IS PHONEME INVENTORY A GOOD PREDICTOR FOR VOCAL TRACT USE IN CASUAL SPEECH?

Christine Meunier and Robert Espesser

Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, France
christine.meunier@lpl-aix.fr, robert.espesser@lpl-aix.fr

ABSTRACT

The aim of this study is to provide a general overview of the expected articulatory positions in French according to phoneme inventory, phonemes in words and phonemes in conversational speech. Three datasets were analyzed and relative frequencies of articulators and articulatory dimensions are given. The results show that tongue body is used most frequently. Palatal and velar dimensions are the most frequent constriction locations. No strong differences are observed between the three datasets. This suggests that the use of articulators and dimensions is correctly balanced within words and more especially within high frequent words in speech.

Keywords: conversational speech, lexicon, phoneme inventory, articulators, dimensions.

1. INTRODUCTION

Each language differs according to the specificities of its phonological inventory. For example, several Arabic dialects are characterized by very back positions (uvular, pharyngeal, or glottal consonants), while Swedish has numerous front vowels (protruded or not). In French there are no very back consonants (except /ʀ/, uvular) and a third of the vowels are front while another third are protruded. Theses particularities provide a general representation of how the vocal tract is used in these languages. Nevertheless, does this representation faithfully describe the set of articulatory positions in vocal tract during the production of casual speech?

Indeed, phoneme inventory attributes the same weight to each phoneme properties while phonemes are diversely present in words (for instance, /ʀ/ is very frequent while /z/ is quite rare). Furthermore, phoneme occurrences within words have to be balanced by token frequency in casual speech. Consequently, to what extent the representation of articulatory positions in phoneme inventories is noticeably modified by the relative frequency of phonemes in words and the use of words in casual speech?

The aim of this study is to provide a general overview of the expected articulatory positions in French according to phoneme inventory, phonemes in words and phonemes in conversational speech.

To achieve this aim, we compare articulatory positions in phoneme inventory to those extracted from a lexical database and from a corpus of conversational speech. To represent articulatory positions, we adopted the terminology given by Articulatory Phonology [1]. AP gestures provide an abstract representation of the use of vocal tract during speech. Acoustic or articulatory features are not enough precise for this study (tongue body and tongue tip are not distinguished). Furthermore, we are more specifically interested by the frequency of some articulatory positions than by the fact that all phonemes have to be distinguished by features. Nevertheless, our study does not provide information on real articulations within speech production and it neither provides information about dynamic articulation. In fact, this is not an articulatory study. Our analyses are based on phoneme frequencies within three datasets. Only expected positions are examined here.

2. SPEECH MATERIAL

Articulatory positions are extracted from three different datasets: the French phoneme inventory, a database of French words and a corpus of French dialogues in conversation.

2.1. French phoneme inventory (Pi)

French phonological system contains 34 phonemes (15 vowels and 19 consonants, Table 1). Vowels are either front or back. /ə/ is the only one central vowel. Back vowels are protruded and half of the front ones are protruded too. Then, rare vowels such as /y/ or /ø/ (front and protruded) are present in French and a third of the vowels are protruded in French. French vowels come in four degree of constriction (open, mid open, mid close and close).

The phonological distinction between mid close and mid open vowels is not robust since their realizations depend on words, syllable structure and regional accents. Only six consonants (three plosives and three fricatives) are voiceless and six phonemes are nasals (four vowels and two consonants). There are six plosives and six fricatives. /l/ is the only one lateral and /ʀ/ represents the backest position (uvular).

Table 1: French vowels and consonants.

Vowels	i y u e ø o ɛ œ ɔ a ɑ̃ ɛ̃ œ̃ ɔ̃ ə
Consonants	p t k b d g f s ʃ v z ʒ m n l ʀ j w ɥ

2.2. French words (Wo)

Phoneme frequencies are calculated on a French word database [2] which contains 339,813 phonemes and 48,150 lemmas. In our study, we used lemmas and did not consider inflected forms since these forms may reflect a nearer representation of word use, and this information is already contained in the corpus of conversational speech. The mean ratio is 7.06 phonemes per word, which suggests that the mean size of words is quite long.

