INVERSE FILTERING

Inger Karlsson
Institutionen för talöverföring, Tekniska Högskolan, Stockholm

Definition: Inverse filtering means that the formants of the speech sound are damped out by filters so that only the voice source signal remains.

My interest in this area stems from my so far unsuccessful attempts to synthesise a female voice. In an earlier study (Karlsson, 1979), I varied all the different parameters that separate male and female voices according to published data, but obtained no definitely human and female voice. As very little data about the voice source is available, my suspicions are that much of the secret is hidden there. The published studies of the voice source mainly pertain to descriptions of methods.

At present considerable research into the voice source and different methods of inverse filtering is being done at the Inst. of Talöverföring: G. Fant (1979) is constructing hard ware filters to be used on speech recorded with an FM tape recorder. In an FM recording the phase of the signal is retained and frequencies down to DC are recorded, the same signal can accordingly be filtered with many different filter settings. G. Fant is using his results to get a mathematical description of the voice source, especially in word endings, stress and phoneme boundaries. J. Sundberg and J. Gauffin (1978) are investigating the voice source of singers and untrained, normal subjects with the aid of the mask for measuring the volume velocity waveform at the mouth constructed by M. Rothenberg (1973). Finally J. Liljencrants has written a computer program for inverse filtering described by him as "an OVE III with formant anti filters replacing formant filters and the speech signal as voice source." This method, as well as G. Fant's, allows the same utterance to be inverse filtered many times and, furthermore, it is possible to alter the frequencies of the anti filters in the computer program every millisecond. I have just started to use this program but have so far no results to present.

In a short study from which I am going to mention some results, I have been using the same experimental set up of the Rothenberg mask as used by Sundberg and Gauffin (1978). Five females and one male participated in the experiment and were asked to repeat the syllable /pɑ:/ with different F_0: high, medium, and low within their normal register and with different voice levels: weak, moderate, and loud (not whisper or scream). Figure 1 shows one speaker's voice pulses for different voice levels at the same F_0. The material is too small to allow for a description of different voice types but I have tried to see how the different parameters defined by Sundberg and Gauffin and by Fant (see figure 2) vary with F_0 and voice level. The only manifest correlation between parameters I have found is between sound pressure level and the derivative at the instant of closing: the peak flow A_G, divided by the offset time T_d. The correlation coefficient between these two parameters varies for the different subjects between 0.83 and 0.96. The two parameters for one subject are depicted in figure 3. Most of the subjects also show a correlation between F_0 and the "base band" formant, $1/T_G$ according to figure 2.

This is only a preliminary study, which I have made principally to get to know the technique. I will now work mainly on getting descriptions of different voices.

References

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

Fant G. (1979). Vocal source analysis - a progress report. STL-QPSR 3-4/1979, 15-24.

Karlsson I. (1979). Försök till utvärdering av akustiska skillnader mellan mans- och kvinnoröster. Könsroller i språket. Rapport nr 75 från avd. för forskning och utbildning i modern svenska (FUMS) 1979, Inst. för nordiska språk vid Uppsala Universitet.

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

Sundberg J. and J. Gauffin (1978). Waveform and spectrum of the glottal voice source. STL-QPSR 2-3/1978, 35-50.

