CROSS-LINGUISTIC DIFFERENCES BETWEEN ACCENTED vs UNACCENTED VOWEL DURATIONS

Paolo Mairano[1], Fabián Santiago[2], Antonio Romano[3]

[1]University of Warwick (UK), [2]University of Paris 3 Sorbonne Nouvelle (France), [3]University of Turin (Italy)
[1]paolomairano@gmail.com, [2]fabian.santiago.ling@gmail.com, [3]antonio.romano@unito.it

ABSTRACT

This study compares durational measures of accented vs unaccented vowels for data of 5 languages (4 speakers per language read translations of the same passage). Results show clear cross-language differences: accented vowels in our data of British English are more marked in terms of duration than in data of Mexican Spanish and Parisian French (data of German and Italian take a somewhat intermediate position). Direct durational measures of accented vs unaccented vowels yield a useful insight in aspects of speech rhythm that are only partially addressed by global measures of V variability (as provided by various popular rhythm metrics). Such results may have implications in rhythm typology and might help understand why stresses are perceived to occur at roughly regular intervals of time in so-called stress-timed languages.

Keywords: duration, accent, rhythm, timing, prominence.

1. INTRODUCTION

Classical literature in phonetics states (on a perceptive basis) that inter-stress intervals tend to occur at roughly regular intervals of time in so-called stress-timed languages, as opposed to so-called syllable-timed languages ([1]). Although this claim has been falsified by several subsequent studies ([17] and [19] among others), differences in perception between these two language types have been reported ([14, 16]). In the last 15 years, some measures of variability at segmental and inter-segmental level have been proposed ([8] and [16] among others) and controversially ([4, 5]) used to classify languages or dialects.

The present study is inspired by [2], who claims that unstressed syllables in English "*are reduced in both quality and quantity to the extent that the resulting rhythmic pattern consists of the stressed syllables alternating with all of the intervening unstressed syllables, i.e. a sort of massive off-beat*" [2:80], while "*accentuation plays a weaker role in Romance phonology*" [2:77] thereby suggesting that the difference between stressed and unstressed syllables is greater in English than in Romance languages. This difference might then stand at the basis of perceived rhythmic differences between languages, since listeners may tend to hear an (inexistent) regularity of stresses (or rather: accents) in English, simply because that is what stands out and provides a pulse in English speech. This hypothesis makes sense in the light of [11]'s findings that listeners have a tendency to superimpose a rhythmic regularity also where it does not necessarily exist (and more so with speech stimuli than with non-speech stimuli).

Obviously, durational differences between accented and unaccented vowels also have a bearing on rhythm metrics measuring V variability (ΔV, varcoV, nPVI, etc.). Yet, results for rhythm metrics are often more controversial for V intervals than for C intervals: [16] themselves claim that ΔV is less apt to discriminate rhythm groups than the other metrics proposed. In effect, while low V variability implies a relatively small contrast between accented and unaccented vowels ("V-V contrast), the opposite is not true: higher V variability does not necessarily imply a greater "V-V contrast (as it may be the product of other phenomena, such as vowel length opposition or final lengthening).

In the present study, we propose to compare durations of accented vs unaccented vowels in 5 languages. (We shall refer here to *accented* vowels as to vowels that are perceptively prominent, not simply lexically stressed.) Other studies have reported measures of accented vs unaccented vowels in specific languages, but direct cross-language comparisons are scant: [7] measured cross-language stressed/unstressed durational differences (but investigated nonsense CVCV words in carrier sentences), while [3] investigated different acoustic parameters for 2 specific degrees of prominence (de-accented and nuclear accented) in 6 languages.

By addressing directly the "V-V contrast, we hope to shed light on cross-language durational differences: our hypothesis is that the durational difference between accented and unaccented vowels is greater for English and German (traditionally classified as stress-timed) than for French and Spanish (traditionally classified as syllable-timed) and Italian (which is usually also associated to syllable-timing).

