The voice source studied by means of inverse filtering

Lindqvist-Gauffin, J.

journal: STL-QPSR
volume: 11
number: 1
year: 1970
pages: 003-009

http://www.speech.kth.se/qpsr
B. THE VOICE SOURCE STUDIED BY MEANS OF INVERSE FILTERING

J. Lindqvist

Abstract

The voice source waveform for three male subjects was investigated both for connected speech and for phonation of a neutral vowel with increasing fundamental frequency and again with decreasing fundamental frequency.

The volume velocity waveform at the glottis was derived for four degrees of voice effort. The speed quotient, the open quotient, and the sharpness of the peak of the waveform as defined according to a triangular approximation of the glottal pulse were plotted as functions of fundamental frequency and speech intensity.

Procedure

The exact nature of the vocal fold vibration is not yet known in spite of considerable research effort by many investigators. To a great extent this lack of precise knowledge is a result of the difficulties associated with studying the vocal folds, per se, and simultaneously controlling the several associated physiological parameters, subglottal air pressure, and so on.

Inverse filtering techniques, Holmes (1962), Lindqvist (1965), and Miller (1959), provide a means for indirectly studying vocal fold movements by deriving the volume velocity waveform at the glottis. Such derivation is accomplished without the use of tracheal punctures, nasal catheters and the like. Subglottal pressure is an important parameter in describing the glottal waveform. In this investigation four degrees of subjective voice effort were used as a parameter, since voice effort and subglottal pressure are related to each other. During the recording sessions a reference for the voice effort was needed. A sound pressure level would not provide an adequate reference since sound pressure is not related to voice effort in a simple manner. Further, checking the level meter would probably have distracted the subject from his major experimental task. Instead, a person trained in phonetics read the text samples to the subject who was asked to repeat the text.

* Some of these data are reviewed in Lindqvist (1965).
sentence by sentence, with the same subjective voice effort used by the phonetician. The distance between the phonetician and the subject was altered in relation to the increase in sound level in order to avoid the unnatural situation of two persons standing near each other and speaking in very loud voices.

It was not practical to choose glottal waveforms covering a wide enough frequency range from the speech samples. Therefore, each subject was asked to produce a neutral vowel with continuously rising and falling pitch immediately following his production of the sentences for each subjective voice effort. The subjects were instructed to do this in a "speech mode" as opposed to singing the vowel.

Three male subjects with normal but varied types of voices were used. One of the subjects was a trained singer (JS). Four degrees of subjective voice effort were used, low, medium, high, and very high. The microphone was kept at a constant distance of 20 cm from the subject's mouth.

Results

Typical voice source waveforms of the three subjects are shown in Figs. I-B-1 - I-B-3. The fundamental frequency is shown along the abscissa and the rms sound pressure level along the ordinate. Zero dB corresponds to 60 dB sound pressure level at a distance of 1 m from the mouth. Each glottal waveform sample has an arbitrary amplitude scale from which the relative amplitudes of the glottal pulses can be determined. These waveforms were taken from the recording with continuously rising and falling fundamental frequency. From these recordings it was possible to choose glottal waveform samples with the same fundamental frequency as found in the text samples. There were no consistent differences between the samples taken from phonation with rising and falling frequency and those taken from the connected speech recordings. Therefore, the waveforms shown in Figs. I-B-1 - I-B-3 can be regarded as typical glottal waveforms in speech at different speech intensities and fundamental frequencies. As an illustration of this, three samples from connected speech are shown in Fig. I-B-4 (subject OK).
Fig. I-B-1. Volume velocity waveforms at the glottis derived by inverse filtering techniques.
Fig. I-B-2. Volume velocity waveforms at the glottis derived by inverse filtering techniques.
Fig. 1-B-3. Volume velocity waveforms at the glottis derived by inverse filtering techniques.
Fig. I-B-4. Typical volume velocity waveforms in speech. Subject OK.
Since vocal tract articulation can influence the voice source in several ways, only open vowels were used in this study. Vowel segments close to voiceless sounds were avoided because the laryngeal opening gesture associated with these sounds may change the positioning of the vocal folds in the vowel. For very close vowels, the vocal tract constriction introduces an air flow resistance which tends to raise the intraoral pressure and thus reduce the pressure drop across the glottis. This effect alone could account for a slight decrease in voice fundamental frequency. It is, however, known that on the average the fundamental frequency is higher for close, i.e. "high", vowels than for "low" vowels, see Mártony (1968), Peterson and Barney (1952). This suggests that there is also a mechanical interaction between the tongue and the larynx. Another effect to consider is the mechanical load of the supraglottal air impedance on the vocal folds. Theoretical considerations suggest that a low first formant frequency may change the conditions for the vocal fold vibrations to a considerable extent. The increase in damping of the first formant at low frequencies caused by the finite vocal tract wall impedance will, however, minimize this interaction.

