LABORATORY VS SPONTANEOUS SPEECH A CASE STUDY OF SONORANTS

M. Chafcouloff and A. Marchal
URA 261, CNRS Parole et Langage
Institut de Phonétique, Aix-en-Provence, France

ABSTRACT

A study was conducted to analyze the acoustical characteristics that distinguish French sonorants in laboratory vs. spontaneous speech conditions. A dialogue was set up to 'elicit' answers from a speaker who uttered lexical words with /jwlr/ in initial and final position in a vocalic context /i,a,u/. Results showed that steady-state duration is significantly shortened in spontaneous speech, whereas transition duration is less affected by changes in speaking rate and stress. No significant differences in F2 values were found across speaking styles, which means that the concept of reduction does not apply to the production of sonorants in French. Results are then discussed in relation to the target undershoot model.

INTRODUCTION

During the past decades, phonetic research has mostly privileged the use of a peculiar kind of speech, namely 'Laboratory speech', i.e. nonsense words or lexical words uttered in isolation or embedded in a carrier sentence, to study the acoustical characteristics of speech sounds. Yet, deceptive results in text-to-speech synthesis and speech recognition systems have led researchers to conclude that the cues extracted from such speech signals were insufficient carriers of 'real' speech. In other terms, and it was a message clearly expressed during the last ICPhS in Aix-en-Provence, it was urgent to move away from laboratory speech to study a more natural speech.

In the study of the acoustic/phonetic characteristics that distinguish laboratory speech from spontaneous speech, much work has concerned vowel reduction; contrary to the results of some previous studies, it was found that short durations due to a faster speaking rate did not necessarily result in formant undershoot, notably in Dutch [10].

So far, most quantitative data about the acoustical characteristics of the sonorants /jwlr/ has been obtained from the analysis of laboratory speech samples [9,5,2]; for a detailed review of acoustic and perceptual work, see [6].The study of the effects of suprasegmental factors as speaking rate and stress, has led to controversial results.Whereas Klatt [8] reported noticeable formant undershoot in English, Chafcouloff [3] found no significant differences in French. As both studies were concerned with the analysis of laboratory speech items, it is of interest to inquire how these sounds behave acoustically in different conditions of speech.

Actually, several questions may be asked:

-Does a change in speaking style drastically affect the formant structure of sonorants in French?

-Does the concept of reduction apply to the production of these sounds?

-Are there any recurrent acoustical characteristics which may allow to distinguish laboratory speech from spontaneous speech?

METHOD AND SPEECH MATERIAL

In order to build up a solid base of comparable data, a controlled elicitation method of spontaneous speech was used. A question-answer dialogue was set up.The recording took place in an anechoic room, where a speaker of southern French was seated in front of the investigator. The role of the latter was to keep the conversation fluent, and to ask questions until the speaker uttered the 'expected' word. Secondly, the same thirty-one lexical words previouly uttered in the dialogue, were read twice by the speaker. The sonorants were found in the initial and final position, e.g.yak vs. paille, in a vocalic context /i,a,u/ and varying stress position, e.g. 'loup vs. lou'bard.

A listening experiment aimed at assessing the naturalness of the utterances was organized. Two speakers met the requirements. Their speech was judged as typical of a spontaneous speech situation, but the third speaker failed the test. Consequently, the results reported here pertain to the data of two speakers only.

A prosodic analysis was conducted for displaying the Fo configurations of the sentences uttered in spontaneous speech. Three main intonative patterns were used:

1. When the speaker was somewhat wavering, his answer was a question for seeking confirmation. In this case, the word lies at the end of the sentence, and a rising intonative pattern is used (62%)
2. When the speaker enumerated several words which might correspond to the answer, the intonative pattern was usually flat (18%).
3. When the speaker was utterly confident of giving the right answer, a declarative falling pattern was used (20%).

The acoustic analysis was based on the use of an editing program. The utterances were digitized using a 10 Khz sampling rate with 12-bit resolution. Speech signals were pre-emphasized to compensate for weak spectral energy of sonorants at high frequencies. Wideband spectrograms were made and formant frequencies were calculated through FFT and LPC analysis.

