



HAL
open science

Rhythm and Rhythmic Variation in British English: Subjective and Objective Evaluation of French and Native Speakers

Anne Tortel, Daniel Hirst

► **To cite this version:**

Anne Tortel, Daniel Hirst. Rhythm and Rhythmic Variation in British English: Subjective and Objective Evaluation of French and Native Speakers. Speech Prosody 4, May 2008, Campinas, Brazil. hal-01363442

HAL Id: hal-01363442

<https://hal.science/hal-01363442v1>

Submitted on 21 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rhythm and Rhythmic Variation in British English: Subjective and Objective Evaluation of French and Native Speakers

Anne Tortel, Daniel Hirst

CNRS Laboratoire Parole et Langage
Université de Provence, Aix-en-Provence, France
anne.tortel@univ-provence.fr;daniel.hirst@lpl-aix.fr

Abstract

This paper presents results from an ongoing research on the evaluation of the prosody of British English spoken by French learners and native speakers. This pilot study examines two potential rhythmic criteria: the analysis of the anacrusis/narrow rhythm unit and that of the pairwise variability index (PVI). The method used is a comparative analysis of French and native speakers' productions: on the one hand a subjective evaluation of the prosody of the speakers including the natives, on the other hand an objective evaluation aiming at correlating acoustic parameters with the level of the speakers. This preliminary study showed interesting results: even though the level of significance was not reached the two rhythmic parameters could be considered as relevant prosodic criteria and are in need of further investigation.

1. Introduction

Research on second language acquisition has been largely developed in the last decades but the examination of prosodic aspects of it is still rather rare. The teaching of prosody had for a long time been neglected. In the middle of the 60's [1] the first attempts were made to find a learning method for prosody. Even though the results of the experiments described in [2] & [3] were not significant, other studies such as those of [4] and [5] illustrated the efficiency of the visualisation of intonation patterns. Recent research has led to the creation of new software providing either a pitch visualiser such as *Winpitch LTL*, [6] or a computer-assisted system for teaching English prosody such as *BetterAccentTutor*, [7] devoted to the prosody of general American or as the *Prosodia* system, [8] providing exercises on British English intonation and rhythm including segmental difficulties.

One thing crucially lacking in all these systems is an objective evaluation based on concrete linguistic information, usable and open to all users. The only feedback found in recent studies is a visualisation of intonation curves and hardly anything has been done with rhythm.

The aim of this study was to run a pilot study on possible rhythmic criteria suitable for an objective evaluation of prosody. Two main rhythmic parameters were selected. The first criterion was the study of the anacrusis/narrow rhythm unit [9] [10], considered as representative of the rhythmic structure of the English language with, unlike in French, considerable reduction of unstressed syllables; indeed one of the most difficult aspects for a French speaker is to produce these vowel reductions. The second experiment was the analysis of the PVI (Pairwise Variability Index) [10]; French being generally classified as a 'syllable-timed' language as opposed to the English which is described as 'stress-timed'. It

was predicted that the PVI of the best speakers would be closer to the PVI rates of the native speakers and that of the model unlike the PVI of low level speakers.

2. Experiment

2.1. Corpus

The corpus was composed of 1720 sentences recorded in a repetitive mode. Forty speakers were asked to repeat 43 sentences. The instruction given was: 'repeat the sentences trying to imitate the way they are produced'. The phrases were taken from the software 'Prosodia' illustrating various segmental and accentual difficulties and different intonation patterns. Among the sentences, three were selected for the analysis of the PVI, two of 7 syllables and one of 16 syllables chosen as they are composed of an alternation between stressed and unstressed syllables.

2.2. Speakers

In order to obtain a more representative sample of different levels of English, four groups were formed according to their academic level: non-specialist speakers in English, first-year university students of English, fourth and fifth year-students and native speakers of British English. Each group is composed of 10 adults aged between 20 and 35. They were all volunteers.