The interaction mechanisms affecting the glottal waveform are generally too small to be discriminated from changes due to slight variations in speech effort. In unstressed parts of speech, where these factors may display a high degree of variability, the source waveform was accordingly variable and often different from those shown in Figs. I-B-1 - I-B-3. At the lowest voice effort the voiced segments tended to be of short duration. This was especially evident between voiceless sounds due to their glottal opening gesture. An example of this is shown in Fig. I-B-5. In Fig. I-B-5a the vocal folds stopped vibration well before the oral closure that resulted from the production of the "k". This clearly demonstrates the fact that the glottal opening associated with the production of the "k" is initiated well before the oral closure.

When the amplitudes of the waveforms in Figs. I-B-1 - I-B-3 are plotted as a function of sound pressure (see Fig. I-B-6), it can be seen that the source amplitude (the measure in Fig. I-B-7) does not change as much as the intensity of the sound. (Note that the
Fig. I-B-5. The voice source in the first vowel in the word "fakta". The duration of the word is approximately the same in all the three cases.
Voice Intensity Re: 60 dB SPL at 1 m

Subject O.K.

Subject S.G.

Subject J.S.

Very high
High
Normal
Low

Voice effort

Source Amplitude

The amplitude scale is arbitrary.
Fig. I-B-7. A typical glottal volume velocity waveform. The speed quotient $S_q$, the open quotient $O_{q'}$, and the peak sharpness $P$ are defined from a triangular approximation of the waveform (dotted lines).

$S_q = \frac{t_1}{t_2}$ \hspace{1cm} $O_{q'} = \frac{t_3}{T}$ \hspace{1cm} $P = \frac{a}{a_i}$
amplitude has a linear scale while sound pressure has a logarithmic scale.) Voice intensity variations in speech are thus not associated only with source amplitude variations. The main intensity factor appears to be the source spectrum shape. The general properties of the vocal tract filter with resonances above the fundamental frequency, as well as the radiation from the mouth, account for a high frequency emphasis. This makes the intensity of voiced sounds more sensitive to changes in the amplitude of the source harmonics than to changes in the fundamental frequency level, Fant and Liljencrants (1962).

In the glottal volume velocity waveform, an increased level of the harmonics is correlated with faster changes in the time function (higher second derivative). We should then expect that, in the straight line approximation (see Fig. 1-B-7) the slope quotient, $S_q$, and the peak sharpness, $P$, should have high values, and the open quotient, $O_q$, and the amplitude, $a$, should have low values for good efficiency of the voice. In the natural voice source this is found to be true only for $a$ and $O_q$. There is a clear difference in the source amplitude for the three subjects. The trained voice JS has an appreciably lower amplitude than the other subjects. This implies that his air consumption is especially low since airflow is proportional to the area under the volume velocity waveform multiplied by the fundamental frequency. The relation between glottal area $A$, pressure drop across the glottis $p$, and volume velocity through the glottis $v$ can approximately be expressed as:

$$\sqrt{p \cdot A} \cong v$$

Taking into consideration that an increase in voice effort implies an increase in subglottic pressure, it can be assumed that the glottal area function amplitude, $a$, varies less than the volume velocity amplitude. At least this appears to be true for the subjects JS and OK. It can also be assumed that their glottal area decreases when the subglottal pressure is raised, except at the highest voice effort. The same effect is sometimes observed in optical glottography. These assumptions, as well as experiments on stop sounds produced by closing a valve put in the mouth, Lindqvist forthcoming article, suggest that changes in the glottal area waveform accompanying
voice intensity changes are not caused by an active adjustment of the
positioning of the vocal folds. It seems more likely that the increased
airflow through the glottis accompanied by an increased Bernoulli
force tends to bring the folds together. On the other hand, when a
stop occlusion is made the loss of transglottal pressure will effectively
separate the vocal folds. In connected speech the increase in funda-
mental frequency due to an increase in voice effort amounts to what
can be expected from an increase in subglottal pressure, Öhman and
Lindqvist (1965), Lindqvist and Öhman (1966). Thus, it appears that
the change in fundamental frequency which is associated with a change
in voice effort, is just a secondary effect of the change in subglottal
pressure.

