Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

# Diverse voice qualities: models and data

Karlsson, I. and Liljencrants, J.

journal: TMH-QPSR
volume: 37
number: 2
year: 1996
pages: 143-146


KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

# Diverse voice qualities: models and data

*Inger Karlsson and Johan Liljencrants,*
*Dept of Speech, Music and Hearing, KTH*

***Abstract***

*The voice source for different voice qualities has been investigated. The glottal signal has been attained using inverse filtering of the speech signal. Two different methods for modelling the results are demonstrated. An acoustic model, the LF voice source model, has been fitted to the inverse filtered waveform and the LF parameters have been used to describe the voice pulse. The parameters of a vocal cord model, the Liljencrants model, has been manipulated to produce glottal pulses similar to the natural pulses. The parameter settings are discussed.*

## Introduction

The voice source for different voice qualities has been analysed by inverse filtering and by comparisons with a glottal model. Two speakers, one female and one male, produced the analysed utterances. The voice qualities varied in loudness, fundamental frequency and tenseness. The most extreme quality, creaky voice, was only produced by the male speaker. The LF voice source parameters were used to get a parametric description of the data. The inverse filtered data for the male speaker have been model using Liljencrants' glottal model (Liljencrants 1996, Karlsson and Liljencrants 1994)

## Speech material

Recordings of a male French speaker and a female English speaker, the reference speakers for the Speech Maps project (Abry, Badin and Scully 1994), uttering the syllable /pœ/ have been used. The material was pronounced with large voice quality variations: breathy, pressed, creaky, loud, soft voice and low and high F0. The female speaker was recorded by C Shadle at DECS in Southampton. The speech signal was recorded under Hi-Fi conditions. The material was sampled at 16 kHz. The recordings of the male speaker was done at ICP in Grenoble under similar conditions and was digitised at 12 kHz at one session, 16kHz at another.

## Inverse filtering

The material was inverse filtered to obtain the glottal source waveform. Automatically measured formant frequencies were fine tuned by hand, using both time and frequency representations of the filtered and the original signal, and zeros and poles were identified and added. For the male speaker, six formants were cancelled up to 6 kHz for the material sampled at 12 kHz and eight formants up to 8 kHz for the material sampled at 16 kHz. For the female speaker, seven formants were cancelled up to 8 kHz. A maximum of two zeros and poles were identified. The analysis was made for one glottal pulse at a time. For each pulse, the LF parameters (Fant, Liljencrants and Lin 1985) were decided by fitting the LF model to the inverse filtered source representation both in the time and the frequency domain. The analysis programmes were written by J Liljencrants.

## Results

The speech material has not been perceptually judged. We are therefore unable to quantify the speakers' accuracy in producing the intended voice qualities. From the results it is obvious that in some cases the two speakers disagree on how a particular voice quality should be produced. This is particularly apparent for the voice qualities pressed and high F0.

The results from the inverse filtering are shown in Table 1. Only a few points will be discussed.

### *Formants and bandwidths*

There are no significant differences in formant values between the different voice qualities. Overall, the bandwidths for the female speaker are somewhat higher. Both speakers show higher bandwidths for breathy voice, and the female speaker also for low level. In both these qualities a glottal leak can be suspected.

### *Fundamental frequency*

The F0 span for the two speakers are fairly similar, with one notable exception: the high F0 for the male speaker is more than one octave higher than his normal voice, while for the

female speaker the difference is only about half an octave.

### *LF parameters*

The parameters used are demonstrated in Fig. 1.

The 'skewness factor' RK is defined as the time from peak flow to excitation divided by the time from opening to peak flow. This means that if the excitation occurs early during the closing time, you can get a low RK for a symmetrical pulse. This can explain the low RK values for the male soft voice and high F0 voice. The high level and the pressed voice for the female speaker show a higher RK than found earlier (Karlsson 1992).

### *Amplitudes*

The amplitude was measured in two ways: on the speech signal giving a value of overall amplitude, A0, and on the inverse filtered waveform giving the excitation strength, EE in the LF model (Fig. 1). It is only possible to compare values within speakers, as some recordings were not calibrated.