2.3. Subjective evaluation

The sentences were evaluated by two French experts in English phonetics and two British English native speakers. The subjective evaluation was presented to the evaluators as a test of perception, comparing the productions with those of the given model. The task was to evaluate the quality of the productions according to two main criteria: (i) general production (quality of repetition) that is, at what level the speaker would be classified and (ii) rhythm, including vowel reduction and the place of the stress. The model was given as reference. A scale of 5 was used going from very bad (1) to excellent (5).

The aim was to test for a correlation between the academic level and the scores obtained for each group. We calculated first the mean of the two scores (production and rhythm) for each speaker and then an overall mean for each speaker. Finally we obtained the mean for the whole group. The results represent a distance between a reference (the model) and the quality of the production. The evaluation showed that the main tendency was respected: on the whole the average obtained for each group corresponded fairly well to their basic level.

Table 1: mean scores for each group

	GR 1	GR 2	GR 3	GR 4
General score	2	2,5	3,5	4,5
Rhythm score	2,5	3	4	4,5
mean score	2,3	2,65	3,775	4,42
Average/group	2	2,5	4	4,5
	LOW	MEDIUM	GOOD	VERY GOOD

3. Analysis: objective evaluation

3.1. Analysis of the anacrusis

Two sentences were analysed and manually segmented into rhythm units (with Praat [11] software) following the model of structure proposed by [9] that is, into anacrusis (ANA) and Narrow Rhythm Units (NRU) where an NRU contains one accented syllable followed by any following unaccented syllables until the end of the word. The ANA corresponds to any syllables not included in an NRU.

The following acoustic measurements were made:

- total duration of the sentence
- % of each anacrusis according to the total duration of the sentence, noted %ANA
- sum of % of anacrusis according to the total duration of the sentence.

3.1.1. Statistical analysis

The measures were intended to test for a correlation between the average score from the subjective evaluation and the calculated percentage of anacrusis. We expected good scores to be correlated with a low percentage of anacrusis for successful productions.

3.1.2. Results

The first sentence containing three anacrusis (underlined parts) is presented below transcribed with the SAMPA alphabet:

-it was a 'heavy 'lorry with a 'full 'load of 'wood- (dive)
 [Itw@z@hevI'QrIwID@fU'I@Ud@v'wUd]

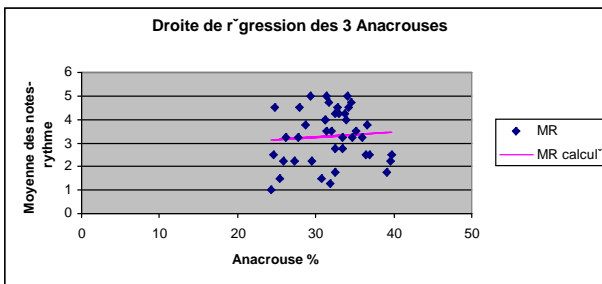


Figure 1: Graph of regression of the anacrusis of the first sentence- $a = 0,020$; $b = 2,662$ $\rightarrow y = 0,02x + 2,662$

Figure 1 shows that the points are concentrated between 28 and 38% and between 3 and 5. The regression line shows that

the higher the percentage, the better the scores. With the detailed study of each anacrusis, it was noted that the anacrusis at the beginning of the sentence shows the same correlation. The two other anacrusis within the sentence follow our expectations, that is, the scores are good when the percentage of anacrusis is low.

The second sentence containing two anacrusis (underlined parts) is presented below transcribed with SAMPA:

-There are 'two 'very 'dirty 'pair of 'brown 'leather 'shoes- (glide up)
 [De@'tu:'verI'd3:tI'p'e@z@f'b'raUn'l'eD@'Su:z]

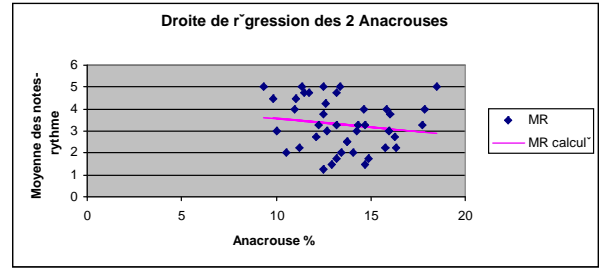


Figure 2: Graph of regression of the anacrusis of the second sentence- $a = -0,078$; $b = 4,354$ $\rightarrow y = 0,078x + 4,354$

