INTRINSIC VOICE SOURCE CHARACTERISTICS OF SELECTED CONSONANTS

Christer Gobl, Ailbhe Ní Chasaide and Peter Monahan
Centre for Language and Communication Studies, Trinity College, Dublin, Ireland

ABSTRACT

This paper presents data on intrinsic voice source characteristics of selected consonants for an Italian and a French speaker. The analytic method involved interactive inverse filtering of the speech pressure waveform, and measurements were obtained by matching the LF voice source model [1] to the output of the inverse filter. Results broadly bear out expectations that the degree of supraglottal constriction is the major determinant of source quality. Over and above this, differences between results for the two speakers suggest that active strategies may also come into play.

INTRODUCTION

Differences in the voice source were studied for the four consonants /l(:) m(:) v(:) b(:)/ in an intervocalic context. This is part of a more general study on the intrinsic voice source characteristics of vowels and consonants.

Our initial expectation here was that voice source effects could be modelled as a passive consequence of the differences in the supraglottal constriction associated with the different manners of articulation of these consonants.

METHODS

The main analysis technique involved inverse filtering of the speech pressure waveform. In order to obtain quantifiable results, a parametric model of differentiated glottal flow (the LF-model, [1]) was matched to the output of the inverse filter. Both the inverse filtering and the matching procedure were carried out using specially designed interactive software allowing optimisation in both the time and frequency domains [2].

From the matched model a number of parameters were subsequently measured. The ones we focused on particularly were EE, RA, RK and RG. EE is the excitation strength and is measured as the negative amplitude of the differentiated flow at the moment of maximum discontinuity. It corresponds to the overall intensity of the signal, so that an increase in EE amplifies all frequency components. RA is a measure of the return phase (dynamic leakage), which is the residual flow (from excitation to complete closure. The acoustic consequence of the return phase is a steeper spectral slope. A large RA corresponds to greater attenuation of the higher frequencies. RK is a measure of the skew of the glottal pulse: a larger value means a more symmetrical pulse shape. RG is a measure that relates to the duration of the opening branch of the glottal pulse. RK and RG together determine the open quotient, and they mainly affect the levels of the lower harmonics in the source spectrum. For a more detailed description of source parameters, see [3].

MATERIALS

The corpus used in this study was taken from recordings of two informants, one French and one Italian. The materials consisted of nonsense words read in similar carrier frames. For the Italian data, disyllabic nonsense words of the form $^{\text{l}}C_1V_1C_2V_2$ were used in the frame *Dico ---- ancora*. As stress typically falls on the second syllable of a French disyllabic word, we used $^{\text{l}}C_1V_1C_2$ monosyllables for the French nonsense words, set in the frame *Dis moi ---- aujourd'hui*. Note that in the French data, the final consonant of the monosyllabic word (C_2) occurred intervocalically in the carrier frame, thus providing a phonetically similar environment to that of the Italian.

C_2, the main object of our study, was each of the consonants /l(:) m(:) v(:) b(:)/. In Italian, C_2 was a long consonant. The first stressed syllable C_1V_1 was /ba/ in both languages. The unstressed V_2 was a vowel of approximately [a] quality in Italian. Five repetitions of each utterance were recorded resulting in a total of 60 utterances (2 speakers X 6 consonants X 5 repetitions).

RESULTS AND DISCUSSION

Figure 1 shows for both speakers the values for EE, RA and RK for the four consonants, and for 100 ms of V_1.

The sonorants /l(:)/ and /m(:)/

For the lateral and nasal consonants, the differences in source parameter values (compared to the surrounding vowels) were relatively small. The slight increase in RA suggests some increase in the spectral slope for both consonants. One difference between the nasal and lateral consonants seems to lie in the tendency of the nasal to have a more symmetrical glottal pulse shape (higher RK).

For the Italian speaker's nasal consonant, there are major perturbations at the V_1C transition in all parameters. There is a momentary drop in the excitation strength, along with a brief increase in dynamic leakage (RA). Concomitantly, the pulse becomes more symmetrical with a longer open quotient. There is little evidence of a similar effect for the French speaker, other than in the return phase.

