

Capítulo 19

English speech rhythm in instructed learners. Its development as shown by VarcoV

Gonzalo Eduardo Espinosa
Universidad Nacional del Comahue
IPEHCS-CONICET
gonzalo.espinosa@fadel.uncoma.edu.ar

Abstract

Speech rhythm is viewed as the product of different phonetic and phonological properties. This work shows the rhythmic acquisition of instructed learners of English as an L2 whose mother tongue is Spanish. Two main variables are considered: levels of language proficiency at university and type of elicitation task. Taking into account that rhythm can be measured acoustically, the production of speech is computed by means of the metric VarcoV, which considers the duration of vocalic intervals. The metric measurements reflect the development of L2 speakers' speech in direction to a more native-like rhythm affected by the type of task.

1. Introduction

Speech rhythm has been conceived of as the result of various phonetic and phonological properties (Bertinetto, 1989; Dasher and Bolinger, 1982; Dauer, 1983, 1987). Spanish is traditionally considered a language with a syllable-timed rhythm that has a simple syllable structure, in which lexical stress has a minor role, and there are no processes of vowel reduction. English, on the other hand, has been classified as a stress-timed language that has a complex syllable structure, in which lexical stress has a major role, and vowel reduction is a recurrent phenomenon.

In order to account for speech rhythm acoustically, rhythm metrics have been developed in order to account for typologically different rhythms (Ramus, Nespors, and Mehler, 1999; White and Mattys, 2007; Wiget et al., 2010). These metrics compute vocalic and consonantal intervals, allowing for a classification of languages according to their values. One of the most implemented metrics is VarcoV (see Section 2). This metric yields high values for languages like English and low values for languages like Spanish. For instance, these two languages have been described with the following VarcoV indices: American English=51.33, British English=62.64 and American and European Spanish=43.75 (Mairano, 2011), British English=64 and European Spanish=41 (White and Mattys, 2007), British English=51.7 (Fuchs, 2016), Argentinean Spanish=49.68 (Gabriel and Kireva, 2014), British English=61 (Tsoi, 2016), and Peruvian Spanish=36 (O'Rourke, 2008). Given that metrics are highly sensitive to different

methodological decisions (such as segmentation or computation criteria), the values for the same language are not always the same, but the tendencies exhibit a clear difference between Spanish and English.

In this work, the development of speech rhythm in instructed second language (L2) speakers of English is measured acoustically by means of VarcoV. The underlying objective is to analyze the way non-native English rhythm is affected by different levels of language proficiency and by two different elicitation tasks.

2. Methodology

2.1 Tasks

For the present study the results of two elicitation tasks are presented. The first one (Task 1) corresponds to a repetition activity in which participants listen to synthesized speech. Similarly to what has been done in morphosyntactic studies (Gass and Mackey, 2012: 27), the purpose of this task is to expose participants to language stimuli that they should understand, retain and produce, filtering what they have listened to with their grammar or mental phonology.

In order to obtain the stimulus for Task 1, a short text that was semantically equivalent in English and Spanish was generated with the Praat synthesizer. The speech rate was set at 115 words per minute with a frequency rate of 44,100 Hz and a 0.03-second pause between words. The samples obtained had no intonation contours or prominence differences among syllables. The English stimulus consisted of 153 syllables that were grouped into 17 fragments, while the Spanish stimulus had 209 syllables grouped in 13 fragments. In other words, each fragment had between 12 and 17 syllables. Participants were first asked to listen to the whole message. They knew that they were listening to a voice message from a robot and that their task was to make it more natural. If necessary, they were allowed to ask questions in order to understand meaning. Then they clicked on each fragment on a screen, listened to it and repeated it as naturally as possible. They were in charge of the transition of fragments.

Task 2 corresponds to a map task that was aimed at obtaining more spontaneous speech. The idea was that participants focus more on the content of the activity rather than on how the message was said (Fuchs, 2016: 103–4). In non-native speakers, this meant that the effect of their metalinguistic knowledge was somehow reduced (Schwartz and Sprouse, 1994: 320). Participants were presented with the map of a city with iconic buildings like a university, a bank, a cathedral, among others. They were asked to say how to get from one building to another, reaching up to five itineraries. Between 10 and 17 tone units were selected per participant, obtaining between 90 and 100 syllables for each of them.

