Waweform and spectrum of the glottal voice source

Sundberg, J. and Gauffin, J.

journal: STL-QPSR
volume: 19
number: 2-3
year: 1978
pages: 035-050

http://www.speech.kth.se/qpsr
III. LOGOPEDICS AND PHONIATRICS

A. WAVEFORM AND SPECTRUM OF THE GLOTTAL VOICE SOURCE

J. Sundberg and J. Gauffin

Abstract

By means of an inverse filter technique ad modum Rothenberg (1972), the glottal voice source is studied with respect to waveform and spectrum in different singer and speaker voices phonating at different pitches, registers, degrees of vocal effort and in different ways. A quantitative relationship is found between (1) the amplitude of the waveform and the amplitude of the source spectrum fundamental and (2) the shape of the waveform and the relative amplitudes of the source spectrum overtones. A phonatory dimension extending between the extremes of phonatory "press" and "flow" is suggested involving a substantial variation in the amplitude and the relative length of the closed phase of the glottogram.

Introduction

A couple of years ago both of the present authors studied the voice source by inverse filtering vowel sounds and by matching vowel spectra on a terminal analogue of the vocal tract, respectively (Lindqvist, 1965 and 1970; Sundberg, 1973 and 1977). At that time both these techniques were very time consuming.

The results of the inverse filtering studies showed how the voice source varied with phonation frequency and sound pressure level in different voices. However, with a reasonable amount of work only a rather limited number of vowel sounds could be studied. Moreover, the interpretation of these results in spectral terms was hardly possible for practical reasons. The results of the source spectrum studies, on the other hand, revealed typical differences in the lower part of the spectrum between differing types of singer voices and between different types of phonation. It was not feasible, though, to interpret these differences in terms of the underlying waveforms, which would be required for a physiological understanding of the results.

After the modifications proposed and explored by Rothenberg (1973), cf. Fig. III-A-1, the inverse filtering method has now become easy to use. Moreover, if combined with a computer offering FFT facilities, it yields information not only on the glottal volume velocity waveform but also on the voice source spectrum. Clearly, this opens up quite new possibilities for the investigation of the voice source.
The aim of the present study was to investigate (1) the relationships between the glottal volume velocity waveform and the low frequency part of its spectrum, and (2) the variations of this waveform associated with varied conditions of phonation including professional opera singing.

Method

Recently Fant devised a new construction of an inverse filter which has been combined with a mask ad modum Rothenberg. At the time of this investigation the equipment (cf. Fig. III-A-2) included two inverse filters only, allowing a study of the voice source up to the frequency region of the second formant. This is sufficient for a study of the lower partials in the source spectrum and the major characteristics of the voice source waveform.

The mask used is designed by Rothenberg, who also kindly provided it. It has been described in detail with respect to theory and design in Rothenberg (1977). For this reason a brief description would be sufficient here. The microphone, on the other hand, was designed at the Dept. of Speech Communication, KTH.

The mask is a standard respiratory plastic mask in which holes have been drilled. The holes are covered with a fine mesh screen. The acoustic resistance of the screen is chosen so as to give a pressure drop which is sufficient for the microphone transducer at normal air flow values and at the same time low enough, so as not to distort the speech.

The microphone fits into the hole for the mask fitting. It is a pressure differential transducer with a metalized mylar membrane having a diameter of 15 mm. One side of the membrane is open to the inside of the mask while the other side is open to the outside. The deflection of the membrane is sensed by an optical system consisting of an integrated light-emitting diode and photo-transistor. The photo-transistor measures the amount of light reflected by the membrane. The temperature drift of this system is inherently low and slow. It would be possible to reduce it, if the light emitting diode voltage is used as a temperature sensor inside the transducer. The membrane has a resonance
Fig. III-A-1. Inverse filtering without (upper series) and with (lower series) a mask. After Rothenberg (1971).