All the perturbations observed in the Italian nasal would be consistent with there being a sudden reduction in the transglottal pressure drop. In the approach to the nasal consonant, the velum is likely to be lowered to some degree, resulting in some anticipatory nasal flow. However, if at the instant of oral closure, the outlet through the velar valve is insufficient for the volume of airflow, a momentary rise in oral pressure could result. The brevity of the perturbations suggests that a sudden, actively controlled, increase in velic aperture may occur soon after oral closure. It is also possible that such a sudden increased aperture could, at least partially, come about passively from the heightened oral pressure. Once the velic opening has increased, the transglottal pressure drop may resume something approaching its previous level, resulting in more efficient voicing.

This kind of explanation begs the question as to why a similar effect is not clearly found for the French speaker. We feel that the difference may have to do with the force of articulation used by the two speakers: the Italian speaker spoke with a relatively louder and more forceful voice, whereas the French speaker had a noticeably soft, lax voice. This is a point we shall expand on below.

The obstruents /v(:)/ and /b(:)/

The changes to the glottal pulse (relative to the surrounding vowels) are much greater for the obstruents than for the sonorants. This would of course be expected, given the greater degree of occlusion in the former.

For the fricative there is a gradual decrease in the excitation strength (EE) of the pulse, as well as a large rise in dynamic leakage (RA). The open quotient increases, and the pulse tends to become more symmetrical (RK), although not in every instance.

The most extreme source effects showed up in the stop. At the time of closure there is a sharp reduction in the excitation strength (EE). Although the reduction during the stop is similar for the two speakers (about 15 dB), the speed of the transition from the preceding vowel differs. For the French