The two speakers show slightly different behaviours; the male speaker shows a large variation in EE compared to the female speaker, while in A0, the pattern is reversed. For the male speaker the variation span is about the same in EE and A0; for the female speaker the variation in EE is far smaller than in A0.

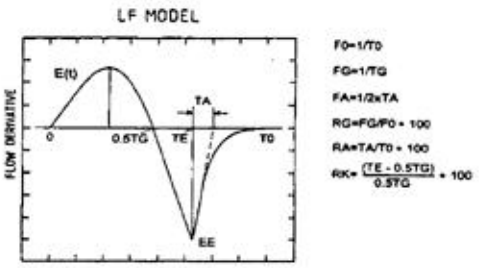


*Figure 1. The LF voice source model. The parameters discussed in the study are indicated in the figure.*

### *Open quotient*

The open quotient (OQ), is here defined as TE+TA in percent of the whole pulse. Again we can see a difference between the two speakers, the female speaker have a longer open quotient, about 10%, than the male. This female-male difference has been found in many studies (Klatt and Klatt 1990, Trittin and de Santos y Lleó 1995, Gobl and Karlsson 1991).

## Creaky voice

The male speaker recorded creaky voice quality together with normal, and low F0 utterances. From this session, samples of eight vowels have been investigated.

*Table 1. F0, FA and FG are given in Hz, EE and A0 in dB, and OQ, RE, RA, RG and RK in percent. Please observe that EE and A0 can only be compared within a speaker, as the absolute sound pressure level is not known for the male voice recordings.*

| | quality | F0 | RE | RA | RK | RG | FG | FA | EE | A0 | OQ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| male | normal | 126 | 52.0 | 2.02 | 37.7 | 132 | 167 | 1001 | 65 | 77 | 54 |
| female | | 246 | 65.8 | 4.96 | 51.7 | 115 | 284 | 810 | 60 | 66 | 71 |
| male | low F0 | 102 | 58.2 | 2.58 | 42.5 | 123 | 125 | 639 | 62 | 74 | 61 |
| female | | 190 | 70.5 | 5.10 | 41.9 | 101 | 192 | 603 | 60 | 61 | 76 |
| male | medium F0 | 131 | 53.4 | 1.50 | 45.0 | 136 | 178 | 1485 | 65 | 76 | 56 |
| female | | 250 | 66.9 | 4.17 | 48 | 111 | 277 | 966 | 61 | 68 | 71 |
| male | high F0 | 288 | 77.0 | 9.84 | 32.1 | 86 | 247 | 478 | 65 | 79 | 87 |
| female | | 360 | 61.6 | 2.95 | 48.7 | 121 | 436 | 1946 | 64 | 75 | 65 |
| male | low level | 129 | 66.3 | 2.72 | 40.7 | 109 | 137 | 785 | 61 | 72 | 69 |
| female | | 249 | 70.1 | 10.50 | 57.1 | 112 | 279 | 379 | 58 | 56 | 81 |
| male | medium level | 127 | 55.5 | 1.86 | 45.0 | 134 | 166 | 1106 | 64 | 76 | 57 |
| female | | 258 | 64.7 | 3.71 | 51.2 | 117 | 303 | 1244 | 62 | 68 | 68 |
| male | high level | 132 | 47.1 | 1.56 | 37.7 | 145 | 183 | 1477 | 68 | 80 | 49 |
| female | | 257 | 61.9 | 1.93 | 52.2 | 123 | 316 | 2123 | 65 | 74 | 64 |
| male | breathy | 131 | 60.5 | 4.57 | 51.0 | 125 | 164 | 461 | 59 | 67 | 65 |
| female | | 254 | 71.0 | 8.12 | 48.3 | 105 | 266 | 503 | 61 | 65 | 79 |
| male | pressed | 128 | 39.2 | 1.31 | 39.5 | 185 | 229 | 1620 | 55 | 72 | 41 |
| female | | 261 | 67.6 | 3.24 | 49.9 | 111 | 289 | 1290 | 61 | 64 | 71 |

*Description of creaky voice*

In creaky voice consecutive glottal pulses are usually different from each other (see Fig. 2). The values for the glottal parameters vary substantially, and a good fit of the LF model is almost impossible. The pulses seem to vary fairly consistently, though, in that every second pulse looks similar but not exactly the same.

