquetry by Marie-Claude Mestral.

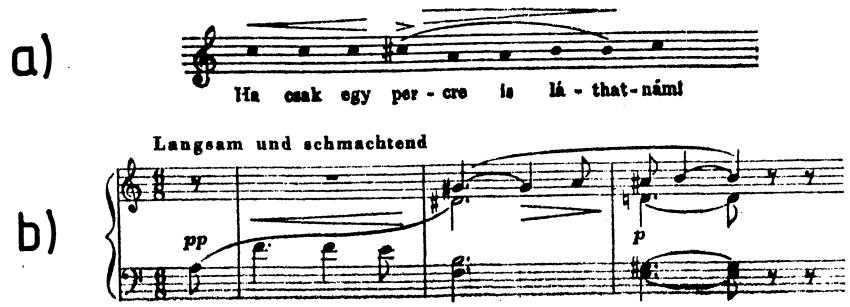


Fig. 10. (a) Musical transcription of the Hungarian sentence "Ha csak egy percre is lathatnam!, If only I could see him for a minute!" as pronounced by a Hungarian actress in a mood of longing. (b) The theme of longing from the prelude of Wagner's opera Tristan und Isolde.

interpreted as a reflection of an inner conflict in the following manner. The weakening of the expiratory effort resulting in the drop of intensity may be an expression of resignation. The relinquishment is contradicted by the increasing laryngeal tension reflected in the pitch rise. Suppressed longing (offbeat) forces its way (slow increase in intensity and pitch frequency). The distant aim cannot be reached (decrease in intensity), the body is extended hopelessly towards the remote object of desire (the rising melodic contour "crosses" the descending intensity curve).

Examples of coquettish luring can be listened to in SOUND EXAMPLE 4. SOUND EXAMPLE 5 gives a musical case collected from Bartok's Miraculous Mandarin. In coquettish luring the quick pitch rises reflect and induce excitement. In terms of melodic gestures they could be interpreted as rapid luring movements masked by a simultaneous drop in amplitude. The sudden switches into head register accentuate the femaleness of the speaker and evoke childish mockery, teasing, playful vexing, sexual provocation. The typical half whisper can be seen as a reflection of concealed passion, and it meets the secrecy of appeal. It may allude to sexual excitement, which is often reflected in imperfect phonation.

The complexity and dramatic character of emotive vocal performance is still more conspicuous in the case of irony. Of course, irony may leave no trace at all on the phonetic level, since the ironic intention is supposed to be hidden. The types of vocal performance to be considered

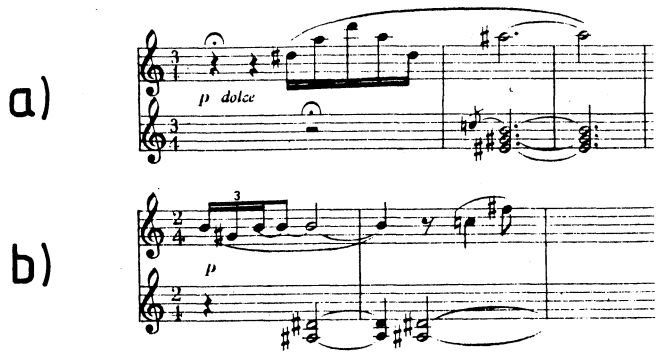


Fig. 11. Coquettish luring of the courtesan in Bartok's opera "Miraculous Mandarin" (cf SOUND EXAMPLE 5).

here corresponds to a somewhat simplified expression of irony, such as can sometimes be heard from the stage in theatres. In reality, the concept of irony covers a wide and highly diversified class of attitudes (see Muecke 1969). According to prosodic and cineradiographic analysis followed by semantic tests based on prosodically different versions of a Hungarian one word sentence (Fónagy 1971b, 1976), the vocal profile of irony can be described by means of three distinct phases, as illustrated in Fig. 12:

(a) an initial phase phonated in creaky voice with a strong constriction of the larynx; the pitch range is very low, and the pitch contour is

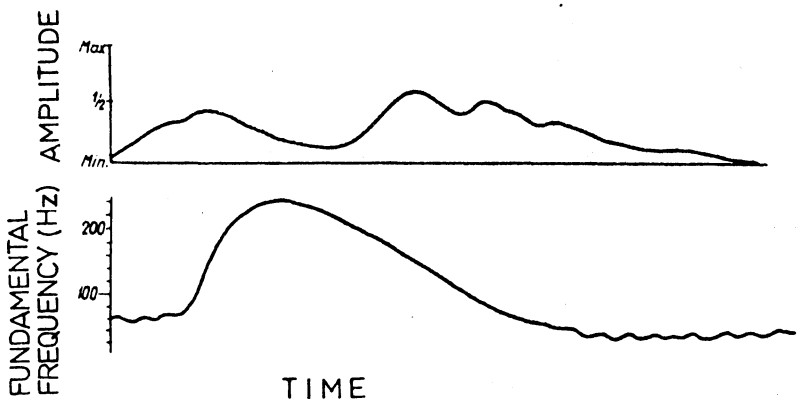


Fig. 12. Intensity and fundamental frequency recordings (upper and lower graphs) of the synthesized Hungarian sentence "Jo, Well". The version in which these curves were used obtained the highest score in the dimension of irony in semantic tests with Hungarian listeners. From Fónagy 1971.

fairly straight; the labial articulation is relaxed and the mouth corners slightly lowered;
(b) a second phase produced in head register at reduced overall intensity, including a rise in pitch towards a very high frequency (250 Hz was measured in a male voice!) and a tense palatalized articulation, i.e. the tongue is displaced forwards-upwards resulting in a clearer and sharper vowel quality, combined with a pharyngeal constriction;
(c) a third phase in which phonation returns to chest register occasionally interrupted by creaky voice phonation, the pitch is returned to the level of the first phase and is immobilized.

In all these three phases the velums are more or less lowered imposing a nasal timbre to the sentence. The tongue is moved backwards in the first phase, is raised and advanced in the second phase, and is retracted again in the third phase, as was clearly shown in the cineradiographic material.

This patterns can be interpreted as a drama in three acts. The changes in tension in the tragedy are physically present in the changes in muscular tension. But what meaning can we derive from this pantomime?

(a) In the protasis of the drama the pharyngeal constriction is a sign of rejection; it typically accompanies the expression of hatred. The resulting "strangled" voice could even contain an allusion to homocide. The low pitch frequency and the chest register may refer to virility and force. Moses (1954) calls this type of phonation "ultra-paternal". The "creak" is a sinister growling. In the slow rate of speech and in the immobility in terms of pitch there is something menacing, something that may suggest a wild animal crouching in order to pounce upon its victim. The delabialization and the lowered corners ot the mouth would express general scorn or disgust. The protasis of our drama is thus dominated by despite, hatred, and menace!

(b) The second phase, or the epitasis of the drama, is marked by sudden changes: shift into head register phonation, pitch rise, decrease of intensity, and palatalization seem to reflect an affectionate, feminine or infantile attitude. This gentleness is partially contradicted by clearly agressive elements, such as muscular tension and pharyngeal

constriction. These antagonistic attitudes create a certain dramatic tension, which prepares for the final solution.

(c) And the drama evolves as expected. The catastrophe is marked by a sudden return to aggression. The mask of sweetness is torn off. The interlocutor, seemingly carried towards the glorious heights, suddenly falls from the clouds, sees himself thrown into the abyss, where he must meet, face to face, the diabolic speaker who is waiting for him.

In this way the pattern of irony, which is an expressive modification of neutral intonation, can be seen as a dramatic performance in three acts, which accompany and differentiate the primary message of the sentence.

The sudden changes of semantic aspects could be roughly represented in the following way:

Character	Phase I	II	III
Masculine	+	-	+
Feminine	-	+	-
Aggressive	+	+	+
Tender	-	+	-

It should be noted here, that a description of these characteristics in terms of average intensity, pitch change, speaking rate etc. would completely blur the essence of the vocal performance with its typical abrupt changes at nearly all levels.

I have tried to emphasize the preconscious or unconscious symbolic dynamic aspects of emotive prosody simply because they are less apparent than the conscious, stylized reproduction of symptoms. This kind of reproduction of symptoms is also typical of emotive prosody and it can be efficiently illustrated by an example taken from the prosody of complaint. In the languages investigated, including the "language" of music, complaint is associated with a recurrent slight (semitone wide) rise and fall of

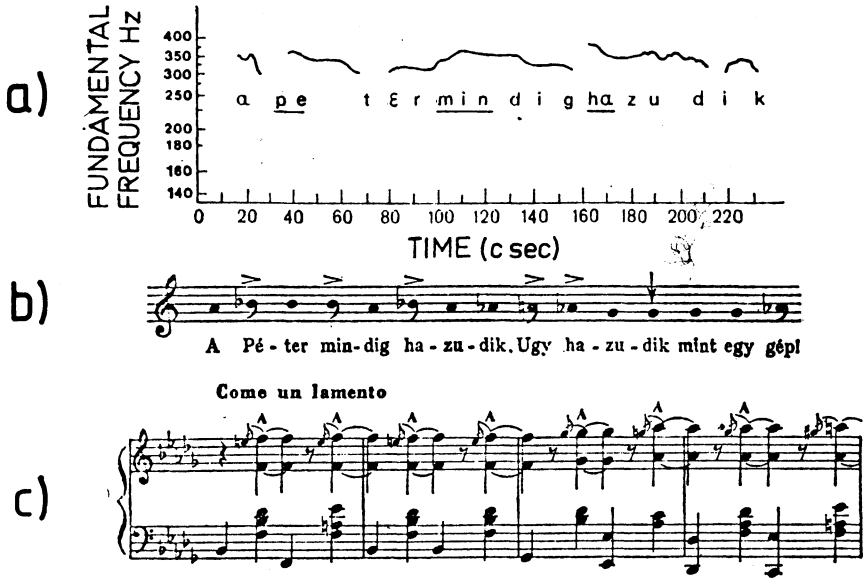


Fig. 13. The Hungarian sentence "A Péter mindig hazudik, úgy hazudik mint egy gép! (Peter is a notorious lyer, he lies like a machine!)" as pronounced by a little Hungarian girl in the mood of complaint. The curves represent (a) fundamental frequency, (b) a transcription of the fundamental frequency to musical notation, and (c) an excerpt of the LAMENTO choir from Monteverdi's opera Orfeo.

the pitch in stressed syllables, as is illustrated in Fig. 13. The observations made were confirmed by means of experiments using synthetic speech. It appeared that the number of responses "complaint" vs. "quarrel" depends strongly on intra- and intersyllabic changes of fundamental frequency. A sudden and deep fall is more likely to be associated with anger and quarrel than a slight up-glide and down-glide in stressed syllables, see Fig. 14. There can be little doubt that such up-glides and down-glides are stylized reminiscents of weeping, where slight pitch rises reflect the contraction of the expiratory muscles (cf Habermann 1955). This constitutes a typical example of symptomatic vocal behavior.

In both symbolic and symptomatic pitch patterns, the expressed emotion is in some way actually present in the prosodic expression. Emotive intonation is expressive only if it reduces a tension by literally expressing, i.e. throwing out the imaginary object which is creating the tension. I think that this is true for all motivated signs: they can be distinguished from purely arbitrary signs by such an intrusion, i.e. the "illegal" presence of the content at the level of expression.

