

Rhythm metrics and the production of English L1/L2

Anne Tortel, Daniel Hirst

Laboratoire Parole et Langage, UMR 6057 CNRS
Aix-Marseille Université, Aix-en-Provence, France

anne.tortel@lpl-aix.fr, daniel.hirst@lpl-aix.fr

Abstract

This study investigates rhythmic parameters in the production of French learners in a dual perspective: (i) to analyse the influence of rhythm of the native language (L1=French) on the target language (L2=English) and, (ii) to provide prosodic evaluative criteria for French speakers' productions. The method used is a comparative analysis of French and native speakers' productions using different rhythm metrics. Based on the analyses of the ANGLISH corpus, the results show that it is possible to foresee the rhythmic tendencies and to distinguish between native and non-native speakers by a combination of rhythmic parameters. A discriminant analysis allows the classification of the speakers into three different levels of group.

Index Terms: rhythm, metrics, English L1/L2, evaluation, French learners

1. Introduction

The literature has begun to confirm the importance of prosodic features in learners' overall intelligibility and perceived comprehensibility [1, 2, 3]. More precisely, [4] analyses the acquisition of the rhythm of English spoken by foreign learners from different backgrounds. The study shows that L2 learners' non-native rhythm is caused by different factors including inappropriate pauses, stress shifting, and insufficient durational differences between stressed and unstressed syllables. This may consequently affect (i) the variability of vocalic interval duration by the lack of difference between the duration of stressed and unstressed vowels and (ii) the variability of consonantal interval duration by the lack of a linking mechanism or an over-articulation of consonant elements. These difficulties seem to be partly due to the different rhythmic structure of L1 & L2 languages in that L1 rhythm may influence the rhythmic production in the target language. This phenomenon may appear to be even more salient when applied to French and English as the two languages are considered to be rhythmically opposed [5,6] and have very different rhythmic structures [7,8,9].

From these observations, various rhythm metrics have recently been proposed [10, 11, 12, 13] establishing objective criteria for classifying languages [10,11,13] or dialects [14,15,16] either as stress-timed, syllable-timed or mora-timed, as [5,6] suggested a few decades ago and failed to demonstrate (see [17] for more details). However, little is known about the evaluation of deviant productions such as those obtained from native/non-native speakers and more particularly in the field of French learners of English. A few rare studies [18,19] have investigated the rhythmic productions of L2 learners and shown that the scores obtained are situated at an intermediate level between the productions

of L1 and those of L2. The aim of this research is therefore to analyse different rhythmic parameters in the production of French learners using different combinations of rhythmic factors and see if any possible rhythmic criteria can be suitable for an objective evaluation of prosody.

2. Experiment

2.1. Rhythm metrics

Among the metrics investigated for this study, we used the most popular ones presented in the literature. The common point of all the metrics is that the calculation is based on (i) the duration of vocalic intervals and (ii) the duration of consonantal intervals. The metrics used are summarised in the following table:

metrics	authors	description
% V	[10]	% duration of vocalic intervals
$\Delta C, \Delta V$	[10]	standard deviation of duration of cons. & voc. intervals
rPVI (c,v)	[11]	raw index of variability between duration of successive cons. & voc. intervals
nPVI (c,v)	[11]	normalised index of variability between duration of successive cons. & voc. intervals
cvC,cvV	[12,13]	coefficient of variation of duration of cons. & voc. intervals

Table 1. Summary of the metrics used in the experiment

Our hypothesis is that since rhythm measurements tend to demonstrate rhythmic differences between different languages and dialects, they can thus also be used to identify rhythmic differences between L1 (French) and L2 (English), hence to distinguish the oral productions of French learners from those of native speakers.