Fig. III-A-2. Block diagram of the experimental setup. $F_1$ and $F_2$ are inverse filters, the output of which is fed into the lowpass filter $LP$. The computer $CD$ stores the signal, computes the spectrum and plots the results on an incremental plotter in terms of the waveform and the spectrum. The equipment in the lower branch measures the sound pressure level, $SPL$. 
at 1.3 kHz, approximately, with a $Q$-value of 5. This resonance is compensated for in the preamplifier which also has means for calibration of the output. The preamplifier is contained in a small box permanently connected to the connection wire of the microphone. Microphone and preamplifier together are designed to give a flat frequency response up to 1.5 kHz.

The actual frequency response of the mask microphone was measured by a recording of sustained vowel sounds in an unechoic chamber. The signals from the mask microphone and from an ordinary Brüel & Kjær 1" microphone were simultaneously recorded on each of the tracks of a two-track tape recorder. The tracks were found to possess identical frequency responses within $\pm$ 0.5 dB in the frequency range of interest. Spectrum samples were then taken at identical places in the recorded vowel sounds ($\pm$ 5 msec) by means of an FFT analysis. The mask microphone spectrum was then corrected with respect to the sound radiation impedance yielding $+6$ dB/octave as well as the radiation characteristic of a point source mounted in a spherical baffle with a radius of 9 cm, cf. Fant (1960, p. 45). The differences in the amplitudes of the spectrum partials were considered only for such partials that were no more than 20 dB below the level of the strongest spectrum partial, so as to avoid errors due to distortion. The average differences from three phonations of each of the vowels [o, α, æ, ø] are shown in Fig. III-A-3. As can be seen in the figure the frequency response is better than $\pm$ 3 dB in the frequency range of 1–1.4 kHz. The response curve is very similar to that published by Rothenberg (1977) referring to the same type of equipment but determined by means of a different method.

The mask microphone signal was branched not only to the inverse filters but also to a derivation circuit which was connected to a voltmeter. In this way the sound pressure level (SPL) of all phonations could be determined.

The inverse filter output was fed directly into the A/D converter of the computer and stored in the memory. In this way, samples of vowel sounds of a maximal duration of 1500 msec could be studied. The signal and its FFT spectrum could be plotted on an incremental plotter.
Fig. III-A-3. Frequency response of the mask microphone used.
Experiment

The voice source was studied in five males. During the experiments the subject was seated in an ordinary room and phonated sustained vowels holding the mask firmly to his face. The vowels [o] and [æ] were used because of their concentration of sound energy in the low frequency region. The experimenter was seated next to the subject and adjusted the inverse filters. He recorded the inverse filter output as soon as the filter settings appeared to be appropriate. At the same time he determined the SPL reading.

The subjects represented different types of voices. Two were highly experienced professional singers from the Stockholm Opera. One was a semiprofessional singer, one had a moderately trained voice, and one had a typically untrained and slightly harsh voice, which, however, did not need any phoniatric treatment.

Only the two professional singers experienced a slight discomfort when they phonated into the mask. The problem was that they heard their own voice in unfamiliar way. This, they claimed, disturbed their feeling of vocal effort to some extent. However, none of the subjects found that the mask influenced on the phonation ability to any appreciable extent. The untrained voice became somewhat strained towards the end of the session, certainly because the subject was not used to phonate at high frequencies and degrees of vocal effort.

All subjects phonated at three different pitches and at least three degrees of vocal effort. In addition, some of the subjects varied the type of phonation. In the case of the moderately trained subject no SPL readings were taken.

Measurements

Fig. III-A-4 illustrates the measures derived from a straight line approximation of the glottogram. Essentially, they were chosen in agreement with Lindqvist (1965).

