# INTERACTIVE VOICE SOURCE MODELLING

*Mats Båvegård and Gunnar Fant*
*Dept. of Speech Communication and Music Acoustics,*
*KTH, Stockholm, Sweden*

## ABSTRACT

Interaction ripple is caused by the non-linear relation between glottal flow and trans-glottal pressure, the latter containing components of vocal tract oscillatory modes evoked during the past history of the phonatory process.

This study deals with the non-linear transformation from glottal area function to glottal flow, with and without constant leakage invoked by a glottal chink. Earlier observations on a certain high frequency boost associated with the glottal chink [1], [2], have been verified. A separate study of the audibility of interaction ripple has been undertaken.

## INTRODUCTION

Interaction ripple is the superposition of quasi-random variations within a voice cycle of the glottal flow which is commonly observed in inverse filtering of the speech signal. A number of aerodynamic-acoustic model simulations have revealed the general mechanism causing interaction ripple, i.e. [2] [3].

The origin is a non-linear perturbation of glottal flow caused by vocal tract oscillatory modes superimposed on the trans-glottal pressure.

The term interaction in a broader sense involves all aspect of a complete supra- and sub-glottal coupling and associated departures from normal phonation. In breathy phonation the coupling causes shifts of formant frequencies and bandwidths and the appearance of sub-glottal formants and noise. These can be studied in a complete articulatory analog.

## FLOW SIMULATIONS

The simulations have been performed with our articulatory speech synthesiser, TRACTTALK [2]. The synthesiser consists of three parts, a sub-glottal system, a model of the time varying glottal opening area modulating the trans-glottal impedance and a supra-glottal system.

The sub-glottal system incorporates three resonance modes and the glottal impedance contains the complete inventory of inductance, kinetic resistance and frictional resistance. The examples Fig. 1-3 pertains to a supra-glottal configuration of a vowel [ɑ:]. For computational details see [4]. The LF-model was used to control the glottal area function $A_g(t)$. The effects of adding a constant leak are also demonstrated in the three examples, Fig 1-3.

### Results

In the zero glottal leak simulation, curve 1 in both Fig 1 and Fig 2, there is an apparent double peak in the positive part of $U_g'(t)$ which is a typical instance of interaction ripple as seen in true speech, similar to what has been discussed in [2]. The main cause of the double peak is the F1 oscillatory component of the trans-glottal pressure. The spectral consequence of a valley and peak around 800 Hz is rather weak. It should be kept in mind that the specific shape of the interaction ripple is highly dependent on the duration of the voice fundamental period which determines the particular phase of the previously excited component upon arrival in the beginning of the next open phase.

Fig. 1 is an example of adding a constant leak to a glottal area function, with return time $T_a$=0.1 ms. Three values of glottal leak have been simulated: no leak $A_{go}$= 0, $A_{go}$= 0.03 cm$^2$, $A_{go}$= 0.075 cm$^2$. The glottal flow derivative $U_g'(t)$ for zero leak displays the typical double peak, which is smoothed out and disappears with increasing constant leak. This is to be expected because of the increased damping and thus the low carry-over of F1 oscillation from the previous period. At the same time we observe an irregularity in the return phase of $U_g'(t)$, a phenomena which has frequently been observed in inverse filtering of real speech and in the simulations, [1], [5]. The mere presence of the leak appears to be sufficient as an explanation.



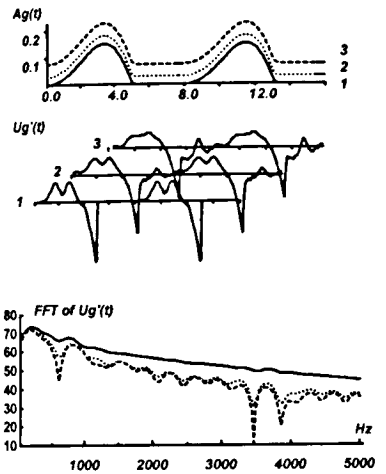


*Fig 1. From the top: Glottal area function $A_g(t)$ with a small return time $T_a$ = 0.1 ms and three values of a constant leak. $T_0$ = 8.0, $T_e$ = 5.0, $T_p$ = 3.5 ms, the resulting differentiated glottal flow and at the bottom the spectrum of the differentiated glottal flow.*
*Curve 1: without constant leak,*
*2: constant leak = 0.03 cm$^2$ included,*
*3: constant leak = 0.075 cm$^2$ included.*

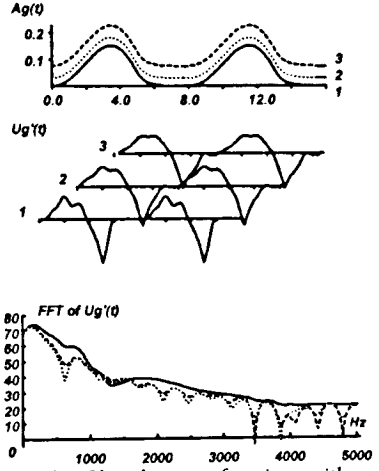


*Fig 2. Glottal area function with a moderate return time $T_a$ = 0.5 ms, else specifications identical as in Fig 1.*

