

VOCALIC HESITATIONS VS VOCALIC SYSTEMS: A CROSS-LANGUAGE COMPARISON

Ioana Vasilescu, Rena Nemoto, Martine Adda-Decker

LIMSI-CNRS, BP133, F91403 Orsay Cedex France

ioana.vasilescu, rena.nemoto, madda@limsi.fr

ABSTRACT

This paper deals with the acoustic characteristics of vocalic hesitations in a cross-language perspective. The underlying questions concern the “neutral” vs. language-dependent timbre of vocalic hesitations and the link between their vocalic quality and the phonemic system of the language. An additional point of interest concerns the duration effect on vocalic hesitations compared to intra-lexical vowels. Acoustic measurements have been carried out in American English, French and Spanish. Results on vocalic timbre show that hesitations (i) carry language-specific information; (ii) whereas often close to measurements of existing vowels, they do not necessarily collapse with them. Finally, (iii) duration variation affects the timbre of vocalic hesitation and a centralization towards a “neutral” realization is observed for decreasing durations.

Keywords: vocalic hesitation, vocalic systems, timbre, duration, centralization

1. INTRODUCTION

Hesitation phenomena can carry different functions in speech and they contribute to the elaboration of the verbal message, although are often viewed as “disfluencies”, thus noisy and irregular events, in opposition with “fluent” or “well formed” speech [7]. According to [3], their main function consists in the initiation of a delay in speaking. Fillers can receive various realizations, as for instance vocalic hesitations (such as *uh/um* in American English, *eh* in French, *eh* in Spanish), lengthened nasal consonants (*mm* in Mandarin [11]) or demonstratives (*ano, eto* in Japanese [10]). We focus here on filled pauses corresponding to vocalic hesitations, widely encountered in world’s languages.

Previous studies on eight languages have shown some evidence about the universal vs language-specific acoustic and prosodic properties of vocalic hesitations [2]. Thus pitch and duration are to be related to some universal patterns [6, 9], whereas timbre exhibits language-specific properties and raises the question of the relation between the hesitation vowel and the vocalic system of a language.

Contrastive studies of vocalic systems across different languages recently received a new impulse thanks to automatic techniques which allow the extraction of intra-lexical vocalic segments and reliable estimations of their acoustic and prosodic properties from large corpora [4]. It has been shown that the vocalic “triangle” obtained with automatically measured formant values describes a shape very close to the one based on controlled laboratory speech with few speakers. An interesting outcome affects the behaviour of the vocalic “triangle” according to segment duration: vocalic triangles established for different segment durations contract towards some central position as duration falls.

In this paper we investigate the link between the vocalic hesitation and the phonemic system of a language in American English, French and Spanish. In [2] it has been emphasized that even though the hesitation vowel generally exhibits a timbre close to a central vowel, this realization does not seem to be a universal “rest position”, but rather a language dependent realization. In the following, we address the questions: (i) does the timbre of the vocalic hesitations correspond to a vowel of the phonemic inventory or (ii) does the vocalic hesitation tend to a given, potentially universal central position or another timbre not attested in the vocalic inventory of the language? Additionally, (iii) we consider the duration effect on vocalic hesitations compared to intra-lexical vocalic segments.

2. CORPORA AND METHODOLOGY

For the current study, we use the following data.

- in French: 20 hours of speech from several national radio and TV broadcast channels (*France Inter, France Info, France2...*);
- in English: 10 hours of similar recordings from different broadcast sources (CNN, VOA, ABC, etc.) distributed by LDC;
- in Spanish: 10 hours of European Parliament debates.

Whereas several dozens of speakers, both male and female, cover the majority of audio data, the findings presented here are based on male speakers only, as

they are roughly twice as frequent as female speakers. Speaking styles in journalistic shows and parliament debates can be considered as comparable, characterized by a rather distinct articulation as the speech is dedicated to a broad audience. It mostly includes (semi)-prepared speech which entails less hesitation marks than in fully spontaneous speech. We hypothesize that observed hesitations tend to be relatively standard and prototypical, which is an interesting property for first contrastive comparisons.

