A Study of Voice Source Interaction

Mats Båvegård
Department of Speech Communication and Music Acoustics,
KTH, Box 70014, S-100 44 Stockholm

ABSTRACT

The present study addresses problems of acoustic source-tract interaction in the production of voiced sounds. In particular we are concerned with changes in the glottal flow, e.g. with respect to superimposed ripple, spectrum shape and formant frequency damping due to the variations in supra- and sub-glottal loads, the choice of glottal area function and the presence versus absence of a constant leak. We have verified that the addition of a constant leak on a glottal area function with gradual decay in the closure region, in addition to formant frequency damping, ripple reduction and a low frequency boost causes a break towards a lesser rate of spectrum fall at higher frequencies. The relations between glottal area function shapes and flow pulse shapes under various conditions and the perceptual significance of the pulse ripple are discussed.

INTRODUCTION

Inverse filtering combined with extraction of LF-source parameters, e.g. Gobl(1988), Karlsson(1992) have provided a knowledge base for improved quality in speech synthesis, Carlsson, Granström and Karlsson (1990). An inherent constraint is the linearity of LF-synthesis (Fant, Liljencrants and Lin, 1985) which assumes independent source and vocal tract characteristics. In a more advanced model taking into account sub- and supraglottal interaction, e.g. that of Lin (1990) the true glottal flow is dependent on the entire system. One of the interaction effects noted already by Ananthapadmanabha and Fant(1982), and by Fant and Lin(1987), Lin(1990) is the presence of ripple superimposed on the glottal flow. In the course of pursuing further work on complete articulatory synthesis we are interested in the perceptibility of glottal ripple and to what extent it adds to synthesis quality. According to earlier experiments, e.g. Nord et al. (1986) the perceptual effects are small. Our interactive synthesis system Tracttalk, (Lin, 1990) provides the facilities to perform more conclusive tests. One of the conditions we have tested is the combination of an LF shaping of glottal area with a constant leakage which provides interesting results.

METHOD AND RESULTS

The LF-model (Fant et al., 1985) of glottal flow is here used as an area function of the glottal area. It is controlled with respect to three waveshape parameters and the fundamental frequency F_0. These are the rise time parameter R_g, the pulse asymmetry parameter R_k and the duration of the residual closure phase T_A. It is also possible to add a constant leak. This area function will then be used to compute the true glottal airflow of the interactive system (for computational details, see Lin, 1990).

Two different voice types are modelled in this study. One symmetrical glottal area function, with different degrees of dynamic and constant glottal leakage and one with asymmetrical glottal area function and different amount of dynamic leakage.

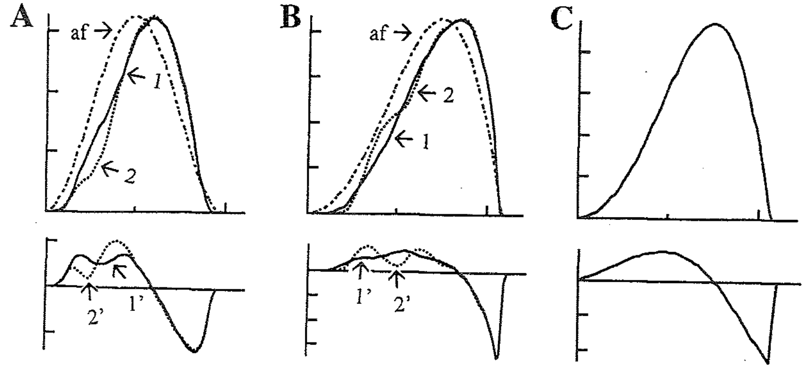

Figure 1. *Two different area functions (af); first glottal flow pulse (1), second glottal flow pulse(2); first and second differentiated glottal flow pulses (1' & 2').*
A: symmetrical area, B: asymmetrical area function, C: linear LF-glottal flow pulse with the same LF-parameters for the flow as for the area in B.

Three different types of interaction effects can be shown from this experiment, see figure 1:
Pulse skewing, due to the total vocal tract and glottal inductance. The glottal flow is skewed to the right compared to the glottal area.
Superposition ripple, originating from short time variations of the transglottal pressure. This is evoked by formant oscillations of the pulse onset and previous glottal periods.
Damping, of the flow oscillation in the open phase of the glottal pulse.

These effects are common for this type of experiments and found in both the symmetrical and the asymmetrical area function simulations.

More interesting is the simulations where a constant leak is introduced. Figure 2 shows simulations of the first formant of an /a/-vowel with a symmetrical LF area function where a constant leak of 0.03 cm^2 is established similar to a glottal chink. The area functions in figure 2, are modelled with T_0 = 8.0 ms, T_p = 2.0 ms, T_e = 4.0 ms and T_A = 1.0 ms, which implies an R_k value of 1.0. The constant lungpressure is 4.0 cm H_2O and the maximum glottal area is 0.15 cm^2.