2.3. French conversational speech (Co)

The CID (Corpus of Interactional Data, [3]) consists of eight hours of audio/video recordings of French conversational speech. Each hour is a recording of a relaxed conversation between two participants. The conversations took place in a studio and the participants were given two topics to discuss: conflicts in their professional environment or funny situations that they had found themselves in. The corpus was transcribed orthographically by two advanced phonetics students who had to specify (with a special code) the deletions, insertions and reductions they could ear. This initial transcription was then processed by a grapheme-to-phoneme converter [4].

The corpus contains 301,918 phonemes and 120,613 words. The mean ratio is 2.5 phonemes per word, which suggests that the mean size of words is clearly shortest than in word database.

These three dataset are obviously different. From Pi to Wo, phonotactics of French and syllabic characteristics change the representation of phonemes. From Wo to Co, word use and frequency lead to an unbalanced representation of words. Most of the words used in conversational speech are monosyllabic function words (*pas* 'not', *mais* 'but', *et* 'and', etc. [5]). This suggests that vocal tract use may be affected by this unbalanced use of words.

3. METHOD

We explain here how articulatory positions are deducted from phoneme properties and how position frequencies are extracted from phoneme frequencies in each dataset.

3.1. From phonemes properties to articulatory positions

AP proposes to express the phonological representation of speech sounds by a set of *tract variables and dimensions* (Table 2). The variables are: Lip Protrusion (LP), Lip Aperture (LA), Tongue Tip Constriction Location (TTCL), Tongue Tip Constriction Degree (TTCD), Tongue Body Constriction Location (TBCL), Tongue Body Constriction Degree (TBCD), Velum (VEL) and Glottis (GLO). The *dimensions* specify the location of the organs: closed, wide, palatal, etc.

Table 2: Tract variable and dimensions used in this study. Tongue Constriction Degree (TTCD and TBCD) and Lip Aperture (LA) are not examined here.

Tract variables	*dimensions*
VEL	closed, wide
GLO	closed, wide
TTCL	alveolar, postalveolar
TBCL	postalveolar, palatal, velar, uvular
LP	protruded

In our study, each phoneme of French is detailed by a set of dimensions for each variable tract (Table 3). Obviously, this description reflects prototypical forms and expected productions but not concrete realizations.

Table 3: Some examples of dimensions affected to French phonemes.

	LP	TTCL	TBCL	VEL	GLO
n	-	alveolar	-	wide	closed
k	-	-	velar	closed	wide
b	-	-	-	closed	closed
ʃ	-	-	postalveolar	closed	wide
u	protruded	-	velar	closed	closed

3.2. Definition of dimension frequencies in each dataset

For each dimension in each dataset, the dimension frequency is defined as follow:

(*occurrence number for a dimension*) / (*number of phonemes in the dataset*) (Pi = 34 ph.; Wo = 339,813 ph.; Co = 301,918 ph.).

The occurrence number for a dimension is obtained by adding values of phonemes for which the dimension is relevant.

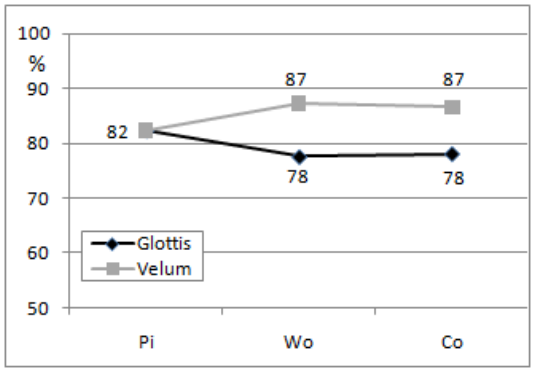
4. RESULTS

Each figure represents the frequency (in percentage) of each tract variable and each dimension according to the three different datasets. To provide comparable figures, values are plotted in a range of 50% (from 0% to 50%, or from 50% to 100%).

