"CAN WE LEARN SOMETHING FROM NONSENSE WORDS?"

Alain Marchal and Sophie Lapierre

CNRS URA 261 Parole et Langage
Aix-en-Provence, France

ABSTRACT - Experiments on speech production using nonsense words allows for a fine control of the phonetic, linguistic and prosodic variables which can interact in a speech sequence; but it is questionnable wether results obtained from these carefully designed experiments bear any significance for the understanding of the processe involved in the production of less artificial speech items. Data for this study has been extracted from the multisensor ACCOR database. We have investigated the production of /l/ by two French speakers in nonsense words, isolated words and sentences. First, the various signals have been annotated using a multitiered phonetic approach whereas articulatory, acoustic and aerodynamic data are marked at major signal discontinuities. These events are then interpreted as landmarks of articulatory gestures. The spatio-temporal organization of these "gestures" for the production of /l/ is compared across speech styles and across speakers. Our results suggest that more data is needed to account for aerodynamic requirements in the production of /l/.

1.INTRODUCTION

Until recently, most of the studies on speech production have relied upon acoustic and articulatory data from nonsense words. This type of speech material allows for a fine control of linguistic and prosodic variables which interact in a speech sequence, but it is questionable whether results obtained from these carefully designed experiments bear any signifiance for the understanding of the process involved in the production of other speech items such as words, sentences... In other words, can we learn something about speech from nonsense words?

2.METHODOLOGY

Data for this study has been extracted from the multilingual EURACCOR database (Marchal et al. (1991); Hardcastle et al. (1992)). This database consists of simultaneous digital recordings of the acoustic soundwave, of the laryngograph signal, of oral and nasal airflow and of linguo-palatal contacts. Multisensor data has been collected for the production of VCV nonsense words, isolated words matching phonetically the nonsense words and the same words embedded in sentences. The speech items have been repeated 10 times at a normal rate. We have analysed here the production by two french female speakers ("ad", "gc"; 20 ~ 25 years old; no sociogeographic marks of pronunciation or speech defect) of the sequence /ulu/ in the nonsense word " oulou "and in " Toulouse ", the french town, as isolated word and " Toulouse " imbedded in the sentence: " La cousine de Vichy épousa un hippie à Toulouse ".

Labelling generally refers to the process of aligning a symbolic description of some speech items to the physical signal itself. Segmentation refers to the division of the acoustic waveform into "segments" of various sizes. However, these processes are problematic since there is no one-to-one correspondence between the physical and the linguistic levels of representation: boundaries between "neighbouring" segments (as defined from a symbolic transcription) are blurred by the speech encoding process and phonemes as discrete abstract units are not represented as such in the speech chain. Various approaches have been proposed attempting to relate the symbolic levels (syntactic, phonemic, phonetic, etc.) to the acoustic signal for a given utterance. However, researchers disagree as to both the optimal number of levels of description as well as the criteria used to identify landmarks in the physical signal.

Speech is the output of a production process which relies for its execution on coordinated gestures. For the purposes of the ACCOR project, we have decided that "segmentation" should reflect articulatory timing rather than specific events in the acoustic signal. Therefore, we should not simply describe acoustic events, but we should interpret them through a kind of grammar specifying the relations between subjacent gestures. What needs to be identified are articulatory events and not segments.

Consequently, we have adopted a multitiered phonetic approach to speech labelling (Marchal et al., 1995), where signal discontinuities are interpreted as the result of coordinated articulatory gestures (Nicolaidis et al., 1993). The various acoustic, aerodynamic and articulatory signals available in the ACCOR database are annotated independently. The present study relies on EPG data only. The following landmarks have been identified: the start of the forward movement of the tongue, the lateral closure, the maximum constriction and the lateral release. They are annotated respectively as ACE, LCE, MCE and LRE. In addition, the beginning and the end of lingual activity is labelled GOE and GEE. The data from these annotation points are used as the basis for the subsequent spatial and temporal analyses. Three phases in the articulation of /l/ can be distinguished. They correspond to intergesture intervals:

Approach to constriction:	/ACE-LCE/
Lateral constriction:	/LCE-MCE/
Release of constriction	/MCE- LRE/

For temporal analysis, durations of the phases are derived from the EPG frames adresses. For spatial analysis, the EPG data is taken respectively from the annotated frames. The annotation of the speech material used in this study has been made by an expert highly trained phonetician. The following criteria have been used to mark the EPG data:

ACE: *Approach to constriction, approach to closure*: The marker is placed at the beginning of the movement of approach to constriction as revealed by the linguo-palatal contact pattern. The segmentation of an ACE is done by a progressive and regressive scanning of the EPG signal, the target consonant must already be known.