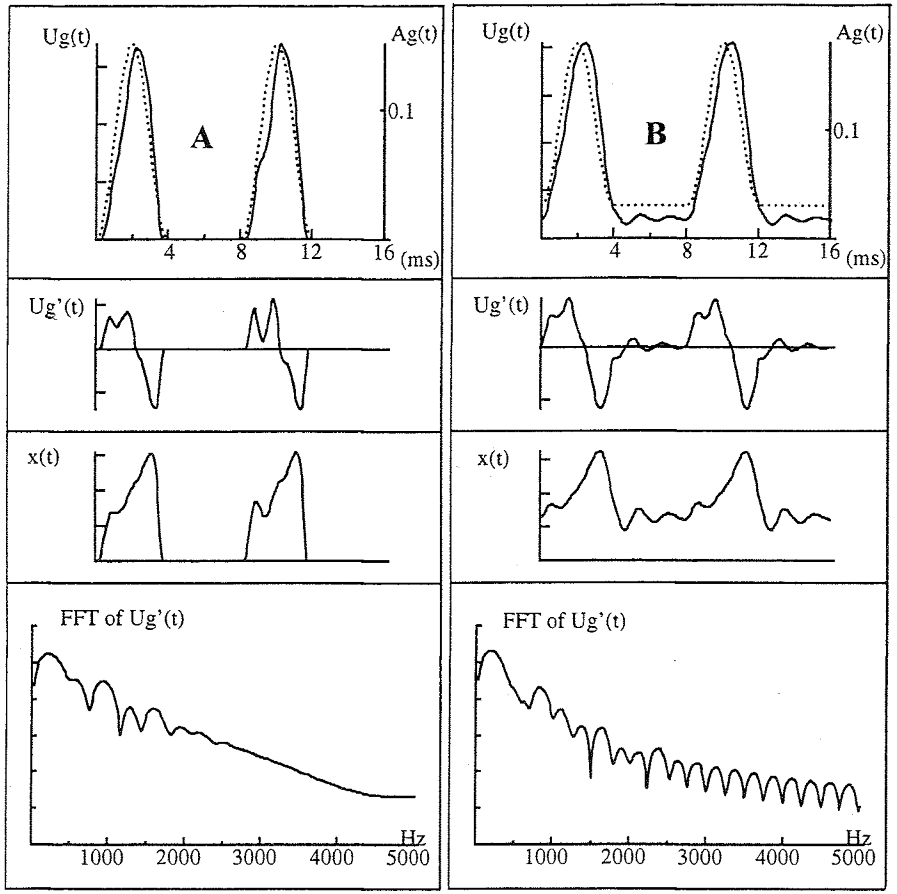

Figure 2. *Two different simulations of an /a/-vowel. From top area function (dotted line) and the glottal flow (solid line), differentiated glottal flow, normalised particle velocity through glottis, and at the bottom spectrum of the differentiated glottal flow. A: Dynamic leak only; B: Dynamic leak and constant leak = 0.03 cm^2;*

Adding a constant leak will increase flow pulse skewing, see figure 2. The corresponding change in R_k is from about 0.6 to 0.4. At the same time an increased T_A and a much reduced ripple effect is observed. T_A changes from about 0.5 ms to 0.7 ms, which means that T_A of the glottal flow heads toward that of the glottal area function (1.0 ms), when the constant leak is introduced.
The spectral tilt is also changed. In the leaky case we see a more flat high frequency spectrum, which is often seen in inverse filtered female speech. The same effect is also observed for large leak values.

DISCUSSION

Simulations with an interactive voice source system demonstrate significant ripple as well as perturbations of the instant of the glottal flow onset, which add to a random fine structure of individual glottal flow pulses. These phenomena which are fairly well understood from earlier modelling (e.g. Lin 1990) are often seen in inverse filtered natural speech (e.g. Gobl, 1988).
We have found that a moderate constant leak in the glottis area, causes an increase of T_A and a decrease of the R_k parameter, and a reduction of the ripple effect, see figure 2. At the same time the interactive system creates a break towards a lesser rate of spectrum fall at higher frequencies, which could be useful in female speech synthesis.
In the flow spectrum domain an increased T_A normally implies more spectral tilt. It is observed in this experiment too, though it is not possible to explain the high frequency spectrum level by the nature of LF-flow parameters. Instead we have to consider a complex interactive system that will have to be further investigated in terms of transforming glottal area function to glottal flow.
Cranen and Schroeter(1993) have made similar observations of the interaction effects of producing flat high frequency spectras with leaky glottis. Lin (1990) also observed a spectrum flattening and proposed that there is an finite interval of increasing T_A, in which the frictional loss is the dominating term of the glottal resistance and if the glottal friction is neglected, the resulting spectrum does not exhibit such high frequency effect.
In our study we have seen that, by introducing varying degrees of constant leaks, the T_A and R_k parameters are changed due to the interaction without affecting the high frequency spectrum stabilisation. More detailed studies of producing flat high frequency spectra are needed.

ACKNOWLEDGEMENTS

I would like to thank Gunnar Fant for advice and support. This work has been funded by ESPRIT project 6975, SPEECHMAPS, financed by NUTEK.

REFERENCES

Fant, G. and Ananthapadmanabha, T. V. 1982. 'Truncation and superposition'. *STL-QPSR* 2-3/1982, pp 1-17.
Carlsson, R., Granström, B. and Karlsson, I. 1990. 'Experiments with voice modelling in speech synthesis'. *Speech Communication.* Vol. 10, pp 481-490.
Cranen, B. and Schroeter, J. 1993. 'Modelling a leaky glottis'. *Proceedings Dept. of language and speech* 16/17, University of Nijmegen. pp 56-64.
Fant, G. and Lin, Q. 1987. 'Glottal source - Vocal Tract Interaction'. *STL-QPSR* 1/1987, pp 13-27.
Gobl, C. 1988. 'Voice Source Dynamics in Connected Speech'. *STL-QPSR* 1/1988, pp 123-159.
Karlsson, I. 1992. *Analysis and Synthesis of Different Voices with Emphasis on Female Speech.,* Ph.D. thesis, STL., KTH, Stockholm.
Lin, Q. 1987. 'Nonlinear interaction in voice source production'. *STL-QPSR* 1/1987, pp 1-12.
Lin, Q. 1990. *Speech production theory and articulatory speech synthesis.,* Ph.D. thesis, STL., KTH, Stockholm.
Nord, L., Ananthapadmanabha, T.V. and Fant, G. 1986. 'Perceptual tests using an interactive source filter model and considerations for synthesis strategies'. *J. of Phonetics* 14, pp 401-404.