The slope of the regression line in figure 2 is reversed. The points are concentrated between 11 and 16% for scores from 1 to 5. Two groups of points can be considered: the first group having scores between 4 and 5 for the lower percentage that is between 10 and 12% and the second group having scores between 1 and 3 for a percentage of 12 to 16%. With the detailed study of each anacrusis, our expectations were only correct for the anacrusis located within the sentences.

3.2. Rhythmic variability-a native/non native comparison: analysis of the PVI

The speakers' productions were labelled and manually segmented into phonemes with the use of the Praat software. A Praat script removed all the boundaries between two adjacent consonants or vowels so that the sentences were segmented into vocalic and consonantal intervals.

The rhythmic variation of vocalic and consonantal intervals was calculated by the Pairwise Variability Index (PVI) which distinguishes the degree of variability between two successive measures. The PVI was calculated using the following equation (example for the vocalic intervals 'V'):

$$100 * \sum_{i=1}^{n-1} \frac{|V_i - V_{i+1}|}{(V_i + V_{i+1})} \quad (1)$$

It was predicted that, first the native speakers and the model should have close PVI rates and secondly that the better the non native speakers' productions and the closer to an English rhythmic structure the rhythm was, the closer to the native speakers' and model' s PVI the rates would be.

3.3. Results

The histogram shows that the vocalic PVI is more relevant than the consonantal one. The vocalic PVI shows a progressive evolution going from group 1 with 26,1 to group 2 and 3 with 30,05 to group 4 (30,9) and to the model with 36,4. The rate increases with the level of the group but the consonantal PVI did not reach the level of significance.

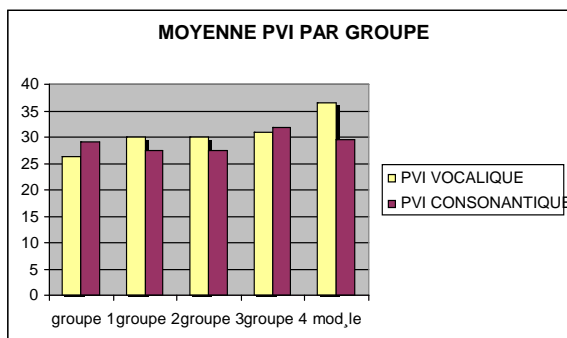


Figure 3: Histogram of the averages of the PVI for each group for the three sentences

Table 2: detailed averages of the PVI for each group

	PVI VOC	PVI CONS
G1	26,1811704	29,0543667
G2	30,0598667	27,4753667
G3	30,0920667	27,5382
G4	30,9481667	31,8381333
MODEL	36,4156667	29,6446667

The detailed results of each sentence show various differences.

For sentence 1 'Tinker, tailor, soldier, sailor, rich man, poor man, beggarman, thief', the consonantal PVI is higher than the vocalic PVI for each group. A difference in both PVI is to be noted between the rates of group 1 with 18,95 (voc.PVI) and 30,7 (cons.PVI) and the native speakers (23 voc.PVI and 33,7 cons.PVI) or the model (35,7). The native speakers and the model have approximatively the same rate of PVI.

Table 3: detailed results of PVI of each group- sentence 1 (K)

moyenne	PVI V	PVI C	ECART TYPE	
G1	18,8851111	30,716	4,400981	4,03126
G2	22,1393	28,7019	4,500318	2,58787
G3	18,9455	30,977	2,471736	3,31840
G4	23,6783	33,6764	3,368078	3,40071
MODEL	23,034	35,694		
Phrase K	n vowels = 16	n cons = 17	(selon modèle)	