2.2. Database: ANGLISH

2.1.1 Materials

The ANGLISH database [20] was created in the research unit *Parole et Langage*. ANGLISH is currently made up of more than 5:30 of spoken English from both L1 & L2 speakers. For this experiment, 4 passages of 5 semantically linked sentences were analysed; this represents 1:30 of readings of the ANGLISH corpus. These passages were selected for the length of the sentences (from 7 to 23 syllables) and for the presence of different polysyllabic words containing vowel reductions such as 'unfortunately' and 'comfortable'.

2.1.2 Speakers

63 speakers were recorded in an anechoic room performing reading and repeating tasks as well as continuous unprepared speech. Three groups were recorded according to their level based on precise selective criteria. The aim was to obtain a representative sample of different levels of English: (i) native speakers of British English (GB), (ii) non-specialist working adult-speakers of English (FR1) and (iii) second-year university students of English (FR2). The GB group is composed of 23 speakers (13 F, 10 M), the FR1 and FR2 groups are each composed of 20 learners (10 F, 10 M). All the speakers were volunteers.

2.3. Methodology

The recordings of the reading part (1260 utterances) of the corpus for the 63 speakers were manually segmented into phonemes and labeled with CVC codes using the Praat software [21]. ‘C’ stands for consonantal segment and ‘V’ for vocalic segment. Traditional segmentation criteria were used in order to obtain a more homogenous manual segmentation [22]. A Praat script was used to calculate the different metrics. The rhythmic variation of vocalic and consonantal intervals was calculated according to the following different metrics:

- %V = percent duration of V intervals compared to C + V and multiplied by 100.
- ΔC = standard deviation of duration of C intervals.
- ΔV = standard deviation of V intervals.
- rPVI_C = raw pairwise variability index of successive C intervals.
- rPVI_V = raw pairwise variability index of successive V intervals.
- nPVI_C = normalised pairwise variability index of successive C intervals.
- nPVI_V = normalised pairwise variability index of successive V intervals.
- cvC = coefficient of variation (= standard deviation divided by mean) of duration of C intervals.
- cvV = coefficient of variation of duration of V intervals.

2.4 Analyses

2.4.1 Discrimination of the groups

The first analyses aim at characterising the different groups of speakers according to 3 levels from a Principal Component Analysis and a discriminant analysis. The idea is to reliably determine to which group the speakers belong (FR1, FR2, GB), using the 9 rhythmic parameters presented above. Figure 1 shows a three-way classification of the speakers into three different groups (from left to right): group FR1 (in green) which is located between [-4 and -2], group FR2 (in blue) between [-2 to 1] and finally group of GB (in red) situated between [1 and 4].

As can be seen from the discriminant analysis (summarised in Table 2), the groups FR1 and GB are particularly well distinguished with a prediction of 72% for GB and 69% for FR1. For both groups only 6% of the sentences are mispredicted. The same tendency is found for FR1.

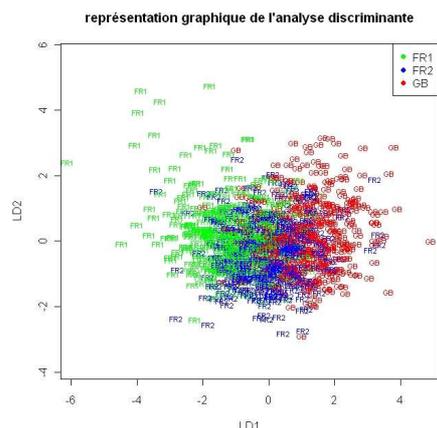


Figure 1. Graphic representation of the discriminant analysis

This can be explained by the fact that some sentences contain fewer consonants and are thus closer to an isosyllabic rhythmic structure. Concerning FR1, it is not so surprising to find 25% of utterances classified as FR2 because some speakers may have a better level of English than the rest of the group. The discrimination of FR2 shows 50% correct predictions. The other 50% are equally shared among FR1 and GB.

real group (in%)	Preditd group		
	FR1	FR2	GB
FR1	69%	25%	6%
FR2	25%	50%	25%
GB	6%	22%	72%

