The development of measurable speech rhythm in Spanish speakers of English

Dellwo, Volker ; Gutiérrez Díez, Francisco ; Gavalda, Nuria

Originally published at:
Dellwo, Volker; Gutiérrez Díez, Francisco; Gavalda, Nuria (2009). The development of measurable speech rhythm in Spanish speakers of English. In: XI Simposio Internacional de Comunicacion Social, Santiago de Cuba, 19 January 2009 - 23 January 2009. s.n., 594-597.

VOLKER DELLWO[1]
FRANCISCO GUTIÉRREZ DÍEZ[2]
NURIA GAVALDÀ[3]

[1]University College London, London, United Kingdom; v.dellwo@ucl.ac.uk
[2]Universidad de Murcia, Departmento de Filología Inglesa, Murcia, Spain; fgut@um.es
[3]Universida de Barcelona, Barcelona, Spain; nuria.gafe@gmail.com

The development of measurable speech rhythm in Spanish speakers of English

Abstract

It has been demonstrated repeatedly that durational characteristics of consonantal (C) and vocalic (V) intervals are robust acoustic correlates of rhythm class (stress-timed, syllable-timed, mora-timed). Here, we investigate how such rhythm measurements change during the acquisition of a second language. In a longitudinal study, 9 native speakers of Spanish were recorded reading a text in English before and after a year of English language training at university level. A control group of 9 native English speakers read the same text. Standard rhythm metrics (%V, deltaC, deltaV, PVI) were calculated for all recordings. Results reveal no significant statistical difference in measurable rhythm pre and post training. The findings are discussed in connection with theoretical issues of rhythm metrics and practical issues of L2 English by Spanish speakers.

1 Introduction

Some languages can be classified into rhythmic types of which the two most prominent are the stress-timed and syllable-timed rhythm classes (James, 1929, Pike, 1945, and Abercrombie, 1967, Bolinger, 1981, Ramus et al., 1999, Grabe and Low, 2002). Behavioural experiments have shown that adult human listeners (Ramus and Mehler, 1999), as well as newborns (Nazzi et al., 1998, Ramus, 2002), monkeys (Ramus et al., 2000, Rincoff et al., 2005), and rats (Toro et al., 2003) can distinguish between languages from different rhythmic classes.

The acoustic correlates of speech rhythm have been unclear until about the 1990s. When the distinction was first suggested it was believed that syllable timing is characterised by quasi isochronous syllabic durations and stress timing by quasi isochronous inter-stress interval durations (hence the terminology stress- and syllable-timing). However, it has exhaustively been demonstrated in experimental research from the early 1960s to the late 80s that such assumptions are not justified (see discussions in Ramus et al., 1999, and Grabe & Low, 2002).

During the late 1990s a new approach based mainly on earlier ideas by Bolinger (1981) looked at the durational characteristics of consonantal (C) and vocalic (V) intervals (C-interval = series of consonants between two vowels, V-interval = series of vowels between two consonants; across syllable and word boundaries, not across pause). This is based on the assumption that the phonological properties contribute to the impression of variable rhythmicity in the speech signal. While syllable-timed languages typically reveal simple syllable structures (e.g. low consonantal complexity and lack of vocalic reductions), stress-timed languages reveal the opposite. This causes C- and V-intervals to be durationally more or less variable.

In two approaches it was demonstrated that C- and V-interval durations vary between languages classified as stress- and syllable-timed . One is Ramus et al. (1999) who reported that the percentage over which speech is vocalic (%V) and the standard deviation of C-intervals (deltaC) are reliable correlates of rhythm class (see Ramus et al., 1999, for the rationale). Another approach by Grabe & Low (2002) calculated the average durational differences between consecutive C- and V-intervals (the Pairwise Variability Index; PVI). They reported that the PVI for V-intervals, in particular, reveals a reliable distribution of languages into rhythmic classes. The measures have been widely discussed and the empirically evaluated. All in all there is general consent that the measures are powerful indicators of rhythm class (White and Mattys, 2007). Other studies pointed out that some of the measures are speech rate dependent (Dellwo & Wagner, 2003, and for a review see Dellwo, submitted), other studies demonstrated that this speech rate dependency can easily be normalised for (Dellwo, 2006, White and Mattys, 2007).