The contention of some generative linguistics (Yorio 1973) interpreting emotive intonation as a surface manifestation of an underlying supersentence, e.g. "I hate that...", "I desire that...", seems to be misleading and anachronistic. The reason is that they confuse two levels of communication, namely communication by means of words, which is based on conceptual thinking, on the one hand, and prosodic communication of pre-conceptual emotive contents by means of motivated prosodic configurations, on the other hand. These are two levels of communication which are rather distant on an evolutional (ontogenetic and phylogenetic)

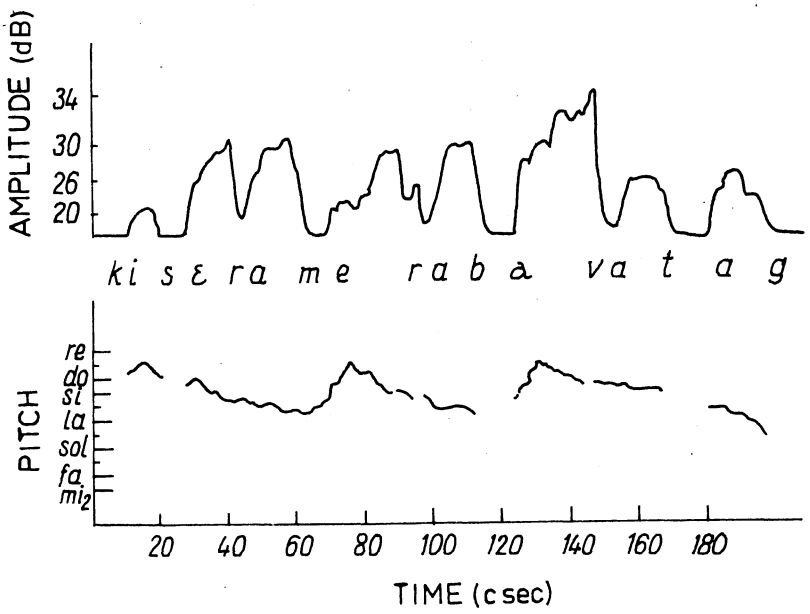


Fig. 14. Intensity and fundamental frequency (upper and lower graphs) of the synthesized pseudo-Hungarian nonsense sentence "Kisera me:ra ba:va-tag" pronounced as a complaint by the majority of Hungarian listeners.

scale. This difference is clearly felt, if we, applying the generative model proposed, try to substitute an outburst of anger with the verbal message "I am angry".

Motivation does not preclude the conventional character of prosodic patterns. In a recent investigation (Fónagy & Fónagy, forthcoming), 18 prosodically differing versions of the pseudo-Hungarian sentence /kisera me:ra ba:vatag/ were presented in a free choice test to 40 Hungarian, 16 American, 20 Czeck, 17 French, and 30 Japanese students. The subjetcs were asked to translate the sentence into their mother tongue, i.e. to provide a sentence which they felt to be identical in meaning with the sentence they had heard. Also, they were asked to determine the emotional attitude of the speaker. The responses were labelled as "correct" if 2 out of 3 judges found them conform with the speaker's intention. We encountered some divergencies in the results. For instance, American and Japanese listeners made certain mistakes in this attempt to interpret an utterance in an non-understood language. However, such errors were also committed by the Hungarian subjects, in spite of the fact that they understood the language. Sometimes the Hungarian subjects even failed more frequently than the others, see Table II. Our previous experiences from similar experiments tell us that such confusions are regularly made in semantic tests. But how to reconcile these errors with the pretended isomorphism of content and expression?

I would be inclined to attribute typical, recurrent identification errors to some analogy between those emotions which the listeners tended to confound. Thus, in semantic tests based on laryngographic recordings of repressed anger or hatred, anger was frequently interpreted as disdain or reproach, see Table III. Undoubtedly, all these affects share an element of aggression. Similarly, tenderness and desire are very close, both being positive libidinal social affects. Joy has been confused with coquetry and insistence: these emotions may occasionally overlap. An example of this is the situation suggested in the text spoken by an actress, who rejoices at the letter she just received from her lover.

Even the striking inaccuracies of subjects identifying fear as reproach or as a mood containing anger are explicable in the Darwinian framework: both anger and fear refer to struggle, the essential difference being

Table II. "Degree of motivation", i.e. the relative frequency of correct answers of foreign listeners (American, Czeck, French, Japanese) as compared to the number of correct answers given by Hungarian subjects in an experiment, where the subjects guessed the attitudes expressed by the different versions of a pseudo-Hungarian sentence /kisera me:ra ba:vatag/.

Attitude	American (N=16)	Czeck (N=20)	French (N=17)	Japanese (N=30)	Average
Anger	.60	.97	.70	1.25	.88
Hatred	1.30	.77	1.15	1.10	1.08
Irritation	1.07	.90	.78	.88	.90
Menace	1.15	.90	.78	.97	.91
Resignation (a)	1.07	.37	.47	1.05	.74
Resignation (b)	1.43	.25	.45	.90	.76
Fear	.65	.80	.93	.62	.75
Joy	.88	.57	.78	.92	.79
Surprise	1.08	.62	1.50	.52	.93
Imploring	.70	1.37	.90	1.09	1.00
Mockery	1.07	.67	.55	.72	.75
Persuasion	.98	.80	.58	1.17	.88
Explication (a)	.70	.38	.55	.78	.60
Explication (b) (insisting)	.93	.65	.58	.35	.63
Reasoning (a)	.65	.80	.93	.62	.75
Reasoning (b)	.50	.52	.30	- -	.44
Meditation	.87	.65	.87	.58	.70
Affection	1.15	.47	.67	.82	.78
Mean values	.93	.69	.73	.84	

that anger expresses the preparation to struggle of the dominant male, while fear expresses the preparation of the weaker one.

The well known analogies between the physiological responses in fear and anger (Cannon 1939, Lindsley 1970, also Arnold 1961) are in agreement with this supposed ancestral relations. The mixup between tenderness and sadness points to the absence of aggressiveness in both these moods (reflected in the reduction of muscular activity).

The semantic tests on emotive versions of the pseudo-Hungarian /kisera me:ra ba:vatag/ also offer some other revealing surprises, as can be seen in Table II (see also Fig.s 3-4). The fact that jubilation is confused with anger is possibly pointing to an underlying common feature, namely excitement or outburst of a violent emotion. The confusion of fear and surprise may suggest that surprise and fear share a common initial phase, namely the response to an unexpected event, which endangers the equilibrium of the organism.

The most unexpected confusion in these results is that of reasoning and argumentation, on the one hand, and anger and fear on the other. These confusions seem to corroborate some hypotheses regarding the instinctual basis of reasoning. For instance, Hermann (1923, 1978, 1979) connects logical and mathematical thinking with aggressive pulsions. Metaphoric expressions such as acute, sharp intellect, clear cut distinction, scientific rigor or the recent semantic development of argument might reflect our preconscious knowledge of such parallels. It might be mentioned here that the Latin word convinco means convince and is derived from vinco, which means 'to defeat'.

In fact, a confusion matrix of semantic tests on emotive speech contains precious indications concerning the psychologic structure of emotive mental states.Prosodic and musical expressions of emotive attitudes reveal much more of the real nature of emotions than conceptual analysis by means of lexical items. Vocal features may uncover some hidden traits giving useful hints for psychologic research.

2. A THIRD DIMENSION: MELODICITY

Looking for similarities between verbal and musical expression of emotions, the basic dissimilarities become very conspicuous. It is certainly more pleasant to listen to the musical outbursts of anger in the aria number 14 of the Queen of the Night in Mozart's "Zauberflöte" than to the expression of even tender feelings in English, French and Hungarian speech. This pleasurable effect is certainly related to the higher degree of organization, which seems to characterize musical expression at all levels. A higher degree of regularity of the intrasyllabic tonal movement distinguishes singing from speaking. The basic component of performed music, i.e., the tone, can indeed be regarded as a piece of music in itself to the extent that it contains a highly organized horizontal as well as vertical structure; there is a regularity in the time domain with respect to the sequence of subsequent cycles and there is a regularity in the frequency domain with respect to the harmonic structure, which can even be said to anticipate polyphony. The pleasant effect of a tone, its euphony, is a consequence of its two-dimensional regularity, which considerably reduces the mental expenditure involved in perceiving it. In a previous article (Fónagy 1960) I attempted to make a rough calculation of the number of measurements which a person, using tuning forks only, would have to perform in order to determine the changes in fundamental frequency constituting a sung /a/ and a noise of identical duration. It appeared that 79 measurements would suffice for the structure of the vowel, while 20.000 measurements were needed for determining the acoustic structure of the noise. I have a limited confidence in these calculations, but it seems to me quite evident, that we perceive, i.e. analyse the tone by means of a highly reduced expenditure of energy as compared with the costs needed for the perception/analysis of noise.

The dicotomy singing vs. speaking is partly determined by cultural habits. The Hopi indians of Arizona distinguish between tawi (singing) and lavayi (every-day speech). Their word ti:ngawa (declaration) is halfway between singing and speech (List 1963). Magdics (1963) investigated the acoustic characteristics of artificial transitions between speaking and singing in Hungarian. Such transitions really exist in all European languages.

Some twenty years ago we attempted to give a more or less adequate description of Hungarian intonation taking into account its variety. Several speech types were analysed, such as lecturing, sermons, street vendors' calls, recitation of poetry, theatre plays, conversation, children's speech. We soon realized that, in order to distinguish children's talk from the speech of adults, sermons from normal conversation, and tender speech from quarrel, we had to postulate a third dimension apart from pitch and time. We called this dimension melodicity and defined it as the perceptual response to the higher or lower degree of regularity /continuity/predictability of the fundamental frequency curve within each syllable (Fónagy & Magdics 1967). It is not easy to find an adequate measure of this particular type of tonal regularity. Standard deviation, auto-correlation, total variation, redundance are possible candidates (Fónagy & Magdics 1963).

The melodicity varies with age. Children's speech is more "melodic" than that of the adult. Adults speaking to a child try to mimic the child's higher degree of melodicity. As regards sex, women's speech is generally more "melodic" than men's speech. Even in everyday banal conversation we can observe occasional switching from a low to a high degree of melodocity. Thus, in Parisian French, playful melodic clichés are frequent in conversation (Fónagy 1981), especially in verbal cliche's such as "mais vous en faites un tete!" "En voila une idée!" "Oh qu'il est mignon!" etc. This seems to be the case also in American English (Ladd 1978), Australian English (Marlene Norst, unpublished), Hungarian (Fónagy & Magdics 1967) and probably other languages as well. Occasionally, melodicity may play a distinctive role in that it contributes significantly to e.g. the characterisation of disjunctive questions vs. disjunctive statements (Fónagy & Be'rard 1980).

The degree of melodicity is much higher in loving (libidinal) emotions, such as tender approach, courtship, desire, longing, coquetry, luring, consoling, than in agressive attitudes such as quarrel, hatred resentment, scolding, ordering, showing rigor. This suggests that a higher degree of regularity is typical of tender emotions on both the inter- and the intrasyllabic level. At the same time, the relative duration of quasi-periodic voiced elements, i.e., vowels and vocoids, is significantly longer in tender speech as compared with angry speech, all other

conditions being equal*. Aggressive emotions appear to be "noisy as compared with more tender emotive attitudes (see Tables III-IV). Similarly, an actor performing the role of a general tends to lengthen the consonants, whilst the same actor lengthens the vowels in the role of an amoroso, cf Table V.

Table III. Duration of vowels (V) and consonants (C) in centiseconds in the speech of three actresses (S1, S2, and S3) simulating two different emotive attitudes (hatred and tenderness).