As regards the source spectra only thepartials below 1000 Hz, approximately, were considered. Apart from the amplitude of the fundamental the amplitude relationships between the overtones were considered.
Fig. III-A-4. Glottogram measures derived by means of a straight-line approximation of the waveform (upper figure), and illustration of how the spectrum slope measure $Mv \Delta$ was determined (lower figure).
As pointed out by Monsen & Engebretson (1977) the source spectrum can only rarely be accurately described in terms of a number of dB/octave. Generally, the envelope falls off in a rather irregular way. Hence, a crude but simple measure of the source spectrum envelope slope seems justifiable. The measure chosen, $Mv \Delta$, gives an idea of the probable correction which should be added to the value that any partial within the frequency range considered would have, if a $-12$ dB/octave slope is taken as reference, cf. Fig. III-A-4. Thus, the procedure for computing $Mv \Delta$ was (1) to apply a $-12$ dB/octave envelope slope starting at the amplitude of the fundamental, (2) to measure the deviations $\Delta$ of each partial from this envelope, and (3) to compute the average of these deviations. It should be pointed out that this measure gives a fair approximation of the amplitudes of the source spectrum partials near .5 kHz even in cases where the spectrum envelope falls off with a constant number of dB/octave. It should also be observed that $Mv \Delta$ is particularly sensitive to the level of the fundamental as compared to the level of the overtones. Thus, an increase of e.g. 5 dB in the level of the fundamental will lower the $Mv \Delta$ by the same amount, other things being equal.

Results

1. General observations

Fig. III-A-5 presents a set of glottograms obtained from the semi-professional singer phonating in different ways. The amplitude scale is the same for all glottograms except those in the bottom series in the figure. For technical reasons the absolute flow values were not recorded. However, zero flow would coincide with the quasi-horizontal baseline of the glottograms except in the case of breathy phonation and possibly also falsetto register.

Ideally, the glottogram baseline should be perfectly straight and horizontal as long as there is no leakage during the closed phase. In practise this was found to be the case only at low pitched phonation. Mostly, the baseline shows a negative slope. The reason for this might be that the vocal folds do not close parallelly. In some cases a ripple is superimposed on the baseline. Such effects may be due to movements on the upper surface of the vocal folds, cf. Lindqvist (1965) and in our case also to the higher formants.
Fig. III-A-5. Glottograms pertaining to various types of phonations as produced by subject J.
The top left pair of glottograms in Fig. III-A-5 demonstrates that the same SPL value is not always associated with the same glottogram characteristics. In pressed phonation, which would imply a high subglottic pressure combined with a high degree of adduction activity in the larynx, the peak air flow of the glottic cycle is low and the closed phase is long. The flow phonation seems to be a sort of phonatory opposite of the pressed phonation, and it gives the glottogram a high peak flow value and a shorter closed phase. The subject felt he had to reduce his vocal effort considerably in order to produce the same SPL value in the flow phonation as in the pressed phonation. When he raised his vocal effort to a degree which he experienced as more normal, the SPL value increased by 8 dB, as seen in the top right glottogram in the same figure.

When the subject phonated with raised larynx position, his glottogram became similar to that pertaining to pressed phonation, and when he lowered his larynx the glottogram became similar to that of the flow phonation. This suggests that the low larynx is associated with the flow phonation in this subject. The reason for this would be that the subject is a trained singer. In male singing a low larynx position is common, and, needless to say, it should be accompanied by a non-pressed type of phonation. In the case of the moderately trained voice the same effect was not as pronounced.

In breathy phonation the glottogram is characterized by the absence of discontinuities, as previously reported by Rothenberg (1973). The falsetto register is generally associated with a short closed phase as compared with phonation in modal register at the same pitch. In vocal fry the main feature seems to be that the glottal pulses do not occur at equal distances in time. As shown in the bottom line of Fig. III-A-5, this distance may even approach zero.

