BACK CLOSE NON-SYLLABIC VOWEL [u] BEHAVIOUR IN EUROPEAN PORTUGUESE: REDUCTION OR SUPPRESSION?

Maria João Barros, Daniela Braga, Diamantino Freitas, João Paulo Teixeira, Vagner Latsch*

Fac. de Eng. da Univ. do Porto, *ESTG – Inst. Politéc. de Bragança

mjbarros@fe.up.pt, dbraga@fe.up.pt, dfreitas@fe.up.pt, vagner@fe.up.pt, joaopt@fe.up.pt

Abstract

In this paper we analyse and describe the phonetic behavior of the non-tonic back closed vowel [u] in European Portuguese (EP) in what respects its quality and relative duration. Slow and faster speech was used. The objective of this work is to contribute with experimental results in order to obtain a more accurate comprehension of this vowel's profile, which can be integrated and extended to other non-tonic vowels' behaviors in the EP non-tonic vowel system. This may represent an important improvement in synthetic speech quality and naturalness, as it concerns acoustical parameters, rhythm and supra-segmental features, which means prosody. This study is inscribed in a project in Speech Synthesis for EP held by an inter-disciplinary group in intimate articulation between the engineering experience and tools and the linguistic approach.

1. Introduction

Back Closed Non-tonic vowel [u] [1] is peacefully accepted and integrated in the Portuguese Sound Pattern by Portuguese Generative Phonology [2]. The same doesn't happen with the Central Closed Non-tonic vowel [ɨ] (that is /@/ in SAMPA), since, for many linguists or phoneticians it isn't more than a "schwa", which means that it has not enough phonological consistence to allow it to be part of the Portuguese phonological frame.

However, from our experience on phonetic labeling [9] and construction of a database for Speech Synthesis, we found out that non-tonic vowels are often reduced, unvoiced, assimilated by their phonetic context or even non- present in continuous oral utterances.

This is a widely spread phenomenon that affects the whole non-tonic vocalic system in natural speech. Despite the irregularity of the phenomenon, the most affected cases are: non-syllabic [ɨ] and [u], then followed by [i] and [ɑ] in a lowest rate of occurrence. Our study will be focused on [u] since this vowel is extremely important to speech synthesis naturalness and because there is still lack of information about it.

We started by assuming that in fact total suppression almost never occurs. This assumption is reinforced when comparing opposite minimal pairs, which only vary in one phone (e.g. "pont**e**/ pont**o**"), which we think we can hear, in a perceptual and cognitive perspective [3]. The phonological distinctive meaning and conceptual image of the heard phonetic material interferes with an objective analysis. Perceptually, it seems that different vowels are heard, because of the syntactic distribution or the semantic association.

Further research made us admit that [ɨ] and [u] non-tonic vowels weren't perhaps very distinctively articulated. Spectral analysis confirmed that we are facing an interesting phenomenon of *neutralization*, in a structuralist theoretical framework [4]. As the Brazilian linguist, Mattoso Câmara, explains it, "neutralizations" occur when "certain significant phoneme oppositions lose their distinctive value" [5]. The consequence of the neutralization phenomenon is the disappearance of one of the opposite phonemes and its replacement by the other in every context of the phonological system.

This event may be included in the general trend of reduction and closing of the non-tonic vowels in EP, started in the XVI century, as many documents and grammarians testify. P. Teyssier [6] refers that "progressive weakness of the EP non-tonic vowels started in the past is an evolution in progress" that is still affecting the non-tonic vocalic system.

2. Methodology

2.1. Linguistic Corpus

The corpus was specifically and carefully built in order to enable the study of non-tonic [u] phonetic distribution. Two sections, one with fast speech and a second one with slow speech, compose it. Each section presents the vowel [u] in a non-tonic syllable (e.g. "**Por**tugal" / "co**po**"), in combination with every consonant of EP phoneme. All the papers we are aware of report studies only at a consonant type level. The design priority of this corpus was the contexts variety, with full coverage, in order to contribute to the broadening of this type of material.

