INVERSION OF THE VOICE SOURCE FOR SOME FRICATIVES

Christophe Vescovi, Eric Castelli
Institut de la Communication Parlée U.R.A. - CNRS N° 368
I.N.P.G./E.N.S.E.R.G. - Université STENDHAL
46, Avenue Félix Viallet, 38031 GRENOBLE Cedex 1 France

ABSTRACT

In this paper inversion of a model of the vocal cords for vowel-fricative-vowel transitions is presented. The robotics approach of the inversion problem based on the forward modelling of the plant has already been successfully studied for the vocal tract and for the voice source in vocalic context. Two items (/pava/ and /paʒa/) are used here to study the validity of the inversion method with extreme source-tract interactions.

INTRODUCTION

The articulatory synthesis takes more and more importance in speech research nowadays. This kind of speech synthesis may be the only way to reflect all the human being variability, but is also very helpful to learn more about the speech production process using analysis by synthesis. Inversion of natural speech is a crucial point in those perspectives as it is the simplest way to provide appropriated commands to speech production models.

ICP's speech production model [1] can be divided in three parts : two acoustic modellings for the lungs and the vocal tract based on the Kelly & Lochbaum model [2] and a physical modelling of the glottis based on the Ishizaka & Flanagan Two-Mass model [3],[4]. Previous studies on the inversion of the voice source [5] have point out that the interaction between the voice source and its environment (mainly the vocal tract) should be taken into account in the inversion algorithm. This inversion method has already given encouraging results for Vowel-Vowel transitions, but has not be tested with more extreme source-tract interactions like in Vowel-Fricative-Vowel transitions.

METHOD

The robotics approach supporting this work is based on a forward model of the plant. The forward model can be defined as a mapping of the system under control, giving proximal to distal relations. In this work a polynomial (result of the analysis of a codebook) is used as a forward model, then a simple error backpropagation algorithm can perform speech to articulatory inversion (Figure 1).

In order to reduce the complexity of the inversion algorithm the dimension of the distal space must be as small as possible. Recent studies on the voice source [6], point out that three parameters could be enough to define the glottal flow of a speaker. Thus, the glottal flow is characterised by the fundamental frequency Fo, the energy of the speech signal E, and a wave shape parameter Rd. Using the classical LF parameters Ee and Uo (minimum derivative flow and maximum flow) [7], the declination ratio Rd can be defined by :

$$Rd = \frac{Uo}{Ee} \cdot Fo$$

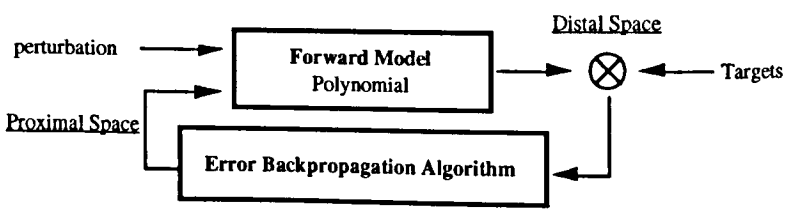

Figure 1 : general description of the inversion method

The proximal space is defined by the commands of the voice source model :
- Ps, the subglottal pressure.
- Lg, the length of the vocal cords.
- Ho, the rest aperture of the vocal cords.

Using the classical Two-Mass Model commands [3], Lg and Ho can be associated to Q and Ag0.

Perturbation of the voice source due to interactions with the sub and supraglottal cavities, must be taken into account in the inversion algorithm. Two parameters are used to define the vocal tract influence on the source, Ivt and Xvt respectively the inductance of the vocal tract and the position of the half inductance in the tract [5].

MEASUREMENTS

Two vowel-fricative-vowel items recorded by a French male speaker PB (/pava/ and /paʒa/) are analysed. The speech signal is first inverse filtered in order to evaluate the glottal flow and the three characterisation parameters are measured on the flow.

Formants trajectories used in the inverse filter are applied to ICP's inversion algorithm of the vocal tract which provides area functions and thus parameters Ivt and Xvt.