2. Waveform and spectrum

In general, no quantitative information on the spectrum can be derived in a simple way from the waveform. A complicated Fourier transform is required for that purpose. In the case of glottograms, however, the physiology imposes certain restrictions as to the variability of the waveform. Thus, in normal cases, a quasi-triangular
deviation from zero flow will occur only once within a cycle. This
may encourage to an attempt to find simple qualitative relationships
between waveform and spectrum characteristics (cf. Flanagan. 1958).

It is clear that discontinuities in the waveform correlate with strong
overtones. The main sharp discontinuity in the glottogram is found in
the closing phase, and the relative length of the closed phase informs
about the duration of the excursions from zero flow in the period.
Hence, it can be assumed that the ratio \( S_C T_0 / T_C \) will exhibit some
correlation with the overtone content of the spectrum. The \( Mv \Delta \), de-
defined in Fig. III-A-3, provides a measure of the dominance of the over-
tones in the source spectrum. Fig. III-A-6 shows the correlation ob-
tained from the entire material representing differences in voice qua-
lity, vocal effort, pitch and type of phonation. Eventhough the scatter
of the data points is substantial, there is a clear linear relationship,
which can be described in the following way:

\[
Mv \Delta \approx 4.5 - 17 \log \left( \frac{S_C T_0}{T_C} \right) \tag{1}
\]

We may conclude that the amplitudes of the overtones in a voice source
spectrum can be estimated using Eq. (1).

In the case of glottograms it is reasonable to assume that the ampli-
tude of the fundamental is dependent on the waveform amplitude, as
suggested by Fant (1960). The case is illustrated in Fig. III-A-7. The
waveforms in the figure were obtained by adding three overtones with
amplitudes decreasing at a rate of -12 dB/octave to a fundamental lying
(a) 12 dB, (b) 6 dB, and (c) 0 dB above the amplitude of the second par-
tial. Moreover, all these four partials reach their positive peak simulta-
aneously. Interestingly, the cases (b) and (c) show a clear similarity
with real glottograms. In the same figure is also shown the relations-
ships which these waveforms display between waveform amplitude and
fundamental amplitude, \( A_0 \). The topmost point and the line represent
the extreme case of a sinewave. It can be seen in the figure that there
is a clear correlation eventhough no one-to-one relationship between
waveform and fundamental peak amplitude.
Fig. III-A-6. Relationship between the spectrum slope measure $Mv \Delta$ and the ratio between closing time and the relative closed phase time in four subjects. The solid line shows the best linear approximation, the equation of which is given in Eq. (1). The corresponding correlation coefficient is given in the frame.
The same stars and the line of Fig. III-A-7 have been combined with data points derived from real glottograms in Fig. III-A-8. It can be seen that there is a fair agreement in all glottograms even though they were derived from widely differing voices, phonation types, pitches, and degrees of vocal effort. The scatter between these data points is small and decreases with increasing glottogram amplitude. Moreover, the data points of Fig. III-A-7 agree well with the rest of the points in Fig. III-A-8 as shown by the stars in that figure. Interestingly, the distance between the data points and the line tends to decrease with rising glottogram amplitude in the figure. This suggests a relation between the waveform and the amplitude of the glottogram. Thus glottograms with high amplitudes tend to be more like (rectified) sinewaves while small amplitude glottograms are more similar to the cases of pressed phonation or raised larynx shown in Fig. III-A-5. Breathy phonations mostly fell around 2 l/sec in our material. The highest glottogram amplitudes were observed at a medium pitch and high vocal effort, as will be shown later. It can be concluded that regardless of overtone content there is a strong correlation between the amplitude of the glottogram and that of the fundamental. The best straight-line-approximation of this relationship is given by:

\[ L_0 = 25 \log A_G + C_1 \]  

where \( L_0 \) is level of the fundamental in dB, \( A_G \) is the glottogram amplitude in l/sec, and \( C_1 \) is a constant.