RESULTS

Temporal characteristics

Measurements made from oscillographic tracings and spectrograms revealed that lexical words were around 15% shorter in spontaneous speech than in laboratory speech. Average word duration pooled over the two speakers was 292ms for spontaneous speech vs.347ms for laboratory speech. This demonstrated that a faster speaking rate was generally used in spontaneous speech (average 6.6syl./sec.) compared to laboratory speech (5.2 syl./sec.).

Figures 1 & 2 illustrate average steady-state and transition duration values for /jwlr/ across speaking styles. One notices that steady-state portions are significantly shorter ($p < 0.1$) in spontaneous speech than in laboratory speech, and that a 2:1 duration ratio is most often observed. This is especially true concerning the /l/-sound which is characterized by the longest steady-states (>100ms) and the shortest transitions (<30ms). However, if the /l/'s duration is relatively constant across speakers, attention must be drawn on the fact that there is a great deal of intra and inter speaker variation for initial /jwr/ which is not reflected on the figure. Steady-state duration of /j/ and /w/ measured from other speech items in spontaneous speech may be as short as 20ms which merely corresponds to 2 or 3 glottal pulses along the time axis; conversely, it may be as long as 80ms when the word is uttered with a strong emphatic stress. Likewise, the steady-state of /r/ varies as a function of the relative duration of a schwa-vowel initial segment; this variety is often found in the allophones of /r/ in southern French [4].

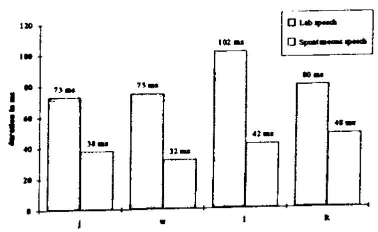

Figure 1. Steady-state durations. Mean duration is pooled over 2 syllable positions, 2 speakers and 3 vowel contexts.

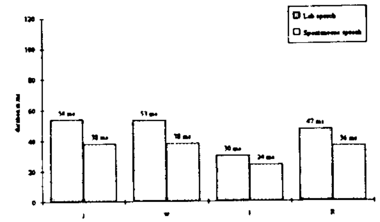

Figure 2. Transition durations. Mean duration is pooled over 2 syllable positions, 2 speakers and 3 vowel contexts.

As far as duration of transitions is concerned, it should be noticed that the transitions of /l/ are shorter than that of /jwr/ in both speaking styles (differences are significant at p< 0.5).Concerning the ratio between steady-state and transition, whereas the steady-state portion of /l/ is usually twice longer than the transition, it turns out that the transitions are of approximately equal duration for /jwr/ in spontaneous speech, which is not the case in laboratory speech.While the lateral's spectrum is essentially static, the glides' is mainly dynamic. Thus, it appears that the transition is affected to a lesser degree than the steady-state by changes in speaking rate. In relation with this, the correlation coefficients are small for /l/ (r (8)= 0.567) and high for /jwr/ (r (8)=0.799, p<0.1).

Spectral characteristics

Differences in terms of vocalic space along a F1/F2 dimensional plane for /jwl/ are illustrated on Figure 3. As the /r/-sound was mostly produced as a fricative allophone with a predominant noise source and no clear-cut formant structure, the results pertaining to the /r/-sound have not been included. As shown by the closeness of the points on the chart, it can be observed that these are clustered in three relatively compact areas, and that the acoustic distance separating the white from the black squares is mostly short. Thus, it seems that neither stress, nor speaking rate exert a decisive influence on the formant frequencies, as no statistically significant differences were found between F2 values for sonorants uttered in different speaking styles.

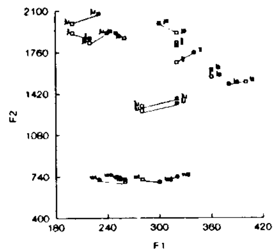

Figure 3. Acoustic distance between /jwrl/ uttered in laboratory (□) vs. spontaneous speech (■).

