IMPACT OF DURATION AND VOWEL INVENTORY SIZE ON FORMANT VALUES OF ORAL VOWELS: AN AUTOMATED FORMANT ANALYSIS FROM EIGHT LANGUAGES.

Cédric Gendrot and Martine Adda-Decker***

*LPP Université Paris Sorbonne Nouvelle CNRS UMR 7018 ILPGA
** LIMSI-CNRS bât. 508, BP 133, 91403 Orsay cedex
cgendrot@univ-paris3.fr, madda@limsi.fr

ABSTRACT

Eight languages (Arabic, English, French, German, Italian, Mandarin Chinese, Portuguese, Spanish) with 6 differently sized vowel inventories were analysed in terms of vowel formants. A tendency to phonetic reduction for vowels of short acoustic durations clearly emerges for all languages. The data did not provide evidence for an effect of inventory size on the global acoustic space and only the acoustic stability of quantal vowel /i/ is greater than that of other vowels in many cases.

Keywords: vowel formants, adaptive dispersion, reduction, quantal theory.

1. INTRODUCTION

The present research focuses on cross-language patterns of variation of vowel spectral quality (frequency of the first, second and third formants, hereafter F1, F2 and F3) as a function of vocalic inventory size and phonetic vowel duration. This investigation makes use of an automated analysis of large speech corpora and aims at contributing to the study of the complex issue of phonetic and phonological universals.

1.1. Theory of Adaptive Dispersion

The Theory of Adaptive Dispersion (TAD, [9]) makes predictions as to the effect of inventory size on the acoustic distribution of elements of vowel systems. TAD claims that the demand for sufficient contrast between elements in a vowel inventory will lead to an "adaptive dispersion" of these elements; in particular, as stated by [4] in her hypothesis, large inventories should tend to expand the acoustic vowel space (e.g., F1/F2). The results of previous studies have been somewhat controversial, some corroborate this hypothesis (for example [3] for Arabic dialects and French while others found less straightforward results ([4], [8]). In this study, data from several languages with differently sized vowel inventories have been obtained in order to shed further light on this question.

According to Lindblom's TAD, the aim of talkers is to produce sounds sufficiently contrastive to promote linguistic comprehension by the listener. The theory has thus been extended to account for within speaker variation. For example, Moon and Lindblom [12] show that under circumstances that require clear speech, a speaker's vowel space will be expanded relative to his or her casual speech vowel space.

1.2. Influence of phonetic vowel duration on formant values

We refer here to the Hyper- & Hypoarticulation model [11], for which hypoarticulated forms are produced with unreached targets ("undershoot") while hyper-articulated phonemes fully achieve their articulatory or acoustic target ("overshoot"), both being respectively related to shorter or longer segment duration.

Our hypothesis can be detailed as follows: in English (and generally in languages with lexical stress such as German, Spanish, Italian and Portuguese) where the vowels of unstressed syllables typically undergo reduction (neutralization to /ə/, with some exceptions, notably for /i:/), the prediction is that length should be strongly correlated with vowel spectral quality (as reflected in the frequencies of the first two formants). Much more so than in a language such as Chinese, where the differences in length across lexical tones (e.g. the brevity of the falling tone, tone 4, as opposed to the rising tone, tone 2) is not reported to have any salient effect on vowel spectral quality. As for Arabic, word stress is not phonemically contrastive but bears a strong relationship to vowel length: there is a tendency for "heavier" syllables to be favoured as stress-bearing syllables (i.e. syllables of longer duration - a closed syllable or a syllable with a long vowel). Since vowel length is phonologically contrasted, we expect fewer reduction phenomena. It is also expected that in French, which has no lexical stress, and is reputedly a "syllable-timed" language, there will be less influence of vowel length on vowel spectral quality and thus less phonetic reduction phenomena.

1.3. Aims and hypotheses

Our study focuses on cross-linguistic similarities in the relationship between the measured vowel length and the frequency of formants, as well as the global acoustic space involved for these languages with differently sized vowel inventories.
According to the Quantal Theory of Speech (QTS) [13], there are regions of stability in the phonetic space corresponding to the "quantal" vowels /i/, /a/ and /u/. Thus this theory predicts that the "quantal" vowels should be in approximately the same locations across all languages, regardless of the vowel inventory size, and that they should show less within category variability than non point vowels.
The main objectives of this study are to assess: (i) whether there is a general tendency of formant variations with phonetic segment duration in the eight languages; (ii) to what extent the observed patterns can be related to the vowel inventory size of these eight languages; and (iii) whether "quantal" vowels /a/,/i/,/u/ are less prone to variation than other vowels.
The present study is made possible by the availability of large audio corpora and powerful automatic tools used for speech alignment.