In most cases the sound pressure level (SPL) of a vowel sound is determined by the amplitude of the partial or pair of partials which lies closest to the frequency of the first formant. If the amplitudes of these source spectrum partials are known, we would be able to predict the SPL value of a vowel sound. Above it has been demonstrated that the amplitude of the source spectrum fundamental, as well as the probable amplitudes of the lower overtones, can be predicted from glottogram data. If we know the first formant frequency we should be able to estimate the SPL even from pure glottogram data. The predictions made were computed by means of the following expression:

\[ \text{SPL}_p = L_0 + Mv \Delta + C_2. \]
Fig. III-A-7. Waveforms derived with differing amplitude of the fundamental \( (A_{f1}) \) but the same spectrum envelope slope of -12 dB/octave starting from the second partial, the amplitude of which was the same in all curves \( (A_{f2}) \). The graph in the right part of the figure shows the resulting relationship between the peak amplitude of the fundamental, \( A_0 \), and the peak-to-peak amplitude of the waveform, \( A_g \). The straight line represents the case that the amplitudes of the fundamental and the waveform increase at the same rate.
Fig. III-A-8. Peak amplitude of the fundamental as function of glottogram (peak-to-peak) amplitude in four voices phonating at various pitches, degrees of vocal effort, and in different types of phonation. The solid line and the stars are the same as in Fig. III-A-7.
where $L_P$ is the predicted SPL value,
$L_0$ was computed using Eq. (2),
$M_v \Delta$ was computed using Eq. (1),
$C_2$ is a constant depending on the ratio between the first formant and the fundamental frequency.

The constant $C_2$ takes into account the fact that in a spectrum falling off at a rate of -12 dB/octave, the amplitude of the $n$th partial is 12 dB stronger than that of the $2 \cdot n$:th partial. If the pitch frequency is lowered by one octave while the first formant frequency is kept constant, the partial underlying that formant will be the $2 \cdot n$:th instead of the $n$:th. This would cause the SPL to drop by 12 dB. However, as the pitch is lowered one octave the density of the partials in the frequency domain is doubled, which reduces the effect by 3 dB, i.e. the total SPL drop will be -9 dB only. Here we have assumed that the source spectrum envelope was falling off at a rate of -12 dB/octave regardless of the pitch. This was required by our definition of $M_v \Delta$, cf. Fig. III-A-4. Fant (1968, p. 201) has shown, that if the shape of the glottal pulse is constant with respect to the absolute duration while the period time is increased, i.e. if the closed phase increases with the cycle time, the SPL would drop by 3 dB only. This is due to the fact that the source spectrum in that case would be increasingly dominated by overtones as pitch drops. By contrast, in the example above, the source spectrum envelope was assumed to be constant. It may be pointed out that the SPL drop accompanying a lowering of the pitch by one octave has been found to be around 10 dB in singers (cf. Sundberg, 1973).

The agreement between observation and prediction is good, as shown in Fig. III-A-9. It should be observed, though, that the slope of the best straight-line-approximation of the data points deviates slightly from 1. The reason for this is that this subject had an $A_0/A_G$ relationship sloping less steeply than the average computed for all subjects and used in the prediction. Thus, the predicted variations in his $A_0$ values became a bit exaggerated.

Summarizing we may conclude that we can use glottogram characteristics in order to predict some principal characteristics of the source spectrum: (1) the amplitude of the fundamental, and (2) the average deviation from a -12 dB/octave slope starting at 0 dB on the fundamental.
Fig. III-A-9. Predicted and observed SPL values. The solid line shows the best straight line fit and the correlation coefficient is given in the frame.
It seems possible to make reasonable estimates of the SPL using these relationships between glottogram data and source spectrum.

2. Pitch and vocal effort

The aim of the present section is to elucidate the dependence of the voice source on pitch, vocal effort and to some extent also type of phonation in singers and untrained voices. As mentioned above there were four trained voices and one untrained voice in the material studied. All data will not be presented in the figures of this section, though. Instead, differences that seem typical in the view of the entire material will be demonstrated by contrasting only one of the singers with the untrained voice.