Table 2. Results (%) of the predicted groups by the discriminant analysis

This is quite interesting as it gives a realistic view of the composition of the students group. Some sentences are particularly far from a native rhythmic production whereas others are closer to a native rhythmic production. The ANOVA confirms this classification, reaching a main effect of group on the principle components. (F (2,1257):69.503;pval<2.2^e-16)

2.4.2 Analysis of the metrics

Following Ramus & al. (1999)’s procedure, the standard error analysis is investigated according to the traditional metrics used in the literature. The idea is to detect the rhythm impact of L1 on the production of L2. Ramus & al. showed that the more complex the type of syllable is, the greater the variability in the number of consonants, resulting in a high ΔC and thus a lower %V. Then our hypotheses were that: (i) ΔC scores of learners would be lower due to the complex syllabic structure whereas ΔV scores would be higher, (ii) %V scores would be relatively higher for French speakers, (iii) PVI scores would be higher if the rhythm is close to that of a native speaker, and (iv) cvC and cvV would be higher if the rhythmic production is correctly produced.

Two types of results are observed. On the one hand, the graphic plots obtained (figure 2) confirm our hypotheses for the normalised metrics such as (nPVI-C, nPVI-V) and (cvV, cvC). Scores are rather high for these parameters.

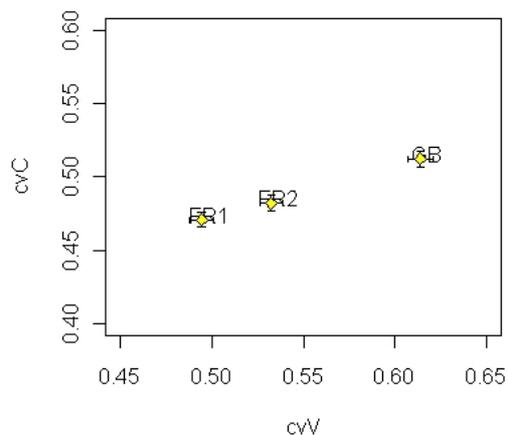


Figure 2. Example of visualisation of standard errors and means for cvV and cvC)

On the other hand, our hypotheses were not validated when looking at the plots of the other parameters, either individually or in combination. %V scores of the French learners (figure 3) were relatively low and even lower than those of the natives. ΔC and rPVI rates of the natives were lower than the scores obtained by the French learners. Non-normalised metrics – as announced in the literature – also do not indicate any rhythm tendencies differentiating native from non-native speakers.

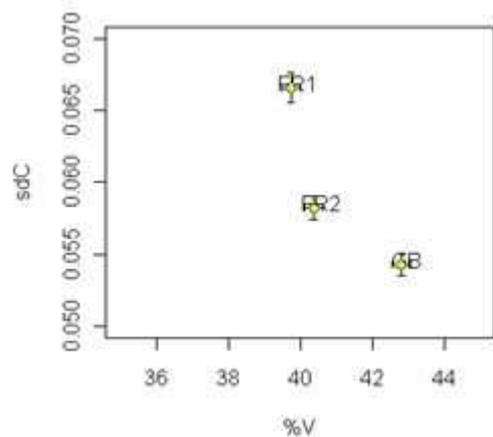


Figure 3. Example of a visualisation of standard errors and means for %V and ΔC)

The discriminant analysis confirms the standard error analysis. The results obtained are summarised in table 3. They show the possibility to distinguish three types of discrimination according to the different metrics. %V and cvC give a better distinction of the learners' productions from those of the natives. The factors ΔC , nPVI_C and nPVI_V as well as the combinations (%V, ΔC); (nPVI_C, nPVI_V) allow a better distinction FR1/FR2. Finally a graduation from FR1 to GB is made possible by the different individual factors: ΔC , rPVI_C, cvV or by the combinations (rPVI_V, rPVI_C); (cvV, cvC); (%V, cvC).