Subject	Attitude	V	C	ptk	fs	bdg	vz	ljmn	w
S1	Tender	9.35	6.06	9.06	9.32	5.20	5.00	4.22	3.00
S2	Tender	8.95	6.92	9.60	9.74	6.25	5.07	6.18	4.38
S3	Tender	11.26	7.33	10.95	11.84	6.50	4.40	4.12	4.17
S1	Hatred	6.36	9.17	13.00	14.60	7.63	6.67	7.30	3.50
S2	Hatred	8.29	11.10	17.07	16.64	10.57	9.60	6.57	5.33
S3	Hatred	7.27	8.65	12.56	12.33	9.67	7.40	5.30	4.00
t-test:	t=	3.92	4.85	3.98	4.06	5.42	4.74	2.45	0.88
	p <	.02	.01	.02	.02	.01	.01	.01	.02
significance		S	S	S	S	S	S	S	S

In a tender mood speakers use vocal and chronemic strategies such as inter-syllabic regularity in the pitch behavior, slow and gradual pitch changes, and lengthening of the "euphonic" elements. These strategies facilitate the decoding and elicite pleasurable feelings by increasing

* The relative length of a speech sound varies with vocal intensity. Higher intensity is generally associated with a shortening of consonants, liquids and nasals excepted, and with a lengthening of vowels (Fónagy & al 1980). In order to determine the influence of emotions on consonant and vowel duration, we have to take such effects into account.

Table IV. Relative duration of vowels ($t_{V/C}$) in the pseudo-Hungarian utterance /kisera me:ra ba:vatag/ as spoken by four Hungarian actresses suggesting 12 different attitudes.

Attitudes		
Mockery (childish)	2.88	4
Complaint	2.06	9
Sadness	2.01	12
Joy	1.85	4
Explanation	1.80	13
Astonishment	1.78	6
Menace	1.72	4
Declaration	1.61	7
Fear	1.59	9
Irritation	1.58	4
Anger	1.50	9
Hatred	1.35	4

Table V. The ratio vowel/consonant duration in the speech of a Hungarian actor reading the same text first in the role of an amoroso and then in the role of a general.

Character	Vowels/p,t,k	Vowels/Consonants
Amoroso	1.08	1.20
General	.74	.87

t(V0)/t(C).100:	Amoroso	General
Observed	120	87
Expected for a random distr.	103.5	103.5

TEST: $X^2 = 5.260$ $p<.02$**

the phonetic redundancy considerably. At the level of vocal gesturing, a high degree of melodicity and the dominance of vocal elements reflect a reduction of laryngeal and articulatory muscular tension. Within the framework of the Darwinian theory, such a behavior corresponds to a peaceful attitude, i.e. to the absence of menace. The mental move away from aggression and towards tenderness seems to bring speech closer to music.

If we may assign an attitude to an artistic type of communication, this seems to place musical expression as the counterpart of aggressive behavior. Phonetic analysis thus seems to rejoin the myth of Orpheus and the Magic Flute by attributing to music the power to tame the wild forces of nature.

REFERENCES

ARNOLD. M. (ed.)(1961): Emotion and Personality II. Neurological and Physiological Aspects, Columbia University Press, New york

CANNON.H. (1979): Bodily Changes in Pain, Hunger, Fear, and Rage, Appleton, New York

CRILE, G.W. (1915): The Origin and Nature of Emotions, Philadelphia, London

DARWIN, C. (1872): The Expression of Emotions in Man and Animals, Murray, London

DUMAS, G. (1948): La vie affective, Presses Universitaires de France, Paris

FAIRBANKS G. & PRONOVOST, W. (1939): "An experimental study of pitch characteristics of the voice during the expression of emotions", Speech Monographs 6, 97-104

FÓNAGY, I. (1971a): "Bases pulsionelles de la phonation, II. Prosodie", Rev. Francaise de Psychanalyse 35, 543-591

FÓNAGY, I. (1971b): "Synthèse de l'ironie. Analyse par la synthès de l'intonation emotive", Phonetica 22, 42-51

FÓNAGY, I. (1976): "La mimique buccale. Aspects radiocinematographique de la vive voix", Phonetica 33, 31-44

FÓNAGY, I. (1978): "A new method of investigating the perception of prosodic features", Language and Speech 21, 34-49

FÓNAGY, I. (1979): "A la recherche de traits pertinents prosodiques du francais. Hypothèses et synthèses", Phonetica 36, 1-20

FÓNAGY, I. (forthcoming): Situation et signification. Prolégomènes à un dictionnaire des énoncés en situation Benjamins, Amsterdam

FÓNAGY, I. & Bérard, E. (1980): "Bleu ou vert? Analyse et synthèse des énoncés disjonctifs", in Melody of Language, (L. Waugh & C.H.van Schooneveld, eds) University Park Press, Baltimore

FÓNAGY, I., FÓNAGY, J. & SAP, J.(1979): "A la recherche des traits distinctifs prosodiques du francais parisien", Phonetica 36, 1-20

FÓNAGY, I., FÓNAGY, J & DUPUY, P. (1980): "Duration as a function of sound pressure level", J. of Phonetics 8, 375-378

FÓNAGY, I. & MAGDICS, K. (1963): "Das Paradoxon der Sprechmelodie", Utal-Altaische Jahrbuecher 35, 1-55

FÓNAGY, I. & MAGDICS, K. (1967): "A beszéd dallama (=Hungarian intonation)", Akadémiai kiado, Budapest

GÅRDING, A. & ABRAMSON, A.S. (1965): "A study of the perception of some American-English intonation contours", Studia Linguistica 19, 61-79

HADDING-KOCH, K. (1961): Acoustic-phonetic studies of the intonation of

Southern Swedish, Gleerups, Lund

HERMANN, I. (1924 & 1978): Psychanalyse et logique, Denoel, Paris

HÖFFE, W.F. (1960): "Ueber Beziehungen der Sprachmelodie und Lautstärke", Phonetica 5, 234-245

HUTTAR, G.L. (1967): "Some relations between emotions and the prosodic parameters of speech", SCRL Monographs 1, Santa Barbara

LADD, D.R. (1978): "Stylized Intonation", Language 54, 517-540

LINDSLEY, D.G. (1970): "The role of nonspecific reticulo-thalamocortical systems in emotion", in Physiologocal Correlates of Emotion, (P. Black, ed.) Academic Press, New York

LIST, G. (1963): "The boundaries of speech and song", Ethnomusicology 7, 1-16

LYNCH, G.E. (1934): "A phonophotographic study of trained and untrained voices reading factual and dramatic material", Archives of Speech 1, 9-25

MAGDICS, K. (1963): "From the melody of speech to the melody of music", Studia Musicologica 4, 325-346

OSTWALD, P.F. (1963): Soundmaking, Thomas, Springfield

SUNDBERG, J. (1963): Röstlära. Fakta om rösten i sång och tal, Proprius, Stockholm

YORIO, C.A. (1973): "The generative process of intonation", Linguistics 97, 111-125

YOUNG, P.T. (1943): Emotion in man and animal. Its Nature and Relation to Attitude and Motive, Wiley, New York

TO PERCEIVE ONE'S OWN VOICE AND ANOTHER PERSON'S VOICE

by professor JOHAN SUNDBERG, Department of Speech Communication Research and Music Acoustics, K T H (RIT), Stockholm

Listening to one's own voice from a tape recording one rarely recognizes the voice, even though one hopefully remembers what was said. On the other hand, one generally recognizes the voice timbre of other persons under the same conditions. This poses the question to be considered in the present article: Why does one's own voice sound so different to oneself as compared with how it sounds to other people?

This question, of course, is important to singers and singing teachers and also to the voice therapist. The ideal voice timbre sounds different to the ears of the speaker than to the ears of the listeners. Thus, the student training his/her voice should not attempt to develop a voice usage which produces the ideal voice timbre in the student's own ears! A singer who sings in such a way that he sounds as CARUSO in his own ears, probably sounds as something far from CARUSO in his listeners' ears. Let us examine a bit closer the factors which determine the voice timbre which the speaker/singer perceives of his/her own voice.

The sound of one's own voice reaches one's ear from two separate avenues. One is the air, which leads the sound from the lip opening to the ear canal. The other is made up by the structures separating the vocal tract from the inner ear. The sound within the vocal tract is of very high amplitude during vowel production, about 20 dB above the sound pressure level which causes pain in our ears! This sound generates vibrations in the vocal tract walls, and these vibrations are transmitted to other parts of the head, including the inner ear. The resulting bone-conducted sound can be perceived only by the speaker/singer, of course. The bone conducted sound does not radiate from the skull so efficiently that it

contributes to any significant extent to the sound reaching outside of the speaker/singer. Here is an important difference between the speaker /singer and the listener. It is only the speaker/singer who perceives any bone-conducted sound of his/her own voice. This is one part of the explanation why the voice timbre is not the same in theearsofthe person using his/her voice and the persons hearing his/her voice.

This is not the only factor contributing to this perception effect, though. There is another factor which affects the air-borne sound. This sound actually consists of two parts. One part travels directly from the lip opening to the ears. Another part reaches the ears after having been reflected one or more times in the room. This part of the air-borne sound reaches the ears with some delay which is considerable if the number of reflections is high. These two parts of the air-borne sound may differ considerably. In the direct sound the low frequency components are transmitted more efficiently than the high frequency components. The reason for this is that the high frequency components, among which we find the singer's formant, cannot radiate backwards as efficiently as the low frequency components. For this reason, the sound which has travelled directly from the mouth to the ears has a somewhat more dull timbre than the sound radiated along the length axis of the mouth. This effect is illustrated in SOUND EXAMPLE 1. It reproduces the same utterance picked up by two microphones simultaneously. One microphone is located just above the left outer ear of the speaker. The second microphone is at exactly the same distance from the lip opening but placed on the length axis of the mouth. The recording was made in an anechoic room, so that there is no reflected sound. There is a quite distinct difference not only in loudness but also in the voice timbre.

The air-borne sound that reaches one's ears does not consist solely of the sound that has travelled directly from the lips to the ears. Some sound leaving the lip opening in other directions might also eventually reach the ears after a smaller or greater number of reflections in the room. Of course, this sound is greatly influenced by the acoustic characteristics of the room. In one room only the low frequency components are reflected. In such a room the sound of one's own voice must be very dull as compared with a room which readily reflects even high frequency components. The bathroom is the classical example of a room of the last-

mentioned type. If a male person sings in the bathroom, or in any room with efficient sound reflections at high frequencies, the singer will hear himself produce a mighty singer's formant, which would perhaps remind him of the sound of the very best male opera singers he has heard. This would explain why men like singing in bathrooms.

Even though formal experiments have not been made (to my knowledge) it seems safe enough to conclude that the acoustic characteristics of the room adds to and influences the timbre which one perceives of one's own voice. If a singer habitually relies heavily on his auditory feedback for controlling his phonation, he would probably be very disturbed, as soon as he sings in a room with short reverberation time. He would try to phonate in such a way that the sound of his own voice in his own ears contains the amount of high overtones which he likes it to contain. But part of these overtones are normally provided by the reverberation of the sound in the room. This singer is likely to resort to "press" phonation under these conditions. However it seems that singers tend to rely less and less on the auditory feedback signal for controlling phonation. Very experienced singers are generally able to sing without any trouble for a long time in an anechoic room, but less experienced singers are often more embarrassed. The very experienced singer probably develops a feedback system which is less influenced by occasional factors such as the room acoustics. Here, vibrations in bone structures caused by bone conducted sound is probably the most important factor.