Fig. III-A-10 shows how the glottogram amplitude varies with SPL at different pitches. The solid line shows the slope which the data points would follow if the glottogram amplitude increased at the same rate as the SPL. This case would imply a constant source spectrum envelope regardless of the SPL. The trained voice exhibits an SPL dependence similar in slope to the reference line. The untrained voice, on the other hand, shows a similar slope as the singer at comparable SPL values only. Thus, the source spectrum overtones become increasingly dominant with increasing SPL in low and moderate vocal effort in the untrained voice, while the source spectrum envelope remains more constant in the case of the trained voice. Note also that pressed phonation is characterized by a small glottogram amplitude and comparatively high SPL, or, in other words, a source spectrum with strong overtones. Note also, that data points lying to the left of the reference line indicate that the same glottogram amplitude is associated with a lower SPL than in the case that the data points lie to the right of the line. It is probably noteworthy that in the case of the untrained voice most of the data points lie to the left of the line, while the reverse is true for the singer. This supports the observation made by Lindqvist (1970) in comparing a trained singer with untrained voices: the glottogram amplitude was smaller in the trained voice for comparable SPL values.

Fig. III-A-11 shows how the glottogram amplitude varies with the fundamental frequency. At low vocal effort the amplitude remains essentially the same for all frequencies studied. At high vocal effort the
Fig. III-A-10. Glottogram amplitude as function of SPL in two voices. The solid line represents the case that the glottogram amplitude increases at the same rate as the SPL. The slope of the dotted line in the right graph shows case that a 10 dB increase in SPL is accompanied by a 4 dB increase in the level of the fundamental.
Fig. III-A-11. Glottogram amplitude as function of fundamental frequency in two voices phonating at a vocal effort ranging from low (p) to high (f), and in the case of the singer also in various ways: F=flow phonation, P=pressed phonation, and FA=falsetto.
amplitude culminates at a pitch lying centrally in the subject's pitch range. The pressed phonation yields an amplitude which is very low. This may lead us to speculate that the smaller amplitude in strong vocal effort in the case of the untrained voice is a sign that his phonation was slightly pressed under these circumstances.

The closing time $S_C$ is shown in absolute measure in Fig. III-A-12. The solid curve demonstrates how the points would fall if the ratio $S_C/T_0$ was constant. The untrained voice tends to have such a constant relative closing time while the singer has a slightly shorter closing time in his lowest pitch than in his middle pitch.

Fig. III-A-13 shows the relative length of the closed phase $T_C/T_0$. In both voices it tends to increase slightly with increasing pitch. However, in other voices examined the length of the closed phase had a tendency to drop somewhat with rising pitch. This was also a general trend in the data reported by Lindqvist (1970).

3. **Vocal efficiency**

It is generally assumed that good and poor voices differ with respect to their ability to convert a certain amount of air to sound of a given SPL. A first rate singer is expected to be capable of sustaining a vowel much longer than an untrained speaker. As glottal leakage during the closed phase was not measured in the present investigation, we can study only if the above holds even if we disregard such leakage. It should also be pointed out that there was no reason to suppose that the subjects differed substantially with respect to leakage. The consumption of air per sec $c$ can be estimated from the glottogram using the expression

$$c \approx \frac{A_G}{T_0} \cdot \left( T_0 - T_C - \frac{S_C + S_0}{2} \right)$$

