ISCA Archive
http://www.isca-speech.org/archive

Models and Analysis of Vocal
Emissions for Biomedical Applications
(MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

ANALYSIS OF INTRAPERIODIC FORMANT MODULATIONS IN SPOKEN AND SUNG VOWELS

Unto K. Laine[*], *Anne-Maria Laukkanen*[**], *and Timo Leino*[**]

[*] Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
PO Box 3000, FIN-02015 Espoo, FINLAND. Unto.Laine@hut.fi
[**] Institute of Speech Communication and Voice Research, University of Tampere, Tampere, Finland.

ABSTRACT

Intraperiodic formant modulations in spoken and sung vowels from four Finnish male opera singers were analyzed by using analytic filtering method and modified covariance type of linear prediction. Both methods revealed clear formant modulations. Differences were seen between spoken and sung samples and between individuals.

1. INTRODUCTION

During the last decade intra-periodic modulation effects in formants have been studied with new mathematical and signal theoretical methods. The so-called truncation phenomenon, i.e. an increased attenuation rate in the amplitudes of the formants, especially that of F1, including a change in their frequencies, has been noted in the open time of the glottis [1][2][3]. This has been explained by pitch-synchronous acoustical changes in the supraglottal resonator: During the glottal closure the supraglottal cavity forms a quarter-wave ($\lambda/4$) type of resonator closed at one end. During the glottal open phase the supraglottal tube is open at both ends and the formant frequencies are shifted. The direction and amount of the frequency shift depends on the acoustic impedance of the subglottal system at the formant frequency.

The quality (amount and rate) of formant modulations bears obviously important information of subglottal, glottal and supraglottal states. Therefore, various modes of voice production differ from each other also in intraperiodic formant modulations. This can be easily seen by comparing spectral representations of vowels produced in normal phonation and in vocal fry. In the latter case the vocal folds are closed most of the time and the subglottal system does not load the supraglottal cavities resulting in high formant Q-factors (a ratio between formant frequency and its bandwidth). The formant frequencies are not modulated. Consequently, the formants are clearly seen in the spectrum. Normal phonation is characterized by a more complicated spectrum with flattened formant peaks and the occurrence of multiple peaks around the formant. Some 'anti-formants' or spectral zeroes may also appear.

There is evidence that the intraperiodic modulations in formants are perceptually relevant [4]. According to Gagné and Zurek [5] a frequency shift of approximately 13 Hz in the first formant and a shift of about 40 Hz in the second formant is already audible. This is not surprising taking that formants have a remarkable role in the auditory perception. Formant modulations may among other things carry important speaker specific information characterizing the personal sound quality.

Detection of intraperiodic events is technically not an easy task. Various methods have been proposed. One novel method has been the use of the Teager-Kaiser energy operator [6]. The operator should be able to give estimates for instantaneous formant frequency and bandwidth. However, it was noted by Laine [7] that the operator works well only when a second order system is in question. Consequently, the formant should be presented alone, without the presence of other formants and without any coupling to the subglottal system. A simpler method based on analytic filtering was proposed with similar results [7].

In the present paper different methods are compared and applied to the analysis of spoken and sung vowels from male opera singers. Operatic singing represents a special vocal technique with strict quality demands. Therefore it is a feasible object in the study of voice production and the acoustical and perceptual characteristics of the human voice.

2. MATERIAL

2.1 Subjects and the samples

Samples from four inter-nationally noted Finnish male opera singers were used as material in the present study. One of the subjects (Subject A) was a bass, two subjects (B and C) were bass-baritones and one (Subject D) was a lyrical baritone. Separately uttered and at a low pitch sung samples of [æ:] were selected from previously recorded material consisting of text reading and vowels spoken separately and sung on a scale. The recordings had been made in a well-damped studio (3,85 x 2,24 x 3,20 cm; reverberation time of 0,4 sec) with a digital recorder (Technics Digital Audio Cassette Recorder SV-P100 and JVC XD-Z1100) using Brüel&Kjær microphone 4165 at 40 cm distance from the subject's mouth. A calibration signal had been recorded together with the samples for calculation of relative sound pressure level (SPL). For the analyses of the present study the samples were digitized with Sound Blaster at the rate of 44.1 kHz (16 bits) using MultiSpeech program (Kay Elemetrics). SPL was measured with a level meter (Brüel & Kjær Frequency Analyzer 2120 A). The integration time used was 30 seconds.

3. METHODS

3.1 Analytic Filtering Method

The analytic filtering method starts with Fourier transformation of the signal to be analyzed. One-sided Gaussian window is placed over the formant to be studied and the formant spectrum is inverse Fourier transformed. Due to the one-sided windowing the resulting signal is analytic: It is complex valued and its real and imaginary parts form a Hilbert pair [7].

The analytic formant (resonator) response can easily be modelled by Equation 1.

$$z[k+1] = \gamma[k]z[k], \qquad (1)$$

where k is time index, $z[k]$ is the analytic signal, and $\gamma[k]$ is a complex valued time-varying coefficient. The instantaneous resonance frequency $\omega[k]$ is given by $\mathrm{Arg}[\gamma[k]]$ while $\mathrm{Abs}[\gamma[k]]$ defines the instantaneous damping factor $r[k]$ of the analytic resonator. Thus the instantaneous resonance frequency and the instantaneous damping factor can be solved by

$$\begin{aligned} \omega[k] &= Arg[z[k+1]/z[k]] \\ r[k] &= Abs[z[k+1]/z[k]] \end{aligned} \qquad (2)$$

Only two samples of the analytic signal are needed to solve for these two variables. The analytic filtering method provides also information of the formant Hilbert envelope in the form of $\mathrm{Abs}[z[k]]$ which is useful when excitation time instant and damping rate of the formant are studied.

3.2 Modified Covariance Method

Linear Prediction (LP) is a conventional method to model signal spectrum with an all-pole transfer function.

Autocorrelation and covariance methods are the most popular ones. They differ in two principles: In the way the signal statistics is handled and in the range of parameter optimization. The differences are depicted in Figure 1. The uppermost frame represents the signal to be predicted (modelled). The three lower frames represent the delayed signals, a weighted sum of which forms the MSE optimal prediction. The auto-correlation method uses all the information depicted and optimizes the model over the entire signal (including the zeroes). The covariance method does not use the data in the areas indicated by the arrows. The modified covariance method is a type of a compromise between these two methods and does not use the data behind the second vertical line (on the right). The reason for this choice is that the model is stable at most of the window positions (almost like the autocorrelation method) while the model gives good estimates even in a relatively narrow window (almost like the covariance method).

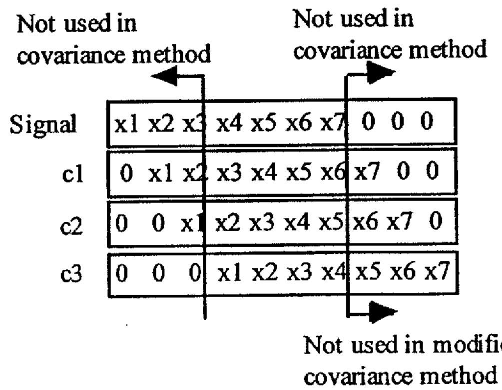

Figure 1. Differences in data statistics and optimization range in autocorrelation, covariance and modified covariance LP methods.

The modified covariance method gives slightly unstable results only when the window is located around the excitation time instant. The order of the LP was 40. Higher order analysis creates 'pseudoformants' which makes the monitoring of the true ones difficult. An analysis frame of 140 samples was used in steps of 11 samples (0.25 ms).

4. RESULTS AND DISCUSSION

Intraperiodic modulation in F1 and F2 can be clearly seen both in speaking and singing samples (Table 1, Figures 2 and 3). In the speaking samples of other subjects except D a decrease in F1 is seen as the glottis opens. For Subject D F1 increases slightly in the glottal open phase. For Subject C an increase is seen after the preliminary fall. In all the speaking samples F2 increases during the glottal open phase. In the samples of Subjects A and B the F1 modulation is little larger than that of F2, while the opposite is true for Subjects C and D.

In the singing samples F1 also tends to fall during the glottal open phase. For Subjects A and D an increase in F1 takes place after glottal closure; for Subject C F1 modulates up and down during glottal open phase. In other subjects except for Subject B the modulation in F1 is larger in singing than in the speaking mode. For Subject B modulation in F1 is more regular in the singing mode. F2 seems to be more stable in all the singing samples.

Variations in F2 and F1 can be mainly explained by time-varying coupling between supra- and subglottal cavities. This coupling seems to be stronger in the singing mode probably due to a change towards the so-called flow phonation characterized by a wider glottal opening. In the singing mode these trained singers seemed (obviously by lowering the larynx) to lower the first two formants to the vicinity of 500 Hz and 1200 Hz. This result fits well to the earlier findings by Sundberg [8] of spectral differences between spoken and sung vowels. Somehow this 'tuning' seems to minimize the frequency modulation of the second formant and also the truncation of the formants during glottal open phase.

An increase in F1 just after glottal closure could be partially explained by increased vertical position of the glottis due to the pushing effect of subglottic pressure. The fact that this is better seen in singing mode would suggest higher subglottic pressure in singing. The further up-going trend is due to the gradual glottal opening. All subjects have a down-going trend in F1 later in the open phase. This is larger and clearer in many of the sung samples. This could, thus, be partially explained by a wider glottal opening. It may also result from the LP method: When the formant energy is damped and the glottal flow pulse is large the method pickes upp the glottal formant F_g instead of F1.

The up-down modulation in F1 in the singing samples of Subjects A and C might indicate a stronger interaction between supra- and subglottic cavities, possibly related to a very wide glottal opening. The F1 modulations are also the widest for these subjects, which could be explained by the strong interaction.

The fact that more modulation could be seen in F1 than in F2 in the speaking samples of Subjects A and B while the opposite observation could be made for Subjects C and D cannot be explained through differences in F0 and SPL or phonatory quality (i.e., different degrees of glottal adduction). Instead, they seem to suggest individual differences in the modulation effects. These differences could be for example related to a longer and wider glottis (Subjects A and B were taller and the measures round their neck seemingly larger).

Figure 3 compares results of the two methods when the speech of Subject A was analysed. Two uppermost frames depict instantaneous F2 values the first given by the analytic filtering method and the second by the LP40 method. The analytic method leads to a much cleaner curve than the LP40. The modulation amplitude is clearly smaller in the analytic case. Both methods produce some artefacts around the glottal closure.

The following two frames depict instantaneous F1 correspondingly. The LP40 results also here in a more noisy signal. The next two frames (FE and FB) show signal samples at the end and at the begining of the LP analysis frame.

