

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

**Glottal waveform parameters
for different speaker types**

Karlsson, I.

journal: STL-QPSR
volume: 29
number: 2-3
year: 1988
pages: 061-067



**KTH Computer Science
and Communication**

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

C. GLOTTAL WAVEFORM PARAMETERS FOR DIFFERENT SPEAKER TYPES*

Inger Karlsson

Abstract

Dynamic variations of the glottal air flow have been studied for several normal speakers. For seven women, a short utterance was inverse filtered and the dynamic glottal parameters defined by the LF-model (Fant, Liljencrants, & Lin, 1985) were measured from the differentiated glottal air flow. Also, for the same seven women, the constant air leakage and the peak air flow of the glottal pulse were measured in an isolated syllable. These seven women's voices have been classified by a speech therapist. The influence of stress on the constant air leakage and peak glottal pulse air flow were studied in a sentence where the stress pattern was varied. Two men and two women served as informants in this experiment. The different glottal waveform parameters have been compared to the classification of the speakers.

1. Introduction

Different aspects of the voice source have been studied for speakers varying in voice types. The aim of the study has been to describe variation in normal, and especially normal female voices. Two different recording techniques have been utilized. Calibrated air flow has been recorded using a Rothenberg mask; the mask is described in Rothenberg (1973). When it was possible, the subglottal pressure (=oral pressure in /p/) was registered. Hi-fi-recordings of speech have been made and the speech pressure wave subsequently inverse filtered to provide the dynamic glottal parameters of the LF-model, Fant & al. (1985).

Differences in voice types such as "breathy", "dark", "normal", etc. will be described using several different glottal waveform parameters such as constant glottal leak, dynamic leak, open quotient, subglottal pressure, peak flow, etc., and also spectral properties of the unfiltered vowels.

2. Recordings and analysis

The study can be separated into three parts. In the first experiment seven normal-speaking women were recorded under Hi-fi conditions. Care was taken to get a phase-true registration of the speech wave. The recordings were used to study the variations of the dynamic glottal parameters. In the second experiment, the same seven female speakers were recorded using a Rothenberg mask to register the air flow from the mouth. At the same time the oral pressure was registered with a pressure transducer connected to a tube with its opening in the mouth cavity. The speech material consisted of syllables and sentences. In the third experiment two males and two females were recorded with a Rothenberg mask and oral pressure tube. The speech material contained a sentence that was pronounced with four different stress patterns.

*Revised version of a paper given at the SPEECH'88 7th FASE Symposium, Edinburgh 22-26 August 1988.

2.1 Experiment 1: Dynamic glottal parameters

Seven women, all without any known voice problems, served as informants. They ranged from speakers with no voice training to a speech therapist and amateur choir singers. Typical examples of fundamental frequencies and formant frequencies and bandwidths for all seven speakers are given in Table 1. The women were recorded under Hi-fi conditions. At this session the informants read a sentences list and an excerpt from a novel. The novel excerpt has been used by a speech therapist to judge and classify the different speakers. For each speaker a rendering of the sentence "ja adjö." (IPA: ja: ajø:) has been inverse filtered using an interactive computer program. After inverse filtering the bandwidth of the signal was 25-4000 Hz. The dynamic glottal parameters used in the LF-model were measured through the whole utterances from the inverse filtered signal, that is from the glottal pressure wave. The LF-model parameters are demonstrated in Fig 1. The mean values of these dynamic source parameters for a 35-msec part of each of the two long vowels /a/ and /ø/ for each speaker are given in Table 2.

Speaker	F ₁	F ₂	F ₃	F ₄	B ₁	B ₂	B ₃	B ₄	F ₀
W1	640	1110	2850	3750	270	170	200	200	207
W2	600	1350	2950	4150	160	120	150	210	230
W3	650	1150	2700	3650	400	200	110	90	198
W4	700	1150	2550	3650	200	60	100	120	199
W5	700	1200	2400	3550	210	190	230	120	188
W6	650	1150	2550	3850	160	100	170	80	204
W7	750	1200	2800	3750	290	290	100	270	237

Table 1. Formant frequencies and bandwidths for the long /a/ vowel and mean fundamental frequencies for three vowels for the seven women informants.

2.2 Experiment 2. Glottal air flow

The same seven women as in Experiment 1 participated in this experiment. The informants read a list containing isolated syllables and sentences. They were asked to pronounce the syllables slowly. The recordings were done with a Rothenberg mask. The oral pressure was recorded simultaneously with the air flow. So far, examples of the syllable /pa:/ have been inverse filtered for each speaker. After inverse filtering the bandwidth of the signal was 0-1500 Hz. The peak flow and the constant leakage flow in the most closed period of the glottal cycle, called dc flow, have been measured from the inverse filtered signal, the glottal air flow. An example of flow measurements are shown in Fig. 2. The means over four consecutive fundamental periods have been calculated. Care was taken not to measure during the beginning of the vowel where articulatory movements can cause air flow. The subglottal pressure was measured in the preceding /p/'s. As /p/ in Swedish is aspirated, the oral pressure during the occlusive part of this consonant is equal to the subglottal pressure.