The radiated sound power can be computed from the SPL values. Fig. III-A-14 shows how the ratio between the square root of the sound power and the air consumption varies with the SPL in four voices. Note that the data points represent differing voices phonating at several pitches; degrees of vocal effort, and with various types of phonation. Only in the case of pressed phonation values are obtained, which
Fig. III-A-12. Closing time as function of fundamental frequency in two voices phonating at vocal effort ranging from low (p) to high (f), and in the case of the singer also in different ways: F=flow phonation, P=pressed phonation, and FA=falsetto. The solid curve represents the case that the closing time is constant relative to the cycle time.
Fig. III-A-13. Length of the closed phase relative to the cycle time as function of fundamental frequency in two voices phonating at vocal effort ranging from low (p) to high (f), and in the case of the singer also in differing ways: F=flow phonation, P=pressed phonation, and FA=falsetto.
Vocal "efficiency", defined as the square root of the radiated sound power (in mW) divided by the air consumption (in l/sec), as function of SPL in four voices phonating at various pitches, degrees of vocal effort, and in different types of phonation. Circled symbols pertain to deviant types of phonation (pressed, breathy, etc.).
deviate substantially from the average. There is a limited scatter and a clear linear relationship between this measure of voice efficiency and the SPL. This shows that voices do not seem to differ to any great extent with respect to their ability to convert air to sound of a given SPL. This is true only if constant leakage of air through the glottis is disregarded. In view of the fact that pressed phonation gives the highest value of vocal efficiency, it is evident that this measure of vocal efficiency is rather misleading. It is possible that a better measure could be obtained if a factor mirroring muscular activity was introduced in the voice efficiency measure, e.g. in terms of the subglottic pressure.

Discussion

Above we have attempted to find phonatory explanations to a considerable voice quality difference between two voices by contrasting their glottograms. Only a few differences have been found. This is not an unexpected finding. As far as the spectrum is concerned male opera singers' voices are known to differ from untrained voices with respect to the "singer's formant" in the first place, i.e. in the high frequency part of the spectrum. The "singer's formant" has been shown to be a resonatory phenomenon involving a clustering of higher formants (Sundberg, 1974). According to Sundberg (1973) the generation of the "singer's formant" does not require more of the voice source spectrum than an envelope slope of -12 dB/octave, thus a slope which can be found in normal speech. Moreover, the "singer's formant" is found in the frequency region 2.5-3 kHz as a rule. Source spectrum partials of such high frequencies can hardly contribute appreciably to the glottogram. Thus, there are several reasons not to expect voice source differences between singers and untrained voices associated with the "singer's formant". On the other hand, the source spectrum envelope slope varies considerably with pitch and vocal effort in untrained speakers (cf. e.g. Fant, 1960). The need for a "singer's formant" would restrict the variability of the voice source in singing. Therefore, we would expect to find differences between singers and untrained voices in the variability of the voice source, i.e., how much it is allowed to vary with, e.g., pitch and vocal effort.

A clear difference of this kind was found in the relationship between SPL and the glottogram amplitude. According to Fant (1960) the
following relationship holds as a rule of thumb: an SPL increase of 10 dB is accompanied by an increase of the level of the fundamental of 4 dB. As can be seen in Fig. III-A-10 this holds as a good approximation for our untrained speaker. For the singers, on the other hand, this is not the case. The level of their fundamental increases at approximately the same rate as their SPL. In other words, the voice source in our singers varies less with the SPL than is normal in untrained speakers.

We may ask how singers learn to avoid source spectrum variations which normally accompany changes in SPL. A closer look at Fig. III-A-10 reveals that the singer does not use the SPL range which the untrained speaker used and where his voice source shows the greatest dependence on the SPL. We may hypothesize that the singer simply discards that part of his available dynamic range just in order to avoid considerable source spectrum variations.

A second, probably important difference between the singer and the untrained voice was found in the dependence of the glottogram amplitude on the pitch. In high vocal effort the singer showed greater peak flow values. When the phonation is pressed this peak flow is reduced, other things being equal. As mentioned before, this may lead us to assume that the untrained speaker used a pressed phonation when he raised his vocal effort. It should also be remembered that the untrained voice became somewhat strained towards the end of the session. A possible interpretation of this is quite trivial: the singer has learnt to phonate with high vocal effort without using a pressed type of phonation. The interesting point would then be that a greater peak air flow value during the glottal cycle may be necessary in order to avoid phonatory press. It should be appreciated that the peak air flow is not identical with air consumption. Nevertheless, the singer’s consumption of air, again disregarding glottal leakage, was somewhat higher than that of the untrained voice at comparable pitch and SPL.