2. DATA AND METHODOLOGY

2.1. The recordings

The data for this experiment consist of 20 recordings of 20 different speakers reading translations of the same text in 5 languages: 4 native speakers of Southern British English (henceforth EN1-4), 4 native speakers of Parisian French (henceforth FR1-4), 4 native speakers of Standard High German (henceforth DE1-4), 4 native speakers of Mexican Spanish (henceforth SP1-4), 4 native speakers of Standard Italian (henceforth IT1-4).

Recordings of speakers DE1, FR1 and EN1 are included in [9]. The other recordings have been produced in the phonetic laboratories of the universities of Paris 3 and Turin.

We are aware of the limitations given by the use of read speech in a fairly controlled situation, but we agree with [18] that, although different speech styles can yield different results, all of them can provide useful insights into speech. Additionally, this approach offers the undeniable advantage of analysing the same text for each speaker (albeit in different translations).

2.2. Data labelling and segmentation

All the recordings were segmented into X-SAMPA phones using a forced-alignment tool ([12]) and then thoroughly checked manually with *Praat* ([6]) by the 3 authors. For languages currently not supported by the tool, the segmentation was done completely manually.

The X-SAMPA annotations were then converted via a simple *perl* script into C (consonants) and V (vowels). The 3 authors discussed the classification of some problematic phones as consonantal vs. vocalic and finally opted for a language-independent solution that includes the following conventions:

- plosives (incl. glottal stops in German/English), affricates, fricatives are C;
- vowels, rising diphthongs and centring diphthongs are V;
- nasals and liquids are C, unless they are syllabic, in which case they are V (NB: cases such as German *ein<u>en</u>* were NOT labelled as V if the 'syllabic nasal' was realised by a simple lengthening of [n], see [15]);
- on-glides ([j, w]) were considered as C (this decision was particularly hard and controversial, so we actually kept two parallel versions with glides classified as C and as V, and finally verified that results changed only slightly);
- adjacent hetero-syllabic vowels (hiatus) were labelled as separate Vs (in contrast to diphthongs, which constitute one only V).

In this study we shall consider exclusively V durations.

2.3. The annotation of accents

In a second step, the data were also annotated by marking accents. This was done independently by authors 1 and 2, who are both trained phoneticians, and are native speakers of Italian and Mexican Spanish respectively. Author 1 speaks all the 5 languages studied, while author 2 speaks all the languages studied except German, so German accents were only labelled by author 1.

The annotation of accents was performed on a purely perceptive basis and does not rely on any theoretic/phonological framework. Authors 1 and 2 listened independently to the recordings and marked all vowels that they heard as accented (=prominent, in a way that a rhythmic pulse is heard). Accents were mostly marked in correspondence of lexical stress (for languages that have it), but this was not a rule; occasionally, two accents were marked on a single word depending on its realisation. Conversely, not all words bearing a lexical stress were marked as accented.

Once the task had been completed independently by authors 1 and 2, both authors analysed their annotations together: during this revision (which lasted several hours), authors discussed each discrepancy and were allowed to make edits and corrections. Vowels for which no agreement could be reached after the revision were excluded from the analysis presented in the next paragraphs. The average agreement between the two authors was 96.48% (σ=1.9) and resulted in the exclusion of 6.12 (σ=3.77) vowels on average per speaker (of English, French, Spanish and Italian data). No vowel was excluded from German data, since German accents were marked by author 1 exclusively.