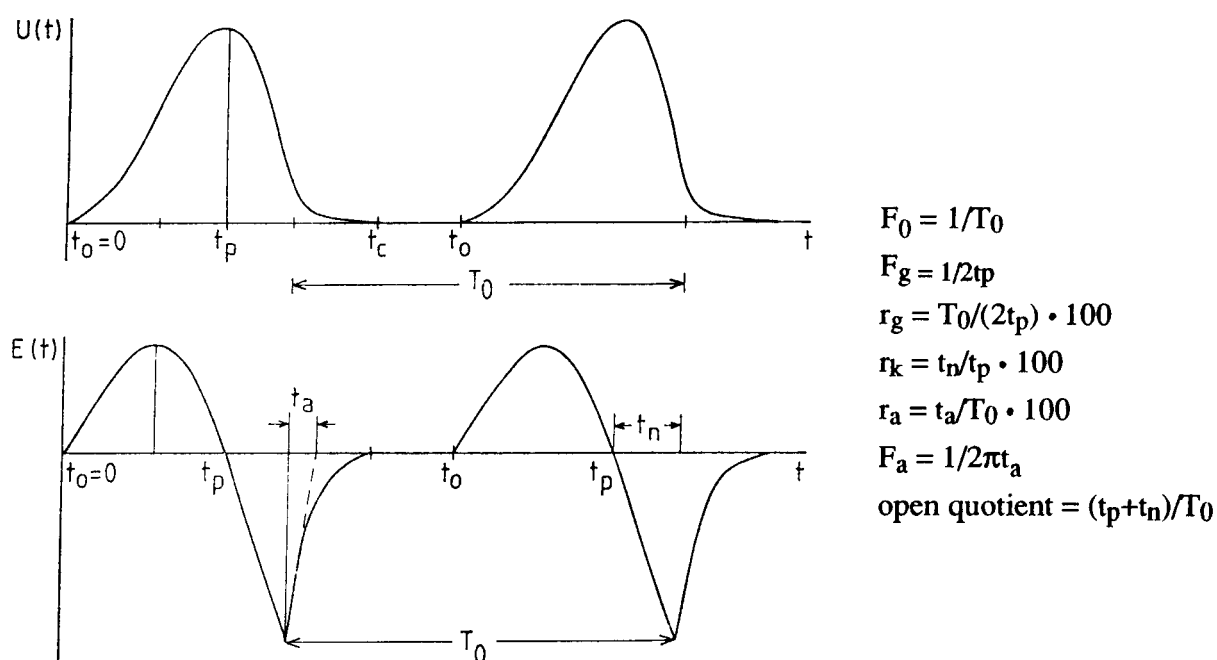


Fig. 1. Dynamic source parameters according to the LF-model (Fant & al., 1986). The abbreviations used in the paper are explained to the right.

Speaker	Vowel	F_0	F_g	r_g	r_k	r_a	F_a	open quotient
W1	a:	215	249	115	39	5	685	60
	ø:	180	176	97	27	5	575	65
W2	a:	234	260	111	44	11	340	65
	ø:	207	250	121	41	13	255	58
W3	a:	208	200	89	24	11	300	70
	ø:	214	271	126	35	6	570	54
W4	a:	178	170	95	24	15	190	65
	ø:	212	215	101	24	16	210	61
W5	a:	177	165	92	25	7	400	68
	ø:	221	270	122	43	9	390	59
W6	a:	198	164	83	22	9	350	71
	ø:	212	215	101	29	7	480	64
W7	a:	202	179	88	19	12	270	68
	ø:	301	312	103	29	12	400	63

Table 2. Dynamic source parameters for all seven speakers. r_g gives F_g in percent of F_0 , r_k is the skewness factor and denotes the time from peak flow to excitation (=maximum closing rate) over the opening phase of the glottal period and r_a , the dynamic leak, is the time between excitation and complete closure in percent of the whole glottal period. F_a , given by T_0 over $2\pi r_a$, denotes the frequency above which an extra 6 dB/octave is subtracted from the spectrum.