According to Sundberg (1971), the major effect of vocal effort on the voice source spectrum envelope in singers is found in the low frequency part of the spectrum. While the envelope slope above 1 kHz
remained essentially the same regardless of vocal effort, the slope below 1 kHz was steeper in low vocal effort than in high vocal effort. This would imply that the Mv Δ, our measure of source spectrum envelope slope, should increase with rising vocal effort. The Mv Δ was found to be related to the ratio between the closing time S_C and the relative duration of the closed phase T_C/T_0. Consequently we would expect S_C to decrease and T_C/T_0 to increase with rising vocal effort. Figs. III-A-11 and III-A-12 show that this is also the case. In this way the results of the present study allow us to interpret Sundberg's source spectrum data in terms of glottogram data.

The notion of pressed phonation has been used several times above. Its major glottogram characteristics are a low peak air flow in the glottic cycle and a long closed period. A hypothetic physiological interpretation of it was given above: a high subglottic pressure combined with a high degree of adduction activity in the larynx. Rubin & al. (1967) has published measurements of subglottic pressure and transglottal air flow for varied types of phonation. They show that a high subglottic pressure and a decreasing air flow may occur when phonation is continued eventhough the reserve of available air in the lungs is almost finished. Their interpretation of this phenomenon is that the glottal resistance is increased. This would be equal to an increase of the adduction activity. It is interesting to note that they term phonations of this kind inefficient. If we instead accept the ratio between the square root of the sound power and the air consumption as a measure of the vocal efficiency we will arrive at the opposite conclusion, as shown. The pressed phonation is the most efficient. Rubin & al had good reasons for speaking of inefficiency in such cases, because this type of phonation imposes strain on the voice organ. Still this point gives a striking illustration of the necessity of separating acoustical and physiological measures. More important, it seems that the present study has found acoustic characteristics of pressed or tense phonation: a relatively weak fundamental and strong overtones. It should also be stressed that the same pitch and degree of vocal effort does not imply only one type of voice source. There is an independent third dimension, the extremes of which we have called pressed phonation and flow phonation. While pressed phonation seems to be avoided in singing, the flow phonation may be used.
Conclusions

The present study has brought about quantitative evidence for the relationships between the waveform and the spectrum of the voice source. Thus, the peak air flow amplitude, i.e. the glottogram amplitude shows a strong correlation with the amplitude of the spectrum fundamental. Thus, if the level of this partial differs between two vowel spectra having the same fundamental frequency and the same first formant frequency, the difference would be due to different values of the peak transglottal airflow during the glottal pulse. The average deviation of the lower source spectrum partials from an envelope slope of -12 dB/octave is correlated with the ratio between the closing time and the relative length of the closed phase in the glottogram. The main voice source difference between singers and an untrained speaker was found in the influence of the SPL, which is greater in the case of the untrained voice. Moreover, when the untrained subject phonated at high vocal effort, his peak transglottal air flow of the glottic cycle did not reach as high values as were observed in singers. As a reduction of this air flow was found to be a characteristic of pressed phonation, this difference should probably be ascribed to a greater degree of phonatory press in the untrained voice. Change of phonation type from pressed phonation to its apparent opposite, i.e. flow phonation, changes the characteristics of the glottal source waveform drastically involving a increase of the peak air flow and a shortening of the closed phase.

Acknowledgments

This work was supported by the following funds: the Swedish Council for Planning and Coordination of Research, the Swedish Council for Research in the Humanities and Social Sciences, the Swedish Natural Science Research Council, the Swedish Medical Research Council, and the Bank of Sweden Tercentenary Foundation.

References