In both sections non-final and final positions in the target word were used. In this study we haven't considered this vowel position towards the tonic syllable, since our purpose was, in one hand, to find out more about the acoustical profile of the permanent reduced back closed non-tonic [u], and in the other hand to determine which influences does this phoneme suffer in fast speech, considering all possible contexts.

2.2. Recording

Speech from two male speakers between 26 and 34 years old and two female between 22 and 26 years old, with college education and little dialectal influence, was recorded so far, with care to obtain an accent as nearer as possible of the EP normative pattern. The number of speakers will be increased in the future. The speakers identification it is made as follows: Speaker 1: G, male, 26 years old; Speaker 2: L, male, 34 years old; Speaker 3: MJ, female, 26 years old; Speaker 4: D, female, 22 years old.

The speech signal is sampled at 11 kHz, 16 bits, mono. The sound recording was done in the Cool Edit Pro® environment. The entire database was recorded in homogeneous conditions,

the same microphone was used and the same quiet room conditions.

2.3. Speech Signal Analysis

The proposed task is to show that in fact the [u] occurrences in fast speech don't suffer total suppression, but instead suffer some kind of transformation that can be a reduction or a deviation toward the other vowels regions. At the signal level, this task is achieved by proving that, although this vowel is often unvoiced or reduced, it is still possible to recognize vowel formants. Perceptual and articulatory observations suggest that the vocal tract configuration still determines F1 and F2 values. The tool used to make these observations was PRAAT [7], an application for phonetics computing.
We limited our study of the spectral analysis to the first and second formants, since it is generally accepted that these segmental parameters are responsible for the vowel quality distinction. The adopted representation for comparison of the obtained values is the well-known " Compared Acoustical vowel Triangle", once it constitutes a reference measure for acoustical definition of vowels in Portuguese.
Word and phone durations were also measured, as well as the phonetic context and co-articulation effects on the vowel in question.

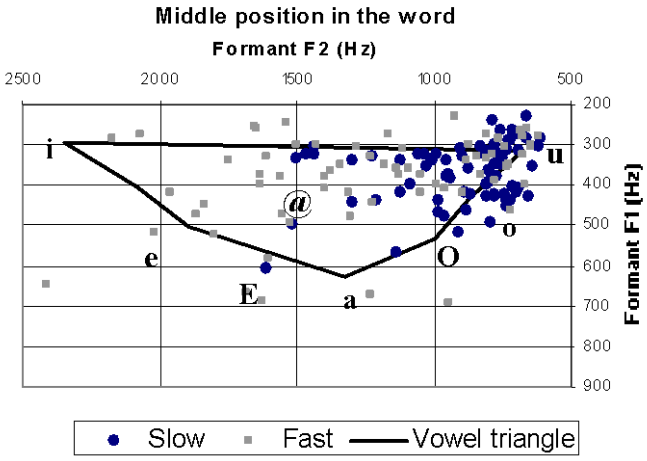

Figure 1 – Representation of non-tonic [u] realizations in word middle position in the F1-F2 plane, for slow and fast speech.

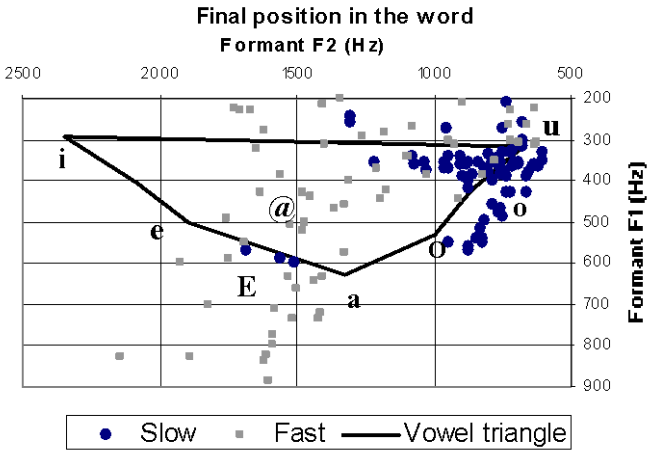

Figure 2 – Representation of non-tonic [u] realizations in word final position in the F1-F2 plane, for slow and fast speech.

