DYNAMIC VOICE QUALITY VARIATIONS IN FEMALE SPEECH

Inger Karlsson

Dept of Speech Communication and Music Acoustics, KTH, Box 70014, S 100 44 Stockholm, Sweden

ABSTRACT

Variations in the voice source for female speakers due to linguistic structure and speaker specificity have been investigated. The study is focused on consonants and transitional segments. The voice source have been analysed by inverse filtering. The consonant source spectra contained less energy in the higher frequency region compared to vowels. For a more leaky voice, transitional segments contained a large amount of noise. Occurrences and origins of zeros in the spectra of voiced speech segments were studied using inverse filtering. For a leaky voice a zero due to the incomplete glottal closure often occurred also in vowels.

1. INTRODUCTION

This study forms part of a project aimed at a complete description of female speech. The investigations have so far been concentrated on the female voice source. Information has been collected about the relationship between emphatic stress and voice source parameters [2] and about voice source variations with place of articulation of vowels [3]. The present study is focused on a description of consonants and transitions between voiced phonemes. Furthermore, the occurrence and origin of zeros in the spectra of voiced speech segments have been investigated.

The voice source was analysed by inverse filtering of the speech wave. A subsequent fitting of the LF voice source model [1] to the inverse filtered wave gave a parametric description of the voice source variations. The voice source parameters used in this study are RK, RG, EE, FA and F0. RK corresponds to the quotient between the time from peak flow to excitation and the time from zero to peak flow. RG is the duration of the glottal cycle divided by twice the time from zero to peak flow. RG and RK influence the amplitudes of the lowest harmonics and are expressed in percent. EE is the excitation strength in dB and FA the frequency above which an extra -6dB per octave is added to the spectral tilt. In addition, the fundamental frequency, F0, is measured.

2. DYNAMIC VOICE SOURCE PARAMETER VARIATIONS.

The present study concentrates on dynamic variations of the voice source. The rate of change of voice source parameters and how these changes correlate with segments and segment boundaries were investigated. For transitions between segments, especially between vowels and occlusive segments, both rate of change and the timing of changes are of crucial importance. In a transition between a vowel and an [l] or a nasal, the voice source parameter values change from typical vowel to consonant values within a few voice pulses. A transition between a vowel and [v] or [j] is much more gradual.

Correlations between the different voice source parameters have also been investigated. RG showed a fairly good correlation with F0 in sentences uttered by different speakers. The correlation coefficient was found to be in the order of 0.75. Deviations occurred for F0 peaks where RG was raised even more, see Figure 1. The remaining parameters did not show any substantial correlation with each other, the variations were more related to phoneme type and prosody. RK showed a large pulse-to-

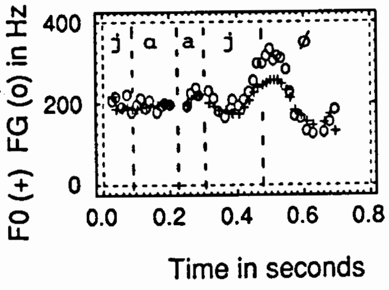

Figure 1. Typical covariation of F0 and FG=F0·RG/100 for a short utterance.

pulse variation. This is due to the uncertainty in defining the exact time for opening. Another source of error is formant frequency differences in the open and the closed interval of the voice pulse. In inverse filtering only one formant value was used in each period. This will result in incomplete cancelling of formant ringings in the open phase and some lack of precision in determining the point of maximum flow. All RK values discussed in this paper are average values, which should minimize these errors.

3. VOICE SOURCE IN CONSONANTS

Voiced consonants in sentences have been inverse filtered, when possible, to achieve a source description. The investigated sentences contained the stops [b, d, g], the voiced fricatives [j, v] that both contain very little noise between voiced phonemes in Swedish, the nasal [n], the sonorants [l, r] and voiced [h]. It was often impossible to inverse filter the stops as they were too weak compared to the background noise. Accordingly, only one joint value was calculated for the stops. [r] was realized as a vowel-like segment in the studied sentences and had voice source characteristics similar to an unstressed vowel. To get a good fit between the LF voice source model and the inverse filtered wave-form for the remaining consonants, it was often necessary to cancel an extra pole/zero pair, especially in [n], [l] and [v].