The present research deals with the question to what degree measurements of rhythm change with competence in a second language. The general argument is that less proficient speakers of a language reveal different measurements of rhythm because they do not produce interval variability in the same way like native speakers (White and Mattys, 2007, Dellwo, submitted). The previous assumption that speakers who switch rhythm class would show different values from speakers staying in the same rhythm class was not confirmed (White and Mattys, 2007b, and Dellwo, submitted). Instead it was found that (a) only measurements of vocalic durational characteristics (%V, VarcoV, nPVI; explanation below) differ between first and second language speakers and (b) that the rhythm class background of the first language has no influence on rhythmic performance in the second language.

It the present research we ask the question whether measurable rhythm changes during the acquisition of a second language. An experiment was carried out with Spanish speakers reading an English text before and after a year of university training. We assumed that when proficiency increases in a second language speech rhythm should approximate speech rhythm performance of native speakers. Here we present the first results from a work in progress.

2 Experiment

2.1 Method

2.2.1 Subjects

Nine Spanish learners of English and seven English native speakers took part in the experiment. The Spanish learners of English were students at Murcia University in Spain. 3 Students were recorded between year 2 and year 3, further 5 students were recorded between year 3 and year 4 of their university degree. All students took part in the study voluntarily and did not receive payment. The seven speakers from the native English control group were taken from the BonnTempo Corpus (Dellwo et al., 2004).

2.2.2 Measurements

Four rhythm measures based on the characteristics of C- and V-intervals were carried out for the present research, (a) %V and deltaC (Ramus et al., 1999) and the vocalic and consonantal Pairwise Variability Index (PVI, Grabe & Low, 2002). In the following we will give a short explanation of these measures, please check the stated sources for more details:

(a) **lsrCV:** Laboratory measurable speech rate in CV-intervals/second (lsrCV). CV-intervals are chosen over the more common measure of syllables/second in order to use the same units for rhythm and rate measurements (compare Dellwo, submitted, for a discussion of possible conflicts when mixing the units).

(b) **%V** is the percentage over which speech is vocalic. It is calculated for the linguistic content of speech only (i.e. excluding any pauses and hesitations).

(c) **VarcoC:** DeltaC is the standard deviation of C-interval durations (Ramus et al., 1999). This measure was demonstrated to be highly speech rate dependant (Dellwo, 2006, White and Mattys, 2007b) which is why the coefficient of variation of deltaC (VarcoC) was used here for rate normalisation as suggested previously by Dellwo (2006).

(d) **VarcoV:** DeltaV is the standard deviation of V-interval durations. Again, this measure was demonstrated to be rate dependant (White & Mattys, 2007b) which is why the coefficient of variation is calculated (VarcoV).

(e) **nPVI** is the rate normalised version of the Pairwise Variability Index. This measure calculates the average difference between consecutive V-intervals in speech (see Grabe & Low, 2002, for the measure and White and Mattys, 2007b, for a discussion).

(f) **rPVInorm:** rPVI is the raw (non speech rate normalised) Pairwise Variability Index which calculates the average difference between consecutive C-intervals (Grabe & Low, 2002). Since rPVI was demonstrated to be highly speech rate dependant it was rate normalised in the same way as suggested for nPVI by Grabe & Low (2002) and is here referred to as rPVInorm.

A total of 25 measurements were produced for each of the above listed measures, one measurement for each text production of each speaker (9 pre, 9 post, and 7 control). When measures were shown to be speech rate dependent they were normalised for this influence as described above. This is of high importance since there are rate fluctuations in the present data which are otherwise likely to produce artefacts.

2.2.3 Procedure

Students were recorded with high end digital recording equipment in a quiet environment. Speakers were asked to read the English text from the BonnTempo Corpus (Dellwo et al., 2004). This text consists of 5 sentences (about 90 syllables), typically realised as 7 intonation phrases. The text was recorded in the BonnTempo fashion in which speakers are asked to produce the text with normal two slow and two fast intended tempo versions, however, only the normal read version was processed for the present research. The text was then phone labelled manually by a human labeller (third author). Using the speech processing software Praat (www.praat.org), the phone labelling was automatically transferred into consonantal and vocalic intervals with Praat scripts specifically written for this purpose. With other Praat analysis scripts from the BonnTempo Tools collection (Dellwo et al., 2004) the CV labelling was statistically analysed and the rhythm and rate measurements described above were processed. The nine Spanish speakers also read a translation of the reading text in their native language, however, the material has not been fully processed at the current stage thus we are unable to present the results for this condition at the moment.