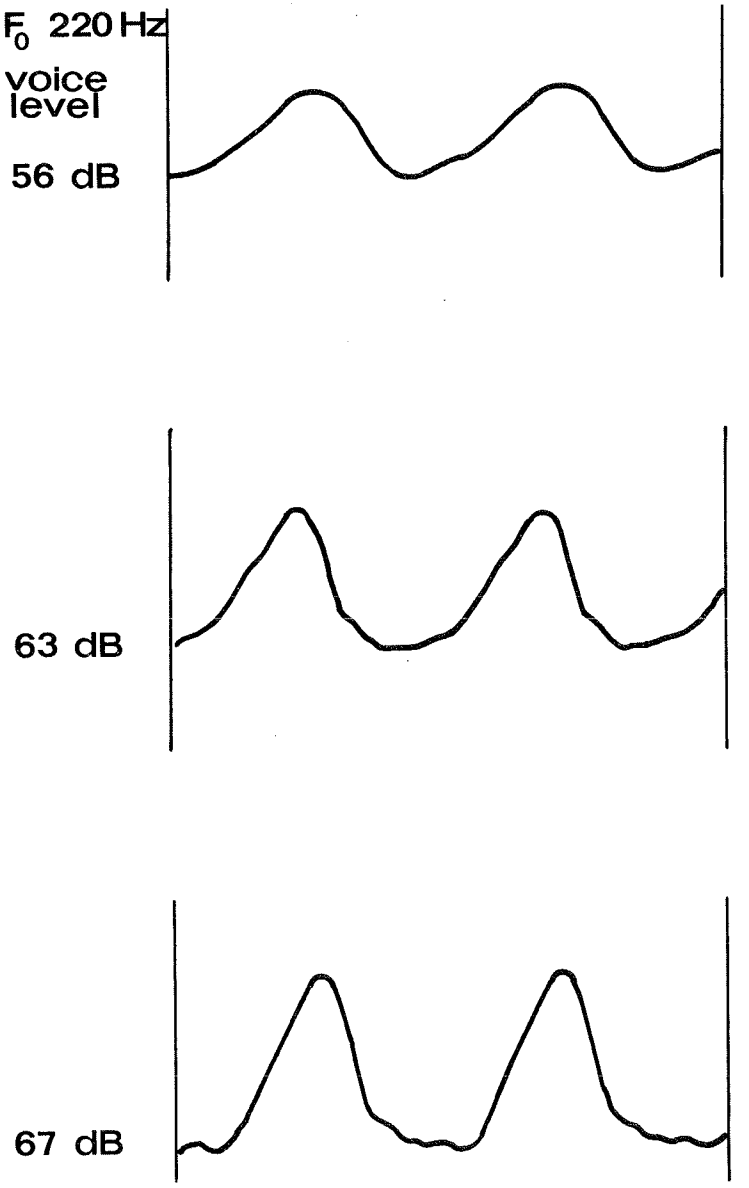

Figure 1. Voice source pulses of the same F_0 for one subject. The voice level is for the top curve 56 dB, middle curve 63 dB, and bottom curve 67 dB.

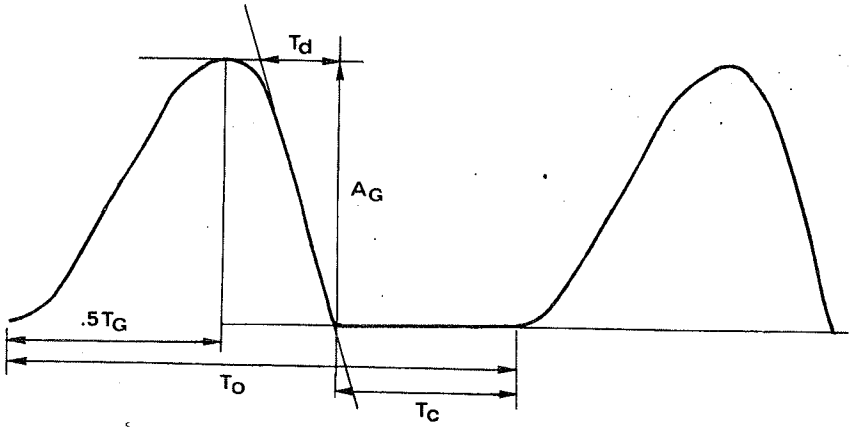

Figure 2. Different parameters used in descriptions of the voice source pulses: T_0 = the duration of the pulse. T_G = two times the pulse rise time, T_C = the closed interval of the pulse, T_d = offset time; A_G = the peak flow through the glottis.

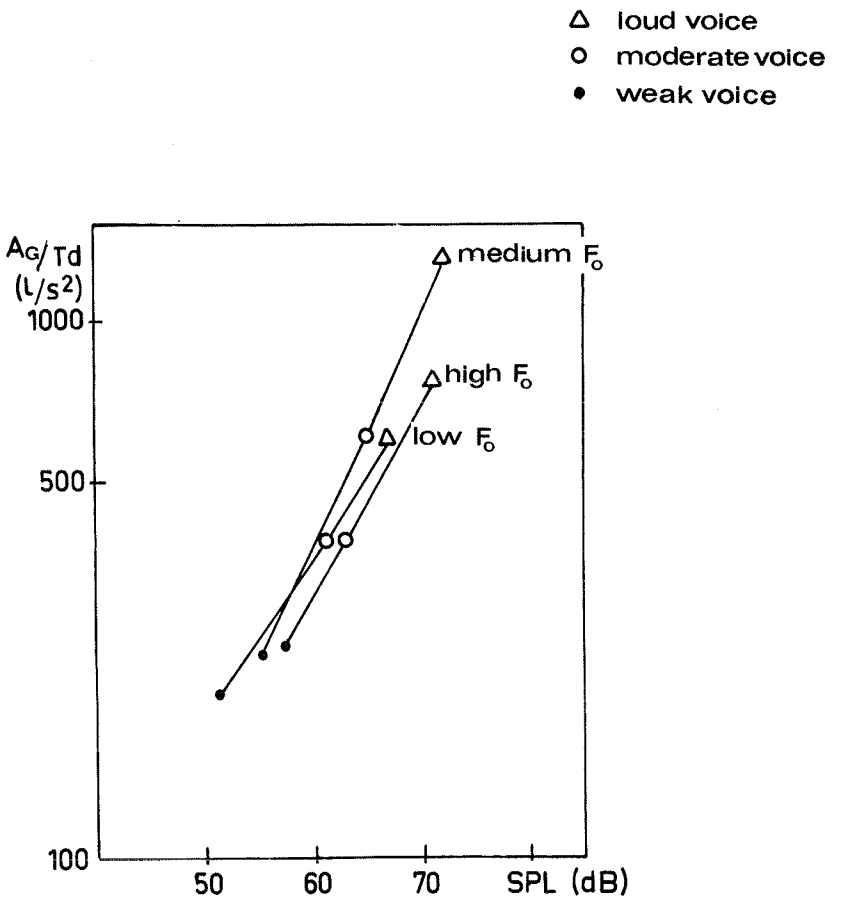

Figure 3. The relation between sound pressure level, SPL, and the derivative of the air flow at the instance of closing for one subject.