2.2 Participants

For the present study the data of 63 participants have been analyzed. There are three groups of non-native English speakers (n=42), one group of Spanish native speakers (n=7) and two groups of English native speakers (n=14). All non-native speakers are university students at a five-year undergraduate program at Universidad del Comahue, Argentina, where they aim to become Teachers of English or English Translators. Their mother tongue is Spanish and they were born in the North Patagonian region, Argentina, brought up by parents whose mother tongue is also Spanish.

Taking into account their stage in the program, their grades¹ and their age, learners have been divided into three main levels of language proficiency: initial (group A), intermediate (group B) and advanced (group C).

Group A (n=14; 3 males) corresponds to students (mean age=20) who have almost completed their first year at university, reaching an A2 level according to the CEFR (Council of Europe, 2001). In relation to phonetics and phonology, these learners have had 10 months of formal instruction on average, obtaining a mean score of 8.39 (out of 10) in their assessment record. Group B (n=14; 5 males) is formed by students (mean age=23) who have completed half of the five-year program. They have obtained a B2 level according to CEFR and have had on average 2 years and 6 months of formal instruction on phonetics and phonology. The mean score obtained in the latter was 7 (out of 10). Finally, group C (n=14; 3 males) consists of students (mean age=26) who are about to finish their undergraduate program. They have developed a C1 or C2 level according to CEFR. The mean period of formal instruction in phonology is 3 years, obtaining a mean score of 7.71 (out of 10).

In order to obtain a point of comparison for the measurement of non-native English, native speakers of Spanish and English are analyzed. The Spanish group (mean age=29; 3 females) consists of speakers born and raised in Patagonia, Argentina. They are all university students whose mother tongue is only Spanish. The English group is formed by two subgroups. The AmE (American English) group consists of English native speakers (mean age=25; 4 females) coming from the northeast area of USA. They are all undergraduate students, and English is their only mother tongue. The BrE (British English) group corresponds to speakers (mean age=25; 4 females) from the southeast area of England. They are graduate and undergraduate students and are monolingual English speakers.

¹ Learners were not tested for the specific purposes of the present study. Instead, all the grades of oral and written exams that students had had in English Language and English phonology were considered. The three groups were selected by taking into account comparable grades. This is considered to be a better way of accounting for their language proficiency because those grades contain the continuous evaluation process along the program, conducted by the same teachers and following similar evaluation criteria.

2.3 Measurement

The whole corpus was manually segmented with Praat (Boersma and Weenink, 2015) following standard segmentation criteria (e.g. Fuchs, 2016; Mairano, 2011; White and Mattys, 2007). The rhythm metric employed in the acoustic measurement is VarcoV (Dellwo, 2006; Dellwo and Wagner, 2003). This duration metric divides the standard deviations of vocalic intervals by their mean duration, controlling for speech rate. It has been demonstrated that this metric is productive to account for typologically different rhythms (Espinosa, 2018; Fuchs, 2016; White and Mattys, 2007). The computations of VarcoV were carried out by means of Correlatore (Mairano and Romano, 2010), following Method B and excluding final intervals as indicated in Espinosa (2018). The output obtained with Correlatore was analyzed statistically by means of ANOVAs and Tukey tests.

All in all, the corpus obtained with Task 1 summed up 10,031 syllables in 1,691 tone units, while with Task 2, the corpus obtained was 6,113 syllables in 883 tone units.

3. Results and discussion

3.1 Native speakers

In order to demonstrate that the metric VarcoV effectively distinguishes two typologically different speech rhythms, the results of native speakers are shown firstly. Fig. 1 contains the results of language effects in the three groups of speakers. It shows the mean values of VarcoV, their standard deviation, the ANOVA analysis and the Tukey post hoc² test.