3. Non-tonic back closed [u] description

3.1. Phonological perspective

Non-tonic [u] occurs in the middle of two consonants or in final word position, when describing its phonological distribution.
This vowel is part of EP vocalic non-tonic phonological framework as we can observe by the phonemic opposition method of minimal pairs in the following Portuguese words: "ponte"/"ponto"/"ponta"/"PONTI" (translations: bridge, point, extremity, sigla for a seasoned theater festival in Oporto).
This structuralist method proves that these vowels are Portuguese phonemes, because they distinguish meanings of four words with the same phonological and phonetic length and phonemic distribution.
The most similar vowels are [ɨ] and [u], within a generative perspective, since they are both high and only differ by the lip rounding.

3.2. Acoustical observations at the spectral level

In figures 1 and 2, by representing each non-tonic [u] utterances F1 and F2 measurements, we can classify the vowel quality and it is also possible to visualize a certain clustering of slow realizations' values around the correct non-tonic [u] position in the chart, when compared to the vowel triangle [8], as well as a second cluster spreading over the regions of non-tonic [u] and extending into the other vowels regions. This second cluster, so to say, is related to faster speech realizations.
These observations are related to articulatory configurations as follows: 1- in slow speech, the phonological distinctive opposition between [u] and the other vowels is kept since the former is maintained in a high back position helped by lip-rounding; 2- in faster speech (continuous) there is a shift forward of the articulation point of [u] towards the other configurations. This centralization is reinforced by the absence of lip rounding, a typical situation in relaxed, continuous and fast speech.

3.3. Acoustical observations at the time scale level

Perhaps the most important observation in this work is the absence of suppression of the vowel in almost any cases. The data suggested this result that brings us a strong contradiction to the commonly spread belief that the non-tonic [u] is affected by a suppression phenomenon, just as has been sustained for the case of the non-tonic central vowel [ɨ].
As a matter of fact we could only observe suppression occurrences in some phrase's final position realizations and never in the middle. The most common behavior in middle positions is the neutralization toward some other vowels, as we can see in Figure 1, with a duration increasing instead of reduction, as we can check in figure 5. To better document this observation, we carefully measured not only the durations of the [u] vowel realizations in the corpus but also the duration of the word, in slow and in faster speech.
From these measurements we tried to extract a behavior trend of the relative vowel duration. The result is that there is not a significant relative shortening of the vowel. On the contrary,

when in final position, in most of the cases there is even a slight increase in the relative duration.
Moreover, an important shift of the [u] quality is observed in this situation, concretely through a movement of formants towards the other vowels regions, which expresses that there is a deviation of the [u] towards these vowels.
This movement can be observed in figure 3 and 4, relative to [u] and [ɨ]. Figures 5 and 6 depict the evolution of durations of non-tonic [u], in middle and final position, respectively. The charts in these pictures represent, for each context, the quotient of the relative variation of duration of the vowel from slow to fast speech to the relative variation of duration of the relevant word [(duration of [u] in fast speech/ duration of [u] in slow speech)/(duration of word in fast speech/ duration of word in slow speech)]. This calculation was designed with the objective of showing only relative variations, which make results less prone to variance due to subject characteristics.
A value of 1 in these charts' categories axis means that the vowel has not changed in relative terms to the word durations, from slow to fast speech.
The average vowel durations are: 164 milliseconds (ms) for male speakers and 179 ms for female, in slow speech and final word position; 77 milliseconds (ms) for male and 91ms for female, in slow speech but middle word position; 104 ms for male and 77ms for female, in fast speech and final word position and 46 ms for male and 33ms for female, in fast speech and middle word position. The average duration of the vowel carrier word in the present male corpus was 766 ms and 1166 ms for final and middle realizations, and for female corpus was 904 ms and 1151 ms. These last values were obtained with slow speech. For fast speech the respective values were 563 ms, 691, 534 and 665 ms, respectively, showing the increase in speed of approximately 50%, in average. The average speed increase in the realizations of the vowel was, in average, 55%, what shows that the hypothetical suppression mentioned in the literature is not confirmed.