Another typical effect is the increase in the effective return time $T_a$ in the flow compared to the $T_a$ of the underlying area function. The $T_a$=0.1 ms of $A_g(t)$ corresponds to a critical frequency $F_a$= $1/(2\pi T_a)$= 1600 Hz. For the no leak case we observe a $T_a$ corresponding to approximately $F_a$ = 850 Hz, with the small leak $F_a$=700 Hz and the larger leak $F_a$= 650 Hz. The corresponding increase in spectral tilt associated with leakage is apparent in the FFT spectrum.

With increasing $T_a$ of the glottal area function and the presence of a constant leak there is a non-uniform shift in the source slope. This is demonstrated in Fig. 2, where $T_a$ of $A_g(t)$ is 0.5 ms. The corresponding $T_a$ of $U_g'(t)$ in Fig. 2 is 0.6 ms with no leak, 1.2 ms for the small leak of 0.03 cm$^2$ and 1.4 ms for the larger leak of 0.075 cm$^2$ equivalent to $F_a$= $1/(2\pi T_a)$ values of 256 Hz, 130 Hz and 115 Hz respectively. The presence of a leak thus causes approximately a doubling of $T_a$ and an octave lowering of $F_a$. Thus, with a leak present the spectral tilt sets in at a lower frequency. However, instead of a uniform slope there is a recovery above 1000 Hz which restores the spectrum level to nearly the same as the no leak case at 3000 Hz. An additional effect of the constant leakage is a reduction of the maximum negative value of $U_g'(t)$, i.e. $E_e$ by about 3 dB.

These effects are even more apparent in Fig. 3 where $T_a$=1 ms in the glottal area function. Above 1700 Hz and with a leak the spectral level is restored to about the same or higher than without leak in spite of the fact that the higher $T_a$ (6 dB) and the lower $E_e$ (2 dB) together would have produced a 8 dB lower level in this range. Because of the higher $T_a$ (lower $F_a$) the presence of a leak causes a steeper spectral slope up to 1000 Hz followed by the restoration. The magnitude is of order 10 dB here.

These observations supporting the earlier findings of [1] and [2], have implication for the overall spectral characteristics of the female voice.

To what extent are these phenomena dependent on the sub-glottal system?

We repeated the experiment in Fig. 3 but with the sub-glottal system short-circuited. The $E_e$ value is not substantially reduced by the introduction of the leakage. The higher $E_e$ without the sub-glottal system is partially compensated by a lower $T_a$ of $U_g'(t)$ and the spectral difference becomes small [4].

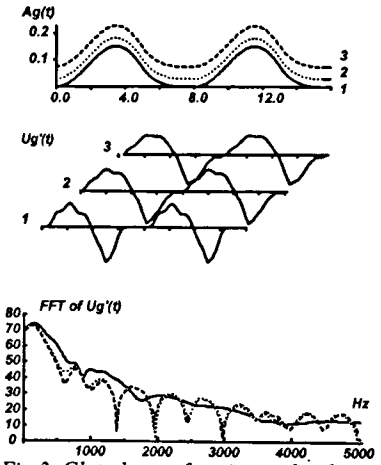


*Fig 3. Glottal area function with a large return time $T_a$ = 1.0 ms, else specifications identical as in Fig 1.*

## PERCEPTUAL TEST

When selecting stimuli for a perceptual test we avoided glottal configurations that produce more complicated interactions. We thus excluded the superimposed constant glottal leak and choose parameters typical for a well developed male voice with normal time constants of the glottal flow return phase. One of the test stimuli was the sentence "Ja adjö", ['jɑ:-a'jø:]. In addition we produced isolated vowels [ɑ:] and [ø:] of about 400 ms duration. These were produced with two different modes of temporal modulation, one with varying $F_0$ and waveform parameters approximating natural speech, the second one with constant $F_0$ and a moderately varying amplitude profile at onset and offset which was expected to enhance the perceptibility of the ripple. Each of these three stimuli types,

(1) Vowels with time varying $F_0$ and waveshape parameters.
(2) Vowels with constant $F_0$ and waveshape parameters.
(3) The complete sentence, "Ja-adjö".

were compared to synthetic versions produced with optimal LF-source in non-interactive formant synthesis.

### Test procedure

The main part of the study was devoted to discriminabilities in AB and ABX tests. The question posed in the AB test was "which do you prefer, A or B?". The ABX-test posed the standard question: "Is X more like A than B?". In addition we ran tests aiming at a categorical rating of perceived differences, within contrasting stimulus pairs of full representations and LF versions.