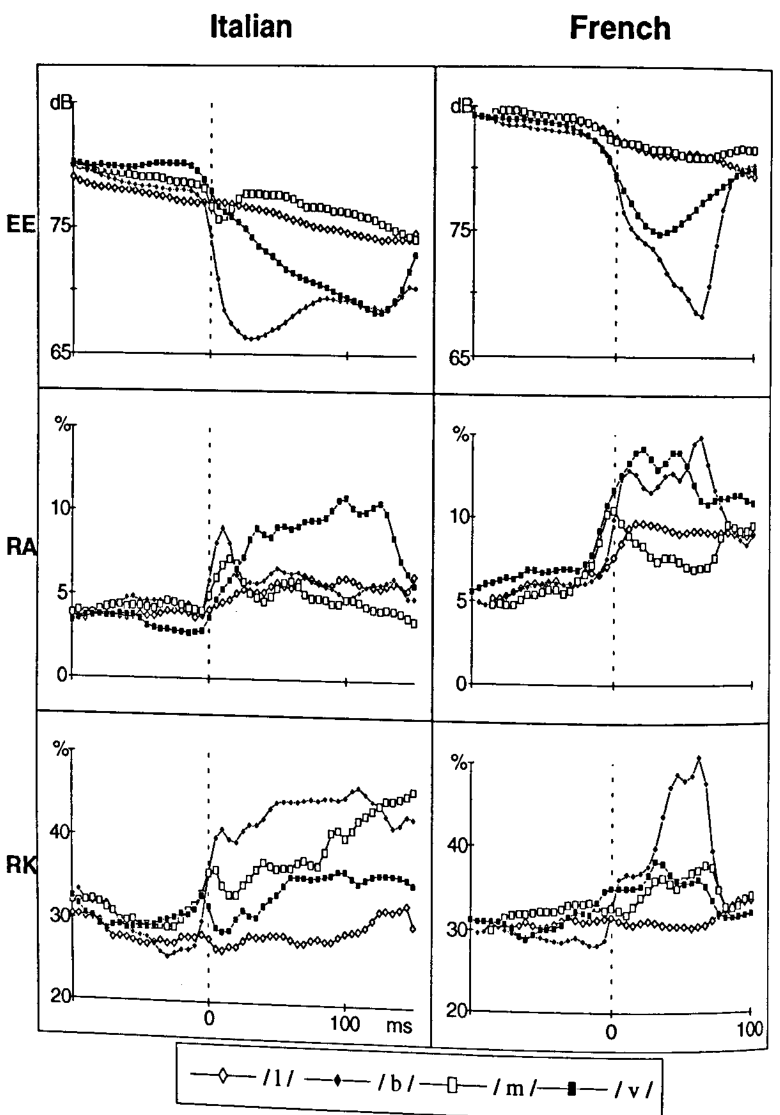

Figure 1. Source values for EE, RA and RK during the four consonants and for 100 ms of V_1. Values aligned to oral closure or onset of constriction for consonant (= 0 ms).

speaker the rate of decay was fairly similar to that of the fricative, whereas for the Italian speaker the decay is much more abrupt than for the fricative. Note also for the Italian speaker that after this sharp initial drop, EE rises somewhat. This is mirrored by a brief rise in f_0 and suggests that some active process is initiated soon after closure in the Italian geminate, which counteracts the other-

wise expected decay in the pulse amplitude. This is presumably some compensatory action to maintain voicing, such as larynx lowering and/or oro-pharynx expansion. Furthermore, for the Italian speaker, there are perturbations around the time of closure, rather like those associated with the nasal.

We would suggest two (possibly complementary) factors might be responsible for these differences in the Italian and French data. First of all, it was mentioned earlier that the Italian speaker spoke in a more forceful style. One would expect a greater force of articulation to involve a more rapid oral closing gesture as well as greater respiratory effort, and a higher rate of flow through the vocal folds. These together should lead to a very rapid decrease in the transglottal pressure drop and a more extreme disruption of the vocal folds' vibratory pattern. This might explain the very rapid fall in EE at closure in the Italian stop, as well as the perturbations to the other parameters. Given this sharp initial drop in EE, the potential for devoicing is greater. Furthermore, the Italian stop here is a geminate: devoicing, due to neutralisation of the transglottal pressure drop is in any case more likely in stops of longer duration. We are therefore hypothesising that the length of the stop and the force of articulation may both conspire to make such an active compensatory adjustment necessary. We should make it clear however that these features are not postulated as necessary features of geminates: in an earlier study, [4] fully voiced geminates of Swedish were found to have a decay pattern more closely resembling the French pattern.

However, in proposing a "force of articulation" difference as being the underlying cause for a number of the differences in the French and Italian data, it remains unclear as to whether this in itself arises out of differences in reading style, possible cross-speaker differences or indeed cross-language differences.

CONCLUSIONS

Results broadly support our initial expectation that source parameters values are directly affected by the degree of supraglottal constriction. Thus, stops show more extreme effects than fricatives, which in turn are more affected than the sonorants. The latter show only minor deviations from the values of surrounding vowels. Of the two sonorants, the lateral was the least affected.

Not all the results can be modelled simply as "passive" source consequences of supralaryngeal occlusion. Some of the effects noted for the Italian speaker suggested compensatory active strategies may sometimes come into play. It is hypothesised that some of the differences between the two speakers may reflect differences in force of articulation.

ACKNOWLEDGEMENTS

This work was supported by Esprit-BRA, no. 6975, SPEECH MAPS.

REFERENCES

[1] Fant, G., Liljencrants, J. and Q. Lin (1985), "A four-parameter model of glottal flow", *STL-QPSR* Vol. 4/1985, pp. 1-13.

[2] Ní Chasaide, A., Gobl, C. and Monahan, P. (1992), "A technique for analysing voice quality in pathological and normal speech", *Journal of Clinical Speech & Language Studies*, Vol. 2, pp. 1-16.

[3] Gobl, C. (1988), "Voice source dynamics in connected speech", *STL-QPSR*, Vol. 1/1988, pp. 123-159.

[4] Ní Chasaide, A. and Gobl, C. (1993), "Contextual variation of the vowel voice source as a function of adjacent consonants", *Language and Speech*, 36, pp. 303-330.