Task	Sp	AmE	BrE	<i>F</i>	<i>p</i> -value	Post hoc ^a		
	<i>Mean</i> (<i>SD</i>)	<i>Mean</i> (<i>SD</i>)	<i>Mean</i> (<i>SD</i>)			Sp	AmE	BrE
1	41.67 (5.73)	52.82 (7.06)	52.04 (4.96)	7.59	0.0041*	a	b	b
2	36.02 (6.32)	58.43 (13.19)	59.55 (7.40)	13.77	0.0002*	a	b	b

Fig. 1. VarcoV values and the effect of language groups in native speakers.

^a Letters in common are not significantly different ($\alpha=0.05$)

* There is a significant difference between groups.

As Fig. 1 shows, the values for the Sp group are lower than the AmE and BrE groups in both elicitation tasks. The differences between Spanish and English are significantly different, as indicated by the *p*-values ($p=.0041$ for Task 1 and $p=.0002$ for Task 2). As expected, the Spanish VarcoV mean values are lower in both tasks (41.67 for Task 1 and 36.02 for Task 2),

² When the ANOVA analysis reveals that there are significant differences in more than two groups, a post hoc test like Tukey shows which groups are similar and which groups are different by assigning letters to each group. If two groups share the same letter (e.g. "a"), then they are significantly similar. If two groups have different letters (e.g. "a" and "b"), they are significantly different.

while these values are higher in the AmE group (52.82 for Task 1 and 58.43 for Task 2) and in the BrE group (52.04 for Task 1 and 59.55 for Task 2). The post hoc test reveals that there are two different types of values: Spanish with letter “a”, on the one hand, and American and British English with letter “b”, on the other. This demonstrates the usefulness of this acoustic measurement to account for typologically different rhythms, allowing for a better analysis of non-native English speech.

3.2 Non-native speakers

After measuring speech rhythm in native speakers and proving that VarcoV distinguishes two different types of rhythm, L2 speakers of English are analyzed in order to observe the way their rhythm develops along their formal instruction at university. Fig. 2 shows the results obtained in L2 learners. It includes the mean values, their standard deviation, the ANOVA analysis and the Tukey post hoc test. In the latter, the results of native speakers are also included for a better view of the data.

Task	A	B	C	<i>F</i>	<i>p</i> -value	Post hoc ^a					
	<i>Mean</i> (<i>SD</i>)	<i>Mean</i> (<i>SD</i>)	<i>Mean</i> (<i>SD</i>)			Sp	A	B	C	BrE	AmE
1	38.03 (5,39)	46.07 (5.60)	46.78 (5.51)	8.96	<0.0001**	a,b	a	a,b,c	b,c	c	c
2	44.63 (5.05)	46.12 (4.92)	50.71 (6.63)	11.06	<0.0001**	a	a,b	b	b,c	c	c

Fig. 2. VarcoV values and the effect of language groups in non-native speakers.

^a Letters in common are not significantly different ($\alpha=0.05$)

** There is a highly significant difference between groups.

Fig. 2 indicates that groups of speakers are highly significantly different ($p<.0001$) in both elicitation tasks. The Tukey post hoc test demonstrates that there are three main shared values among groups. In Task 1, these shared values correspond to groups Sp, A and B (letter “a”), groups Sp, B and C (letter “b”), and, finally, groups B, C, BrE and AmE (letter “c”). This means that there are some properties of L2 speakers’ speech that are shared with their mother tongue and with the target language. More precisely, groups A and B are similar to the Spanish group, while groups B and C are similar to the English groups.

Task 2, on the other hand, has shown that there are some shared values only in the non-native speakers, as indicated by letter “b” of the Tukey test in Fig. 2. This is viewed as the particular properties that an L2 may have, which are not found in the mother tongue or in the target language. Nevertheless, group A shares some values with the Spanish groups (letter “a”) and the only non-native speakers that share values with the target language is group C (letter “c”).