The intra-oral pressure (recorded simultaneously with the speech signal) is low-pass filtered as to give a approximation of the aerodynamic supraglottal pressure. Thus, the subglottal pressure Ps can be estimated in the consonant /p/ (the glottis is opened). ΔPg, the pressure drop at the glottis during the fricative can also be estimated by subtracting the intra-oral pressure to the estimated subglottal pressure (Table 1).

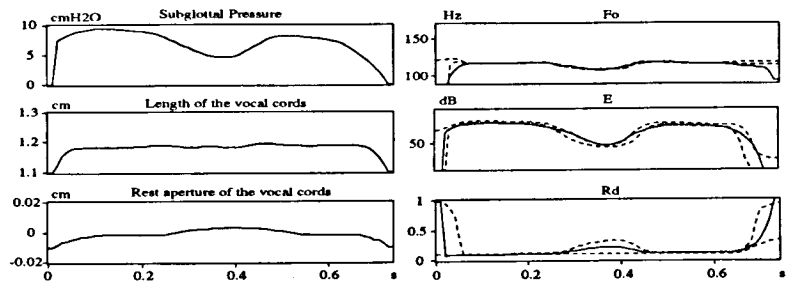

Figure 2 : results of the inversion for /pava/.
left : proximal space Ps, Lg and Ho.
right : distal space Fo, E and Rd
dashed lines = targets, solid lines = estimation of the forward model

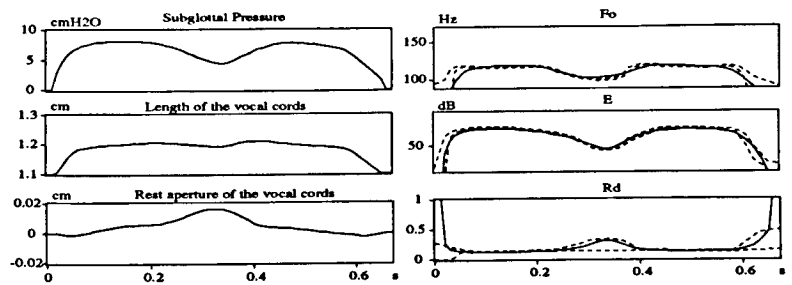

Figure 3 : results of the inversion for /paʒa/.
left : proximal space Ps, Lg and Ho.
right : distal space Fo, E and Rd
dashed lines = targets, solid lines = estimation of the forward model

Table 1 : measurements of subglottal pressure and pressure drop at the glottis in the fricative. (cmH2O)

	Ps	ΔPg
/pava/	9.2	3
/paʒa/	7.6	2

RESULTS

The measured flow characterisation parameters Fo, E and Rd are applied as targets to the inversion algorithm, whereas the parameters Ivt and Xvt are applied as perturbations. The results of the inversion for /pava/ and /paʒa/ are shown on Figures 2 & 3. For both items, the targets have been reached by the forward model without any normalisation of the distal space (especially for the wave shape parameter). this outstanding point can be explain by the specificity of the Rd parameter which does not reflect a geometrical property of the flow wave shape (like the traditional open quotien). Thus, Rd is a more "universal" wave shape parameter, unable to characterise speaker's differences, but very useful in our study to avoid speaker normalisation.

Trajectories in the proximal space are very similar for both items : one can think that glottis control strategies are the same for different voiced fricatives.

Trajectories of Lg and Ho seem to be quite realistic. There is no major actions on the control of pitch, Lg is quite constant during the utterance. Ho increase slightly in both fricatives as a way to maintain voicing [8].

On the contrary, Ps trajectory is more doubtful. In natural speech, the subglottal pressure varies very slowly, which is not the case in the result of the inversion.

The validity of the forward model can be easily checked by measuring Fo, E and Rd values on synthesised speech and comparing this values to those predicted by the forward model. Figure 4 shows that the predicted values are very close to measured ones, which means that the error might occurs in the speech production model. This is confirmed by the measurement of the simulated pressure drop at the glottis during the fricative, when the simulation is run using the inversion results. The simulated value of ΔPg (2.8 cmH_2O for /v/ and 2.2 cmH_2O for /ʒ/) are very close to the measured ones (Table 1). This result means that the two-mass model provides the wanted flow with the right glottis pressure drop, but those values are obtained for lower subglottal pressures (4.5 vs 9.2 for /v/ and 4 vs 7.6 cmH_2O for /ʒ/).