The extraction methodology makes use of automatic speech transcription techniques [5]. Timbre analysis with first, second and third formants F1, F2, F3 has been conducted using Praat [1]. Two formant filtering strategies have been adopted in order to eliminate potential detection errors. The first filtering is based on the voicing ratio. For each segment a voicing ratio has been computed as the ratio between the number of voiced frames and the total number of frames (the frame rate is 5ms). Segments with a voicing ratio > 0.4 are selected. The second filtering follows the method described in [4] and is based on canonical formants values described in the literature. Rejected segments represent less than 5% of the data. Pitch and formant measurements correspond to segment centers and have been computed as a mean of the 3 central voiced frames. Segment durations have also been stored. The analysis considers only oral monophthong vowels.

3. TIMBRE OF HESITATION VOWEL VS VOCALIC SYSTEMS

In this section we first investigate the timbre of hesitation vowels in terms of F1 and F2 measures across languages. Next, for each language the hesitation vowel is examined with respect to the intra-lexical vowels of each system. Automatic alignment yielded approximately 1300, 2000, and 1700 fillers in the French, English and Spanish corpora respectively. Presented results correspond to male speakers, although similar tendencies have been measured for female speakers.

3.1. Timbre of hesitation vowels

As already highlighted in the introduction the transcription of vocalic hesitations varies across languages, for example *uh/um* in American English, *eu*h in French or *eh* in Spanish suggesting differences in their perception by native speakers. In order to evaluate if those empiric observations can be correlated with acoustic measures, formant values have been extracted using Praat. Figures 1, 2 and 3 present mean F1/F2 values obtained for American English, French and Spanish, showing distinct

timbres for the hesitation vowel (encoded as "&") in each language and confirming that it does not necessarily result in a central realization [9]. Hesitation vowel timbre moves from mid-open in English to mid-closed in Spanish with a central realization in French. The measured F1/F2 values are consistent with the expected formant values of the orthographic transcriptions of hesitation vowels in each language. They also show differences between hesitations across languages. Inter-language specificities of vocalic hesitations have been statistically evaluated with one factor ("language") ANOVA: we found a strong effect of language on the timbre of vocalic hesitations both for F1 (F(102.8), $p < 0.0001$) and F2 (F(83.8), $p < 0.0001$). English vocalic hesitations are the most open and also the most posterior on F2 axis, compared to French and Spanish, the latter exhibiting similar degrees of mid-closed opening. Acoustic measures and subsequent statistical analyses support the hypothesis of a language-dependent timbre of vocalic hesitations.

3.2. Link between hesitation vowel and vocalic systems

In the following, we evaluate the timbre of vocalic hesitations in the framework of their vocalic systems. To do so, formant values have been extracted and average values computed for all intra-lexical vowels in the three languages. Figure 1 shows the resulting vocalic system of French in a F1/F2 space. The vocalic hesitation ("&") is thus represented in the F1/F2 space together with the other vowels of the language inventory. This representation displays an objective average distance between hesitation vowels and the phonemic vowels. Similarly Figures 2 and 3 show hesitation vowels in the vocalic systems for English and Spanish.

The comparison between vocalic hesitations and the vocalic systems highlights a common pattern: the vocalic hesitation is close or almost collapses with an intra-lexical segment. We checked this point statistically with ANOVAs with one factor "segment" (using as nominals "hesitation" vs "intra-lexical" segment) and we considered as intra-lexical segments to be compared the central and/or open vowel(s) with similar timbre in each language. In French the timbre is close to a central position and could be associated to /ə/ or /œ/, however the hesitation vowel is significantly more open (F(261.4), $p < 0.0001$) and more anterior (F(151.4), $p < 0.0001$) than its intra-lexical neighbour (we considered /ə/ and /œ/ together). Whereas perceptually the vocalic hesitation and intra-lexical vowel /e/ coincide in Spanish, statistical analysis shows that vocalic hesitation is both more open (F(47.5), $p < 0.0001$)

and more central on F2 axis ($F(465.0)$, $p < 0.0001$) than the intra-lexical /e/. The open character of hesitation vowels observed in the three languages might be related to duration effects (see next section), whereas the strong centralizing movement on F2 axis observed in Spanish might potentially be related to a “neutral position pressure” which affects hesitation vowels in different manners in terms of acoustic realization. Finally in American English, the average timbre of the hesitation vowel is close to /ʌ/ and /æ/, i.e. an open/central position. However, differences among the three vowels are statistically weaker compared to French and Spanish.

The relation between the vocalic hesitation and language inventory varies thus across languages. The average timbres of vocalic hesitations are close to a central, mid-open realization in French and American English, but close to a front mid-closed vowel in Spanish. They also exhibit language-dependent differences when compared with intra-lexical vowels with similar timbres.