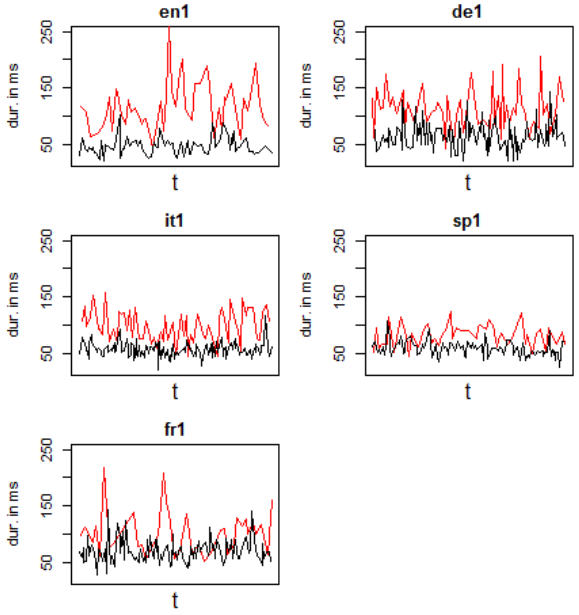
3. RESULTS

The duration of each vowel was extracted from the recordings, and then analysed and visualised in different ways.

In figure 1, all accented vs. all unaccented vowels are plotted in charts that clearly show differences: speaker EN1 differentiates clearly between very long (up to 250 ms) accented vowels (red line), and shorter unaccented vowels (black lines). The distinction becomes progressively smaller for DE1, IT1, FR1 and SP1. This first result is in line with the hypothesis that stresses are more prominently marked in so-called stress-timed languages (English

and German) than in so-called syllable-timed languages (Spanish, Italian); FR1 shows mixed trends, with accented vowels being generally not much longer than unaccented vowels, but with a few very remarkable exceptions (all of them corresponding to sentence-final accented vowels).

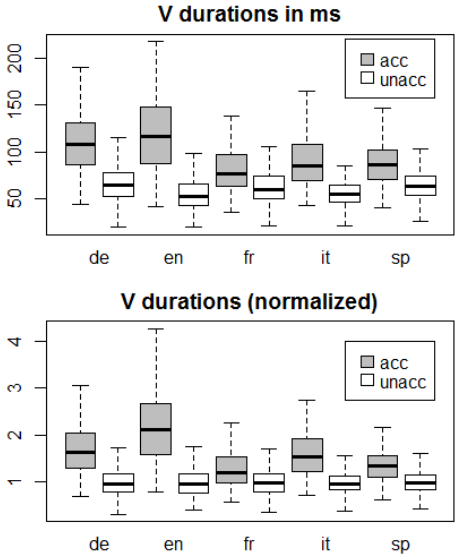
Figure 1: Durations in ms for unaccented (in black) and accented V (in red), plotted sequentially as they occur in the recordings. Final V (whether accented or not) were excluded from this plot. Only one speaker per language is shown, due to room constraints.



In figure 2, boxplots illustrate the duration of accented V durations vs unaccented V durations grouped by language. In this case, vowels bearing the last accent of a sentence and all unaccented vowels that follow have been discarded from the computation (for the present study) since phenomena of final lengthening are well known and documented for all the languages analysed. In the boxplot above, durations are expressed in ms. In the boxplot below, durations have been divided by the mean duration of unaccented vowels for each speaker, i.e. they are expressed as a proportion of the average unaccented vowel. We want to stress that although all charts report results *by languages*, the normalisation was carried out *by speaker*, with the intention to normalise possible differences in speech rate *across speakers*.

Normalised data was analysed using linear mixed effects models, with accented vs. unaccented and language as predictor variables, random intercepts, and slopes for subjects. The contribution of each predictor variable was estimated using a model reduction and likelihood ratio tests (χ^2).

Figure 2: Durations of accented (in gray) and unaccented (in white) vowels, grouped by language. Durations are expressed in ms in the boxplot above, while in the boxplot below they are normalised with respect to the average unaccented vowel for each speaker. Outliers are not plotted.