3 Results

3.1 Speech Rate

The results for laboratory measurable speech rate (lsrCV; measured in CV-intervals/second) for the Spanish pre- and post-training group and the English control group are presented in the box-plot in Figure 1/left (the black line indicates the median, the upper and lower lines of the boxes the inter-quartile range and the whiskers the range). The results show that the Spanish native speakers produced English slower (median about 9 CV-intervals/sec) than the English native speakers (about 11 CV-intervals/sec). An ANOVA with ‘language’ (pre-training, post-training and English control) as a fixed factor and lsrCV as the dependent variable revealed a significant effect ($F[2,24]=24,2$; $p<0.001$). Post-hoc analysis (Tukey) confirmed this observation by showing that there is a highly significant difference between the English controls and any of the Spanish English speaking groups ($p<0.001$). Descriptively there appears to be a slight trend for the post-training group to increase in lsrCV towards the value of the English controls. This effect, however, was not significant in the post-hoc analysis ($p=.84$).

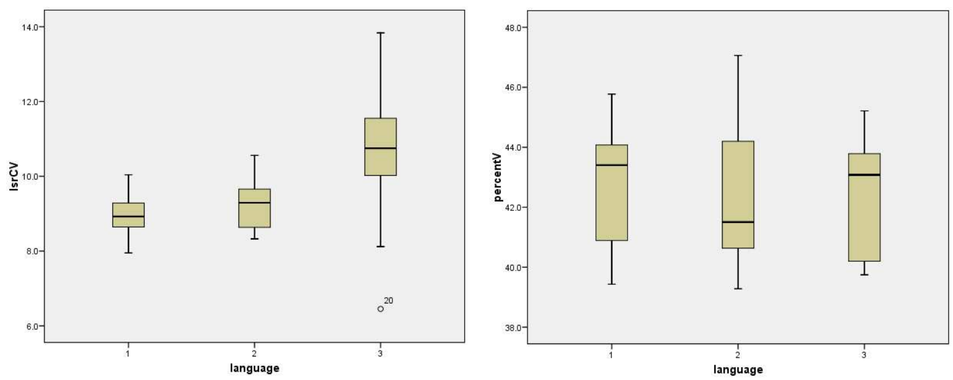

Figure 1: *Box-plots showing speech rate (left) and %V (right) for the Spanish English learner pre- and post-training groups (1 and 2) and the English control group (3).*

3.2 Speech Rhythm

The results for %V are presented in Figure 1/right for the pre (1), post (2) and English control group (3). Descriptively the graph reveals that there is no difference between the three groups. An ANOVA confirms this impression showing no significant effect.

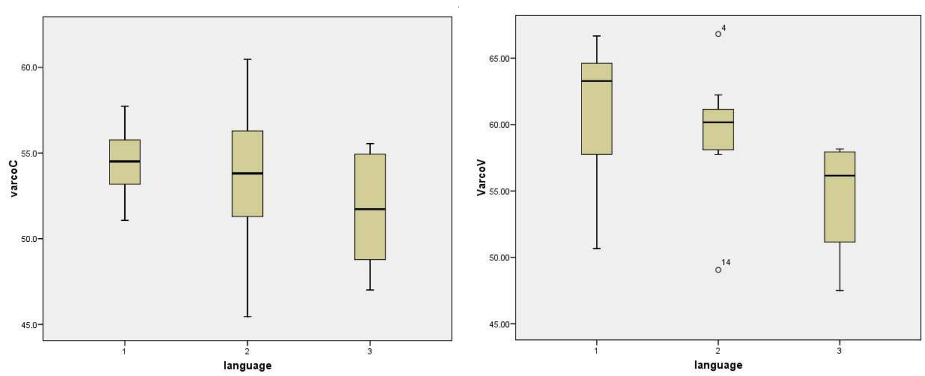

Figure 2: *Box plots showing VarcoC (left) and VarcoV (right) for the Spanish English learner pre- and post-training groups (1 and 2) and the English control group (3).*

VarcoC and VarcoV are presented in Figure 2. It is visible that in both cases the English control group (3) has lower values than both Spanish English groups (1 and 2). Also, judged mainly by the median values (and in particular for VarcoV) the impression arises that the Spanish post-training group has slightly lower values than the pre-training goup. Both observations, however, are merely descriptive trends and are not confirmed by inferential analysis. Two ANOVAs (language*VarcoC and language*VarcoV) do not show a significant effect (($F[2,24]=.15$; $p=.86$ and $F[2,24]=1.3$; $p=.29$).