There is one interesting difference between the voice source in
singing and speech; intensity is changed with nearly constant source
spectrum (Sundberg, personal communication), and variation in the
amplitude during singing. During speech, on the other hand, ampli-
tude is nearly constant and the spectrum changes with an increase in
speech intensity.

Fig. I-B-8 shows the "peak sharpness" \( P \) as a function of funda-
mental frequency. The value of \( P \) has a clear tendency to increase
up to about 180 Hz which is the upper limit of fundamental frequency
in normal speech for these subjects. The correlation with speech ef-
fert and intensity, on the other hand, is very low. The glottis impe-
dance, which is at its minimum when the glottis opening is at its
maximum, will cause some of the source energy to dissipate down
into the trachea. Therefore, it is possible that speech intensity is
less dependent on the shape of the peak in the source waveform.

The open quotient \( O_q \) as a function of frequency is shown in
Fig. I-B-9. This measure and the amplitude \( a_1 \) are related to the air
consumption \( V \) in the following way:

\[
a_1 \cdot O_q \approx k \cdot V
\]

Provided the distance between the microphone and the mouth is given,
k is a constant. The factor \( k \) can be determined by simultaneous
measurement of \( a_1 \), \( O_q \), and \( V \). This was not done in this study. How-
ever, the relative measure \( a_1 \cdot O_q \) can be used to compare the three
subjects. By selecting typical values for \( O_q \) and \( a_1 \) for normal voice
Fig. 1-B-8. The peak sharpness plotted as a function of voice fundamental frequency.
Fig. 1-B-9. The open quotient plotted as a function of voice fundamental frequency.
intensity, this product is approximately 11 for subject JS, 36 for subject SÖ, and 19 for subject OK. The air consumption for subject SÖ is accordingly three times as high as that for the trained voice of subject JS. The subjective quality of their voices seems to correlate highly with this measure.

The open quotient has a tendency to increase with the fundamental frequency which is in agreement with findings in optical glottography, Sonesson (1960). Also, it has been observed in optical glottography that the open quotient decreases when the intensity is raised. The tendency to do this is evident in Fig. I-B-9, but it is not consistent. For the two higher voice efforts, the reverse relation is sometimes seen. A major difference between the volume velocity waveform and the area function derived with optical glottography or high speed cine film is that the speed quotient, $S_q$, generally is above one for the volume velocity but approximately one for the area function.

For the subjects in this study, $S_q$ is approximately 2 (see Fig. I-B-8). An $S_q$ value less than one is seldom observed in the volume velocity waveform. One possible explanation for this fact is the influence of the mass reactance of the glottis opening. Theoretical calculations, Crystal (1965), support this hypothesis.

The $S_q$ value has a tendency to decrease with increasing frequency. If the influence of the mass reactance was the only explanation, the opposite should be expected, even when the reduced depth of the glottis at high $F_0$ levels is taken into account. In the laryngeal structures, factors in addition to the glottal area and pressure drop determine the sound output. Besides the influence of the phase difference between the upper and lower edges of the vocal folds, on the flow mechanical movements of the structures surrounding the glottis may have some influence on the acoustic output, Lindqvist (1965). The amplitude of these movements may be small but the area of the moving structures can be much larger than the glottis area. The tendency for $S_q$ to decrease with increasing frequency has not been observed in optical glottography, Sonesson (1960). The highest $S_q$ values of the volume velocity waveform are found for the low frequency end of the voice range where the vocal folds are known to be rather thick and soft. The upper and lower edges of the folds.
are then vibrating with different phase which also may contribute to the high $S_q$ value.

There is a low correlation between voice intensity and the speed quotient. Apparently, it is mainly the open quotient, the amplitude of the waveform and the fine structure of the volume velocity waveform at the closure of the glottis that influences speech intensity. The latter factor is lost in the straight line approximation of the waveform used in this study.

References:


