Spectral Consequences of a Time-varying Glottal Impedance

B. Cranen and L. Boves
Nijmegen, the Netherlands

1. Introduction

Virtually all research in modern acoustic phonetics relies on the linear source-system model of acoustic speech production (Fant, 1960; Flanagan, 1972). The alleged linearity of the source-system cascade permits the separate development of accurate models of both the source and the system. The assumption that the glottal impedance is very high enables one to treat the vocal tract as an acoustic tube that is closed at the glottal end. Since all articulators have an appreciable mass they are constrained to relatively slow motions, and consequently the vocal tract resonances can be supposed to vary so slowly that they may be considered as constant during time intervals of 10-30 ms. This assumption is at the basis of all present day speech technology, notably including all LPC techniques.

The ubiquitous experience of the machine-like quality of much synthetic speech, however, has motivated a re-examination of the customary source-system model. However unimpeachable the assumption of an infinite glottal impedance may be during the closed glottis interval of a fundamental period, during the open glottis interval the termination impedance at the glottis may well assume values of the same order of magnitude as the input impedance of the vocal tract.

Although it is recognized that changes in the glottal termination impedance may affect the formants of speech signals, little is known about the way in which specific changes in glottal termination affect the formants of various vowels. The present study atttempts to derive techniques with which the effects of the termination impedance can be determined so that it becomes possible to estimate their contribution to the naturalness of synthetic speech. To this end we combined measurements on real speech with model simulations.

2. The Wakita-Fant model

Unlike the customary models of speech production, the Wakita-Fant approach (henceforth WF-model) includes a model of the subglottal system. In the WF-model it is dimensioned so as to represent the measurements of Ishizaka, Matsudaira, and Kaneko (1976) on the input impedance of the subglottal system, who found resonances at 640 Hz, 1400 Hz, 2150 Hz, and 2850 Hz, with bandwidths between 200 and 400 Hz. In our implementation the supraglottal system consisted of six sections of unequal elliptical cross-sections and unequal lengths. We have studied the model for seven vocal tract configurations representative of the vowels /a,o,u,i,ɨ,e,ə/ Cross-sectional areas and lengths of the sections were adapted from the data in Fant (1960). Subglottal pressure was varied from 4 cm water up to 10 cm water in steps of 2 cm, and glottal area(Ag) was varied between 0 and .2 cm^2 in steps of .01 cm^2. In order to assess the importance of the subglottal system, all calculations have been repeated with a model in which the glottal termination consisted of an R-L circuit modeled after the properties of the air plug in the glottis.

The results of these model calculations can be summarized as follows: Formant frequencies and bandwidths appear to behave as monotonic (though not necessarily linear) functions of subglottal pressure and glottal area. The effect of changes in Ag appear to outweigh those of variations in subglottal pressure by far. This permits us to restrict our discussion to the results obtained for the 'typical' condition of 8 cm water for Psg.

For the model with the subglottal system the frequency of F1 increases by approximately 10% as Ag increases from 0 to .2 cm^2. Although the actual rate of change is quite vowel dependent, all vowels show variations which are clearly greater than 3%, the JND for formant frequencies according to Flanagan (1972). Except for the /a/-vowel, variations in F2 do not exceed the critical value of 3%, nor do the higher formants of any vowels.

Formant bandwidths in the complete model increase by approximately 400% in F1, 150% in F2 and 50% in F3 as Ag increases from 0 to .2 cm^2. The variations are strongly vowel dependent: F2 of /i/, for instance, increases by no more than 18%, which is clearly less than the JND of 50% (Flanagan, 1972). F2 of /o/ and /ə/ on the other hand, increase by more than 230%.

No simple conclusions can be drawn with respect to the results obtained for the model without a subglottal system, except that the dependence on vowel type is much greater than was the case with the complete model. Interestingly F1 of /i/ decreases slightly with increasing Ag in the model with the R-L glottis. The rate of change of formant bandwidth as a function of Ag is between two and three times as great for the simplified model.

Since the WF-model assumes stationary conditions and since the glottal impedance changes with a frequency which cannot be considered as very low with respect to the frequencies of the formants, a verification of the results of the model is in order. This has been done by studying real speech and the output of another model.

3. Formants measurements on real speech

Five subjects participated in an experiment in which simultaneous recordings were made of the electroglottogram, subglottal pressure, and the acoustic speech wave (Boves and Cranen, 1982). The electroglottogram is a reliable indicator of the moment of glottal opening and closure. Thus the signal provides the information needed to obtain the open and closed glottis intervals from each glottal cycle. The subglottal pressure signal enables us to determine the resonance of the subglottal system. Using a linear (time domain) averaging technique combined with cepstral smoothing spectra of the subglottal pressure waves were computed. The results consistently show resonances around 500, 1200, and 1900 Hz. Thus, we appear to find resonance frequencies of the subglottal system which are consistently lower than those reported by Ishizaka et al. More often than not clear spectral dips are present between these resonances.

Estimation of the parameters of the supraglottal system was attempted using a covariance LP analysis. In view of the fact that a (rectangular) analysis window -the left boundary of which is positioned at the beginning of the closed (or alternatively open) glottis interval- is not allowed to extend into the following open (or closed) interval, window lengths had to be constrained to at most some 30 samples. The extremely short-time LP analysis appears to yield stable and credible results for the closed glottis intervals. The open glottis intervals, on the other hand, consistently give very unstable results in the sense that variations from period to period are very great. Also, more often than not F1 and F2, or F3 and F4 are no longer resolved.

4. Formant measurements made on the Ishizaka-Flanagan model

The speech production model as described in Ishizaka and Flanagan (1972) (henceforward IF-model) was implemented using the 1130 Continuous System Modelling Program (developed for use on IBM 1130 and 1800 computers but adapted to run on a DG Eclipse S/200). Seven vowels were generated using the same vocal tract configurations as in the experiments with the WF-model described above. The parameters of the two-mass model of the vocal folds were given 'typical' values. Our implementation of the IF-model did not include a subglottal system. Outputs of the model comprise the speech signal radiated at the lips and the volume flow at the glottis. From the latter signal the boundaries of the open and closed glottis intervals can be established unequivocally.

A covariance LP analysis carried out on the synthetic speech confirmed the results obtained with real speech. Formant estimates during closed glottis intervals are stable and in very good agreement with the results of the WF-model, whereas estimates derived from the open glottis intervals often fail to resolve all formants.

5. The failures explained

Merging of the formants can be accounted for by an increased damping in the open glottis interval. Formant bandwidths may become excessively large and may thereby prevent neighbouring peaks from being resolved.

This increased damping may be due to a decrease in glottal impedance during the open glottis interval. This effect is apparent in the (stationary) WF-model for e.g. the vowel /a/.

A time-varying glottal area might, however, contribute to a merging of formants in yet another way: the time derivative of a uniformly growing conductance acts like an inertance. Ananthapadmanabha and Fant (1982) have shown that the source filter interaction during the open glottis interval will give rise to both an upward shift in frequency and a broadening of the first formant, in addition to the introduction of a number of spectral zeros.

Fourier analysis of glottal flow signals obtained by inverse filtering of real speech and as an additional output of the IF-model shows that the source signals indeed tend to contain clear and sharp spectral dips. As can be seen from figures 1 and 2 one dip invariantly lies very close to the F1-frequency of the closed glottis interval as predicted by Ananthapadmanabha et al. (1982).

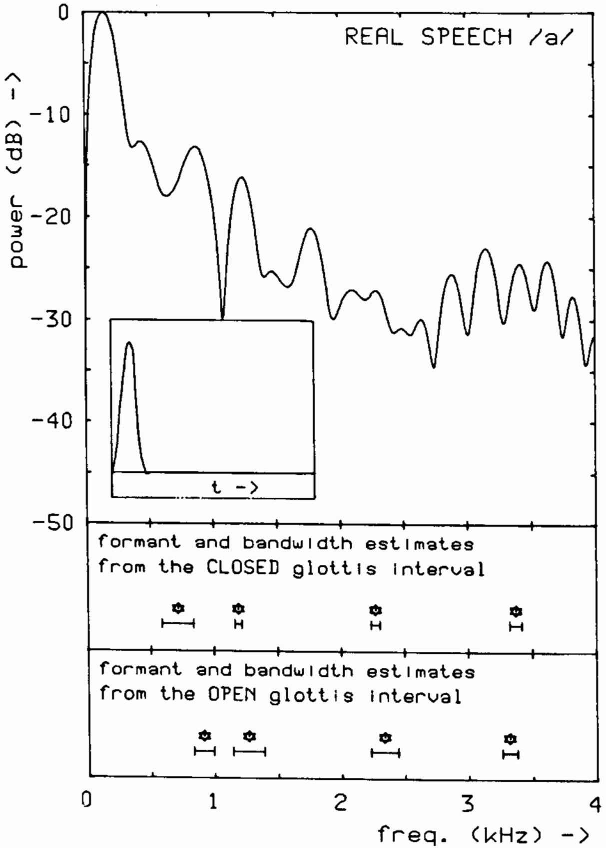

Figure 1. 1024 point FFT-spectrum of (the time-derivative of) a reconstructed glottal pulse (REAL SPEECH, vowel /a/) and formant/bandwidth estimates of the speech signal in the closed, respectively open glottis interval. The pulse of which a 25 ms time segment is shown in the inset, has been obtained by inverse filtering of the speech signal. The inverse filter has been estimated in the preceding closed glottis interval by means of an LP-analysis (covariance method, rectangular analysis window = 24 (closed)/36 (open) samples, prediction order = 10 (closed)/12)open)).

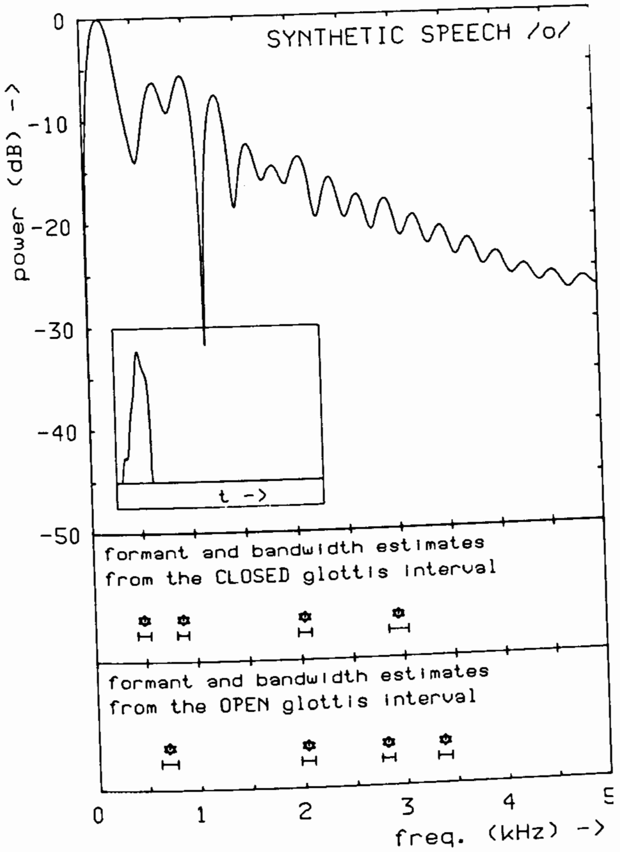

Figure 2. 1024-point FFT-spectrum of (the time-derivative of) a glottal pulse (IF-MODEL, vowel /o/) and formant/bandwidth estimates of the speech signal in the closed, respectively open glottis interval. The pulse of which a 25 ms time segment is shown in the inset, has been generated by the IF-model. The formant estimates have been obtained by LP-analysis of the corresponding speech segments (covariance method, rectangular analysis window = 24 (closed)/36 (open) samples, prediction order = 10 (closed)/12 (open)).

6. Conclusions

From the experiments with the WF-model it is clear that within-cycle variations in the termination impedance at the glottis may well have effects on formant parameters of such an order of magnitude as to be perceptually relevant. Thus, such variations may account for (part of) the machine-like character of much synthetic speech.

Attempts to verify the predictions of the WF-model by measurements on real speech and on the output of the IF-model confirmed the importance of within-cycle variations, but failed to yield numerical results since formant estimations in open glottis intervals appeared to be virtually impossible. To a large extent this failure can be explained by the presence of clear spectral zeros in the voice source.

The outcomes of our measurements, combined with the theoretical work of Fant and his coworkers, raise the question whether it is at all possible and sensible to use the formant concept in describing the speech signal during the open glottis interval.

Acknowledgement

This research was supported by the Foundation for Linguistic Research which is funded by the Netherlands Organization for the advancement of pure research, ZWO.

References

Boves, L. and Cranen, B. (1982). Evaluation of glottal inverse filtering by means of physiological registrations. *Conference Records on the International Conference on Acoustics, Speech, and Signal Processing ICASSP-82,* 1988-1991.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Ananthapadmanabha, T.V. and Fant, G. (1982). Calculation of true glottal flow and its components. *Speech Communication* **1**, 167-184.

Flanagan, J.L. (1972). *Speech analysis, synthesis, and perception.* Berlin: Springer.

Ishizaka K. and Flanagan, J.L. (1972). Synthesis of voice sounds from a two mass model of the vocal cords. *Bell System Technical Journal,* **51**, 1233-1268.

Ishizaka, K., Matsudaira, M. and Kaneko, T. (1976). Input acoustic impedance measurement of the subglottal system. *Journal of the Acoustical Society of America,* **60**, 190-197.

Wakita, H. and Fant, G. (1978). Toward a better vocal tract model. *Speech Transmission Laboratory-QPSR,* **1**, 9-29.