The auditory image which bone conduction generates of one's own voice is determined by the characteristics of the transformation of acoustic energy in the vocal tract to vibration energy in the bone structures of the skull. The frequency characteristics of bone conduction has been analysed by Tonndorf (1972). The results show that bone conduction is most efficient in the neighborhood of 600 Hz and that the efficiency decreases somewhat more steeply towards lower than towards higher frequencies (9 dB/octave and 6 dB/octave, approximately) as shown in Fig. 1. Thus, the voice components which are close to 600 Hz and have been converted into skull vibrations are more efficiently transferred in the skull structures than other components.

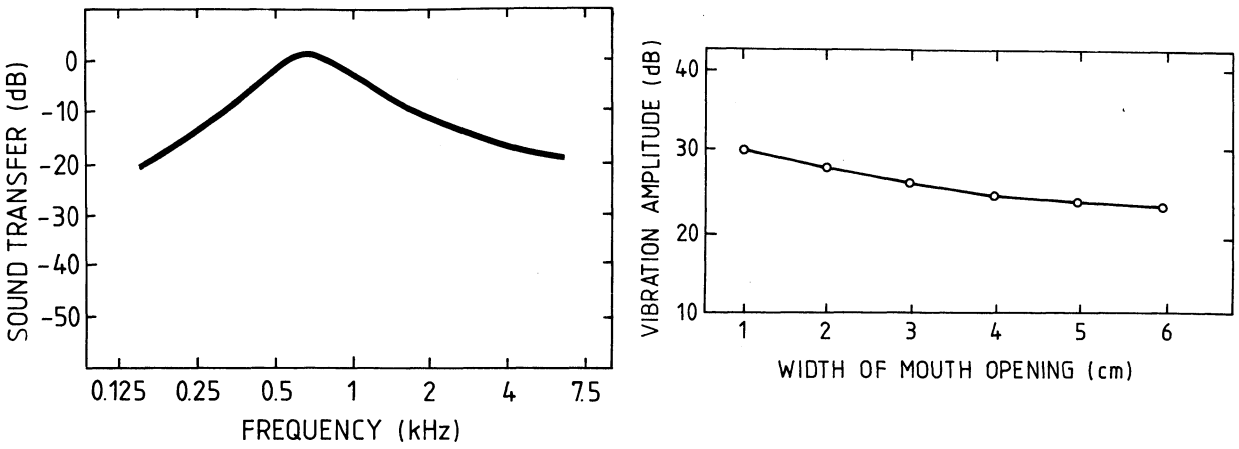


Fig. 1 (left graph). Frequency characteristics of bone conducted sound. After Tonndorf (1972).
Fig. 2 (right graph). Relationship between the skull vibration amplitude and the width of the jaw opening. After Kirikae & al (1964).

An equally relevant question is how sound energy in the vocal tract is transformed into vibration energy in the vocal tract walls. Kirikae & al (1964) measured systematically the vibrations caused by phonation on different parts of the skull. They observed that the wider the jaw opening, the sm ller the amplitude of the vibrations measured on top of the skull, as is illustrated in Fig. 2. As shown by e.g. Lindblom & Sundberg (1972) the first formant frequency is particularly sensitive to jaw opening, so we may assume that it is the frequency of the first formant which is the most important factor here. If this is true, we must postulate that the conversion of sound energy to vibration energy in the vocal tract walls is more successful at low frequencies than at higher frequencies. Thus, even though bone conduction is most efficient near 600 Hz, the bone conducted sound of one's own voice does not seem to contain very much sound in this frequency range, because the the sound in the vocal tract is not very efficiently converted into vibrations at these frequencies. (Here we have assumed that the vibrations on top of the skull are representative of the vibrations in the inner ears.) This agrees with the reports from persons who have suddenly lost their ability to hear air-borne sound, e.g. rupture or puncture of the eardrums. They find the sound of their own voice very dull as if there were no high frequency components present.

Returning to Fig. 2, we may also observe that different vowels generate vibrations differing considerably in amplitude, and that nasalized sounds generate vibrations with high amplitudes. Fig. 3 shows vibration amplitudes in the forehead during the pronunciation of various vowels. As observed in the case of the skull, the vibration amplitude is highly vowel dependent. Vibrations can easily be picked up by means of accelerometer microphones, and listened to. SOUND EXAMPLE 2 is a recording of my own voice as picked up by means of such an accelerometer which was applied on my forehead.

To summarize, how we perceive our own voice depends on four different factors: (1) the frequency dependent ability of sound to travel backwards from the lip opening to the ears; (2) the frequency dependent ability of the walls, floor and ceiling of the room to reflect sound, i. e. the room acoustics; (3) the frequency dependent ability of the sound in the vocal tract to transform into vibrations in the vocal tract wall structures; and (4) the frequency dependent ability of the bone structure of the skull to transmit vibrations from the vocal tract walls to the inner ear. From these points it is quite evident that the timbre of one's own voice is dependent of the room acoustics, among other things, and this makes the auditory feedback a whimsey judge of the quality of one's own phonation.

Considering the above, it is natural that singers need to develop a more reliable feedback system for phonatory control. From Kirikae & al. (1964) it is clear that vibrations in the skull are the results not only of phonation but also of articulation, as the amplitudes of the skull vibrations are vowel dependent, cf Fig.s 2 and 3. Thus, they do not relate in a simple way to phonation, which is generally independent of articulation. Kirikae & al. studied the vibrations caused by phonation at no less than 40 different locations on the body surface, and some results are shown in the same Fig. 3. It can be seen that, as soon as we go below the level of the vocal tract, the influence of the vowel at various places on the vibration amplitudes become smaller. On the thyroid cartilage and on the sternum the vowel dependence is small. This is not surprising as the vowel differentiation is achieved in the vocal tract.

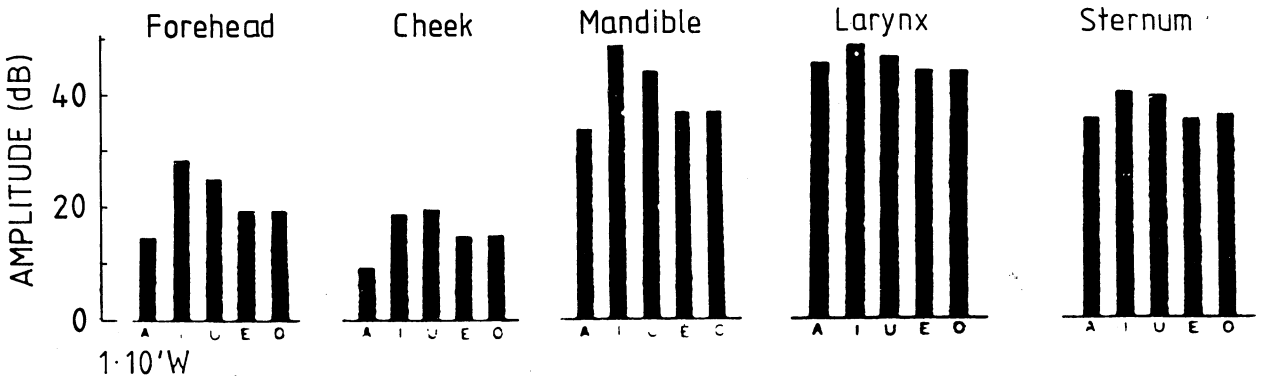


Fig. 3. Vibration amplitudes at various places on the body during phonation of different vowels. From Kirikae & al (1964)

When I worked for IRCAM (Institut de Recherche et Coordination Acoustique/Musique) in Paris I studied the chest wall vibrations in singers in detail. An accelerometer, which can be described as a (stone deaf) microphone sensing vibrations only, was fastened to the chest wall of each of the seven singers. The vibration amplitude and the amplitude of the vowels produced were measured at different phonation frequencies. Fig. 4 shows the results. The vibration amplitude varies with the phonation frequency. The main trend is that the displacement amplitude decreases by about 9 dB/octave rise of the phonation frequency. This holds across singers as a first approximation. It implies that at high phonation frequencies, such as those produced by female singers the vibration displacement amplitude is very low. Of course, there is rather noisy traffic of blood in this part of the body. The rsulting noise has an amplitude which appoaches the amplitude caused by phonation at phonation frequencies in the neighborhood of the pitch of C5.

The relevant question now is what vibration amplitudes can be perceived, or in other words where the threshold of vibration sensation is in relation to the vibration amplitudes caused by phonation. The heavy line in Fig. 4 represents an estimate of this threshold. It is based on threshold measurements obtained by vibrating the chest wall from the outside, by means of a vibrator. For the lowest frequencies the smallest perceptible vibration amplitude occurs at approximately the same rate as the phonatory chest wall vibrations. The maximum sensitivity is observed at the

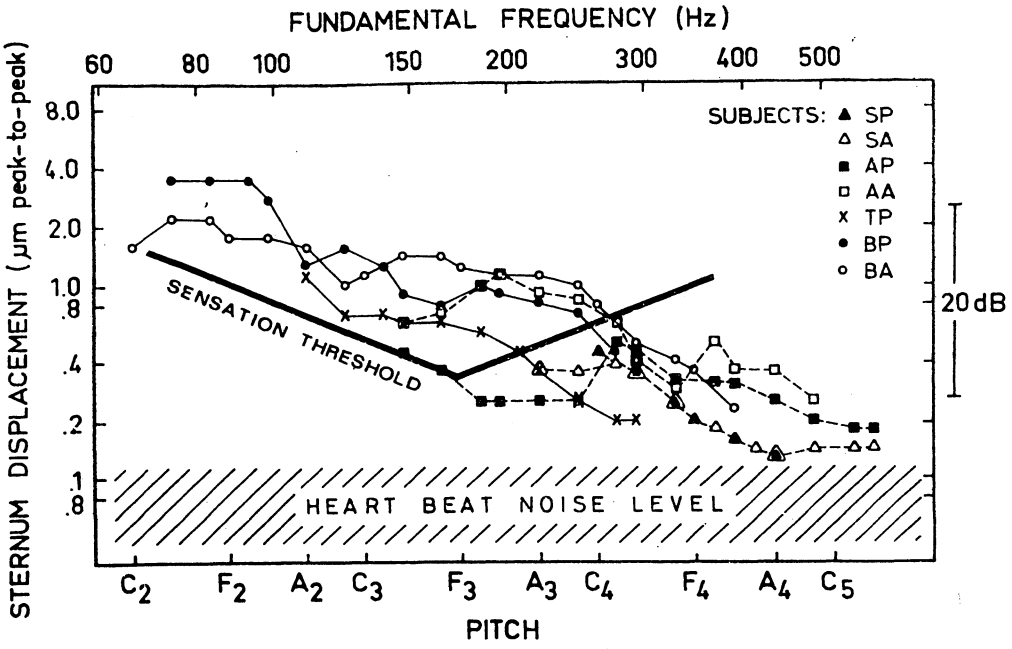


Fig. 4. Amplitude of the chest wall vibrations in different singers singing at various pitches. The heavy solid line shows the approximate threshold for sensing chest vibrations when the chest is vibrated from outside. From Sundberg (1979a)

pitch of F3, or about 175 Hz. Above this frequency the threshold starts to rise with increasing frequency, so that at the pitch of D4 (close to 300 Hz) the threshold is higher than the vibration amplitude caused by phonation. This implies that it would be difficult for a singer to sense any chest wall vibrations above this phonation frequency. Interestingly, the register embrasing the lowest phonation frequencies in the female voice is referred to as the "chest" register.