2. METHOD

2.1. Corpus selection

Speech data come from annotated corpora of broadcast speech in eight languages: Arabic, Mandarin Chinese, American English, French, German, Italian, European Portuguese, and Spanish. All corpora gather around 25000 oral vowels each (approx. 2 hours); they correspond to radio and TV journalistic shows: articulation, without being emphasized, remains quite distinct, so that speech can be understood by a broad audience. Reduced vowel phenomena are undoubtedly less important in these corpora than in conversational-style spontaneous speech.
The LIMSI speech transcription system was used for corpus alignment [2] and formant extraction makes use of the Burg algorithm implemented in the PRAAT software. F1, F2 and F3 frequencies were extracted automatically (following a procedure set out in [7]).

The considered languages vary in terms of numerous factors. In particular concerning vowel inventories, some have contrastive vowel rounding, length and/or nazalization. Tones of Mandarin Chinese's vowels have been pooled as their effect on formant frequencies was judged negligible for our study. Let's recall that Italian is usually considered as having seven vowel phonemes: /a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/, but the pairs /e/-/ɛ/ and /o/-/ɔ/ are often confused.

Table 1: Number of vowels according to the language separated into peripheral and others.

language	peripheral vowels	other vowels	Tot. number
Spanish (Eur)	i u o a e		5
Italian	i u o a e		5
Arabic (litt)	i u a	ɪ ʊ æ	6
Mand. Chinese	i y u o a ɛ		6
Portuguese (Eur)	i u o ɔ a ɛ e	ə ɯ	9
French	i y u o ɔ a ɛ e	œ ø	10
English (US)	iː uː o ɔ ɑ ɜː ɛ	ɪ ʊ ə æ	11
German	iː yː uː eː øː ɛː oː ɔ aː	ɪ ʏ ʊ a ø œ	15

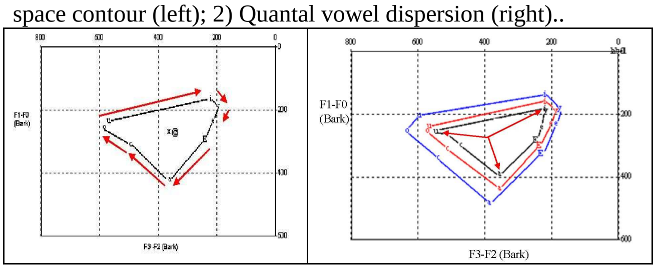
2.2. Normalisation procedures

As pointed by Disner [6], it is not advisable to use normalisation procedures in order to compare markedly different phonological systems, especially with procedures that put emphasis on the mean and standard deviations (like Nearey's and Lobanov's).
Raw measurements in Hertz seem reasonable to use for analysis as soon as large groups of speakers and data are available: inter-speaker and more particularly male/female differences are expected to average out on large corpora.. Nonetheless, Syrdal & Gopal's normalisation procedure (1986) is used as different male/female ratios are found in our languages' corpora. Formant frequencies are first converted into Bark units, using Traunmüller's formula [15]. Secondly, it is based on information contained within a single vowel (intrinsic) but employs relations between spectrally adjacent (fundamental and) formant frequencies: F1-f0 for aperture dimension (in place of F1) and F3-F2 for the front-back dimension (in place of F2 supposed to be more reliable than F2). According to Adank [1], it performs well at reducing anatomical/physiological differences. In our opinon, it also allows to take into account the third formant variations ($F2^{prime}$ has also been measured showing similar results) which is specifically non negligible for high front vowels /i/ and /y/ in several languages (e.g. French and German).

2.3. Acoustic measurements and durations catégories

- To check whether the acoustic space expands with an increasing vocalic inventory sizem we propose to measure the vocalic space contour based on peripheral vowels[1]. It is illustrated by Figure 1 and is measured as follows: Euclidean distances between adjacent peripheral vowels in the acoustic space were measured and summed. This measure depends both on the shape and on the size of the acoustic vowel space and thus is expected to depend on the inventory size. It is also assumed to be more accurate than a triangle contour in more square shaped acoustic vowel spaces such as English or German.

Figure 1-2: illustration of introduced measures: 1) Vocalic space contour (left); 2) Quantal vowel dispersion (right)..



- *A measurement of* "degree of dispersion" *distance from acoustic center* (is illustrated in Figure 2. This measurement was realized using a technique applied in [5]: the acoustic center of gravity for the vocalic diagram is detected as the mean of F1 and F2 values (respectively F1-f0 and F3-F2 in our case) of the "quantal" vowels /a/, /i/ and /u/. The euclidean distance from each vowel's acoustic value to the acoustic center of gravity is finally meaned.

