JSLHR JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH

The role of native language and fundamental design of the auditory system in detecting rhythm changes

Journal:	*Journal of Speech, Language, and Hearing Research*
Manuscript ID	JSLHR-S-18-0299.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	15-Nov-2018
Complete List of Authors:	Ordin, Mikhail; Basque Center on Cognition Brain and Language, Spoken Language; Ikerbasque, Polyanskaya, Leona; Universita degli Studi di Padova, Department of Linguistic and Literary Studies Gomez, David; Universidad de O'Higgins, Institute of Educational Sciences Samuel, Arthur; Stony Brook University, Psychology; Basque Center on Cognition Brain and Language, n/a; Ikerbasque, n/a
Keywords:	Prosody, Speech perception, Central auditory processing, Cognition, Psychoacoustics

The role of native language and the fundamental design of the auditory system in detecting rhythm changes

RUNNING HEAD: L1 and cognitive mechanisms in rhythm perception

Mikhail Ordin[1,2], Leona Polyanskaya[1], David Maximiliano Gómez[3,4], Arthur G. Samuel[1,2,5]

[1]Basque Centre for Brain, Language, and Cognition, Paseo Mikeletegi 69, Donostia, 2009, Spain
[2]Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 3, Bilbao, 48013, Spain
m.ordin@bcbl.eu
[3]Institute of Educational Sciences, Universidad de O'Higgins, Avenida Libertador Bernardo O'Higgins 611, Rancagua, Chile
[4]Center for Advanced Research in Education, Universidad de Chile, Periodista Jose Carrasco Tapia 75, Santiago, Chile
[5]Department of Psychology, Stony Brook University, NY, USA

Corresponding Author:

Mikhail Ordin
Basque Centre for Brain, Language, and Cognition,
Paseo Mikeletegi 69,
San Sebastian, 20009, Spain
Tel: +34 943 309 300 F: +34 943 309 300
Email: mikhail.ordin@gmail.com, m.ordin@bcbl.eu

Conflict of Interest: The authors report no relevant conflicts of interest related to this manuscript.

Funding: The authors acknowledge support from the Spanish Ministry of Economy and Competitiveness (MINECO) Grant # PSI2017-82563-P (to AGS), from the 'Severo Ochoa' Programme for Centres/Units of Excellence in R&D (SEV-2015-490), and from the Basque Foundation for Science (IKERBASQUE). DMG was supported by Grant PIA/Basal FB0003 from the Chilean Research Council (CONICYT). LP was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) via Juan de la Cierva fellowship.

Abstract

Purpose: We investigated whether rhythm discrimination is mainly driven by the native language of the listener or by the fundamental design of the human auditory system and universal cognitive mechanisms shared by all people irrespective of rhythmic patterns in their native language.

Method:

In multiple experiments, we asked participants to listen to two continuous acoustic sequences and to determine whether their rhythms were the same or different (AX discrimination). Participants were native speakers of four languages with different rhythmic properties (Spanish, French, English, German), to understand whether the predominant rhythmic patterns of a native language affect sensitivity, bias and reaction time (RT) in detecting rhythmic changes in linguistic (Experiment 2) and in non-linguistic (Experiments 1 and 2) acoustic sequences. We examined sensitivity and bias measures, as well as RTs. We also computed Bayes factors in order to assess the effect of native language.

Results: All listeners performed better (i.e., responded faster and manifested higher sensitivity and accuracy) when detecting the presence or absence of a rhythm change when the first stimulus in an AX test pair exhibited regular rhythm (i.e., a syllable-timed rhythmic pattern) than when the first stimulus exhibited irregular rhythm (i.e., stress-timed). This result pattern was observed both on linguistic and non-linguistic stimuli and was not modulated by the native language of the participant.

Conclusion: We conclude that rhythm change detection is a fundamental function of a processing system that relies on general auditory mechanisms and is not modulated by linguistic experience.