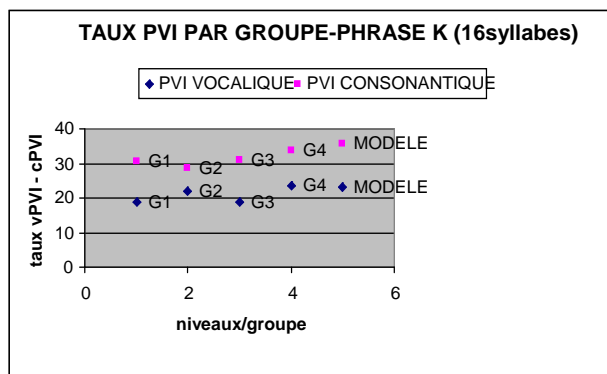


Figure 4: PVI rates for each group - sentence 1 (K)

For sentence 2 "have you got an empty cup", according to the figure the vocalic PVI is higher and more relevant for all the groups than the consonantal PVI which is more variable and does not follow the level of the groups. The vocalic PVI increases following three stages: group 1 with about 23, group 2, 3, 4 with about 30/31 and the model with 41,5. There is a difference again between group 1 and the native speakers and even more with the model.

Table 4: detailed averages of the PVI for each group

moyenne	PVI V	PVI C	ECART TYPE	
G1	23,0085	25,1913	7,139286	9,363950
G2	30,3941	19,9734	6,430944	4,367113
G3	31,7411	20,6905	6,578820	4,272206
G4	30,1319	27,8909	6,50265	7,740311
MODEL	41,503	22,214		
Phrase A	n vowels = 7	n cons = 8	(selon modèle)	

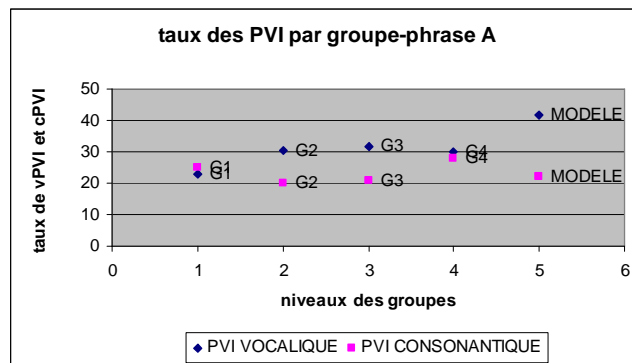


Figure 5: PVI rates for each group - sentence 2 (A)

For sentence 3 'he said cheese to the waiter', the vocalic PVI is higher than the consonantal one and once again more relevant than the consonantal PVI. A distinct difference is to be noted with the vocalic PVI between group 1 (36,6) and the native speakers (39,03) and even more with the model (44,71). Within these extremes the graphic shows a progressive increase of the vocalic PVI which also follows the level of the groups going from a low rate for a low level to a

higher rate for the best levels. The consonantal PVI is not relevant and variable according to the group.

Table 5: detailed results of PVI of each group – sentence 3 (B)

moyenne	PVI V	PVI C	ECART TYPE	
G1	36,6499	31,2558	4,576214	3,475018
G2	37,6462	33,7508	5,552858	4,545146
G3	39,5896	30,9471	4,927473	5,600734
G4	39,0343	33,9471	3,387276	5,160099
MODEL	44,71	31,026		
Phrase B	nvowels = 7	ncons = 7	(selon modèle)	

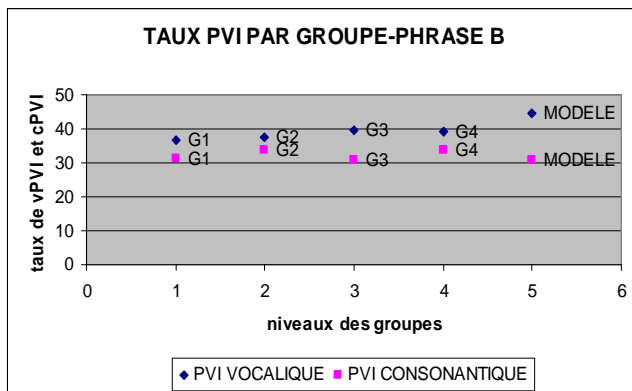


Figure 6: PVI rates for each group-phrase B

4. Discussion

The study made on the anacrusis showed that the results corresponded to our hypothesis with the anacrusis situated within sentences but not with the anacrusis located at the beginning of sentences. The lower the rates of the anacrusis, the better the scores on the subjective evaluation were. It can be deduced that the anacrusis within the sentences are pronounced quickly enough and on the whole correctly for the groups which obtained better marks, that is, groups 3 and 4. Although this experiment showed preliminary positive results, it needs to be studied thoroughly with for example the analysis of the phoneme durations and formants to check the quality of the vowels (reduction or not).