Mean VarcoV values reveal that non-native speakers show intermediate values towards a more stress-timed rhythm, similarly to what has been found in the literature (e.g. Ordin and Polyanskaya, 2014, 2015; Pellegrino, He, and Dellwo, 2017; White and Mattys, 2007). This is shown in Fig. 3 as the dotted circle indicates.

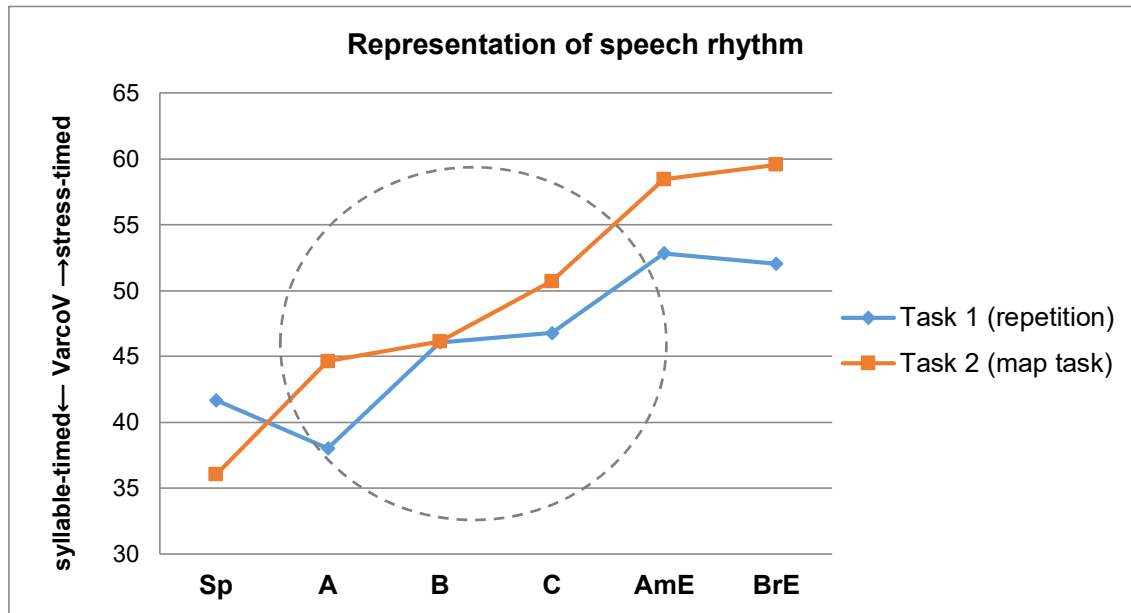


Fig. 3. Representation of speech rhythm in native and non-native speakers.

Another aspect that can be analyzed in the data is related to speech rhythm as affected by the type of task. Group B is the only group of L2 speakers who do not show differences in their rhythm. Perhaps this could be explained by the amount and type of formal instruction that this group has received. Learners in this group have gained recent knowledge on English phonetics and phonology, which could allow them to keep a more steady rhythm in different elicitation tasks. Group A has received little formal instruction and this might explain its similarity to the Spanish group. Group C corresponds to the learners that have reached a more similar value to the English native groups. This could be interpreted as the consequence of having had more years of instruction at university.

However, when comparing the two elicitation tasks, there are differences in some non-native groups.³ The mean values in native speakers of English is 52.43 VarcoV for Task 1 (repetition) and 58.99 VarcoV for Task 2 (map task). Group A has not shown major differences with respect to native speakers: 14.4 in Task 1 and 14.36 in Task 2. This means that their rhythm deviates from native values in the same way in both tasks. Groups B and C have indicated that their speech rhythm is less stress-timed in the map task. In Group B, there is a 6.36-VarcoV

³ Native groups also show a difference between the two elicitation tasks. In the Spanish group this difference is not significant ($F=0.07$; $p=.7948$). In the English groups there is significant different between tasks ($F=7.87$; $p=.0094$), but this analysis is beyond the scope of the present study. More research needs to be done in order to explain why English native speakers have shown a difference in their speech rhythm according the type of elicitation task.