Voice source parameters for consonants and for some unstressed vowels in the same sentences are given in Table 1. The values are averaged over at least 4 period. Compared to vowels in the same sentence the consonants tend to have higher RK values, i.e. more energy in the lowest harmonics. The excitation amplitude, EE, was slightly lower for [r, j, n, h] than for vowels. For [v] and the stops, EE was often 10 dB weaker, the stops showed a rapid fall in EE through the sound. FA showed considerably lower values for all consonants with the exception of [r], [h] and for one speaker [j]. A possible reason for the high FA is discussed below under "Noise excitation". FA was only slightly higher than F0 for the remaining consonants. This means that the voice source contains less high frequency energy for these consonants than for vowels.

Table 1. Voice source parameters for voiced consonants and unstressed vowels for two female speakers. The last column gives the number of occurrences of the phoneme in the investigated sentences. F0 and FA are given in Hz and RK and RG in percent. EE is given in uncalibrated dB so only comparisons within a speaker is possible.

speaker W1			consonants			
	F0	EE	RK	RG	FA	n
[v]	185	50	48	109	215	1
[l]	191	55	48	116	356	8
[n]	270	53	51	141	297	1
[r]	253	57	42	120	657	1
[j]	161	56	43	101	849	1
[h]	216	57	51	107	492	1
stop	232	49	58	101	354	3
unstressed vowels						
[a]	212	61	45	124	867	4
[I]	218	55	38	104	500	2

speaker W2			consonants			
	F0	EE	RK	RG	FA	no
[v]	237	48	59	119	362	2
[l]	209	57	48	111	322	8
[n]	186	56	37	104	208	1
[r]	222	57	43	115	678	3
[j]	196	57	43	100	255	1
[h]	250	57	72	123	600	1
stop	220	51	66	112	343	2
unstressed vowels						
[a]	213	62	35	96	763	4
[I]	277	60	28	90	480	1

4. VOICE SOURCE ZEROS

Zeros in voiced speech segments can have different origins. They are either a personal trait, often due to a leaky voice source, or a segment related feature, especially in consonants, where it is due to the configuration of the vocal tract. Both these types of zeros have been investigated.

4.1 Zeros in consonants.

The investigated sentences contained consonants whose transfer functions contained zeros: [l] and [n]. For [l] and [n] the zero and the connected pole are normally due to the geometry of the vocal tract. Zero/pole pairs found in [l] and [n] for two female speakers are given in Table 2.

The zero sometimes detected in [v] as well as a low zero, about 900 Hz, sometimes found in [l], is presumably due to a more leaky voice source and consequently a coupling to the subglottal system in these consonants. This could be due either to an overall leaky voice or to a personal variation for these particular sounds. These zero/pole pairs are also listed in Table 2.

4.2 Voice source zeros in vowels

Normally, while inverse filtering vowels, only anti-formant filters cancelling the vocal tract resonances were used. For more leaky voices an additional pole/zero pair often had to be cancelled to achieve a good fit to the LF-model. The origin of this pole/zero pair is presumably a coupling to the subglottal system as for some consonants discussed above. The speakers who showed a zero/pole pair had a comparatively large amount of constant air flow during phonation in recordings with a Rothenberg mask [5]. This implies an incomplete vocal cord closure and a coupling between the sub- and the supraglottal cavities. The frequency values of the pole/zero pair, a zero at about 800 Hz and a pole at about 1500 Hz, compares well with known values for subglottal poles and zeros for women [4]. In Figure 2 an example is shown of a vowel that has been inverse filtered using or not using an extra zero/pole pair.