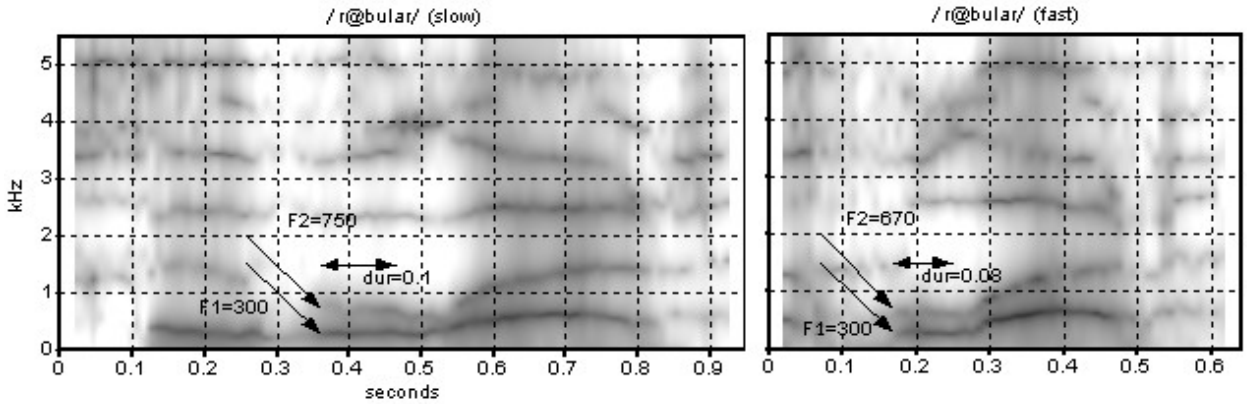

Figure 3 – Illustration of two utterances at different speeds, showing the non-existence of relative reduction of [u].

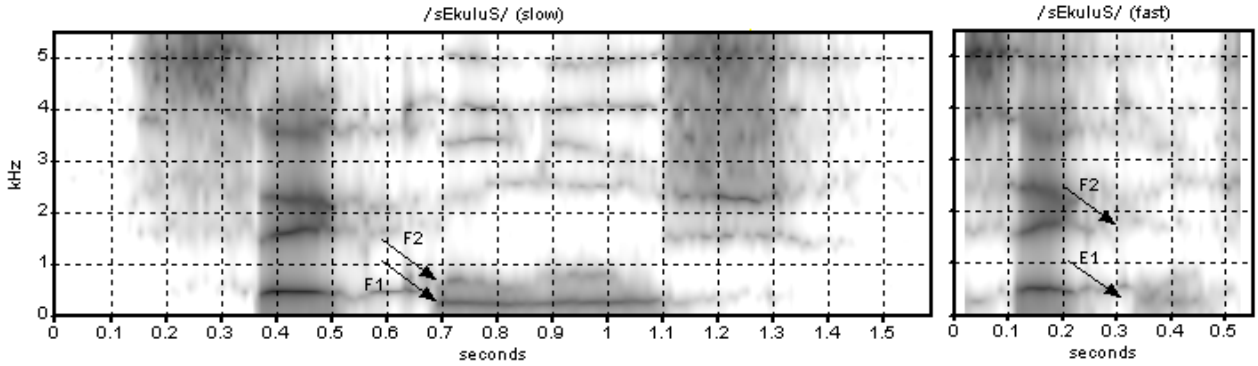

Figure 4 – Illustration of a second set of two utterances at different speeds, showing the neutralization of [u] and][ɨ].