4.1. Glottis and Velum

Glottis and Velum were relevant for all phonemes. For both of them the specification was either "wide" or "closed". The "close" position was the most frequent one for Glottis and for Velum (82% in Pi). The frequency of close Glottis slightly decreased within Wo and Co datasets (Fig. 1). This means that unvoiced consonants are quite frequent in the effective speech use. At the opposite, close Velum was more frequent in these two datasets, suggesting that nasal phonemes are less used.

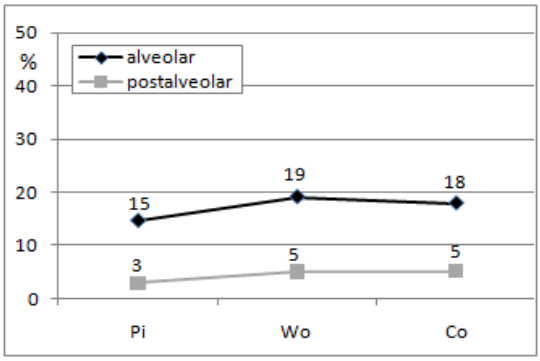
Figure 1: frequency (in %) of close position for Glottis and Velum in the three datasets.



4.2. Tongue Tip Constriction Location

Tongue Tip represented 18% of gestures in Pi. It increased in Wo (24%) and in Co (23%). Alveolar dimension was more frequent than postalveolar one in the three datasets for TTCL (Fig. 2). We also observed that both positions increased in Wo and Co. This may be due to the high frequency of /t/, /s/ and /l/ within these datasets.

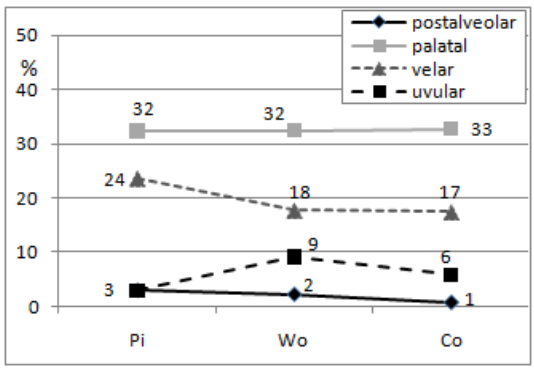
Figure 2: frequency (in %) of TTCL dimensions in the three datasets.



4.3. Tongue Body Constriction Location

Tongue Body was involved very frequently in our three datasets of French (62% in Pi, 61% in Wo and 57% in Co). Four dimensions were specified for Tongue Body with unequal frequency. Palatal was the most frequent one and it did not change through the three datasets (Fig. 3). Velar dimension is also frequent but a decrease is observed in Wo and Co. The high frequency of these dimensions was due to vowel articulation (either palatal or velar). Postalveolar dimension was the less frequent position. Uvular has a very low frequency in Pi but increases in Wo (9%) and, to a lesser extent, in Co. Uvular position is represented by only one phoneme, /ʀ/, which is the most frequent phoneme in Wo and Co.

Figure 3: frequency (in %) of TBCL dimensions in the three datasets.

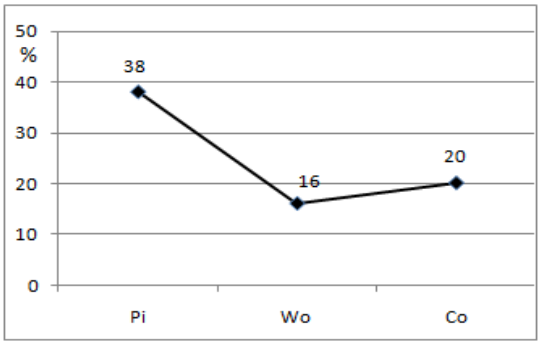


4.4. Lip Protrusion

Lip protrusion provided the greatest changes between Pi and both other datasets (Fig. 4). Protrusion was highly present in Pi, because a third of all French vowels are protruded. Nevertheless the most frequent vowels used in French are front and non protruded vowels (/a/, /e/, /i/). Consequently, the high representation of protrusion

in French inventory (38%) is strongly minored in lexicon and speech use.