In these conditions, is it to say that there are no changes at all concerning the production of sonorants? As a matter of fact, a closer examination reveals subtle modifications in the acoustic spectrum. As mentionned above, /j/ and /w/ may be uttered in spontaneous speech as very brief segments. From our data, it appears that /j/ may be produced as a voiced obstruent especially in the /i/ context which is not the case in laboratory speech. Moreover, /j/ is characterized by asynchronous movements in the F-pattern especially at the third formant level. This observation brings credit to earlier remarks from other authors who have noticed the complex temporal evolution of formants. This is true for /j/ [1], for /l/ [7], and for /w/ [2]. Contrary to laboratory speech, the release of the /l/-sound is often characterized in spontaneous speech by a transient click in the upper part of the spectrum. This remark is consistent with the earlier observations of Dalston [5] who suggested that a noise transient associated to a rapid release of the apex away from the alveolar ridge might be an important cue for the identification of this sound. In addition, we have found several cases of formant continuity/discontinuity in the /l/'s spectrum both as a function of position and speaking style. Lastly, as the /r/-sound is most sensitive to contextual

effects [4], it is evident that any changes in speaking styles should be followed by subsequent changes in its acoustical structure. Observations have revealed changes in terms of predominance of an harmonic spectrum vs. a noise spectrum as a function of coarticulatory effects. As the /r/'s duration is shortened, the number of flaps across its stationary portion may be similarly affected. Finally, a retroflex allophone of /r/ was found in spontaneous speech, especially in a back-vowel context.

CONCLUSION

In the state of research, it must be acknowledged that the results presented here are limited in that our data pertain to a single utterance of thirty-one speech items uttered spontaneously by two speakers. Nevertheless, preliminary results indicate that the temporal characteristics of sonorants undergo changes as a function of different conditions of speaking rate and stress. However, these durational changes do not result in any systematic differences in formant frequencies especially at the F2 level. The fact that the sonorants' acoustic targets remain essentially unchanged, implies that the concept of reduction does not apply to the production of sonorants in French. This statement is not necessarily at variance with Klatt's findings who reports significant neutralization of formant target cues for /wjlrh/ in English [8]. Because of the basic tense-lax opposition between the two languages, one should expect that English sonorants tend to be more reduced than their French counterparts.

Moreover, as no frequency differences were found, despite the fact that that segmental duration was generally shorter in spontaneous speech, we may conclude that our results do not support the target undershooot model and its refined versions. Instead, they agree with the results of Van Son and Pols [10] who found no measurable relation between vowel duration and F2 frequency values in normal and fast speaking conditions. Thus, it seems that this model may not apply to all speech sounds across languages and also may not be valid for all speaking styles, especially in spontaneous speech.

REFERENCES

[1] Carlson,R. and Nord, L. (1990) "Analysis and Synthesis of Swedish sonorants-Part 2", Phonum (1), 70-73, University of Umea

[2] Chafcouloff, M. (1980) "Les caractéristiques acoustiques de /jwlr/ en français", Travaux de Phonétique de l' Institut de Phonétique d'Aix, 7, 7-56

[3] Chafcouloff, M. (1982) "L' effet du débit de parole sur les caractéristiques de /jwlr/", Travaux de Phonétique de l' Institut de Phonétique d'Aix, 8, 163-187.

[4] Chafcouloff, M. (1986) "Quelques variantes de /r/ en français méridional" Travaux de Phonétique de l' Institut de Phonétique d'Aix, 10, 101-120.

[5] Dalston, R.M. (1975) "Acoustic characteristics of english /wrl/ spoken correctly by young children", Journal of the Acoustical Society of America 57, 462-469.

[6] Espy-Wilson,C. (1992) "Acoustic measures for linguistic features distinguishing the semivowels /jwlr/ in American English", Journal of the Acoustical Society of America 92 (2), 736-757

[7] Fant, G. (1960) "Acoustic Theory of Speech Production", Mouton, the Hague.

[8] Klatt, D.H. (1974) "Acoustics characteristics of /jwlrh/", Journal of the Acoustical Society of America, 55 Suppl.1, Paper G11, p. 397.

[9] Lehiste, Ilse (1964) "Acoustic characteristics of selected English consonants", International Journal of American Linguistics 29, 1-197.

[10] Van Son, R.J. and Pols, L.C. (1992) "Formant movements of Dutch vowels in a text read at normal and fast rate", Journal of the Acoustical Society of America 92 (1), 121-127.