levels of distinction	learners vs natives	FR1 vs FR2	Graduation (FR1-FR2-GB)
1 parameter	% V	ΔC	ΔC
	cvC	nPVI_C	rPVI_C
		nPVI_V	cvV
combination	% V	%V, ΔC	rPVI_V, rPVI_C
	cvV	nPVI_C, nPVI_V	cvV, cvC
			%V, cvC

Table 3. Summary of the different metrics according to the discrimination task

2.4.3 Rhythm characterisation by SVM

A support vector machine analysis was used to determine which combination (from 1 to 4 factors) is the best predictor of rhythmic differences. The best predictions are presented below (table 4).

variable	score
cvV	45,79%
ΔC , cvC	62,30%
%V, sdV, cvV	63,09%
%V, sdV, cvV, nPVI_V	66,74%

Table 4. Summary of the best scores obtained by SVM for combinations from 1 to 4 factors

Results confirm the graphic observations made with the standard-error analysis. The greater the number of parameters combined, the better the prediction. However the prediction (62.30%) of the combination (ΔC , cvC) is nearly as good as that obtained with 3 (63.09%) or even 4 factors (66.74%).

3. Discussion

The results found with the discriminant analysis and the standard error analyses enable us to distinguish between native and non-native speakers as well as different levels within the group of French learners. In the literature, we have seen that non-normalised metrics did not indicate interesting results. The different results found with some of these metrics may be explained (i) by the rhythm of the text itself [17], and (ii) by an effect of hyperarticulation produced by the French speakers. Indeed, in our analyses, the same text was used for all the speakers, so this has the effect of neutralising the difference of rhythmic structure between the two languages. Facing the phonotactic difficulties of the sentences, the speakers may have produced an over-articulation which may explain the high values for the factors ΔC and rPVI.

It has been shown that it was possible to correctly classify the speakers by means of 9 parameters with a global prediction of 69.5%. This discrimination shows a major distinction between FR1 and GB. We observe three types of discrimination according to the different metrics i.e.

- (i) a distinction learners/native speakers,
- (ii) a distinction within the French learners (FR1/FR2),
- and (iii) a graduation indicating 3 levels from FR1 to GB.

The two rhythmic parameters (ΔC , cvC) appear to be the most discriminating combination in predicting the rhythmic

tendencies of the productions. If we compare this combination with those proposed in the literature, none of these binary combinations are found in the list of the best combination calculated by the SVM ((%V-ΔC) = 46.58% / (nPVI_V-rPVI_C) = 48.65% / (%V-cvV) = 49.68%, (%V-cvC) = 46.34%). These results are far from the 62.30% obtained with the combination ΔC,cvC. This suggests that such combinations could be used as an objective evaluation of the rhythmic productions of French learners and could be integrated into computer-assisted systems for teaching English prosody.

For a complete automation of this task, we hope to implement an algorithm such as that described by [24] in order to obtain automatic estimations of the relative durations of different vocalic and consonantal segments of native and non-native speakers' productions.

To pursue this study, other metrics, such as [23] for example need to be tested. It would also be interesting to test the result of the application of the combination ΔC,cvC to productions of spontaneous speech. We also intend to investigate other corpora presenting data such as pathological disabilities or synthetic speech.

4. Conclusions

It has been shown that it is possible not only to make a fairly reliable distinction between L1 and L2 rhythm productions but also to classify different levels of non-native speakers' productions. This research opens perspectives for a considerable number of further studies and needs to be applied to the spontaneous speech part of the ANGLISH database. The final aim is to standardise an automatic evaluation metric for non-native, and more generally non-standard, speech rhythm.