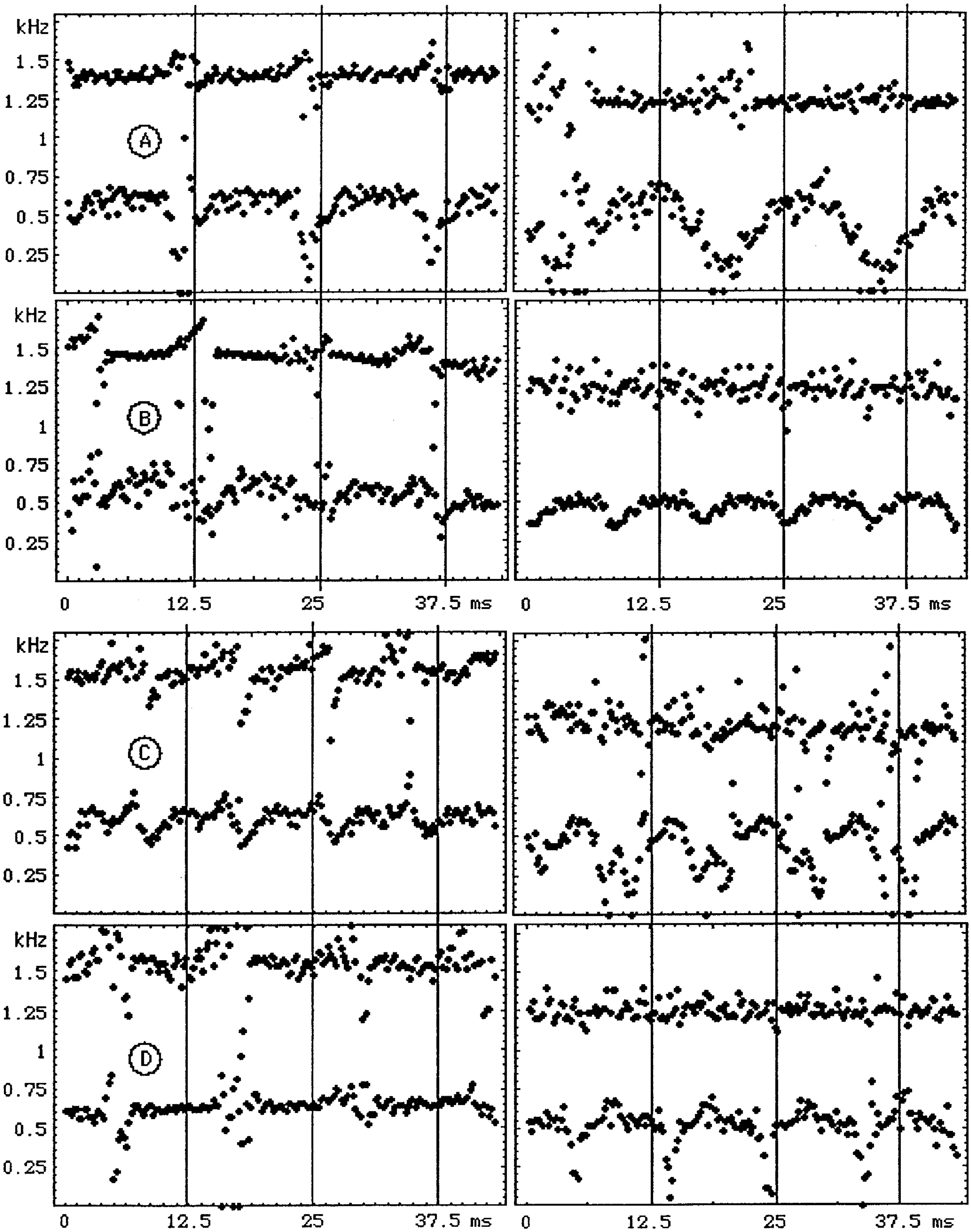

Figure 2. Spectrogram-like representation of the instantaneous formant frequencies of the spoken (left column) and sung (right column) vowels [æ:] of Subjects A, B, C, and D (modified covariance LP method of order 40, frame length = 140 samples, frame step = 11 samples 0.25 ms).

Table 1.
F0, SPL and modulation in F1 and F2 for spoken and sung samples."closed/open" refer to the phase of the glottal cycle.

	Spoken [æ:]			Sung [æ:]		
	F0	**SPL**		**F0**	**SPL**	
Subject A	86 Hz	85 dB		70 Hz	80 dB	
	closed	**open**	**change**	**closed**	**open**	**change**
F1	550 Hz	650-500? Hz	+100/-50? Hz	500 Hz	650-400? Hz	+150/-100 Hz
F2	1410 Hz	1500 Hz	+90 Hz	1200 Hz	1300 Hz	+100 Hz
Subject B	86 Hz	72 dB		130 Hz	73 dB	
	closed	**open**	**change**	**closed**	**open**	**change**
F1	500 Hz	650 Hz	+150 Hz	490 Hz	550-350? Hz	+60/-140? Hz
F2	1450 Hz	1550 Hz	+100 Hz	1300 Hz	1300? Hz	0? Hz
Subject C	111 Hz	60 dB		100 Hz	70 dB	
	closed	**open**	**change**	**closed**	**open**	**change**
F1	550 Hz	600, 750? Hz	+200? Hz	500-620 Hz	400? Hz	-100-220? Hz
F2	1550 Hz	1700 Hz	+150 Hz	1200 Hz	1250? Hz	+50? Hz
Subject D	82 Hz	63 dB		104 Hz	77 dB	
	closed	**open**	**change**	**closed**	**open**	**change**
F1	620 Hz	750? Hz	+130? Hz	500 Hz	600-650? Hz	+100/+150? Hz
F2	1550? Hz	1750 Hz	+200 Hz	1300? Hz	1300? Hz	0? Hz

The LP analysis results are synchronized with the end point of the frame (see the vertical lines). Figure FB indicates the position of the analysis frame the results of which are found at the leftmost vertical line. All the vertical lines are positioned close to the point where the new excitation starts. The LP40 method has clear difficulties to estimate formant frequencies in frames where the new excitation is coming in.

The two lowest frames depict the damping factors of the formants two and one created by the analytic filtering method. The corresponding results of the LP40 method were too noisy for any detailed analysis. Each new excitation causes a rapid increase in the damping factors which exeed over the value one (instable resonance). This indicates a rapid growth in the energy of the resonator. There is a small secondary excitation seen close to the time instant of the glottal opening. Surprisingly, r1 indicates lower damping during the open glottis period. This may be interpretated as a problem to distinguish between the true damping and the new excitation caused by the glottal opening.

To conclude, the analytic filtering method seems to work reasonably well when compared to the LP method. The quality of the result depends on the bandwidth used in the frequency domain filtering. The best results were gained when the bandwidth of the Gaussian filter was somewhat broader than the bandwidth of the formant under analysis.

5. CONCLUSIONS

Analysis of intraperiodic formant modulations is a challenging task. In the present study analytic filtering method and modified covariance type of linear prediction were applied leading to somewhat inconsistent results. However, both methods revealed clear formant modulations. Strict interpretation of the results is not possible at the moment. Methodology will be elaborated further and results will be tested with simulated and physical models.