The results for the measures nPVI and rPVInorm are presented in Figure 3 and show clearly lower values for the English controls (3) compared to the Spanish pre- and post-groups (1, 2). Training, however, does not seem to have had an effect on any of the measures since the distributions for pre- and post-training groups mainly overlap in both cases. Two ANOVAs (language*nPVI, language*rPVInorm) confirm this observation by showing highly significant effect in case of the nPVI ($F[2,24]=24.2$; $p<0.001$) and a significant effect for rPVInorm ($F[2,24]=5.1$; $p<0.014$). Post-hoc pre and post groups revealed no effect (nPVI: $p=.99$; rPVInorm: $p=.94$) while the English controls compared to the Spanish groups show effects of $p<.05$ in case of the nPVI and $<.001$ in case of the rPVI.

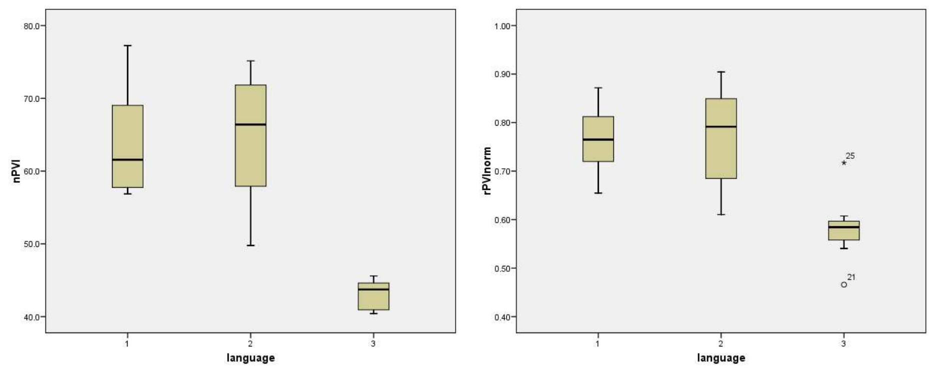

Figure 3: *Box plots showing nPVI (left) and rPVInorm (right) for the Spanish English learner pre- and post-training groups (1 and 2) and the English control group (3).*

4 Discussion

The main findings for the present study are on two levels, (a) the measures distinguishing English native speakers from the Spanish English speakers best are the vocalic and consonantal Pairwise Variability Index (nPVI, rPVInorm) and speech rate (lsrCV), and (b) our Spanish learners of English did not change in measurable speech rhythm before and after a year of university training. Both findings are rather surprising and will be discussed in the following:

Re (a): Previous studies showed that %V is a powerful indicator to distinguish between L1 and L2 speech (Dellwo, submitted) while other studies point out VarcoV (White and Mattys, 2007a) or a complementation between VarcoV and %V (White and Mattys, 2007b). The latter two studies are most interesting as comparison studies to the present research since they also focussed on Spanish learners of English (Dellwo, submitted, looked at Germans speaking English and French). The fact that our Spanish English speakers reveal measurements of %V and VarcoV that are identical with the ones of native English speakers may raise the assumption that these speakers were native like speakers of English. However, while their English proficiency as university students of English is possibly higher than that of the average Spanish English speaker, auditory impressions of the recording material revealed highly characteristic Spanish second language features on a segmental and supra-segmental level. So there should clearly be a measurable difference between our Spanish English and English English speakers. Such differences are obtainable: for example, the fact that the Spanish English speakers produce English significantly slower than the natives and the varying consonantal and vocalic PVI demonstrate that there are objectively measurable differences between the groups. Interestingly, these measurable differences, in particular rPVI, are not the ones that were previously pointed out to be the indicators which distinguish rhythmical performance between first and second language speakers. White and Mattys (2007) and Dellwo (submitted) showed that consonantal rhythm metrics do not reveal differences between L1 and L2 speech. The results of the present study, which indicate that nPVI and rPVI can be the best parameters to distinguish L1 and L2 speech and %V and VarcoV do not distinguish at all, are surprising and suggest that the situation is probably more complex than previously assumed. Measures of vocalic durational characteristics such as %V and VarcoV are not necessarily best in distinguishing L1 form L2 speech. The reasons for this divergence between the present and previous studies are at present unclear and will receive special attention in the future work of this project.