Figure 4 : frequency (in %) of protruded position in the three datasets.



5. DISCUSSION

5.1. Articulators and dimensions in French

The first aim of this study was to highlight the relative frequency of articulatory positions in French. Our results suggest that tongue body is more frequently used (about 60%) than tongue tip (about 24%) in French.

Palatal (in TB) is the most frequent dimension, followed by velar dimension (TB) and alveolar one (TT). These articulatory tendencies underlie different configurations of the vocal tract: alveolar dimension is only produced with consonants while palatal and velar dimensions concern both consonants and vowels. These configurations will be examined in a further study including Constriction Degree. Other dimensions are rarer: even if TT and TB are merged, the frequency of postalveolar dimension remains low (about 6%). Uvular dimension is special since /ʀ/ is the only one phoneme concerned. Nevertheless, its high frequency in words (9%) suggests that tongue body is quite often in back position during speech production.

Constriction Degree is examined through Glottis and Velum. The close position is clearly the most frequent position for both articulators. Obviously, these are “expected articulation” and the effective realizations may show different patterns.

5.2. Articulators in lexicon and speech use

Surprisingly, and except for lip protrusion, and to a lesser extent for velar dimension, the general articulatory tendencies observed in phoneme inventory are preserved in words and conversational speech. Indeed, we expected that phonotactics and syllable structure would increase articulator over- or under- representations in word database. Furthermore, we expected too that the over-representation of some short and very frequent function words (such as *pas* ‘not’) in speech conversation would increase the changes observed in word database. This was actually the case but not in the dimension we expected. This suggests that, even if the frequency of phonemes in words is unbalanced, articulators and dimensions are sufficiently balanced through words to keep stable the use of vocal tract. This suggests too that phoneme inventory is a good predictor for the use of vocal tract in real speech. Of course, this has to be checked with studies on other languages.

Surprisingly again, no strong differences are observed between words and conversational speech for articulator frequencies. We expected that the few short words extremely present in conversational speech, would provide a high weight to their phonetic and articulatory composition. In fact, this may be the case, suggesting that phoneme inventory may be distributed in a well-balanced way within frequent words. This may explained that in some cases (see figures 1, 3, 4) values of Pi are closer to those of Co than to those of Wo.

This study is a first step of a more general one in which different languages will be compared with the same procedure. We expect to evaluate phoneme inventories as predictors of vocal tract use during speech in real conditions.

6. REFERENCES

[1] Browman, C.P.; Goldstein, L. 1992. Articulatory phonology: an overview. *Phonetica* 49 (3-4), 155–180.

[2] Vanrullen, T., Blache, P., Portes, C., Rauzy, S., Maeyhieux, J.F., Guénot, M.L., Balfourier, J.M., Bellengier, E. 2005. Une plateforme pour l'acquisition, la maintenance et la validation de ressources lexicales. *Proceedings of the TALN Conference*, June 2005, Dourdan, France, 511-516.

[3] Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. 2008. Le CID. Corpus of Interactional Data. Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues*, 49 (3), 105-134.

[4] Di Cristo, A., and Di Cristo, P. 2001. Syntaix, une approche métrique-autosegmentale de la prosodie, *Traitement Automatique des Langues*, 42 (1), 69–111.

[5] Meunier, C., & Espesser, R. Vowel reduction in conversational speech in French: The role of lexical factors. Journal of Phonetics (2011), doi:10.1016/ j.wocn.2010.11.008.