The amplitudes of the chest wall vibrations shown in Fig. 4 were recorded under normal phonation conditions. However, the amplitudes can be varied considerably by varying the phonation characteristics. Particularly a change of phonation along the phonatory dimension ranging from "pressed"

over "flow" to "breathy" has a dramatic effect on the chest wall vibration amplitude. "Press" phonation results in small vibration amplitude and "flow" and "breathy" phonation generates large vibration amplitudes. According to the results of my investigation the reason for this is that the chest wall vibrations mirror the amplitude of the voice source fundamental, which depends on the type of phonation in terms of the "pressed/breathy" dimension. And note that this concerns the voice source, so it is practically independent of vowel articulation.

This is of course quite interesting. If one changes phonation from "flow" type to "pressed" type, the amplitude of the chest wall vibrations will decrease. Using myself as a subject I measured a drop of 17 dB in chest wall vibration amplitude under these conditions without changing the acoustic amplitude of the sung vowel at all! Thus, the chest wall vibration amplitude will mirror this aspect of phonation quite efficiently.

So far we have spoken about male voices mainly. With respect to female voices the situation is a bit different. Since they phonate at higher fundamental frequencies than males, they can hardly take a systematic advantage of chest wall vibrations. On the other hand it is interesting that Kirikae & al. found the nose dorsum vibrations to be reasonably independent of vowel articulation, so it may be that such vibrations are related to the voice source. But we do not know anything about this yet, we can only formulate questions for future research.

So far we have considered the case of listening to one's own voice, while little have been said about how we hear other people's voices. In previous seminars, in this series, on the other hand, aspects of this have been discussed, e.g how the "singer's formant" is generated and how it affects the listener's auditory impression; what sounds "in tune" or "off pitch" etc, and articles complemented by sounding illustrations have been published (see Sundberg, 1977a and 1979b). Here two other aspects will be considered, namely (1) the perception of vowel identity of sung vowels, and (2) the perceived/inferred musicality of the singer.

The identification of a sound as a specific vowel can normally be related to the frequency location of peaks in the spectrum envelope of this sound. These peaks occur at the frequencies where the sound transfer

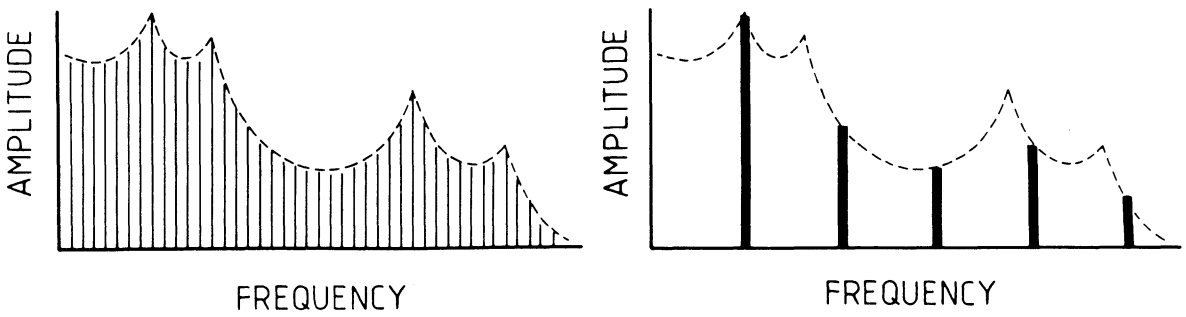


Fig. 5. Schematical illustration of the difficulties to determine the formant frequencies when the fundamental frequency is high. The two spectra were actually generated with the same formant frequencies, as suggested by the envelope, but the fundamental frequency difference is three octaves.

ability of the vocal tract culminates, i.e. at the so called formant frequencies. But composers demand rather high fundamental frequencies from female singers. In such cases the frequency distance between adjacent spectrum partials is great, so that it is impossible to detect the formant frequencies. The situation is illustrated in Fig. 5. The perceptual consequence of difficulties to locate formant frequencies from the spectrum seems to be difficulties in identifying the sound as a specific vowel. It seems reasonable that such difficulties will increase as the fundamental frequency is raised, i.e. when the pitch is raised. Hence, we see here a possibility to understand why we find it more difficult to identify vowels sung in high-pitched tones.

Fig. 6 is taken from an investigation of this phenomenon made by Morozow (1965). He had subjects guess what the syllable was when produced at different pitches by professional opera singers. Some unexpected observations can be made from the figure. A score of 100 % correct was obtained for no fundamental frequency. Thus even the male singers caused ambiguities, and not only at their top pitches. However, it should be kept in mind that this investigation concerned syllables, not vowels; it may be hard enough to hear if the singer sings "do" or "to". The most important information conveyed by this figure is the abrupt decline of the curve

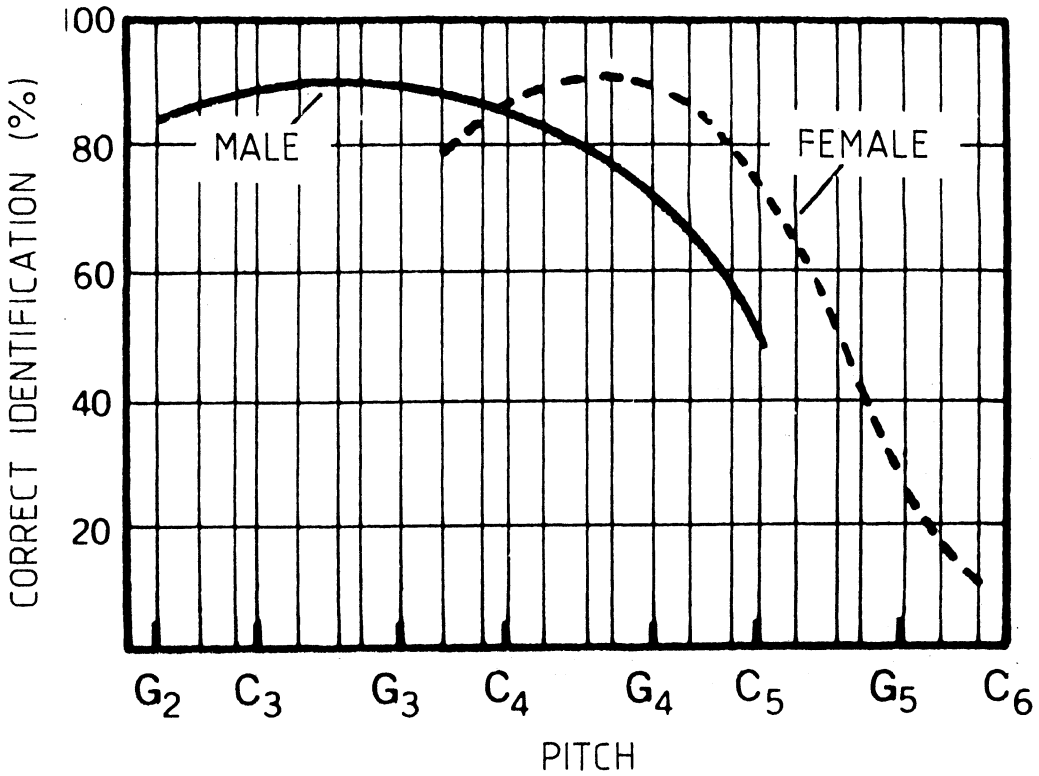


Fig. 6. Averaged intellegibility of sung syllables sung by professional singers at different pitches. After Morozov (1965)

for female singers above the pitch of C5. At the pitch of A5 the situation seems perfectly chaotic; the syllable is correctly identified in one case out of ten only!

These effects certainly depend to a great extent on the singing technique, as has recently been elegantly demonstrated by Smith & Scott (1980). Their results are shown in Fig. 7. They had a professional opera singer produce vowels in various ways, and presented these vowel sounds to listeners, who were asked to identify the vowels. From the figure we can see that there is a great loss of intelligibility above the pitch of C5 in isolated vowels. If the vowel is sung with a deliberately raised larynx, the intelligibility is considerably better up to the pitch of A5. If the vowel is preceded and followed by the consonant b, the intelligibility is much better, and in this case the improving effect of a raised larynx is rather small.

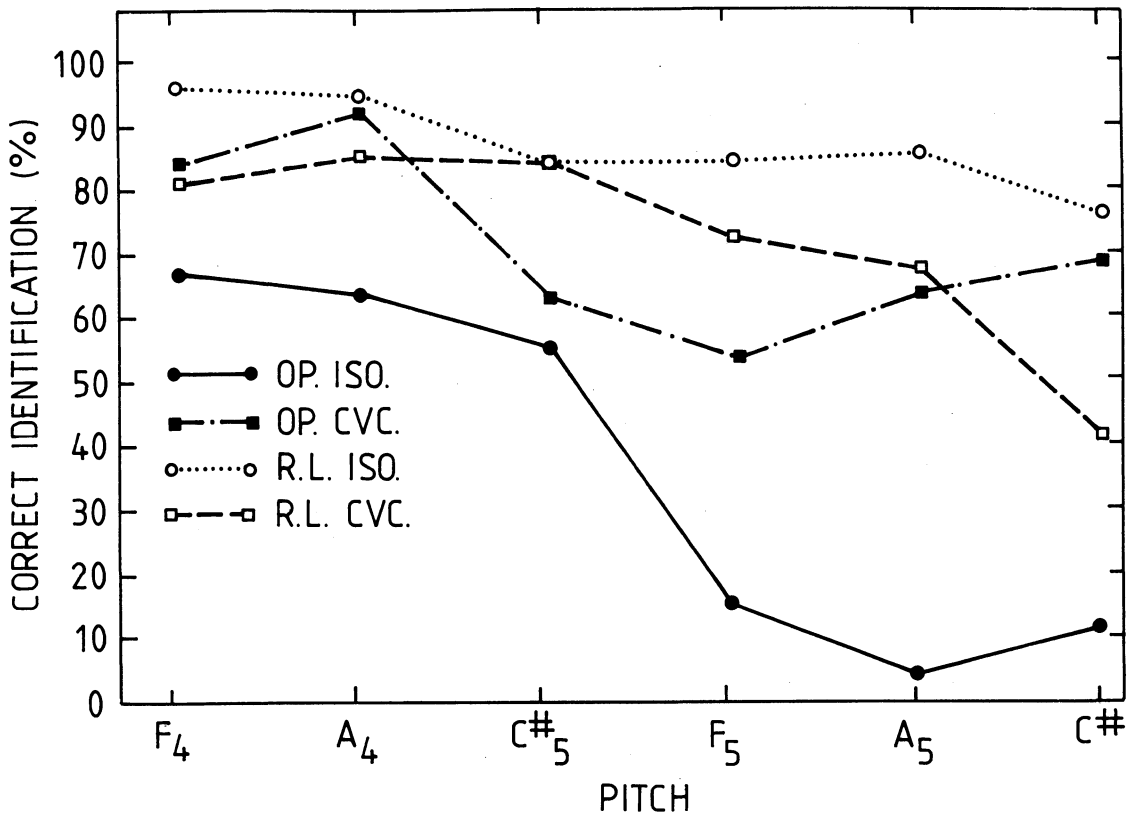


Fig. 7. Intellegibility of vowels sung by a soprano singer. ISO refers to vowels sung in isolation and CVC refers to vowels preceded and followed by the consonants b and d, respectively. OP and RL pertains to singing with normal and raised larynx, respectively. After Smith & Scott (1980).