Concerning the second study, it was noticeable that the vocalic PVI in most cases was more relevant than the consonantal PVI. It showed that the hypothesis predicting that the better the production (at a rhythmic level), the higher and also the closer to the native speakers the PVI would be, was verified by the vocalic PVI. With the detailed study of the three sentences, two showed significant differences with the vocalic PVI and especially between groups 1 and 4 (the native speakers); the third sentence showed this tendency with the consonantal PVI. It could be agreed that two extreme groups were more relevant than the in-between groups. This study which turned out to be an interesting starting point will be enlarged to other methods using different parameters such as [12] or [13] and reinforced once again with the analysis of the vowel durations and formants to check for vowel reduction.

5. Conclusion

This pilot study opens perspectives for a considerable number of further studies. Indeed, this work needs to be reinforced by other parameters: on one hand the project will be further developed as it will be extended to intonation and on the other hand other rhythmic parameters will be investigated much more thoroughly and with much larger quantities of data. The search for prosodic criteria in order to develop an objective prosodic evaluation for the students' productions needs consequently to be pursued actively, the ultimate goal being to standardize an automatic evaluation software for speech prosody.

6. References

- [1] Guberina, P., 1965. La méthode structuro-globale audiovisuelle. *Revue de Linguistique Appliquée* vol.1, 11-30.
- [2] Lane, H.; Buiten, R., 1969. A Self-Instructional Device for Conditioning Accurate Prosody: Trends. *Language Learning*, Valdman, A. New York: Academic Press, 159-174.
- [3] Vardanian, R., 1964. Teaching English through Oscilloscope Displays. *Language Learning* 3-4, 109-117.
- [4] James, E., 1977. The Acquisition of a Second-Language Intonation Using a Visualizer. *Canadian Modern Language Review*, 33-4, 503-506.
- [5] De Bot, K., 1983. Visual Feedback of Intonation I: Effectiveness. *Language and Speech*, vol.26-4, 331-350.
- [6] Martin, P., 2005. Winpitch LTL, un logiciel multimédia d'enseignement de la prosodie. *Alsic*, vol.8, 95-108.
- [7] Kommissarchik, J.; Kommissarchik, E., 2000. Better Accent Tutor – Analysis and visualization of speech prosody. *Proceedings of INSTIL 2000*, Dundee, Scotland, 86-89.
- [8] Herry, N., 2001. Evaluation objective et subjective de la prosodie de l'anglais parlée par des français : apport de l'enseignement assisté par ordinateur. *Doctoral Thesis*, Université de Provence.
- [9] Jassem, W., 1983. *The phonology of Modern English*, Panstwowe Wydawnictwo Naukowe.
- [10] Hirst, D.J. & Bouzon, C. 2005. The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English). in *Proceedings of Eurospeech/Interspeech 2005*.
- [11] Low, E.L.; Grabe, E.; Nolan, F., 2000. Quantitative characterizations of Speech Rhythm: syllable-timing in Singapore English. *Language and Speech*, 43-4, 377-401.
- [12] Boersma, P.; Weenink, D., 1992-2007. Praat, a system for doing Phonetics by Computer. <http://www.praat.org>
- [13] Ramus, F., 2002. Acoustic correlates of linguistic rhythm: perspectives. *Proceedings of Speech Prosody 2002*. Aix-en-Provence. 11-13 April 2002.
- [14] Gibbon, D.; Gut, U., 2001: Measuring speech rhythm. *Eurospeech-2001*, 95-98.