*Comparison with other qualities*

There were no differences in formant frequencies and bandwidths between creaky voice and the other qualities. Two pole-zero pairs were identified in the normal voice quality. One of those was found also in the low F0 voice but no zeros were visible in the spectra for the creaky voice. As the glottal pulses in the creaky quality deviated considerably from the LF model, it is hard to compare glottal parameter values. The EE parameter is easily identified, though. For the creaky voice, EE varies about 10 dB between pulses and the stronger pulses are about 6 dB weaker than for the normal and low F0 voices. FA has similar values for the three qualities.

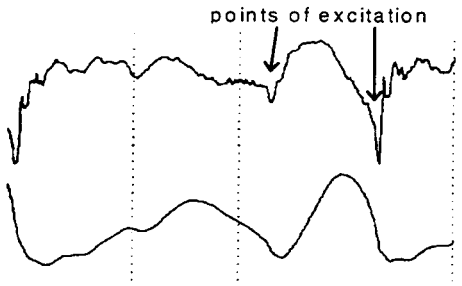


*Figure 2. Voice pulses for creaky voice. The upper curve depicts the inverse filtered speech signal, the lower the glottal air flow, retrieved by integrating the inverse filtered speech. As can be seen from the upper curve, the excitations are very different in amplitude*

# Modelling

The different voice qualities produced by the male speaker have been modelled using Liljencrants' vocal cord model (Liljencrants 1996, Karlsson et al. 1994). Different model parameters were varied to get good fits between the glottal flow retrieved by inverse filtering of the natural waveform and the glottal flow produced by the model, both in the time and the frequency domain. Care was taken that the parameter variations were physiologically feasible and in accordance with known facts of the different voice qualities.

Glottal waveforms for the different voice qualities produced by our male speaker and the model are shown together in Figure 3.

*Model parameters*

Only parameters that were found to be crucial for changing the voice quality will be discussed here.

The model parameters can be divided into different groups depending on origin. Some parameters pertain to the sub- and supra-glottal cavities: the area of the trachea, the lowest tracheal resonance, the first formant of the oral cavity, the damping factor in the trachea, the damp factor of the oral cavity and the subglottal pressure. There are also parameters describing the properties and the state of the vocal folds: vocal fold length, the gyration radius, the damping of the movements during the time the glottis is open, the damping when the folds close, the longitudinal tension in the vocal folds, the quotient between the adduction force and the subglottal pressure, the rotary to translational resonance frequency ratio, the width of the remaining glottal opening in the closed part, the maximal width of the remaining abduction angular gap, and the angle of the deviation between the upper and lower parts of the vocal folds when the lower parts close.

*Parameter settings and discussion*

The vocal fold length is constant except for creaky voice where they are longer.

The gyration radius, that is the tendency for the vocal folds to rotate, does not need to be altered to vary F0 or level. It is an important factor though for producing the more extreme qualities. For breathy voice it was increased, resulting in a lessened rotation tendency. For the pressed and the creaky qualities it was decreased.

The damping during the open part was increased for the varying F0 and the pressed quality, decreased for the creaky voice. The damping at closure was largest for the creaky voice and also large for the low level voice. This indicates that a large part of the energy is lost in the collision. For the more stiff vocal folds in pressed and breathy voice, more energy is carried over into the next period.

The pressure parameter is primarily used to regulate loudness. It was also necessary to raise it for the high F0 and the pressed qualities. Both these qualities were relatively loud in the natural samples (see Table 1). The modelled high F0 has come out a bit stronger than the natural (see Figure 3), but the pressure is still very high

compared to, for example, the high level sample. Note also the low pressure in the creaky voice.