Summarizing the intelligibility aspects, then, it seems that we have to accept poor intelligibility in vowel sounds produced at pitches considerably higher than the pitch of C5. Scotto di Carlo (1972) notes that several composers seem to realize this, when they let the singer repeat the same text even at low pitches. On the other hand many composers seem to fail in this respect so that they keep hiding important passages of the text parts by presenting them at too high pitches.

Another phenomenon apparently related to the intelligibility of high-pitched sung vowels is the vibrato. It is often assumed that the vibrato helps vowel identification. The background of this assumption is illustrated in Fig. 8. It is related to three basic facts. (1) At any moment, the spectrum of a vowel is practically harmonic, i.e. the frequencies of

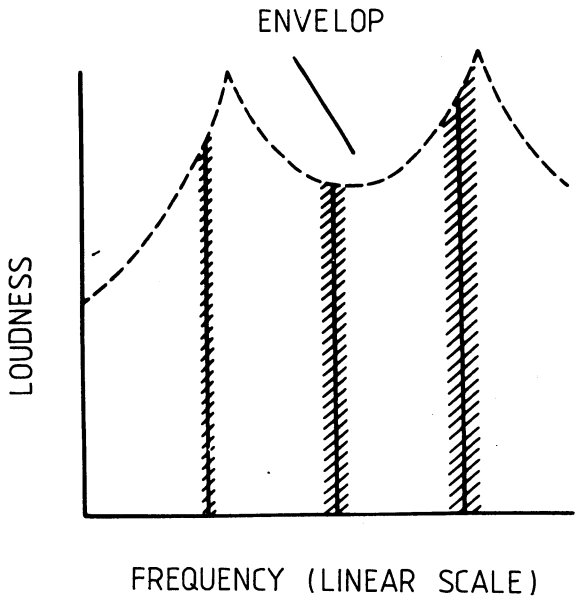


Fig. 8. Schematical illustration of the spectral effect of a vibrato.

the partials form a harmonic series ($f_n = n\, f_1$). The second fact is that the vibrato corresponds to a regular low-frequency modulation of the fundamental frequency. Thus, if the vibrato makes the fundamental frequency vary between 500 and 550 Hz, the frequencies of all partials sweep accordingly, so that, at any moment, they form a harmonic series. At the moment when the fundamental frequency is 525 Hz, the frequencies of the overtones will be 1050, 1575, 2100 Hz, and so on. (3) The amplitude of a partial depends on the frequency separation between the partial and the formant which is closest to it in frequency. Thus, if the frequency of the closest formant is lower than that of a given partial, the amplitude of this partial will drop, as soon as the frequency of that partial is raised. If the frequency of the closest formant is higher than that of a partial, the amplitude of this partial will increase, as soon as the frequency of that partial is increased. In this way, then, the vibrato will help us to locate the frequencies of the formants; if there is a formant just above the frequency of a partial, the amplitude of that partial will increase, when the vibrato makes the frequency of this

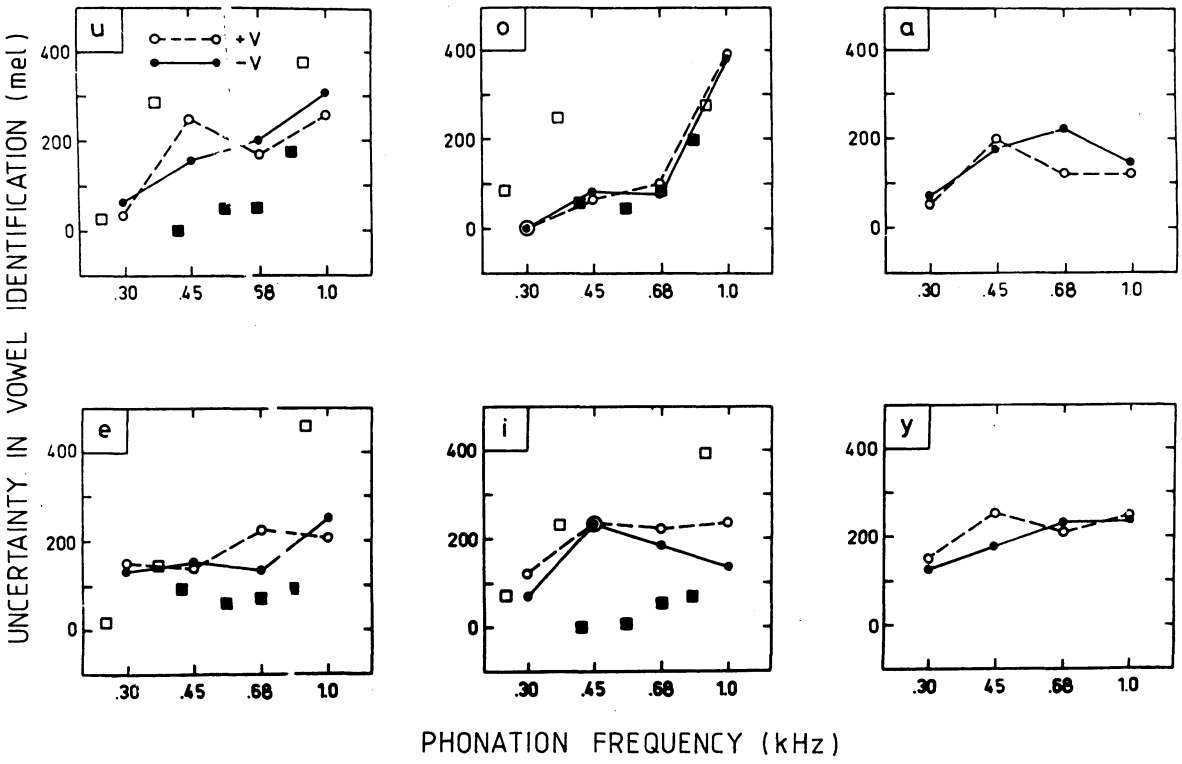


Fig. 9. Scatter of responses obtained in attempts to classify high-pitched synthetic vowel sounds as specific vowels, which had and lacked vibrato (solid and dashed curves). The filled and open squares represent data published by Stumpf (1926). They were obtained for natural vowels sung by a professional and two non-professional singers, respectively. From Sundberg (1977b)

partial increase, and vice versa. Therefore, we would obtain more information about the frequency location of the formants from a vibrato vowel than from a vibrato-free vowel, and hence it would be easier to recognize the identity of a vibrato vowel than that of a non-vibrato vowel.

This last mentioned assumption was studied a couple of years ago (Sundberg 1977b). Vowel sounds synthesized on a singing machine (MUSSE) were presented with and without vibrato to a group of phonetically trained subjects, who were asked to identify the vowel. The fundamental frequency was varied between 300 and 1000 Hz. Fig. 9 shows the scatter of the responses obtained for the various patterns of formant frequencies used.

The different graphs in the figures reveal that there is a very small influence on vowel intelligibility from vibrato, as is clear from the small differences observed between the cases with vibrato (dashed lines) and without vibrato (solid lines).

The filled and open squares refer to natural vowels rather than synthetisized vowels. The filled squares refer to a professional soprano singer and the open squares refer to two amateur soporanos. The data have been adapted from an investigation by Stumpf (1926). We may note that in the case of the vowel /o/ the singing machine produces vowels that are as intelligible as those of a professional soprano singer. However, as a rule, the synthetic vowel sounds give values which are just between those of the professional soprano and of the two amateur sopranos. This tells us that the data obtained from the synthesizer are not unrealistic. With respect to the main question. i.e. the importance of vibrato to vowel intelligibility at high pitches, the results allow us to conclude that this importance is very small or even non-existent. If there is a reason for the presence of the vibrato in singing, it can hardly be to increase vowel intelligibility.

My guess is that the vibrato does indeed serve a purpose in sung music, namely to allow the performer some freedom as regards fundamental frequency. In a perfectly vibrato-free, accompanied singing, the demands on accuracy in fundamental frequency are very high indeed, because every deviation from just intonation will lead to beats with the accompaniment. The vibrato eliminates such beats, so that the singer may use the fundamental frequency, or rather short deviations from expected pitches as a means for artistic expression. Thus, perhaps the vibrato opens up a channel for musical communication, namely deviations from expected pitches.

This leads us over to the next aspect on singing, namely expression. After having listened several times to the synthetic "performance" in SOUND EXAMPLE 3, we probably all feel a bit embarrassed. The singer has a good voice, but he seems to be a poor performer in the sense that he does not want to convey anything of importance to his listeners. He seems to think of very prosaic things while he is singing. The result is that no musical communication is established. This is the typical reaction of

musically experienced people listening to this type of performance, where nothing is added to the pitches and durations specified in the notation. Thus, it seems that a sequence of tones is not necessarily music, something "extra" must be added to the music scetched by the composer in the notation. Using a computer program for conversion of written text into speech my colleagues R. Carlson and B. Granström and I made an attempt to find out what was lacking (see Sundberg 1978).

The strategy was to study a professional singer's performance of a song, try to generalize the observations in terms of performance rules, and then to try these rules in a synthetic performance. By listening to SOUND EXAMPLE 4 we can explore the musical effect of various performance rules.

The starting point was rather neutral from a point of view of musical expression, as can be heard in SOUND EXAMPLE 3. One observation that was turned into a rule was that a following lower pitch was approached from below, not from above, as we might expect. The effect of this rule is demonstrated in SOUND ILLUSTRATION 4a. The singer made a small crescendo-decrescendo gesture on each quarter note. When converted into a rule the effect is as in SOUND ILLUSTRATION 4b. The effect is almost comical. It sounds as if the singer believes that he he is singing something like a military march. In other words, the emphasis on the individual quarter notes has now become far too strong. In order to sound more muscially acceptable, it is necessary to copy another feature of the real performance of the song, namely phrasing. The singer seemed to achieve this by letting each phrase describe a simple dynamic pattern: 3 notes level + 1 note decrescendo + 1 note crescendo + 1 note decrescendo. The sounding result plus a piano accompaniment is given in SOUND ILLUSTRATION 4c. Most musically experienced listeners prefer this performance to the previous one, as they infer an attempt on the part of the (imagined) singer to shape phrases rather than to represent a chain of unrelated tones.

This last mentioned observation is interesting. It is safe to postulate that musically experienced listeners are capable of telling where the structural units called the phrases begin, culminate, and end, regardless of how the music is being performed. Still, it is apparently very annoying to listen to a performance which does not announce this phrase structure. Perhaps the disturbing factor in this situation is that absence of

phrasing or impossible phrasing tells the listener that the musician has poor musical talent, which is an undesired and thus disturbing information when listening to a piece of music.

Apart from these purely structural points of view on perfomance, there are certainly others which are equally important. One would be the sound events leading the listener's associations to experiences of emotions, motion by running or walking, heart rate, breathing .... These are aspects of music of which we know very little or nothing explicitely but very much intuitively. Let us hope that future research fill this curious and challenging gap!