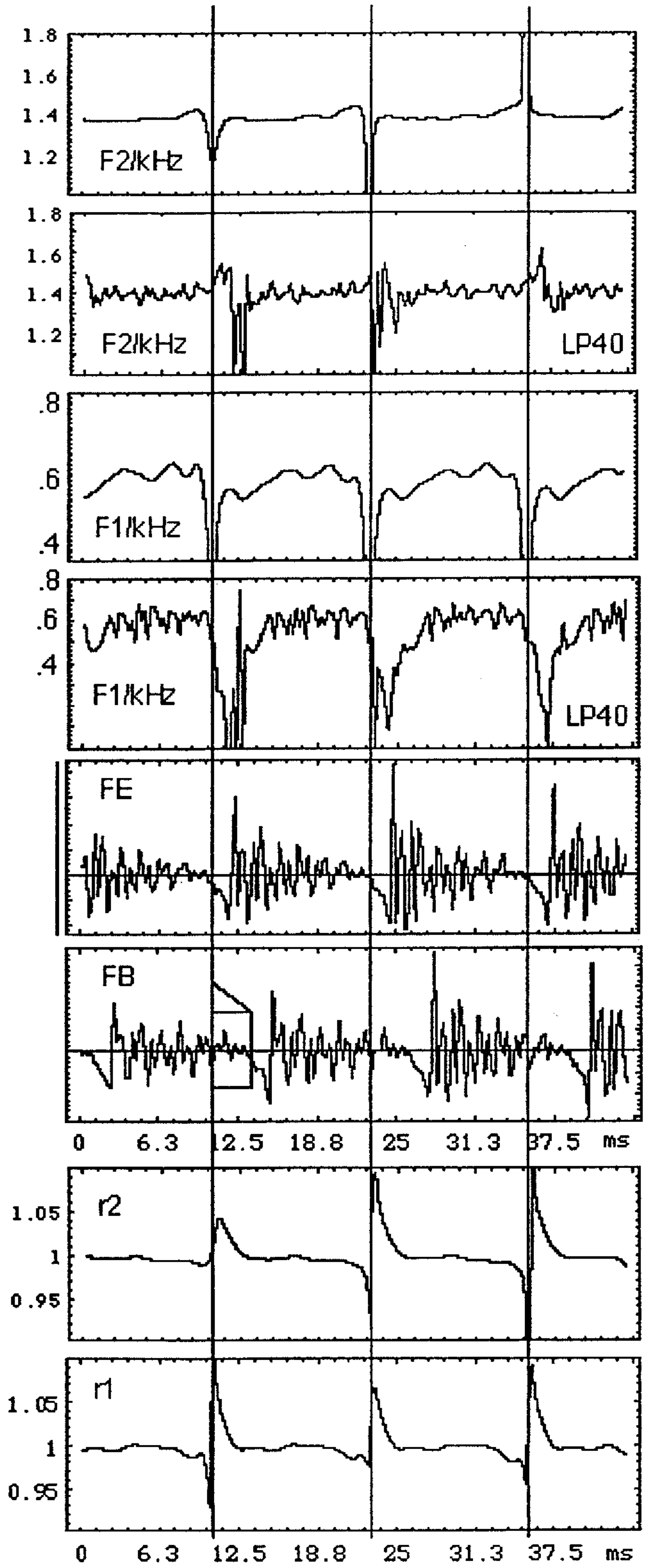

Figure 3. Comparision of results given by the analytic filtering method and the modified covariance method (LP40). The vertical lines indicate the time instances of the maximal excitation at the glottal closure. Subject A speech.

REFERENCES

[1] Fujimura O., and Lindqvist J. (1971), "Sweep-tone measurements of vocal-tract characteristics," JASA 49 (Part 2), 541-558.

[2] Ananthapadmanabha T. V., and Fant G. (1982), "Calculation of true glottal flow and its components," STL-QPSR 1.

[3] Fant G. (1985), "Acoustic parameters of the voice source," In Proceedings of the Stockholm Music Acoustics Conference 1983, vol. 1, A. Askenfelt, S. Felicetti, E. Jansson & J. Sundberg, eds., Stockholm 1985, pp.11-26.

[4] Fant G., and Liljencrants J. (1979), "Perception of vowels with truncated intraperiod decay envelopes," STL-QPRS 1.

[5] Gagné J., and Zurek P. M. (1988), "Resonance-frequency discrimination," JASA **83**, 2293-2299.

[6] Maragos P., Kaiser J. F., (1993) Quatieri T. F., "Energy separation in signal modulations with applications to speech analysis", IEEE Trans. on Signal Processing, vol. 41, pp. 3024-3051, Oct. 1993.

[7] Laine U. K. (1998), "Analysis of pitch-synchronous modulation effects by using analytic filters," Proc. of EUSIPCO'98, Rhodes, vol. 2, 1145-1148.

[8] Sundberg J. (1978), "Articulatory interpretation of the 'singing formant'," JASA **55** 838-44.