The vocal fold tension is used primarily to regulate the fundamental frequency, the high F0 value is considerably higher than the rest of the values. It is noteworthy that the tension parameter is low for both the pressed and the creaky voice. The quotient between the adduction force and the subglottal pressure is equal to 1 when the folds are neither pulled apart nor pressed together. A value larger than 1 means that the folds are pulled apart. This parameter is especially important for the production of a pressed voice, where the folds are pressed together, and creaky voice, where they are pulled apart. For the creaky voice this is counteracted by the negative value for the residual gap parameter. Otherwise, non-complete closure is needed to model the high F0 voice and breathy quality where a triangular constant opening is specified.

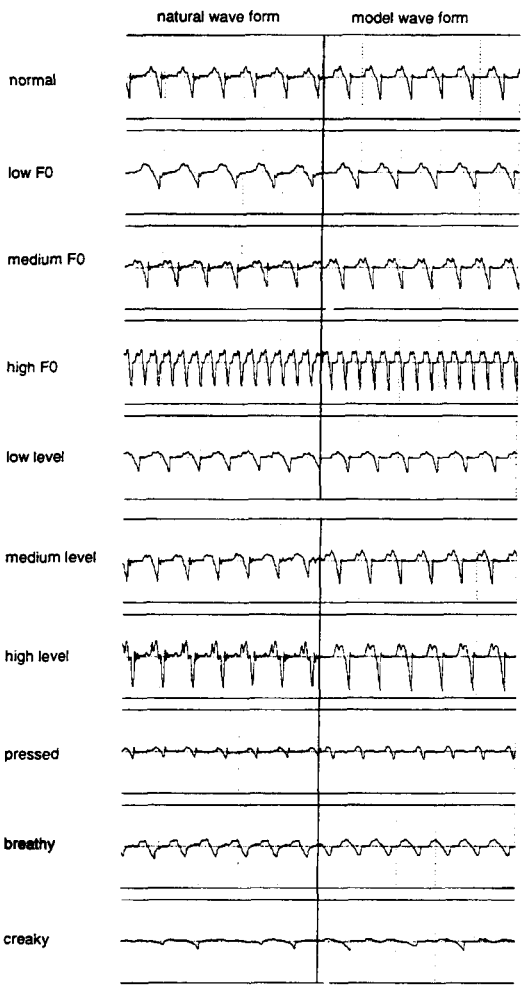


*Figure 3. Glottal wave forms for different voice qualities. The inverse filtered natural speech is shown on the left, the wave form produced by the glottal model on the right.*

The frequencies and q-values of the lowest resonance above (that is F1) and below the glottis are crucial for achieving a good match in the open part of the glottal pulse. The relation between the resonances also determines the position of the zero pole pair in the glottal wave form spectrum. For our male speaker this zero is found at about 900 Hz and the pole at about 1600 Hz.

The modelled glottal wave forms have been used to produce synthetic vowel samples that have been compared to the intended natural voice qualities. In a very informal listening test, the results were judged to be very good in nearly all cases.

## Acknowledgements

This research has been supported by ESPRIT BR no. 6975 Speech Maps, the Swedish Technical Research Council (TFR) and KTH.

## References

Abry C, Badin P & Scully C (1994). Sound-to-gesture inversion in speech: The *Speech Maps* approach. In: Varghese K, Pfleger S & Lefèvre JP, eds, *Advanced speech applications*, Berlin: Springer Verlag, 182-196.

Fant G, Liljencrants J & Lin Q (1985). A four-parameter model of glottal flow. *STL-QPSR* 4/1985: 1-14.

Gobl C & Karlsson I (1991). Male and female voice source dynamics. In: Gauffin J & Hammarberg B, eds, *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, San Diego: Singular Publishing Group Inc, 121-128.

Karlsson I & Liljencrants J (1994). Wrestling the two-mass model to conform with real glottal wave forms. *Proc ICSLP 94*, 1: 151-154

Karlsson I (1992). Modelling voice variations in female speech synthesis. *Speech Communication*, 11: 491-495.

Klatt D & Klatt L (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *JASA*, 87: 820-857.

Liljencrants J (1996). Analysis by synthesis of glottal airflow in a physical model. *Ibid.*

Trittin P & de Santos y Lleó A (1995). Voice quality analysis of male and female Spanish speakers. *Speech Communication*, 16: 359-368.