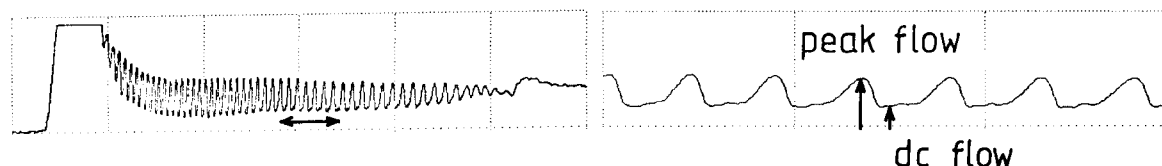


Fig. 2. Glottal air flow in the syllable /pa:/. The right part is a time enlargement of the section that is marked in the left part. This is the part where the measures were made. Peak flow and dc flow are indicated in the right part.

Speaker	dc flow	peak flow	dc/peak flow	subglottal pressure
W1	60	200	0.30	5.7
W2	100	220	0.45	6.2
W3	130	330	0.41	9.8
W4	70	150	0.46	7.4
W5	70	220	0.33	3.6
W6	40	150	0.29	7.3
W7	100	170	0.60	5.6

Table 3. Glottal air flow in ml/sec and subglottal pressure in cm H₂O for seven female speakers.

The different air flow measures together with the subglottal pressure for each speaker are given in Table 3. As can be seen from Table 3, all speakers showed a dc flow. This agrees well with findings by Södersten & Lindestad (1987) in a study of women's vocal fold movements during phonation using fibre optics.

2.3 Experiment 3: Stress variation

Two females and two males participated in this experiment. One speaker of each sex was a church choir singer; the two remaining speakers had no singing or speech training. One of the sentences they were asked to say was /papa ka:par aspar/ (=father cuts aspens.). This sentence was read with four different stress patterns, neutral stress and with focus on each of the three words. The test sentence is not well suited for an investigation of absolute amount of constant air leakage as the consonants are aspirated. This will presumably raise the dc flow. The interest here, though, has been in comparing behaviour among speakers which makes the absolute amount of dc flow of less interest. The consonant /p/ also makes it possible to measure the subglottal pressure and gives a good indication of zero flow. The air flow wave was later inverse filtered and for each vowel the peak and dc flow in the mid-portion was measured as an average over four periods. The oral pressure in the /p/ consonants was measured in the same utterances. The air flow values for all four sentences are shown for one male speaker in Fig 3. As can be seen, no consistent differences could be detected between stressed and unstressed variants of the same word. This was true for all four speakers. Accordingly, the mean peak flow and percent dc flow were calculated and are given in Table 4.

Sentence:...	pa-	ppa	ka:-	par	as-	par
Speaker						
MaleT	475/.41	530/.50	340/.31	320/.20	450/.46	180/.01
MaleS	410/.31	400/.41	270/.22	200/.18	380/.46	300/.23
FemaleT	370/.34	420/.55	190/.13	200/.20	240/.48	110/.40
FemaleS	370/.41	330/.35	220/.16	280/.18	340/.39	200/.29

Table 4. Mean peak flow and dc flow quotient in sentences. Peak flow in ml/sec is given to the left of the slash (/), and dc/peak quotient to the right.

The subglottal pressure shows, for three of the speakers, a tendency toward higher values in the focally stressed words while the fourth speaker had the same subglottal pressure for all renderings of the same word, see Table 5. This speaker, the female choir singer, showed the highest average subglottal pressure even though the variation between speakers was small.

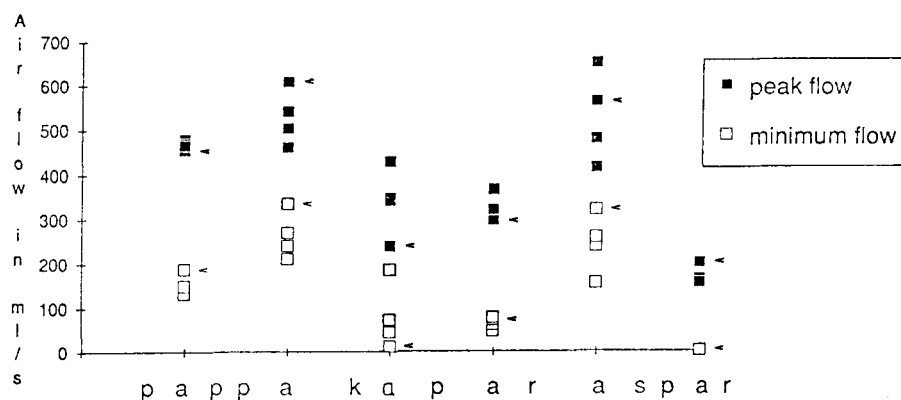


Fig. 3. Peak and dc flow in the sentence /pappa kɑ:par aspar/ for a male speaker, MT. The vowels in the focally stressed words are marked with <.