- *Durations of vowels are classified into 3 categories*: short (strictly below 60ms,), intermediate (between 60 and 80ms) and long (strictly above 80ms). These intervals have been determined so as to obtain the most regular distribution of all duration intervals for all languages (cf. Table 2).

Table 2: Proportion of vowels according to the established duration categories (in %).

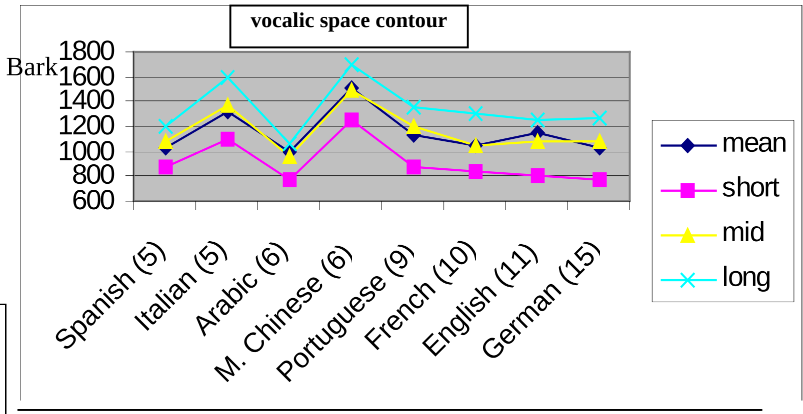
dur	Arab	Eng	Fr	Ger	It	Chn	Por	Sp
short	49	27	50	45	59	31	50	44
mid	33	28	30	33	25	35	23	35
long	18	45	20	22	17	34	28	21

3. RESULTS

Figure 3 shows the sum of the Euclidean distances in Bark between the peripheral vowels plotted as a function of inventory size. It can be seen that vocalic spaces contours do not display any coherent tendency to grow with inventory size.

[1] Vowels are considered as peripheral when allowing a maximum vocalic acoustic space while minimizing the space contour.

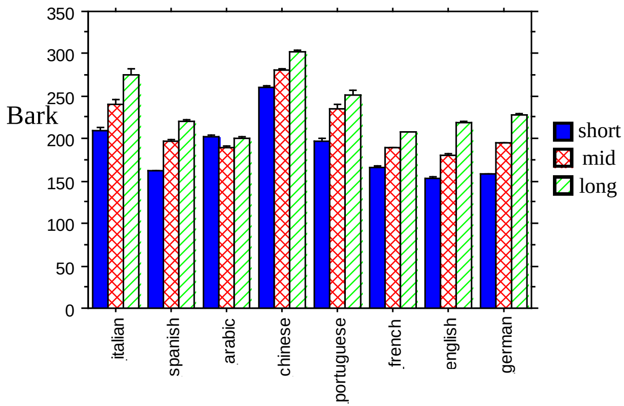
Figure 3: Vocalic space contour of each language for all duration categories and mean values. Between brackets stands the number of oral vowels in the language.



We can notice however identical tendencies for each language : long vowels show a larger vocalic space contour than vowels of short duration and the mean value is very close to the intermediate category.

The measurement of vowels' distance from acoustic center (vowel dispersion) is illustrated in Figure 4 and shows that contrary to expectations, vowel reductions are observed significantly and with a comparable degree in 7 languages (French, Spanish, German, European Portuguese, Italian and English). This suggests that reduction is not an exclusively linguistic phenomenon, but admits also explanations of a physical or physiological nature. Variations have also been measured for Arabic although not showing a tendency to reduction for vowels of short duration as expected in the introduction.

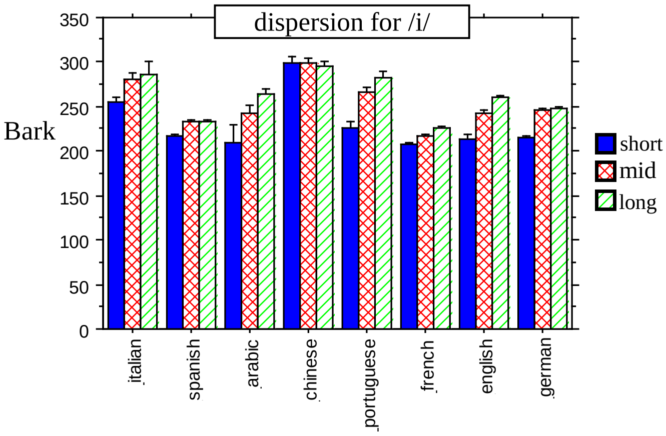
Figure 4: Distance (dispersion) from acoustic center for each language for all duration categories (all vowels pooled).



As can be seen on Figure 5, the reduction phenomena observed from vowels of long duration to vowels of short duration are of less magnitude for the vowel /i/ for (compared to vowels /a/ and /u/, see Figures 14 and 15 in appendix-document joined and compared to all vowels meaned, see

Figure 4). This is specifically the case for French, Spanish, Chinese, Italian and German.