5. NOISE EXCITATION

In inverse filtering and model fitting the model parameters tend to include the noise excitation since the inverse filter time window is one fundamental period. Accordingly, in a spectral section, no harmonics are visible and it is impossible to separate voice and noise excitation. This means that often a breathy segment will give quite high FA values contrary to theory. The high FA-values for [h] and [j] in Table 1 are presumably due to this effect. To avoid this type of error, spectrograms of the utterances were studied. When a simultaneous voice and noise excitation could be suspected, partial inverse filtering was applied: all formants except one were damped out. The excitation pattern of the remaining formant showed if noise was a major excitation source. In Figure 3 an example of measured FA variations for a breathy voice and a more sonorant voice are shown. FA is highest during the transition from consonant to vowel for the breathy voice while FA is higher in the vowel for the more sonorant voice. The high FA values during the transition for the breathy voice turned out to be due to high noise content. We are presently trying to find a method to separate the two kinds of vocal tract excitations, this will be discussed further at the congress.

Table 2. Zeros and corresponding poles in the voice source and vocal tract transfer function for some consonants. Zeros and poles are measured by inverse filtering. Zn denotes a zero frequency and BZn its bandwidth. Pn denotes the corresponding pole and BPn its bandwidth. All are given in Hz. * denotes presumed voice source zeros and poles.

		Z1	BZ1	Z2	BZ2	P1	BP1	P2	BP2
W1	[l]	*940	300	2050	300	*1450	150	2500	200
W2	[l]	*990	450	2200	300	*1600	200	2550	200
W1	[n]	750	450			1900	100		
W2	[n]	860	150			1700	100		
W1	[v]	*700	250			*1650	150		
W2	[v]	*840	350			*1600	200		
W1	[h]	1850	600			3000	500		
W2	[h]	2300	600			2600	450		

6. ACKNOWLEDGEMENTS

This project has been supported in part by grants from the Swedish Board for Technical Development (STU) and Swedish Telecom

7. REFERENCES

[1] FANT, G., LILJENCRANTS, J & LIN, Q. (1985): "A four-parameter model of glottal flow", *STL/QPSR 4/85*, 1-13

[2] GOBL, C. & KARLSSON, I. (1989): "Male and female voice source dynamics." *Proc. of Vocal Fold Physiology Conference, Stockholm.*

[3] KARLSSON I. (1990): "Voice source dynamics for female speakers" *Proc. of the 1990 Int. Conf. on Spoken Language Processing, Kobe.* 69-72

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

[5] ROTHENBERG M (1973): " A new inverse filtering technique for deriving the glottal air flow waveform during voicing", *JASA 53*, 1632-1645

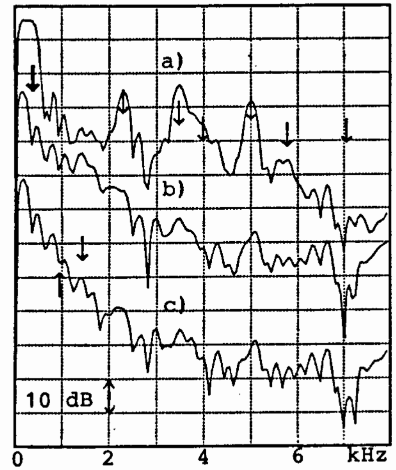

Figure 2. Spectra for one fundamental period of a vowel before and after inverse filtering. From top to bottom the spectrum of a) an unfiltered voice pulse, b) the same pulse with the formants cancelled and c) with an extra pole/zero pair cancelled as well. The formant anti-filters are marked by down-pointing arrows in the upper part of the figure and the extra pole/zero with arrows in the lower part.

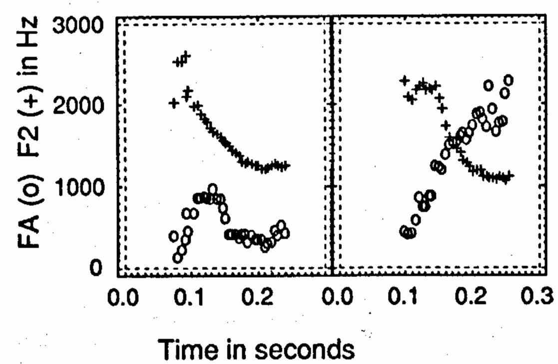

Figure 3. FA variations in a transition from [j] to [ɑ]. F2 is plotted to illustrate the transition. The left half shows a leaky voice, the right part a more sonorant voice.