Speaker:	MT	MS	FT	FS
Word:				
pa-	6.4/5.8	3.9/5.3	6.8/4.6	6.6/6.7
ppa	7.4/7.4	7.6/5.9	7.1/6.4	8.2/8.7
kɑ:par	6.9/6.2	7.1/6.4	6.6/6.4	7.3/7.5
aspar	7.1/6.0	5.6/4.9	6.6/5.4	6.2/6.1

Table 5. Subglottal pressure, in cm H₂O, in stressed and unstressed words. The value for the focally stressed word is given to the left of the slash (/), the mean value for unstressed words to the right.

3. Speaker classification

3.1 Speakers in Experiments 1 and 2

The voices of the seven women that were recorded in Experiments 1 and 2 have been judged by a speech therapist. The sentence that was inverse filtered and the novel excerpt, which was between 45 and 70 sec long, was used as speech material for the classification. All seven speakers were classified as normal speakers without voice problems, so the given judgements should be read as if preceded by "a tendency towards".

The results were:

Speaker W1: Normal, somewhat tight, sonorous.

Speaker W2: Thin, not breathy, lacking sonority, young, high pitch.

Speaker W3: Dark, slightly coarse, sonorous, swollen vocal cords.

Speaker W4: Thin, slightly breathy, restrained, lacking sonority.

Speaker W5: Dark, sonorous.

Speaker W6: Not breathy, grating, creaky.

Speaker W7: Thin, breathy without turbulent noise, lacking sonority, high pitch.

3.2 Speakers in Experiment 3.

The two men have similar voices. They can both be described as having ordinary, non-leaky voices of middle range. One, MS, has some choir singing experience and his voice is slightly more sonorant than MT's, the other male. One female, FS, the choir singer who is also a speech therapist, has a sonorant voice of middle range. The other woman, FT, has a somewhat breathy, husky voice.

4. Result and discussion

4.1 Results from Experiments 1 and 2.

One aim of these experiments was to try to find correlations between the perception of certain voices and the glottal waveform parameters. The speakers described as "dark," W3 and W5 have low fundamental frequency and formant values, while W2 and W7, who have high F_0 and formants, are described as having high pitched voices (see Table 1). Two of the voices, W4 and W7, were described as breathy. Both show a high amount of dc flow (see Table 3) and high value for the dynamic leak, r_a (see Table 2). W7, who is said to be breathy with no turbulent noise, has an especially high amount of dc flow. W2 also has high values for both types of leak, but she is not judged to have a breathy voice. The parameter where a difference can be detected between W4 and W7 on one hand and W2 on the other is the skewness measure, r_k . W2 has a higher value for this parameter. All three speakers were classified as lacking in sonority, which accordingly seems to correlate with the leak parameters, while a low r_k value seems to be needed to give a breathy impression. A low r_k value could mean that the excitation occurs early in the closing phase when the glottis still has a large opening. The speaker classified as somewhat tight, W1, has least dynamic leak. The coarse voice, W3, shows higher values for subglottal pressure and for both peak and dc air flow than do the rest of the speakers. The grating voice, W6, has a high skewness factor, r_k , and low formant bandwidths.

4.2 Experiment 3.

In Experiment 3, the relation between stress and both glottal air flow and subglottal pressure was studied. There were clear differences among the speakers in the amount of average glottal air flow: the two males show higher values, while the relative amount of dc flow seem to be similar for all speakers. A difference can be seen between the breathy female voice, FT, and the other speakers. In the last vowel, FT has a higher de-

gree of dc flow. This might be one of the features of the perceived breathiness in her voice. The subglottal pressure did not vary much among these speakers, no difference could be detected between the males and the females.

4.3 Further work.

One important aspect of the glottal waveform parameters has not been touched upon in this paper. The glottal wave also contains noise to a varying degree. It is well known that the amount of noise in the glottal pulse is an important factor for the perceived naturalness of especially a synthetic female voice (see Klatt, 1987). It is mainly the higher formants, F3 and higher, that are sometimes excited by noise. The noise can be seen in a partially inverse filtered signal where only one formant remains. The problem of how to describe and quantify the noise still remains to be solved.

References

- Fant, G., Liljencrants, J., & Lin, Q. (1985): "A four-parameter model of glottal flow", STL/QPSR 4/1985, pp. 1-13.
- Klatt, D. (1987): "Acoustic correlates of breathiness: First harmonic amplitude, turbulent noise, and tracheal coupling", J.Acoust.Soc.Am. **82**, Suppl.1, p. S91 (A).
- Rothenberg, M. (1973): "A new inverse filtering technique for deriving the glottal air flow waveform during voicing", J.Acoust.Soc.Am. **53**, pp. 1632-1645.
- Södersten, M. & Lindestad, P.-Å. (1987): "Vocal fold closure in young adult normal-speaking females", Phoniatrik and Logopedic Progress Report 5, pp. 12, Dept of Logopedics and Phoniatics, Huddinge University Hospital, Karolinska Institutet, Stockholm.