### Results

The AB-tests, Fig 4, show that the majority of the listeners preferred the interactive source in all vowels and in the sentence "Ja-adjö". For vowels the average score was of the order of 70%, i.e. well above chance level, and not significantly different for the time varying and constant parameter settings. In the sentence test the score was higher, 85% which was to be expected in view of other shortcomings in the overall matching to the human reference.

The ABX test, Fig 5, supported the general findings from the AB test. Apart from the low discriminability of the vowel [ɑ:] which could be explained by its initial placement in the test sequence without previous training, the tendencies appear similar to those of the AB test and with higher test scores.

The perceived difference test supported the view that the differences in interactive and LF synthesis are small or very small. Only 13% of the votes were in the category of a large difference. In the sentence test, on the other hand, as much as 55% votes were in the large difference category [4]. However the "large" assignment is a relational rather

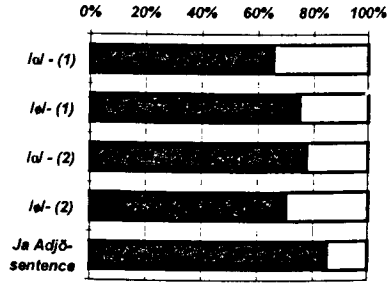


*Fig 4. The AB-test, (1) Vowels with time varying $F_0$ and waveshape parameters. (2) Vowels with constant $F_0$ and waveshape parameters.*



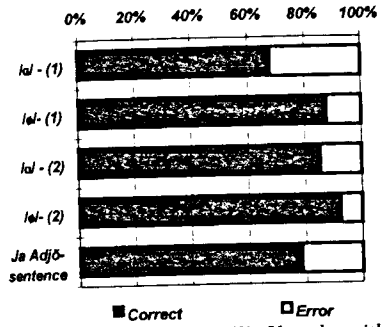


*Fig 5. The ABX-test, (1) Vowels with time varying $F_0$ and waveshape parameters. (2) Vowels with constant $F_0$ and waveshape parameters.*

than an absolute judgement. The synthetic version is a rather good approximation to the human utterance. This overall impression is supported by the no difference votes of approximately 20% in the ABX test.

## CONCLUSIONS

Interaction ripple is a non-linear perturbation superimposed on glottal flow. It originates from residuals of vocal tract oscillatory modes, largely those evoked in the preceding fundamental period, adding to the instaneous value of the trans-glottal pressure drop which has a second power relation to glottal flow. The occurrence of ripple does not require a glottal leak or the influence of the sub-glottal system. On the contrary a constant leak adds to the damping of formant oscillations and thus smoothes out the ripple as is encountered in phrase final abduction.

One recurrent observation is that the return time $T_a$ of the glottal flow tends to be larger than the $T_a$ of the glottal area function and that it increases with increasing glottal leakage. A glottal leakage also reduced the excitation amplitude $E_e$ but only when the sub-glottal network is retained.

We have verified the tendency observed in [1] and [2] that a combination of a constant leakage as with a glottal chink, and a finite $T_a$ of the glottal area function causes a high frequency boost that partially inhibits the spectral drop above 1000 Hz. This effect is of the order 6-12 dB at 2500 Hz. It is somewhat reduced when the sub-glottal impedance is short-circuited. This phenomena has implications for the interpretation of female voice source spectra. The presence of a glottal chink may also induce an irregularity in the closing phase as earlier observed in [1] and [5].

Our ABX test provided data on the detectability of interaction ripple while the preference was judged by an AB test. The main outcome is that the presence of interaction ripple adds a weak but detectable quality which is prefered by a majority of the listeners. This was found in both stimulus category (1) and category (2), as well as for the sentence "Ja-adjö". Although the synthetic version of the sentence was a fairly good replica of the spoken version the perceived difference was judged to be substantial.

Our perceptual studies have been directed to one aspect of interactive synthesis only, that of the ripple. A more important object for future research will be to gather experience of quality gains associated with complete articulatory synthesis and thus a more complex source-filter interaction. For this purpose we are now considering a more flexible glottal area function than the LF-model which initially was intended for glottal flow only.

## ACKNOWLEDGEMENTS

This work has been funded by ESPRIT/BR project 6975, SPEECH-MAPS, in part financed by NUTEK

## REFERENCES

[1] Cranen, B. and Schroeter, J. (1993): "Modelling a leaky glottis". *Proc. Dept. of Language and Speech* 16/17, University of Nijmegen. pp 56-64.
[2] Lin, Q. (1990): "*Speech Production Theory and Articulatory Speech Synthesis*", Ph.D. Thesis, Dept. Speech Com. and Music Acoust., KTH, Stockholm.
[3] Fant, G. (1986): "Glottal flow: models and interaction," *J of Phonetics*, 14 Nos (3/4), pp.393-399.
[4] Båvegård M., Fant G. (1994), "Notes on voice source interaction ripple", *STL-QPSR* 4/1994, pp 63-77.
[5] Karlsson I., Liljencrants, J. (1994): "Wrestling the two mass model to conform with real glottal wave forms", Proceedings ICSLP'94, Yokohama, pp 151-154