difference in Task 1, while a 12.87-VarcoV difference in Task 2. In Group C, this difference corresponds to 5.65 VarcoV in Task 1 and to 8.28 VarcoV in Task 2. The greater difference that the intermediate and advanced groups show in Task 2 with respect to native speakers of English might be due to the fact that participants were not able to monitor their language, because the task encouraged them to give accurate instructions in relation to the map. When eliciting spontaneous speech, participants focus more on the content than on the way language is being transmitted (Fuchs, 2016: 103–4) and more implicit knowledge can be elicited (Ellis, 2015). In other words, the VarcoV indices indicate that groups B and C have obtained more similar values to those of native speakers, but their rhythm is less native-like in more spontaneous production.

4. Conclusion

In this study speech rhythm has been conceived of as a consequence of vowel reduction, stress and type of syllable, which is measurable by means of acoustic cues such as duration. By means of the metric VarcoV, the development of English speech rhythm in instructed learners of English has been accounted for in two elicitation tasks. The initial group shares more rhythmic characteristics with their mother tongue in both elicitation tasks. The intermediate and advanced groups get more native-like values, especially in the repetition task. With the map task, however, these two groups have produced a more syllable-timed rhythm, probably due to the lack of self-monitoring.

Overall results indicate that the metric VarcoV can demonstrate the gradual development of learners' speech rhythm, which starts from a more syllable-timed rhythm (as indicated by similar values to Spanish) to a more stress-timed rhythm (as indicated by more similar values to English native speakers). This means that learners' phonology is developing towards a more stress-timed syllable, especially in speech in which speakers can monitor their pronunciation. Knowing how, if ever possible, non-native speakers can reach native values is a matter of more research. As far as further research in L2 phonology is concerned, the productivity of rhythm metrics can serve as a useful tool for measuring the production of non-native speakers who receive a different type of instruction.

Bibliography

- BERTINETTO, Pier Marco. 1989. 'Reflections on the Dichotomy "Stress" vs. "Syllable-Timing"'. *Revue de Phonétique Appliquée* 91–93: 99–130.
- BOERSMA, Paul, and David WEENINK. 2015. 'Praat: Doing Phonetics by Computer (Version 4.5) [Computer Program]'. *Version 5.4.08*, [Http://www.Praat.Org/](http://www.praat.org/).
- COUNCIL OF EUROPE. 2001. *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council of Europe.

- DASHER, Richard, and Dwight BOLINGER. 1982. 'On Pre-Accentual Lengthening'. *Journal of the International Phonetic Association* 12: 58–69.
- DAUER, Rebecca M. 1983. 'Stress-Timing and Syllable-Timing Reanalyzed'. *Journal of Phonetics* 11 (1): 51–62.
- DAUER, Rebecca M. 1987. 'Phonetic and Phonological Components of Language Rhythm'. In *11th International Congress of Phonetic Sciences*, 447–450. Estonia: Tallinn.
- DELLWO, Volker. 2006. 'Rhythm and Speech Rate: A Variation Coefficient for DeltaC'. In *Language and Language-Processing*, edited by Pawel KARNOWSKI and Imre SZIGETI, 231 – 241. Frankfurt am Main: Peter Lang.
- DELLWO, Volker, and Petra WAGNER. 2003. 'Relations between Language Rhythm and Speech Rate'. In *Proceedings of International Congress of Phonetic Science*, edited by Maria Josep SOLÉ, Daniel RECASENS I VIVES, and Joaquin ROMERO, 471–474. Barcelona: Casual Productions.
- ELLIS, Rod. 2015. 'Form-Focused Instruction and the Measurement of Implicit and Explicit L2 Knowledge'. In *Implicit and Explicit Learning of Languages*, edited by Patrick REBUSCHAT, 417–441. Amsterdam: John Benjamins. <https://doi.org/10.1075/sibil.48.17ell>.
- ESPINOSA, Gonzalo Eduardo. 2018. 'El Ritmo Del Español y El Del Inglés. Repercusiones Metodológicas Sobre Su Medición Acústica'. *Quintú Quimün*, 2018.
- FUCHS, Robert. 2016. *Speech Rhythm in Varieties of English: Evidence from Educated Indian English and British English*. Singapore: Springer.
- GABRIEL, Christoph, and Elena KIREVA. 2014. 'Prosodic Transfer in Learner and Contact Varieties'. *Studies in Second Language Acquisition* 36 (02): 257–281. <https://doi.org/doi:10.1017/S0272263113000740>.
- GASS, Susan M, and Alison MACKEY. 2012. *Data Elicitation for Second and Foreign Language Research*. New York: Routledge.
- MAIRANO, Paolo. 2011. 'Rhythm Typology: Acoustic and Perceptive Studies'. PhD thesis, University of Turin.
- MAIRANO, Paolo, and Antonio ROMANO. 2010. 'Un Confronto Tra Diverse Metriche Ritmiche Usando Correlatore'. In *La Dimensione Temporale Del Parlato (Proc. of the V National AISV Congress, University of Zurich, Collegiengebäude)*, edited by Stephan SCHMID, Michael SCHWARZENBACH, and Dieter STUDER, 79–100. Zurich, Switzerland.
- ORDIN, Mikhail, and Leona POLYANSKAYA. 2014. 'Development of Timing Patterns in First and Second Languages'. *System* 42 (1): 244–257. <https://doi.org/10.1016/j.system.2013.12.004>.