LCE: *Lateral closure*: This mark corresponds to the first frame that has any of the four midsagittal electrodes activated in the first four rows.

MCE: *Maximum closure or constriction*: This point marks the first EPG frame showing the largest number of electrodes touched for a given constriction or closure.

LRE: *Lateral release*: This mark corresponds to the last frame with constriction across the four midsagittal electrodes.

This study deals with the spatio-temporal organisation of gestures for the production of /l/. EPG patterns at ACE, LCE, MCE and LRE have been analysed as an indication of the amplitude of the tongue tip gesture. The temporal organization of the gestures is given by the durations between these marks. They correspond to the following phases: approach (ACE-LCE); closure (LCE-MCE) and release (MCE-LRE). EPG data has been statistically analyzed using the paired t-test and the ANOVA linear regression method.

3.RESULTS

3.1.Spatial Organization

The amplitude of the articulatory gesture has been estimated from the number of contacts for each electrode row. We can observe in the figure 1 how the contacts are distributed on the hard palate during MCE for the three speech context. This representation gives a global view of the linguo-palatal contacts repartition for ten repetitions (Fig.1):

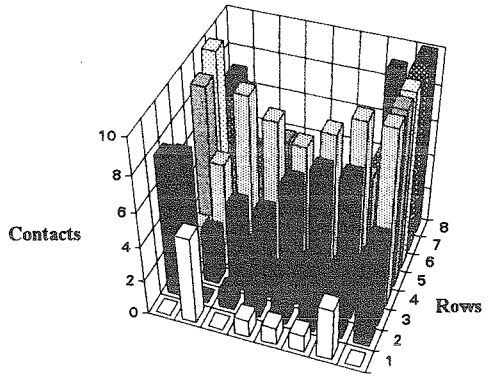

Figure 1:Total of activated electrodes at first frame of maximum linguo-palatal contact (MCE) for the production of ten repetitions of /i/ by speaker "ad" in nonsense words.

On a hyper- to hypo-continuum and in the framework of the H & H theory, the following prediction can be made: the amplitude of the lingual gesture is expected to be larger for the nonsense words than for real words, and it should be the smallest for the sentence context (Farnetani, 1991). The amplitude of the tongue gesture can be estimated from the number of activated electrodes. Measurements were made at the point of maximum contact MCE and at ACE and LCE for the three speech types: Total number of linguo-palatal contacts, number of contacts by rows and by palatal areas (A=alveolar; B=prepalatal; C=palatal; prevelar, as shown in Fig. 2):

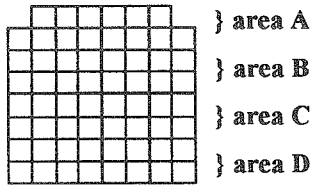

Figure 2: Area delimitations of the EPG frame

As far as the general spatial organization is concerned, two remarks can be made: 1) There is no significative difference between the number of linguo-palatal contacts between nonsense words, words and sentences for both speakers in each context. This means that they both

reach similar spatial targets in terms of general amplitude of the lingual gesture.; 2) Concerning the contexts, the difference which is observed between the mean contact number for each articulatory landmark is more important for the nonsense words than for the other speech contexts. For example, the mean difference between the number of contacts of ACE and LCE in sentences for "gc" speaker (Fig. 3) is 7.3 contacts against 11.7 in nonsense words (23.3 cts - 16 cts against 23.1 - 11.4 cts), but, contrary to Farnetani (1991), these differences are however not significative.

When we consider the number of contacts in the various palatal areas, the same tendancy can be observed for the alveolar and prepalatal regions. We note that the gesture amplitude for the nonsense word is not significantly different from the other contexts, but it is suggested that the nonsense context differs most from the sentence context and suggest the following decreasing order of gesture amplitude: nonsense word / word / sentence contexts.

3.2.Temporal Organization

The first part of the temporal analysis consisted in comparing the total duration of /l/ as a function of the given contexts (Fig.3). The ANOVA analysis of variance indicates a significative difference between the total duration of the articulation for three contexts, $F(27.2)$, $p < .001$ for "ad" speaker, $F(26.7)$, $p < .001$ for "gc" speaker. As could have been expected, we observe the shortest duration for /l/ in the sentence, then by increasing order in the real word and in the nonsense word.