4. HESITATION VS VOCALIC SYSTEM: DURATION EFFECTS ON TIMBRE

In this section we address the question of duration effects on hesitation vowels compared to intra-lexical vowels. According to [4] vocalic systems tend to a more central configuration as the duration of the segments decreases. Conversely, they tend to be more peripheral on canonical positions with increasing durations.

Vocalic hesitations are generally described as perceptually and objectively beyond the average duration of intra-lexical vocalic segments [6], [8]. Whereas intra-lexical vowels do not exceed an average value of 60ms, hesitation vowels exhibit contributions in the range of 150 to 250ms thus beyond the upper limit of intra-lexical vocalic segments (about 150ms). Data have been divided into four duration ranges, and average formant values have been computed for each duration range (Table 1).

The movement of hesitation vowels towards a central realization is indicated with an arrow in Figures 2 and 3. Observations made by [4] about the vocalic systems of 8 languages are confirmed here. Figure 4 details the average centralization tendency of the vocalic system and the hesitation vowel in Spanish for different duration ranges (concentric curves correspond to decreasing duration ranges, as given in Table 1). All the measured mean formants values occupy the vocalic space in an organised manner. The centralization patterns are different from one language to another, but the general

Table 1: Distribution in percentage of vowel V and hesitation & segments for English, Spanish and French according to duration. **XShort:** [30-40ms] for V, [30-120ms] for &; **Short:** [50-60ms] for V, [130-190ms] for &; **Medium:** [70-90ms] for V, [200-340ms] for &; **Long:** [100+ms] for V, [350+ms] for &; V: lexical vowel; &: hesitation vowel.

		Rate of segments per duration			
	Type	XShort	Short.	Medium.	Long
En	V	24	21	22	33
	&	1	13	50	36
Sp	V	42	27	17	13
	&	15	20	32	33
Fr	V	27	30	24	19
	&	12	14	34	40

behaviour of the vocalic systems seems to be universal.

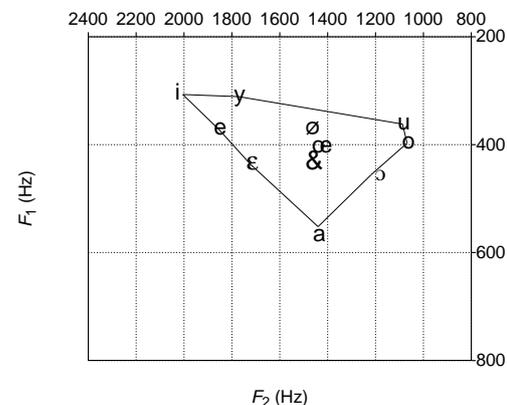


Figure 1: French F1/F2 mean values for intra-lexical vowels and hesitation vowel (“&”).

A centralisation movement. is also observed for hesitation vowels according to decreasing durations. The general tendency is to reach a central position but the pattern is different across languages. For French, the realizations are quite stable for the vowels &, /ə/ and /œ/ as they are positioned in the center of the triangle. In Spanish, the hesitation vowel describes a movement towards a central position on F2, whereas the F1 axis is not exploited. The timbre of the hesitation vowel is also more central than its closest neighbor /e/. However, the vocalic hesitation does not seem to reach a “neutral” position even for particularly short segments. Finally, the hesitation vowel in American English exhibits the most evident centripetal movement. This movement is probably related to its open character and suggests that the more hesitation vowels are peripheral the more important is the centralization effect. In [4] a strong statistically significant effect of duration on intra-lexical vowels has been observed: the

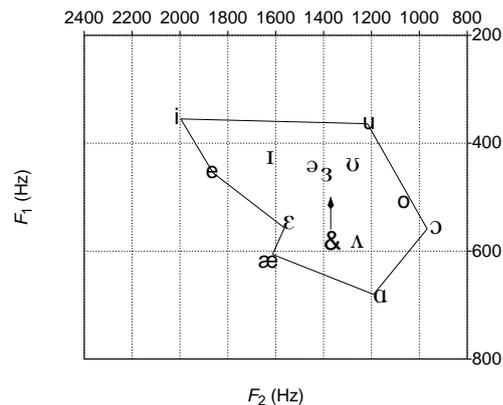


Figure 2: English F1/F2 mean values for intra-lexical vowels and hesitation vowel (“&”).