REFERENCES

CARLSON, R. & GRANSTRÖM, B. (1975) "A phonetically oriented programming language for rule description of speech", in Speech Communication, Vol. II (G. Fant, ed.), Almqvist & Wicksell, Stockholm

KIRIKAE, J., SATO, T., OSHIMA, H. & NOMOTO, K. (1964): "Vibration of the body during phonation of the vowels", Revue de Laryngologie Otologie-Rhinologie 85, 317-345

LINDBLOM, B. & SUNDBERG, J. (1972): "Acoustical consequences of lip, tongue, jaw, and larynx movement", Journ. Acoust. Soc. Amer. 50, 1166-1179

MOROZOW, V. P. (1965): "Intelligibility in singing as a function of fundamental voice pitch", Soviet Physics Acoustics 10, 279-283

SCOTTO DI CARLO, N. (1972): "Etude acoustique et auditive des facteurs d'intelligilité de la voix chantée", Univ. de Provence Travaux de l'Inst. de Phonétique d'Aix 1, 115-125

SMITH, L.A. & SCOTT, B.L. (1980): "Increasing the intelligibility of sung vowels", Journ. Acoust. Soc. Amer. 67, 1795-1797

STUMPF, C. (1926): Die Sprachlaute, J. Springer, Berlin

SUNDBERG, J.(1977a): "Singing and timbre", in Music Room Acoustics (Publications issued by the Royal Swedish Academy of Music #17, second printing 1980), 57-81

SUNDBERG, J. (1977b): "Vibrato and vowe identification", Polish Archives of Acoustics 2:4, 257-266

SUNDBERG. J. (1978): "Synthesis of singing", Swedish Journ. of Musicology 60:1, 107-112

SUNDBERG, J. (1979a): "Chest vibrations in singers", Speech Transmission Laboratory Quarterly Progress adn Status Report 1/1979, 49-64

SUNDBERG,J. (1979b): "Rent och falskt i klingande praxis", in Vår Hörsel och Musiken (Publications issued by the Royal Swedish Academy of Music #23), 78-101

TONNDORF. J. (1972): "Bone conduction", in Foundations of Modern Auditory Theory, (J. V. Tobias, ed.), Academic Press, New York

SINGING AND THE HEALTH OF THE VOICE

by docent BJÖRN FRITZELL, Department of Logopedics and Phoniatrics, the Karolinska Institute, Huddinge Hospital, S-14186 HUDDINGE

Introduction

Singing makes you healthy, and singing is a sign of good health. The present paper will deal with not only the healthy voice but also the unhealthy voice, and for this reason it seems appropriate to start out by paying honors to Manuel Garcia and giving him thanks from all of us who are involved in the medical care of the vocal folds. Manuel Garcia was born in Spain and became a renowned teacher of singing in Paris and London. One of his students was the famous Swedish soprano Jenny Lind, and in 1848, Garcia also became a member of the Royal Swedish Academy of Music.

Garcia's great achievement with respect to the medical care of the vocal folds was to introduce the laryngeal mirror. This simple device gave the medical doctors an instrument for observing the vocal folds which still is one of the most important ones in use. Almost all of the information which I will present here has been collected by means of the laryngeal mirror.

In my presentation, I will discuss acute inflammation of the vocal folds, i.e. laryngitis, nodules of the vocal folds, the influence of hormones on the singing voice, singing during growth, and finally I will spend a few words on singing and psyche. These topics will be dealt with more or less thoroughly, and most of them less thoroughly than I would wish, unfortunately. This is due to the fact that I do not know very much. On the other hand others do not seem to know much either. Thus, there is still much research to be done in this area.

Acute laryngitis - acute inflammation of the vocal folds

The main symptom of this disease is sudden hoarseness, and the main finding in a visual inspection is a reddening of the vocal folds. Under normal conditions, the vocal folds are white or yellow-whitish, when examined by means of the laryngeal mirror. It is the thickness of the mucous membrane which determines the colour. During inflammation the mucous membrane is swollen because an increased circulation of blood, and hence, the vocal folds appear red. The swelling of the mucous membrane interferes with the vibrations of the vocal folds. When the inflammation is fierce, the swelling is considerable so that it prevents the vocal folds from vibrating. Then, the patient loses his voice and becomes aphonic.

The cause of acute laryngitis is an infection with viruses or bacteria, and the laryngitis is usually part of an upper respiratory infection. The most common course is aphonia for a couple of days, whereafter the voice gradually return. After a week it is more or less restored to normal conditions. For singers in particular, it is very important to know that the vocal folds are very vulnerable during acute laryngitis. Small hemorrhages and swellings arise easily, and misuse of the voice, i.e. too much use of the voice may lead to long lasting, deleterious consequences.

The most important component in the treatment of acute laryngitis is voice rest, and preventing coughing is also urgent. If there is a tendency to cough, the patient should be given expectorants and cough-depressing drugs. Occasionally, if the patient is a singer scheduled for a very important performance which cannot be postponed, a slight or moderate laryngitis can be treated locally in order to reduce the swelling of the vocal folds mucous membrane. This will allow the singer to sing for a few hours. It is thought, however, that this kind of therapy may delay the recovery, and it is not often used.

A kind of singer's tracheitis has been recognized in Stockholm during recent years (Leanderson 1980). It often starts as a common cold, but it may also occur without any such signs. The symptom is mainly a kind of singing fatigue: the patient can sing for 5 to 10 minutes, and then the voice loses its quality, and the intensity decreases. After a voice rest

of 1-2 hours, the singer can sing again at the usual level of performance, but only for a short while. The main, and often the only finding, is a reddened mucous mebrane of the trachea, i.e. what can be seen below the vocal folds during examination with the laryngeal mirror. The duration of this disorder is usually 10-14 days, but it may last for several months. It has been hypothesized that the cause is an easily developing edema of the subglottic mucosa, i.e. a swelling of the mucous membrane of the lower surface of the vocal folds. Antibiotic therapy with vibramycin is effective in most cases.

Finally, it should be pointed out that singers should avoid medication with acetylsalicylic acid during upper respiratory infections. This drug increases the tendency for bleeding, and its use may cause the development of a submucous hematoma of the vocal folds during singing.

Nodules of the vocal folds

Vocal fold nodules is a term used for swellings of the edge of the vocal fold in the middle of its membranous portion. This is where the amplitude of the vocal fold vibration is maximal. It is commonly believed that there are mechanical reasons for the development of vocal fold nodules. Since such nodules are mainly seen in subjects who habitually use high vocal pitch and intensity, it has been hypothesized (Sonninen & al. 1972) that the high vocal fold tension and the forceful clapping together of the vocal folds due to the so called Bernoulli effect lead to microhemorrhages or bleedings in the mucous membrane, and the consequence of the repeated traumas is a constant swelling.

Nodules of the vocal folds are frequently seen in children who shout much. Among singers, it is well known that sopranos and tenors are liable to develop nodules. According to Lacina (1972), sopranos tend to develop nodules gradually, whereas in tenors they most often arise acutely. Lacina (1972), in his study of 80 opera singers, also found singers' nodules in 4 altos (cf Table I). In childhood, nodules are more common in boys than in girls, whereas in adulthood, nodules are uncommon in men.

Table I. Occurrence of vocal fold nodules in opera singers (from Lacina, 1972).

Voice type	Number of singers examined	Number of singers with nodules	Percentage
Soprano	24	8	33
Mezzosoprano	5	1	20
Alto	10	4	40
Female total	39	13	33
Tenor	16	7	43
Baritone	14	0	0
Bass	11	0	0
Male total	41	7	17

The main symptom of singers suffering from nodules, is difficulty with piano singing in the upper part of the range. The diagnosis of nodules is based on mirror examination of the larynx. Most often insufficient closure of the vocal folds during phonation is also found, as demonstrated by Moore & al (1979) by means of ultra high speed filming. However, the laryngeal examination is carried out with the patient phonating under the most unnatural conditions (wide jaw opening and the tongue pulled maximally anteriorly). Of course, observations of the vocal fold function made under such cir cumstances is of limited value.

The most important treatment of "young" nodules is voice rest. The vocal behavior which has lead to the development of the nodules should be changed. In singers this maight mean a change of singing teacher. Old, firmly established nodules might require surgical treatment.

Hormones and the singing voice

The sex hormones have a crucial influence on the singing voice. The history of surgical excision of the sexual glands in boys before puberty for production of castrato singers is well known.

During childhood, there is no difference in the development of voice between boys and girls. With puberty, the male voices undergo a rapid change. Naidr & al (1965) studied the development of voice in 100 boys over a 5-year period, between the ages of 11 and 15. The first voice change observed was a lowering of the upper range limit. The main voice changes occurred during the first part of puberty. The average length of the mutation of voice was 13 months, i.e. from the first observed changes until the voice showed a seemingly stable function in an adult male range. It has been shown that the male voice changes also after puberty. Shipp & Hollien (1969) demonstrated that listeners can identify the age of unknown male speakers amazingly well, and the average fundamental frequency differs between age levels; it exhibits its lowest level in the interval of 40 and 50 years of age, and it rises in old age (Hollien & Shipp 1972)

If androgens (male sex hormones) or anabolic steroids be given to women, there is a virilisation of the voice. The voice becomes deeper, and some women even experience a mutational voice change, like boys during puberty. Anabolic steroids were given as strengthening medication in the 1950s and 1960s, and androgens were used in drugs for women in order to reduce symptoms of menopause. Unfortunately, this voice change is in most cases irreversible, so that the voice does not return to normal, when the medication is ended (Heinemann 1976). For this reason, the above mentioned uses of these drugs have been stopped.

Voice changes with the menstruation cycle

Lacina (1968) made an enquiry among 100 female singers and students of singing about voice changes with the menstruation cycle. Among these 80 gave a positive answer. Twentyfive experienced voice problems only during the days preceding the menstruation, while 37 experienced such problems

only during the menstruation, and 19 experienced voice problems both before and after the menstruation. The type of these voice problems was described in a great variety of ways, the most common complaint being that the voice became husky and breathy, and that it did not possess its normal ability to "carry". Mirror examinations of the larynx during the menstruation period revealed insufficient closure of the vocal folds during phonation.

A related study was carried out by Flach & al (1969), who sent a questionnaire to 136 female singers in order to obtain information about the influence of menstruation and pregnancy on the voice function. The answers indicated voice changes in relation to the menstruation period in 77%, which is in good agreement with the 80 % that Lacina found. Four fifths of these singers reported negative changes. However, one fifth reported a positive voice change during menstruation, in that the felt their voices to become softer. Out of the 136 female singers, 61 had born children. 40 of them had observed clear voice changes during the pregnancy. Among them, 21 had had very positive experiences; they found that their voices became richer in quality and more mature. Five of the singers had experienced voice problems during the pregnancy, but the authors noted that these women also reported problems of different kinds such as feeling sick and vomiting. It is often stated that the voice returns to its pre-pregnancy status after the birth of the child. However, I know of no systematic investigation of this issue.

Contraceptive pills have been suspected of influencing the voice function of female singers. Dordain (1972) studied the matter, but failed to find any support for this suspicion.

Finally, it should be noted that other hormones may also influence the voice, and here the thyroid gland should be mentioned. It is well known that hypothyroidism, i.e. when too little thyroid hormone is produced, leeds to voice changes. The voice pitch is lowered and the quality becomes coarse. If compensatory medication is given, the voice returns to normal, so that the voice change is reversible. The opposite condition, thyrotoxicosis, i.e when too much thyroid hormone is produced, is not known to leed to any direct voice changes. Indirectly, however it

certainly does, because of general and psychic symptoms, which may be very severe.

Singing during growth

I believe that singing should start at an early age. One might raise the question, of whether there is an optimal age for learning to sing. For all other kinds of learning there seem to be optimal ages, so I think it is justified to infer that this is true for singing as well.