5. References

- [1] Jilka, M., *The contribution of Intonation to the Perception of Foreign Accent*, Doctoral Dissertation, Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS), 6 (3), University of Stuttgart, 2000.
- [2] Hahn, L.D., Primary Stress and Intelligibility: Research to Motivate Teaching of Suprasegmentals, *TESOL Quarterly*, 38:201-223, 2004.
- [3] Field, J., Intelligibility and the Listener: the Role of Lexical Stress, *TESOL Quarterly* 39:399-423, 2005.
- [4] Adams, C., *English Speech Rhythm and the Foreign Learner*, The Hague: Moulton, 1979.
- [5] Pike, K.N., *The Intonation of American English*, Ann Arbor, University of Michigan Press, 1945.
- [6] Abercrombie, D., *Elements of General Phonetics*, Edinburgh, Edinburgh University Press, 1967.
- [7] Delattre, P., A comparison of Syllable Length Conditioning among Languages, Off-print, *IRAL*, vol.4 (1), p. 183-198, 1966.
- [8] Wenk, B.J. (1985) Speech Rhythm and Second Language Acquisition, *Language and Speech*, 28 (2), p.157-175, 1985.
- [9] Bolinger, D., *Two kinds of Vowels, Two Kinds of Rhythm*, Bloomington IN, USA.: IULC Publications, 68p., 1981.
- [10] Ramus, F., Nespors, M., Mehler, J., Correlates of Linguistic Rhythm in the Speech Signal, *Cognition* 72:1-28, 1999.
- [11] Grabe, E., Low, E.L., Durational Variability in Speech and the Rhythm Class Hypothesis, In Gussehoven C., Warner, N. (eds), *Laboratory Phonology 7*, coll. *Phonology and phonetics*, 4-1, Berlin, Allemagne: Mouton de Gruyter, p.515-546, 2002.
- [12] Dellwo, V., Rhythm and Speech Rate: A Variation Coefficient for Δ C., *Language and Language-Processing*, 231-241, 2006.
- [13] White, L., Mattys, S., Calibrating Rhythm: First Language and Second Language Studies, *Journal of Phonetics*, 35 (4): 501-522, 2007.
- [14] Fernandez-Cruz, R., *Analyse phonologique et phonétique du portugais brésilien parlé par les communautés noires d'Amazonie*, thèse de doctorat, Université de Provence, 2000.
- [15] Low, E.L., Grabe, E., Nolan, F., Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English, *Language and Speech* 43:377-401, 2000.
- [16] Barkat-Defradas, M., Hamdi, R., Pellegrino, F., De la caractérisation linguistique à l'identification automatique des dialectes arabes, 2004.
- [17] Hirst, D., The Rhythm of Text and the Rhythm of Utterances: from Metrics to Model, *Interspeech 09 Proceedings*, Brighton, 2009.
- [18] Lin, H., Wang, Q., Vowel Quantity and Consonant Variance: a Comparison Between Chinese and English, Between Stress and Tone conference, Leiden, June 2005, 2005.
- [19] Grenon, I., White, L. Acquiring Rhythm: a Comparison of Canadian English and Japanese, In H.Chan, H. Jacob & E. Kipia (eds.), *BUCLD, Proceedings of the 32nd annual Boston University Conference on Language development*, Somerville, Massachusetts: Cascadilla Press, 32: 155-166, 2008.
- [20] Tortel, A., ANGLISH: Base de données comparatives de l'anglais lu, répété et parlé en L1 & L2, *Travaux Interdisciplinaires du Laboratoire Parole et Langage (TIPA)*, 27 :111-122, 2008.
- [21] Boersma, P.; Weenick, D., *Praat: a System for Doing Phonetics by Computer*, version 5.1.04., www.praat.org, 2009.
- [22] Peterson, G.I.; Lehiste, I., Duration of Syllable Nuclei in English, *Journal of the Acoustical Society of America*, 32:693-703, 1960.
- [23] Bertinetto, P.M., Bertini, C., On Modeling the Rhythm of Natural Languages, *Proceedings of the Fourth International Conference on Speech Prosody*, Campinas, Brazil, 2008.
- [24] Loukina, A.; Koschanski, G.; Shih, C.; Keane, E.; Watson, I. Rhythm Measures with Language Independent Segmentation, *Interspeech 09 Proceedings*, Brighton, England, 2009.