Figure 5: Distance (dispersion) from acoustic center for each language for all duration categories (for "quantal" vowel /i/).



As can be seen in Figures 5 to 12 in Appendix (joined document), all the measured vowel formants of the corpus occupy the vocalic space in an organised manner: for shorter durations the vocalic triangle shrinks progressively resulting in "concentric" surfaces. The vocalic surfaces undergo a centripetal movement towards some central vowel position. The formant values for central vowels remain somewhat stable whatever the duration variations.

4. DISCUSSION AND CONCLUSION

The presented study allows us to put forward answers to the questions listed in the introduction: (i) a tendency to reduction for vowels of short duration clearly emerges for all languages, with notably less magnitude for Arabic. (ii) The present measurements have not provided clear evidence either for an effect of inventory size on the global acoustic space between peripheral vowels as predicted by TAD. This negative result could be interpreted such that languages use additional means for accommodating elements in crowded vowel spaces. It has been shown for consonants [10] that large systems tend to use new articulatory dimensions and it seems reasonable that large vowel systems are also based on analogous mechanisms using dimensions such as nasality, diphthongisation or voice quality. Indeed diphthongs, which are quite common in languages like Mandarin, English and German, but not French or Italian, were not included in our measurements. The exact relationship between inventory size and utilisation of more elaborate vowel types hence remains to be explored. It is also likely that diachronic factors should be taken into account. (iii) As for the validation of a higher acoustic stability for "quantal" vowels /a/, /i/ and /u/ as postulated by QTS, the results presented here only suggest a higher stability for vowel /i/ in five languages. /a/ is reputedly a variable vowel, but the acoustic variability found for /u/ might be best understood if we consider it a central articulatory constriction as stated in particular by Vaissière [16], the stability of the low second formant being mainly due to lips rounding.

5. REFERENCES

[1] Adank, P.M. (2003). Vowel normalization: A perceptual-acoustic study of Dutch vowels. *Ph.D. thesis*, U. of Nijmegen.

[2] Adda-Decker, M., Boula de Mareüil, P., Adda, G., Lamel, L. (2005). « Investigating syllabic structures and their variation in spontaneous French », *Speech Comm.* 46 (2),119–139.

[3] Al-Tamimi, J. & Ferragne,E. (2005) Does vowel space size depend on language vowel inventories? Evidence from 2 Arabic dialects and French. *proc. Interspeech*, Lisbon, 2465-2468.

[4] Bradlow,A.R.(1995). A comparative acoustic study of English and Spanish vowels. *J. Acoust. Soc. Am.*, 97 (3), 1916-1924.

[5] Bradlow, A.R., Torreta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and finegrained acoustic-phonetic characteristics. *Speech Comm.*, 20, 255-272.

[6] Disner, S.F. (1980) Evaluation of vowel normalization procedures. *J. Acoust. Soc. Am.* 67(1): 253-261.

[7] Gendrot, C. & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Proc. Eurospeech* Lisbon, 2453-2456.

[8] Jongman, A., Fourakis,M. & Sereno, J.(1989). The acoustic vowel space of Modern Greek and German. *Lang.Speech* 32, 221-248.

[9] Lindblom B (1986) Phonetic universals in vowel systems. In: Ohala JJ, Jeager JJ (eds). *Experimental Phonology*. Orlando, FL: Academic Press, pp. 13-44

[10] Lindblom, B. & Maddieson, I (1988) Phonetic universals in consonant systems. In L.M. Hyman and C.N. Li (eds.), *Language, speech and mind. Studies in honour of Victoria A. Fromkin*. pp. 62-78. London: Routledge

[11] Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H & H theory. In W. Hardcastle & A. Marchal (eds.). *Speech Production and Speech Modelling*. Dordrecht: Kluwer.

[12] Moon, S-J. & Lindblom, B. (1989). "Formant undershoot in clear and reference speech: a 2nd progress report," *Speech Transm. Lab. Quar. Prog.. and Stat. Report* 1/89 pp. 121-123.

[13] Stevens KN (1989) On the Quantal nature of speech. *J Phonetics* 17: 3-45

[14] Syrdal, A.K., & Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels *J. Acoust. Soc. Am.* 79, 1086-1100.

[15] Traunmüller H (1990) Analytical expressions for the tonotopical sensory scale, *J. Acoust. Soc. Am.*, suppl 1, vol. 88, pp 97-100, 1990

[16] Vaissière (in press) Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic and perceptual properties of sounds across languages In 'Experimental Approaches to Phonology'. Oxford University Press. Maria Josep Sole, Pam Beddor & Manju Ohala editors.