The two-mass model is not the cause of the error, but the modelling of the tract aerodynamic pressure used in the plant is not valid for small constrictions.

This problem is classical with the Kelly & Lochbaum modelling. This model is an acoustic modelling which is applied to low frequencies aerodynamic pressure by the introduction of specific losses. Those losses are subject to controversy because thy are based on very simple assumptions on the air flow through the tract that are not valid for most consonant production.

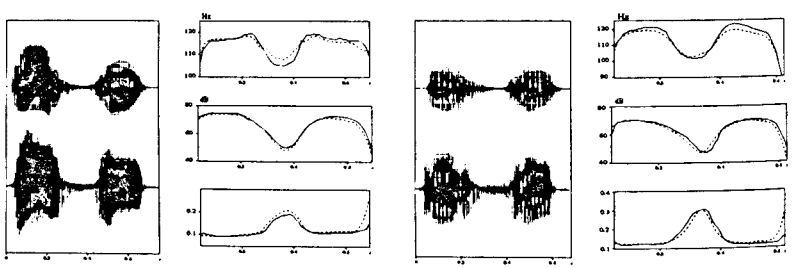

Figure 4 : comparison of estimated and measured values of the flow characterisation parameters for synthesised speech (left /pava/, right /paʒa/).
For each item (left to right and top to bottom) : synthesised speech, natural speech, Fo, E, Rd.
dashed lines : estimated values ; solid lines : measured values.

CONCLUSION

This study shows that the inversion method proposed for the voice source is able to deal with strong source-tract interactions as in voiced fricatives production. The results of the inversion are coherent with traditional knowledge on voiced fricative production and with measurements made on the speaker during the items production.

However, the aerodynamic pressure simulated in the plant, does not correspond to the measurement, and our modelling must be corrected before testing the method on other VCV, especially with unvoiced consonants, where there is a crucial contribution of the source-tract interactions to devoicing.

ACKNOWLEDGEMENT

This work has partially be funded by the European ESPRIT/BR project Speech Maps.

REFERENCES

[1] Bailly G., Castelli E., Gabioud B. (1994) *Building prototypes for articulatory speech synthesis.* Proceed. of the 2nd ESCA/IEEE Workshop on Speech Synthesis, New York, 9-12.

[2] Kelly J.L. & Lochbaum C.C. (1962) *Speech Synthesis* . in Proc. Stockholm-Speech Communications Seminar - R.I.T. 127-130. and 4th Int. Congr. Acoust., G42.

[3] Ishizaka K. & Flanagan J.L. (1972) *Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords.* B.S.T.J., 51, 1233-1268.

[4] Pelorson X., Hirschberg A., Van Hassel R.R., Wijnands A.P.J., Auregan Y. (1994) *Theorical and Experimental Study of Quasi-Steady Flow Separation within the Glottis during Phonation. Application to a modified two-mass model.* J. Acoust. Soc. Am., 96, 3416-3431.

[5] Vescovi C., Castelli E. (1994) *Gestural Supervisor for the Vocal Cords of a Speaking Machine.* In Proceedings of the Fifth Australian International Conference on Speech Science and Technollogy, December 1994, Perth, Vol.2, 613-618.

[6] Fant G., Kruckenberg A., Liljencrants J., Bavegard M. (1994) *Voice source parameters in continuous speech. Transformation of LF-Parameters.* Proceed. of the ICSP, September 18-22, 1994, Yokohama, Japan, Vol. 3, paper S25-4, 1451-1454.

[7] Fant G., Liljencrants J., Qi-Quang L. (1985) *A Four-parameter Model of Glottal Flow.* STL-QPSR 4/1985, 1-13.

[8] McGowan R.S., Koening L.L., Löfqvist A. (1995) *Vocal tract aerodynamics in /aCa/ utterances : Simulations.* Speech Communication, Vol 16, No 1, 67-88.