- ORDIN, Mikhail, and Leona POLYANSKAYA. 2015. 'Perception of Speech Rhythm in Second Language: The Case of Rhythmically Similar L1 and L2'. *Frontiers in Psychology* 6 (316): 1–15. <https://doi.org/10.3389/fpsyg.2015.00316>.
- O'ROURKE, Erin. 2008. 'Speech Rhythm Variation in Dialects of Spanish: Applying the Pairwise Variability Index and Variation Coefficients to Peruvian Spanish'. In *Proc. Fourth Conf. on Speech Prosody*, 431–34. Campinas, Brasil. <http://isle.illinois.edu/sprosig/sp2008/papers/id173.pdf>.
- PELLEGRINO, Elisa, Lei HE, and Volker DELLWO. 2017. 'Computation of L2 Speech Rhythm Based on Duration and Fundamental Frequency'. In *Elektronische Sprachsignalverarbeitung 2017*, edited by Jürgen Trouvain, Ingmar Steiner, and Bernd Möbius, 246–53. Dresden: University of Zurich.
- RAMUS, Franck, Marina NESPOR, and Jacques MEHLER. 1999. 'Correlates of Linguistic Rhythm in the Speech Signal'. *Cognition* 73: 265–92. [https://doi.org/10.1016/S0010-0277\(99\)00058-X](https://doi.org/10.1016/S0010-0277(99)00058-X).
- SCHWARTZ, Bonnie D, and Rex A SPROUSE. 1994. 'Word Order and Nominative Case in Nonnative Language Acquisition: A Longitudinal Study of (L1 Turkish) German Interlanguage'. In *Language Acquisition Studies in Generative Grammar: Papers in Honor of Kenneth Wexler from the 1991 GLOW Workshops*, edited by T HOEKSTRA and Bonnie D SCHWARTZ, 317–368. Amsterdam: John Benjamins.
- TSOI, Wai-chuen. 2016. 'Acquisition of English Speech Rhythm by Hong Kong Cantonese Learners'. MA thesis, Pokfulam, Hong Kong SAR: Hong Kong University.
- WHITE, Laurence, and Sven L MATTYS. 2007. 'Calibrating Rhythm: First Language and Second Language Studies'. *Journal of Phonetics* 35 (4): 501–522. <https://doi.org/10.1016/j.wocn.2007.02.003>.
- WIGET, Lukas, Laurence WHITE, Barbara SCHUPPLER, Isabelle GRENON, Olesya RAUCH, and Sven L MATTYS. 2010. 'How Stable Are Acoustic Metrics of Contrastive Speech Rhythm?' *The Journal of the Acoustical Society of America* 127 (3): 1559–1569. <https://doi.org/10.1121/1.3293004>.