Rhythm Perception. Inter and intra-speech variation

Victoria Marrero Aguiar (vmarrero@flog.uned.es)
Universidad Nacional de Educación a Distancia
Luis E. López-Bascuas (lelopezb@ucm.es)
Universidad complutense de Madrid
José Luis Martín Fernández (joseluisjmartin@gmail.com)
Universidad complutense de Madrid

Introduction

Linguistic rhythm is not directly or automatically derivated of physical events that occur at objectively equal or similar time intervals in the speech signal. It is the human mind which perceives certain physical cues as forming a rhythmic patterns. Decades of research have shown that it is necessary to perform a number of interpretative tasks to hear rhythm (Auer, Couper-Kuhlen y Müller, 1999).

But the fact is that rhythmic perceptions have proven to be very robust to distinguish among languages (Ramus & Mehler, 1999 Ramus, Dupoux y Mehler, 2003; White & Mattys, 2007), or dialects (White & Mattys, 2007; Deterding, 2001; Nolan & Asu, 2009; O'Rourke, 2008; Toledo, 2008, etc.), even in newborns (Ramus 2002). Some studies (Eriksson and Wretling, 1997, Leeman, Kolly y Dellwo, 2014) suggest that linguistic rhythm could even be a personal trait, similar to other motor behaviours, as finger or leg movements, with a high level of intra-individual stability and important inter-individual differences. If it is the case, rhythmic patterns could play an important role in the speaker characterization and identification.

Our aim in this study was to check this hypothesis and to try to find out the acoustic metric more related to the perceived personal linguistic rhythm.

Stimuli and procedure

We obtained eigth of the most extensively used rhythm metrics (%V, ΔV, ΔC, VarcoC, VarcoV, rPVI, nPVI-V y nPVI-C; Ramus, Nespor y Mehler, 1999; Low, Grabe y Nolan, 2000, White y Mattys 2007, among others) in order to characterize the speech rhythm of 12 male speakers, selected from AHUMADA database (Ortega-Garcia, Gonzalez-Rodríguez y Marrero-Aguiar, 2000), in three different reading sessions. All the metrics were extracted from two sequences of six syllables each, the one with a simple CV structure, and the other predominantly complex, CVC. And then, the greatest contrast between two different sessions of the same speaker, and two different speakers have been selected in 20 pairs of matched stimuli.

The stimuli were manipulated by means of a Praat script (Lahoz 2012) and re-synthesized with Mbrola (Dutoit et al. 1996), in order to turn all the consonants into /s/ and all vowels into /a/, matching also the differences in pitch (*f*0) and intensity. The geolectales features and rate of speech were also controlled (differences < JND, Martin Fernandez Lopez and Marrero Bascuas Aguiar, 2014).

c) Task: 2IAX paradigm, four combinations of each pair (aa, ab, ba, bb) showed repeated 8 times in random order (in two separate blocks, CV and CVC), totaling 640 trials. The test lasted 45-50 minutes (depending on the response time, that was free).

d) Subjects: 10 university students living in Madrid, 4 men and 6 women

Results and conclusions

Overall rate of correct identification: 69%. When the stimuli were the same, correct rejections reached 84% of presentations, but when they were different, only 54%. As can be seen in Table 1, results in CV sequences are always better than CVC and show the expected pattern: good discrimination for different stimuli coming from different speakers, and poor discrimination for those from the same speaker (differences statistically significant). On the contrary, the stimulus sequences obtained with CVC structure have success rates below 50% in all cases, and even better discrimination for intra-speaker differences than inter-speaker (differences not significant). %V is the metric that fits better with our

Table 1: General results. Rate of hits by metrics

CV			
INTER-SPEAKER	%	INTRA-SPEAKER	%
%V	93,75	%V	26,25
VarcoC=nPviC	81,88		
ΔV	84,38	ΔV	55,13
VarcoV	75,00	nPvi	58,13
		ΔC=VarcoC= rPvi=nPviC	66,25
rPvi	51,88	VarcoV	66,88
ΔC	41,25		
nPvi	39,10		
CVC			
%V	61,25	%V	45,00
ΔC=VarcoC	59,38	ΔC=VarcoC	39,74
		ΔV=VarcoV=nPvi	39,38
ΔV=VarcoV=nPvi=nPviC	36,54	rPvi=nPviC	26,25
rPvi	29,38		

The results included in each common cell do not show significant differences between them.

In sum, individual rhythm seems to be detected in simple syllabic sequences, and % V is the metric that best fit with our hypotheses.