Performing the calculation above described of relative variation of duration, this time for the total average values, a result of –13% for the vowel in non-final position in the carrier word and a value of +12% for the final position were found.
An ensemble average value of 99% relative duration variation is then obtained.
It must be stated however that, although not relevant in ensemble average, the phenomenon of reduction of the non-tonic back closed [u] is still very important in function of vowel position and in relation with the consonantal context. This can be inspected from figures 5 and 6.
The apparent lengthening of the vowel in the latter case not only contradicts the claimed suppression but also indicates that the vowel is relatively more prolonged. This phenomenon is associated to the neutralization that happens in this case, as explained above.

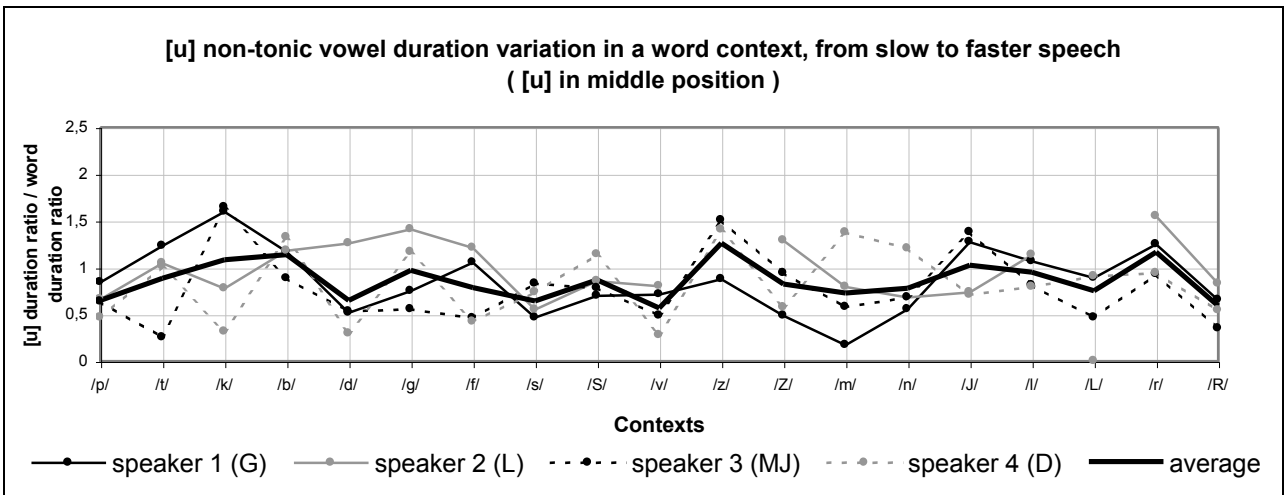

Figure 5 – Representation of non-tonic [u] realizations in word middle position in the F1-F2 plane, for slow and fast speech.