Re (b): The reasons why our Spanish learners of English do not show a change in measurable rhythm are further unclear. The thought may occur that rhythmic measures of the type discussed here are no reliable indicators for changes in L2 proficiency. However, the metrics (here n and rPVI) do reveal measurable differences between L1 and L2 proficiency, so why should they not reveal differences between high and low proficient L2 speech. The more likely explanation is therefore that the differences in rhythmic proficiency between our two Spanish groups are not large enough to be measurable with the metrics we applied here. The question that arises here is: How much did the pronunciation, in particular characteristic rhythmic features of English, improve over the course of the one year training? The speakers did not receive a particular training in applied speech rhythm (we are not aware that such content is generally being taught at universities) but we assumed that a general training in pronunciation would also lead to a better performance in speech rhythm. In order to address the question as to how much pronunciation and in particular rhythmic features improved we are planning to have the English performance of our speakers (pre and post training) evaluated in perception experiments by English native speakers. Should there be a notable improvement in rhythmical features of English the value of rhythmic measurements for L2 proficiency will have to be questioned.

References:

Abercrombie, D. (1967). "Elements of General Phonetics," Edinburgh: University Press.

Bolinger, D.L. (1981). "Two kinds of vowels, two kinds of rhythm," Bloomington, Indiana: Indiana University Linguistics Club.

Boersma, P. (2001). "Praat, a system for doing phonetics by computer." Glot International 5:9/10, 341-345.

Dauer, R.M. (1983). "Stress-timing and syllable-timing reanalyzed," Journal of Phonetics 11, 51-69.

Dellwo, V. (submitted). "Influences of speech rate on acoustic correlates of speech rhythm: An experimental investigation based on acoustic and perceptual evidence." PhD thesis, Bonn University, Germany.

Dellwo, V. (2006). "Rhythm and Speech Rate: A Variation Coefficient for ∆C," Pawel Karnowski & Imre Szigeti (eds.) Language and Language-processing. Frankfurt am Main: Peter Lang, 231-241.

Dellwo, V., and Wagner, P. (2003). "Relations between Language Rhythm and Speech Rate," D. Recasens, M.J.Solé and J. Romero (eds.), Proceedings of the 15th ICPhS, Barcelona, Spain, 471–474.

Grabe, E. and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis," C. Gussenhoven and N. Warner (eds.) Papers in Laboratory Phonology 7, Berlin, New York: Mouton de Gruyter.

James, A. L. (1929). "Historical introduction to French Phonetics," London: ULP.

Pike, K.L. (1945). "Intonation of American English," University of Michigan Press: Ann Arbor.

Ramus, F. (2002). "Language discrimination by newborns," Annual Review of Language Acquisition 2, 85-115.

Ramus, F., Hauser, M.D., Miller, C., Morris, D., and Mehler, J. (2000). "Language discrimination by human newborns and cotton-top tamarin monkeys," Science 288, 349-351.

Ramus, F., and Mehler, J. (1999). "Language identification based on suprasegmental cues: A study based on resynthesis," J. Acoust. Soc. Am. 105(1), 512-521.

Ramus, F., Nespor, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," Cognition 73, 265-292.

Rincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., and Mehler, J. (2005). The role of speech rhythm in languages discrimination: further tests with a non-human primate. Developmental Science 8(1), 26-35.

Toro, J.M., Trobalon, J.B., and Sebastian-Galles, N. (2003). "The use of prosodic cues in language discrimination tasks by rats," Animal Cognition 6(2), 131-136.

White, L. and Mattys, S. (2007a) “Rhythmic typology and variation in first and second languages. In P. Prietro, J. Mascaró, and M. J. Solé (Eds.), *Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series)*. Amsterdam, Philadelphia: John Benjamins.

White, L. and Mattys, S. (2007b) "Calibrating rhythm. First language and second language studies," J. Phonetics, doi: 10.1016/j.wocn.2007.02.003