An old study by Flatau & Gutzmann (1908) concerns the range of the child's singing voice. They maintained that the range is very limited during the first years of life and increases very slowly. Thus, at the age of seven it is still less than an octave. This classical study was disputed by Hartlieb (1957), who found that children, already during their first years of life, could use their voice within a pitch range of no less than 3 octaves. The discrepancy between these two studies may perhaps be explained by different methodsof observation and examination.

A classical question concerning singing during growth is the following: Is it advantageous for the development of good voice function to sing in children's choirs? No doubt, the answer depends of the type of voice usage practised in the particular choir. There are great dangers! Hess (cited by Weiss, 1950) claimed that 60 - 70% of good voices in boys were ruined by unwise choir leaders. And Laumer (also cited by Weiss, 1950) found that only 2% of all the boys in the famous German and Austrian boys' choirs became singers at an adult age. This is particularly remarkable in view of the excellent musical education which is given to these boys in connection with their choir singing. This is not to say that these results can be applied to choir singing in any country, but the risk of joining choirs in early age should not be neglected. Leaders of children's choirs should be very observant and follow the development of each child's voice carefully.

Let us now turn to the question of singing during puberty. I am convinced that this is the most dangerous period, at least in boys. The mutional voice change has been dealt with in a classical and comprehensive review

by Weiss (1950). The fierce dispute on this topic between Manuel Garcia and the first British laryngologist, Morell Mackenzie, is legendary. Apparently, Garcia had been singing during puberty, and he thought that his voice had suffered damage from this singing. Thus, he strongly recommended abstinence from singing during puberty. Mackenzie, on the other hand, strongly believed that singing wisely, without strain was good and that it was a natural way to stimulate the development of the voice even during puberty.

In Swedish schools, instruction in singing is provided for all children over a number of years. However, some children have obvious problems with their singing. Some years ago, an investigation was made in Göteborg (Persson, 1964) of students described as "weak in singing", growlers, and monotoners (sångsvaga, brummare, monotoner). These labels appear a bit queer, but they seem to be very resistent. Thus, I often meet them among my adult patients. I ask all patients, who come to see me for voice problems, if they like to sing. Many of them answer: "No, I cannot sing." When I ask them: "How do you know?", they report that they were told this in the first grade of the school. Thus, several people have learned at school not how to sing but rather that they cannot sing, and this seems to prevent them from really trying again! This is obviously very unfortunate. I am convinced that most people can learn to sing and have a lot of pleasure from their singing. It is important that we realize that learning is a process which takes some time. Nobody can ride a bicycle, read, or play an instrument at the first attempt, and some children may be slow learners. It would be perfectly impossible that the teacher tells a child after the second or third lesson of reading that he cannot read, so that the best thing to do is to keep quiet when others are reading, or go and find something else to do. Still, it seems that this attitude towards learning is practised with regard to singing even in our time.

It seems, though, that more and more attention is being paid towards the importance of helping all pupils to learn singing. For instance, in Stockholm, so called singing clinics (sångkliniker) have been established in a number of elementary schools. In these clinics, children with singing problems receive special training. The results have been most rewarding, not only in terms of singing, but also with respect to the self-

confidence and the attitude towards school has improved (Olsson-Ekström, 1974).

There seems to exist a relationship between lack of musical talent and voice disorders. Eisenson & al (1958) tested 90 patients with voice disorders, and 87 subjects in a control group for their ability to discriminate pitch and loudness using subtests of the Seashore Measures for Musical Talent (Seashore & al 1960). The results indicated a significantly lower score for subjects with voice disorders in the pitch discrimination test, but not in the loudness test. Out of these subjects suffering from voice disorders 15 were retested at the end of a 15-weeks course in voice therapy and ear training with the emphasis laid on pitch discrimination. The retest indicated a statistically significant improvement in pitch discrimination ability.

Bergendal (1976) also used Seashore Measures of Musical Talent and compared the scores from this test with the results of voice training and voice therapy in a group of 27 students of logopedics and 7 patients with voice disorders. It turned out that students and patients with high Seashore scores responded well to voice training, whereas little improvement was observed in those having low scores. May be early training would have changed the situation for those scoring low on Seashore's test at an adult age.

Singing and psyche

One of the classical songs for male choir from the 19th century ("Dåne liksom åskan, bröder!") contains some lines that are often quoted. In my translation they read: "Singing creates noble feelings; the key of the heart is singing". I think we can all agree with these statements. Singing makes you happy. Thus, it seems that singing influences the mood favorably. On the other hand, if your voice is not functioning as smoothly as it usually is, the influence from singing on your mood is in the opposite direction; you feel uncomfortable and unhappy. If the voice does not function at all, so that you cannot phonate, or if absolute voice rest is requested, e.g. because of a laryngeal operation or an acute laryngitis, this may have a negative or even unbearable influence on your

psyche and your relations to the environment. A complete silence may easily be interpreted as a silent criticism. It also occurs, curiously enough, that people react towards such a silence by starting to talk slower and more distinctly, as if they believed that this would help you to share their conversation!

The relationship between singing and psyche appears to be reciprox. If you are psychically out of balance, if you are tense and under stress, your voice reveals your mood; you do not sound well. Psychic factors have a heavy impact on a singer's performance. Self-confidence and assurance are apparently very important to a singer's ability to perform. For this reason, the task of the voice expert treating a singer is often mainly a psychotherapeutic one. He uses his mirror and examines the vocal folds, but the essence of the treatment is to encourage, support, and reassure the singer that everything is fine, and that he will do very well.

These last statements are, of course, entirely based on my personal experience. I do not know of any systematic investigations of these particular aspects on the effects on psychic factors on voice function and vocal performance.

Final comments

The question was raised by our chairman, whether one can "heal one's voice by singing". Claims have been made that you can successfully treat voice diseases and disorders such as common cold, asthma and stuttering by means of singing, but statements of this kind seem to mirror personal beliefs rather than facts. I would restrict myself to support the view that singer's nodules may disappear, if the technique of singing is corrected. But there is no doubt singing is good for your vocal and general health. Training in singing gives us a better command of our voice, an awareness of what is going on in our larynx, and an improved auditory perception. With this we are better equipped to meet the vocal stresses and strains of today's noisy life.

To sum up, there is much research to be done in the area of singing and the health of the voice. Just to mention one example, what happens to

girls' voices during puberty? As far as I know, nobody has ever published any study of this truly basic issue.

To end this presentation I would like to modify an old Latin proverb. We all know that "navigare necesse est" (sailing is necessary). Now, as singing means sailing on the wings of the song, it seems justified to conclude that "NAVIGARE ET CANTARE NECESSE EST"

REFERENCES

BERGENDAL, B.-I. (1976): Musical talent testing used as a prognostic instrument in voice treatment, Folia Phoniat. 28, 8-16

DORDAIN, M. (1972): Etude statistique de l'influence contraceptifs hormonaux sur la voix, Folia Phoniat. 24, 86-96

EISENSON, J., KASTEIN, S. & SCHNEIDERMAN, N. (1958): An investigation into the ability of voice defectives to discriminate among differences in pitch and loudness, J. Speech & Hearing Disorders 23, 577-582

FLACH, M., SCHWICKARDI, H. & SIMON, R. (1969): Welchen Einfluss haben Menstruation und Schwangerschaft auf die ausgebildete Gesangsstimme?, Folia Phoniat. 21, 199-210

FLATAU, T. & GUTZMANN, H. (1908): Die Singstimme des Schulkindes. Archiv fuer Laryngologie 20, 327-348

HARTLIEB, K. (1957): Der Umfang der Jugendstimme, Folia Phoniat. 9, 225-239

HEINERMANN, M. (1976): Hormone und Stimme, J. A. Barth, Leipzig

HOLLIEN, H. & SHIPP, T. (1972): Speaking fundamental frequency and

LACINA, O. (1968): Der Einfluss der Menstruation auf die Stimme der Sängerinnen, Folia Phoniat. 20, 13-24

LACINA, O. (1972): Das Vorkommen von Stimmlippenknötchen bei den Sängern, Folia Phoniat. 24, 345-354

LACINA, O. (1975): Eine Hypothese ueber eine der möglichken Ursachen von Stimmlippenknötchen bei Altstimmen, Folia Phoniat 27, 321-324

LEANDERSON, R. (1980): Personal communication

MOORE, G.P., CANNON, K.A. & WILSON, L.I. (1979): Vocal fold vibration in the presence of vocal nodules, in Transcripts, 8th Symposium on the Care of the Professional Voice June 1979, part III, The Voice Foundation, New York, 24-31

NAIDR, J., ZBORIL, M. & SEVCIK, E. (1965): Die perturbalen Veränderungen der Stimme bei Junge in Verlauf von 5 Jahren, Folia Phoniat. 17, 1.18

OLSSON-EKSTRÖM, I. (1974): Sångklinik ger hjälp för mer än rösten. - Några reflexioner angående olika typer av sånghämning, Musikkultur nr 6/1974

PERSSON, J. (1964): Studier av röstomfång och sångförmåga hos barn i grundskolans 1:a, 3:e och 5:e årskurser, Thesis work in pedagogics, Göteborg University

SEASHORE, C.E., LEWIS, D. & SAETVEIT, J.G. (1960): Seashore measures of musical talents, Manual, rev. ed., The Psychological Corporation, New York

SHIPP, T. & HOLLIEN, H. (1969): Perception of the aging male voice, J. Speech & Hearing RTesearch 13, 703-710

SONNINEN, A., DAMSTE, P.H., JOL, J. & FOKKENS, J. (1972): On vocal strain, Folia Phoniat. 24, 321-336

WEISS, D.A. (1950): The pubertal change of the human voice, Folia Phoniat. 2, 126-159

SOUND EXAMPLES

Side A

BENNETT: SINGING SYNTHESIS IN ELECTRONIC MUSIC

Track I 1. Synthesized phrase from Gesualdo's "Moro Lasso"

Track II 2. Excerpt from the tape part of "Aber die Namen der seltnen Orte und alles Schöne hatt' er behalten", by G. Bennett

Track III 3. Natural versus synthesized sopranos (three singers)

4. Three synthetic scales with SPL-dependent timbre

5. Two synthesized examples of vibrato evolution

6. Regular vs. random variation of the fundamental frequency

Track IV 7. Excerpt from "Winter (1980)" for tape, by G. Bennett

Side B

JOHAN SUNDBERG: THE VOICE AS A SOUND GENERATOR

Track I 1. Examples of (a) breathy, (b) "flow", (c) normal, and (d) pressed phonation

IVAN FONAGY: EMOTIONS, VOICE AND MUSIC

Track II 1. Anger and tenderness in three languages

2. Three synthesized versions of an emotionally ambiguous sentence, interpreted as a) a rebuke
b) an attempt to convince
c) consoling

Track III 3. Excerpt from W.A. Mozart:"Die Entführung aus dem Serail"
(Osmin's aria "Drum beim Barte des Propheten...")

Excerpt from W.A. Mozart:"Don Giovanni"
(Ottavio's aria no. 10)

4. Coquettish luring in English and French

5. Excerpt from B. Bartok:"The Miraculous Mandarin"
(luring of the courtesan)

JOHAN SUNDBERG: TO PERCEIVE ONE'S OWN VOICE

Track IV 1. Effect of microphone position on voice quality:
a) in front of, and b) to the side of the speaker

2. Speech as picked up by an accelerometer placed on the speaker's forehead

Track V 3. Singing synthesis "off the sheet"; no performance rules added

4. Singing synthesis, with performance rules accumulating:
a) initial pitch dipping
b) crescendo-decrescendo on each note
c) simple phrasing and piano accompaniment