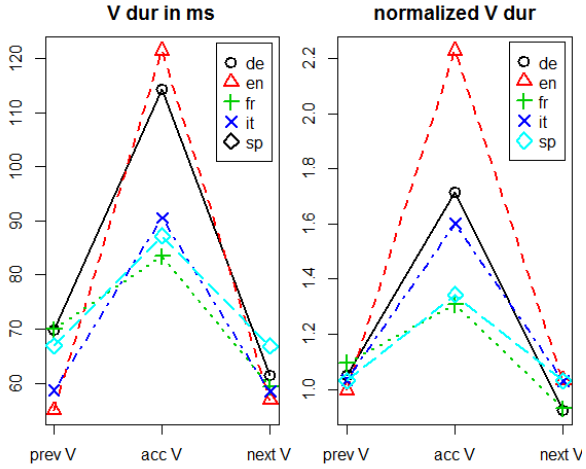
Results confirm (if that was ever needed) that accented vowels tend to have significantly longer durations than unaccented vowels for data of all languages studied ($\chi^2(1)=1107.5$, $p<.0001$). But, more interestingly, they also show that the proportions are not the same across languages: different models evaluating the interaction of language and the accented condition show that accented vowel durations in English significantly differ from the accented condition in any other language studied (German: $\chi^2(5)=13.553$, $p<.001$, French: $\chi^2(5)=21.597$, $p<.0001$, Italian $\chi^2(5)=12.759$, $p<.01$, Spanish: $\chi^2(5)=22.164$, $p<.01$). In fact, an accented vowel in our English data is on average twice as long as an unaccented vowel, while an accented vowel in our Spanish and French data is on average just <1.4 times as long as an unaccented vowel (see figure 2, plot below).

When comparing German with the other languages, we found a similar tendency: accented vowels are proportionally longer in German than in French ($\chi^2(5)=11.351$, $p<.01$) and Spanish ($\chi^2(5)=10.768$, $p<.05$), but not with respect to Italian ($\chi^2(5)=1.717$, $p<.886$). Finally, French and Spanish data behave similarly and with no statistical difference ($\chi^2(5)=.739$, $p=.980$).

Our initial hypothesis seems to be supported by the results obtained until now for English, Spanish and French; German and Italian do not contradict the hypothesis, but show milder trends than expected.

We shall now look at each accented vowel and its adjacent unaccented vowels. By *adjacent unaccented vowels* we mean the unaccented vowel preceding (prevV) and the unaccented vowel following (nextV) a given accented vowel (accV) – in most cases across a consonant or a consonantal cluster. The reason for observing this specific context is that an accented vowel will presumably be heard as prominent mainly in relation to the surrounding unaccented vowels, rather than in relation to just *any* unaccented vowel. Such duration values are given in figure 3 (NB: the last accented vowel before a pause has been discarded from calculations).

Figure 3: Mean durations of prevV, accV and nextV by language. Durations are expressed in ms (left) and normalised by the average duration of unstressed V for each speaker (right).



Cross-language differences in our data are, once more, eye-catching. At one end, French and Spanish show a relatively small difference between the mean duration of accV vs prevV and nextV. At the other end, English shows greater durational difference between accV vs prevV and nextV. Italian and German lie somewhat in an intermediate position (see normalised durations on the right), the former leaning slightly towards its syllable-timed peers, the latter leaning towards stress-timed English.

It is also worth mentioning that prevV and nextV for DE, EN, IT and SP do not seem to be significantly shorter or longer than the average unaccented V (which would score 1 in the right chart of figure 3). But interestingly, French prevV seems to be longer than nextV, with prevV being slightly longer than average and nextV being slightly shorter than average. This result confirms findings by [13] and should be interpreted in the light of the fact that accents in French mostly occur at (and mark) prosodic boundaries [10]. The proximity of a prosodic boundary may in effect explain the increased duration of prevV, while nextV is not affected as it comes after the prosodic boundary.

4. CONCLUSION

This study has revealed remarkable cross-language differences in the duration of accented vs unaccented vowels in our data of 5 languages (4 speakers per language): although accented vowels tend to be longer than unaccented vowels in all languages studied, the proportions are not the same. As initially hypothesised, the durational difference of accented vs unaccented vowels in languages traditionally classified as stress-timed (English and, to a lesser extent, German) tends to be greater than for languages traditionally classified as syllable-timed (French and Spanish – but less so in Italian). If we see the results on a continuum, our English data can be regarded as the most "accenting" (in terms of duration), followed by German data; Spanish and French data are the least "accenting", preceded by Italian data.