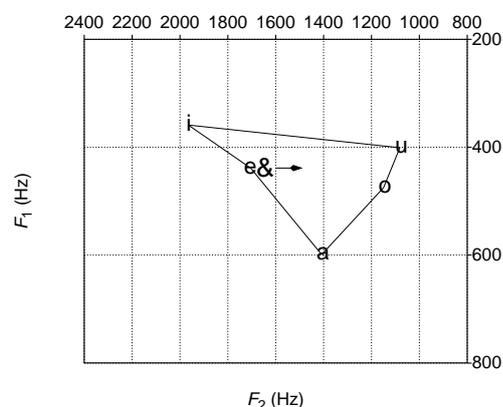


Figure 3: Spanish F1/F2 mean values for intra-lexical vowels and hesitation vowel (“&”).

more the duration decreases the more the vowels are moving towards a central target. We evaluated the same effect on hesitation vowels: centralization, even though present, is weaker than for intra-lexical segments and the duration factor does not generate a statistically significant centralization effect.

5. DISCUSSION

This paper focused on acoustic characteristics of vocalic hesitations in a cross-language perspective. We aimed at exploring the relation between the vocalic hesitations and the languages' vocalic inventories. Results on timbre quality show different realizations of vocalic hesitations across languages and support the hypothesis of language-specific realizations which do not necessarily favour a central “neutral” position. Vocalic hesitations exhibit also similarities with some intra-lexical vowels but the degree of similarity is language-dependent. Finally, their timbre varies according to duration and the vocalic surfaces contract towards a more central position as segment duration falls. This movement is similar but less prominent to the one observed for intra-lexical vocalic segments. Further studies can be carried

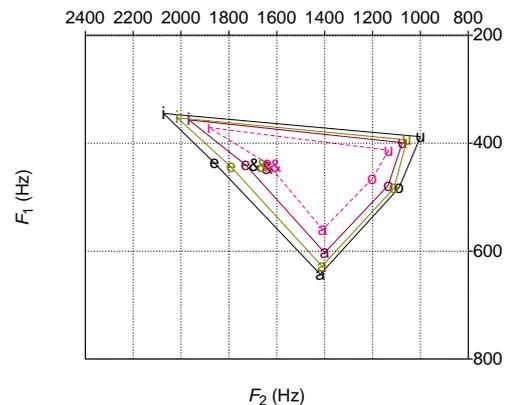


Figure 4: Spanish F1/F2 mean values for intra-lexical vowels and hesitation vowel (“&”) according to duration.

out in the future including other languages and accents, other speaking styles for a more in-depth understanding of hesitation phenomena.

6. REFERENCES

- [1] Boersma P., Weenink, D. 1999. Praat, a system for doing phonetics by computer *Institute of Phonetic Sciences of the University of Amsterdam*, 132–182.
- [2] Candea, M., Vasilescu, I., Adda-Decker, M. 2005. Inter- and intra-language acoustic analysis of autonomous fillers *In DiSS-05*, 47–51.
- [3] Clark, H. H., Fox Tree, J.E. 2002. Using *uh* and *um* in spontaneous speaking *Cognition* 84, 73–111.
- [4] 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 *Proceedings of Eurospeech-Interspeech*, Lisbon.
- [5] Gauvain, J.L., Lamel, L., Adda, G. 2002 The LIMSI Broadcast News Transcription System *Speech Communication* 37(1-2), 89–108.
- [6] Shriberg, E.E. 1994. Preliminaries to a Theory of Speech Disfluencies *PhD thesis*, University of California at Berkeley.
- [7] Maclay, H., Osgood, C. E. 1959. Hesitation phenomena in spontaneous English speech *Word*, 15, 19–44.
- [8] Vasilescu, I., Adda-Decker, M., 2006. Language, gender, speaking style and language proficiency as factors characterizing autonomous vocalic fillers in spontaneous speech *Proceedings of ICSLP*, Pittsburgh, USA.
- [9] Vasilescu, I., Adda-Decker, M., 2006. A cross-language study of acoustic and prosodic characteristics of vocalic hesitation In: NATO Security Through Science Program, *"The Fundamentals of Verbal and Non-verbal Communication and the Biometrical Issue"*, IOS Press BV.
- [10] Watanabe, M. et al., 2005. The effects of filled pauses on native and non-native listeners speech processing *In DiSS-05*, 169–172.
- [11] Zhao, Y., Jurafsky, D., 2005. A preliminary study of Mandarin filled pauses *In DiSS-05*, 179–182.