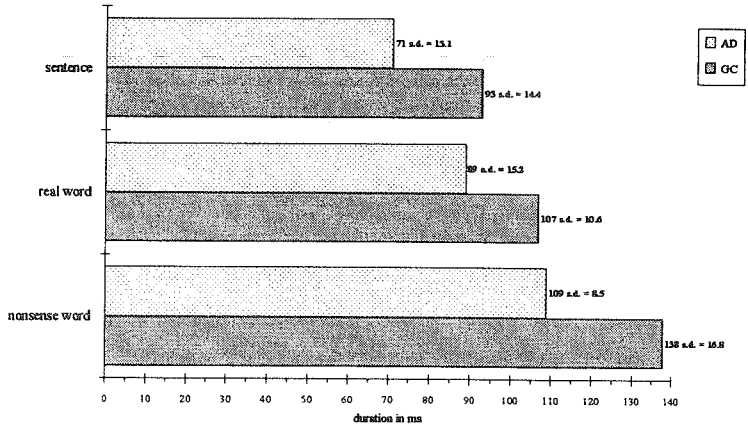

Figure 3: Mean total duration of /l/ in French for speakers "ad" & "gc" from EPG data

The question arises to know if the variation of the total duration which was observed as a function of context affects equally or not each phase. If the ratio duration of phases / total durations of /l/ is kept constant, this would imply that the internal organisation of the various gestures involved in the production of /l/ are not altered as a function of the context. The duration of each phase, being proportionally increased or decreased from sentences to nonsense words. Since the amplitude is not affected (previous observations), this would imply

that there exists a saturation effect and that the intended lingual gesture is in fact masked by competing demands on the articulator. An alternative hypothesis would explain the observed facts as an internal reorganization of the various phases. This is indeed what our analysis reveals for speaker "ad": the relative duration of phase 1 and phase 2 is primarily concerned. There is a clear shortening of the approach to constriction in the sentence context for speaker "ad" but not for speaker "gc".

Table I: Analysis of variance of phases duration as a function of contexts and speakers

Mean total duration of phases in ms for speaker "ad"

	Sentence	Real word	Nonsense word
ACE-LCE	12.7	23.9	22.9
LCE-MCE	20.7	28.5	36.8
MCE-LRE	38.2	36.8	49.7

Mean total duration of phases in ms for speaker "gc"

	Sentence	Real word	Nonsense word
ACE-LCE	29.7	33.7	38.9
LCE-MCE	31.5	35.6	47.4
MCE-LRE	32.6	37.8	52.5

p values for raw data between the two speakers

	Sentence	Real word	Nonsense word
ACE-LCE	.002	.015	.015
LCE-MCE	.020	.033	.036
MCE-LRE	.329	.617	.774

p values for normalized data between the two speakers

	Sentence	Real word	Nonsense word
ACE-LCE	.009	.129	.079
LCE-MCE	.359	.475	.602
MCE-LRE	.051	.110	.144

In order to examplify more clearly this difference in articulatory organization, we have adopted a phase representation. This representation (Fig. 4 & 5) uses phase/total duration ratio. Each angle indicates the relative timing of each phase translated into degrees. The circumference corresponds to the total duration of the consonant. A degree representation is adopted as a mean to represent on the same graph the total and relative intergestural durations for /l/, and it shows in more obvious way the differences between the speech contexts and the speakers.

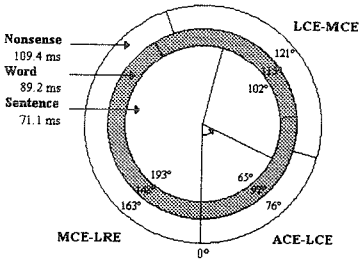

Figure 4: Degree representation of articulatory phases for /l/: speaker "ad"

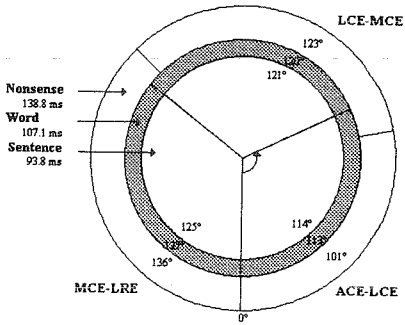

Figure 9: Degree representation of articulatory phases for /l/: speaker "gc"