We think that this difference contributes to explain the perceived isochrony of feet in English [1, 14] and other stress-timed languages. However, at the moment this remains a hypothesis because our results (simply based on the analysis of produced durations) certainly do not allow us to draw conclusions on perception: the fact that accentual durational differences exist does not imply that such differences stand at the basis of perceived isochrony of inter-stress intervals. Experiments addressing perception are of course needed to confirm this, and will hopefully be carried out in the future. Moreover, perceived prominence does not simply depend on duration, as is well known. In the future, pitch and intensity will be taken in consideration; for now, some cross-language differences have emerged even focusing exclusively on durations.

As a final note, we would like to remind that recordings consisted of a comparable passage read in 5 languages. We do not claim that our findings can be directly generalised to other types of data. Duration is a parameter that depends on various factors (e.g. speech rate) and that can easily and heavily be modified by speakers to achieve various effects. Further analyses will tell whether similar results also apply cross-linguistically in other types/styles of speech.

5. ACKNOLEDGEMENTS

This study was supported by the French *Investissements d'Avenir - Labex EFL program* (ANR-10-LABX-0083).

6. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh University Press.

[2] Allen, G. D. 1975. Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3, 75-86.

[3] Andreeva, B., Barry, W., Koreman, J. 2014. "A Cross-language Corpus for Studying the Phonetics and Phonology of Prominence". *Proc. of the 9th LREC Conference*, Reykjavik, 26-31 May 2014.

[4] Arvaniti, A. 2009. Rhythm, Timing and the Timing of Rhythm, *Phonetica*, 66, 46-63.

[5] Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S. & Kostadinova, T. 2003. "Do rhythm measures tell us anything about language type?". *Proc. of the 15th ICPhS*, Barcelona, 2693-2696.

[6] Boersma, P., Weenink, D. 2014. *Praat: doing phonetics by computer* [computer programme]. Version 5.4.0.4, retrieved on 28 Dec 2014 on http://www.praat.org/.

[7] Botinis, A., Banert, R., Fourakis, M., Pagoni-Tetlow, A. 2002. Prosodic effects and crosslinguistic segmental durations. *Proc. of Fonetik, 77-80*.

[8] Grabe, E., Low, E.L. 2002. Durational variability in speech and the rhythm class hypothesis. In: C. Gussenhoven, N. Warner (eds), *Papers in Laboratory Phonology* 7, Berlin: Mouton de Gruyter, 515-546.

[9] IPA 1999. *Handbook of the International Phonetic Association.* Cambridge: Cambridge University Press.

[10] Jun, S.-A., Fougeron, C. 2002. Realizations of accentual phrase in French intonation. *Probus*, 14, 147-172.

[11] Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.

[12] Goldman, J.P. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *Proc. of InterSpeech*, Sept. 2011, Florence, Italy.

[13] Michelas, A., D'Imperio, M. (2012). When syntax meets prosody: tonal and duration variability in French Accentual Phrases. *Journal of Phonetics*, 40, 816-829.

[14] Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics*, 12, 75-83.

[15] Pompino-Marschall, B., Janker, P.M., 1999, Production and perception of syllabic [n] in German. *Proc. of the 14th ICPhS*, San Francisco, 1079-1082.

[16] Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73/3, 265-292.

[17] Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In: D. Crystal (ed.), *Linguistic controversies*, London: Edward Arnold, 73-79.

[18] Wagner, P., Trouvain, J., Zimmerer, F. 2015. In defense of stylistic diversity in speech research. *Journal of Phonetics*, 48, 1-12.

[19] Wenk, B. & Wioland, F. (1982) Is French really syllable-timed? *Journal of Phonetics*, 10, 193-216.