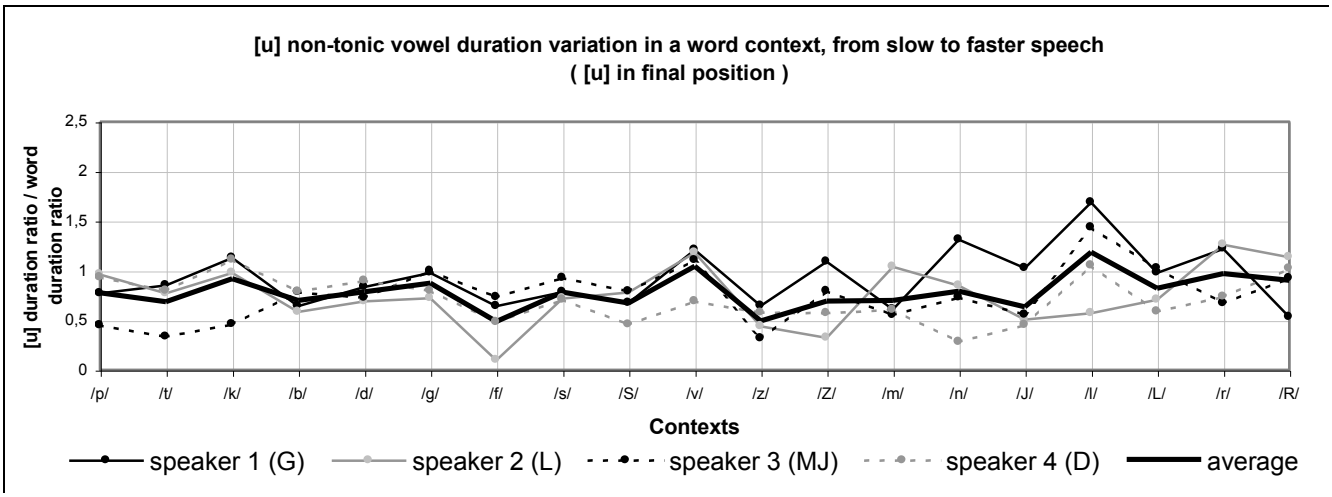

Figure 6 — Representation of non-tonic [u] realizations in word final position in the F1-F2 plane, for slow and fast speech.

4. Conclusions

From a perceptual point of view it has been defended [3] that subjects re-build vowels that they don't listen since they establish confusion with some other voiced realization, which may be explained in a psycho-cognitive framework.

However, when looking for experimental evidence of these situations, in particular for the less studied [u], we found that, in fast speech, there is not a confusion, in the way of saying that the listener could think that heard something that was not there, but instead, a confusion of phonemes takes place, in general, for fast speech, because a deviation can be observed with a spreading of [u] realizations into other vowel fields. In some cases there's even a neutralization towards the non-tonic central vowel [ɨ]. In slow careful speech the vowels seem to be reasonably preserved, not only in spectral quality but also in relative duration. We claim that these observations are very important for the EP TTS prosodic rules programming, in order to increase naturalness of synthetic speech.

5. Aknowledgments

The work reported in this paper is partly founded by EUREKA/IC-PME Project Antígona.

6. References

[1] IPA symbols. SIL Encore Fonts, 1992-1998 Summer Intitute of Linguistics;
[2] Mateus, M. H., *Aspectos da Fonologia Portuguesa*, Publicações dos Estudos Filológicos, Lisboa, 1975.
[3] Mateus, M.; Martins, R., *"Contribuição para o estudo das vogais átonas* [ɨ] *e [u] no Português Europeu"*, in, Biblos, Universidade de Coimbra, Portugal, vol.LVIII, 1982.
[4] Martinet, A., *Éléments de Linguistique Générale*,
[5] Câmara, Mattoso, *Problemas de Linguística Descritiva*, Vozes, Petrópolis, 1971.
[6] Teyssier, P., *História da Língua Portuguesa*,(portuguese tranlation), Sá da Costa, Lisboa, 1987.
[7] PRAAT, Copyright 1992-2001, by Paul Boersman and David Weenink, www.praat.org.
[8] Deller Jr, J., et al, Discrete-time processing of speech signals, Macmillan publishing company, New York, 1993.
[9] Teixeira, J. P., Freitas, D., Braga, D., Barros, Mª. J. and Latsch, V., "Phonetic Events from the Labeling the European Portuguese Database for Speech Synthesis, FEUP/IPB-DB".
[10] Martins, M.R, *Ouvir Falar,* Ed. Caminho, Lisboa, 1988.