Keywords: rhythm perception, rhythm discrimination, rhythm processing, speech rhythm, linguistic experience

1 INTRODUCTION

Rhythm perception in general and discrimination of rhythmic patterns in particular are essential skills for speech and language processing and for language acquisition in infancy (Langus, Mehler & Nespor, 2018). Rhythmic patterns differ between languages (Gervain et al., 2008; Grabe & Low, 2002; Payne et al., 2012; Ramus & Mehler, 1999; White & Mattys, 2007) and non-native rhythm is a salient aspect of L2 (second language) speech (van Maastricht et al, in press; Ordin & Polyanskaya, 2015; White & Mattys, 2007). Non-native (Polyanskaya, Ordin, & Busa, 2017; Tajima, Port, & Dalby, 1997) or pathological (Kent et al., 1989) rhythmic patterns affect speech accentedness and comprehensibility by disrupting inter-speaker entrainment via speech rhythm (Borrie & Liss, 2014; Peelle, Gross & Davis, 2013). These observations suggest that rhythmic patterns in speech might be processed via the phonological filter of the native language. Alternatively, rhythmic perception could rely on a fundamental neurophysiological mechanism (Gitza, 2011; Greenberg & Ainsworth, 2004; Hickok et al., 2015; Howard & Poeppel, 2012) that is shared by all humans irrespective of their native language. In fact, this type of basic neurophysiological mechanism could underlie rhythm discrimination by animals (Tincoff et al., 2005; Toro, Trobalon, & Sebastian-Galle, 2003) and pre-linguistic babies (Nazzi & Ramus, 2003; Ramus, Nespor & Mehler, 1999) as well.

The existing literature is in fact consistent with two plausible and reasonable hypotheses: Either (a) linguistic experience (primarily, one's native language) shapes rhythm processing, or (b) prosody in general (and rhythmic structures in particular) in natural languages is shaped by the general design of the auditory system, cognitive mechanisms, and neural physiology. The objective of this study is to pit these two hypotheses, both logically coherent and plausible according to prior empirical evidence, against one another. The importance of addressing this question lies in the fact that deficits in rhythm perception are related to dyslexia (Molinaro et al., 2016; Muneaux et al., 2004); various speech disorders are also linked to rhythm abnormalities (Liss, White, Mattys, et al., 2009). It would be beneficial to know to what degree the clinical solutions are language-independent, versus linked to a particular native language of the affected person.

1.1 SPEECH RHYTHM

The word *rhythm* implies some degree of periodicity and isochrony (as in music, where certain patterns re-occur at regular intervals). Considering these properties led James (1940), Pike (1945), Abercrombie (1967) and Ladefoged (1993) to divide languages into stress-timed (in which intervals between stressed fragments are of roughly equal duration, e.g., German, English, Dutch, Russian), syllable-timed (in which syllables are perceived as equally long, e.g., French, Italian, Spanish) and mora-timed (with equal moras, e.g., Japanese, Finnish, West Greenlandic, and Austronesia languages such as Gilbertese and Hawaiian). This distinction was initially based only on auditory impressionistic analysis, and acoustic measurements in later studies failed to fully support this claim (Dauer, 1983; Pamies Bertran, 1999; Roach, 1982).

However, despite the elusive nature of acoustic cues underlying rhythmic differences in speech, research has shown that rhythmic differences between languages are indeed perceived (e.g., Gervain et al., 2008; Ramus & Mehler, 1999). Recent psycholinguistic data show that discrimination of language-specific rhythmic patterns is based on timing cues that influence the durational variability of speech constituents: segmental length contrasts, compensatory lengthening and shortening, final lengthening or initial strengthening at the edges of prosodic units (White et al., 2012). Also, languages may differ in regard to the presence or absence of long-short vowel contrasts, vowel harmony, flexibility of stress placement, and the relative contribution of duration and pitch to linguistic prominence. In addition, the degree of vowel reduction that occurs in unstressed syllables, stress-induced lengthening of vowels, and phonotactic constraints that allow longer and shorter consonantal clusters may further enhance or inhibit variability of speech intervals. Such phonetic properties lead to cross-linguistic differences in durational variability of vocalic intervals (V), consonantal clusters (C) and syllables (S), as well as differences in the proportion of vocalic and consonantal material in speech. Languages with more properties that enhance durational variability are those that are traditionally labeled as stress-timed on the basis of their auditory impression (Dauer, 1983; Schiering, 2007). This understanding is based on the idea that rhythm is a serial arrangement of time intervals, and timing relations define the organization and temporal structure of the auditory scenes around the listener at different timescales. The differences in timing organization are captured by various rhythm metrics. In Table 1, we summarize the most widely used metrics that have been proposed

to capture the patterns of durational variability. Higher %V and lower values of the other measures signal a higher degree of syllable-timing.

The durational cues to perceived speech rhythm are not binary but continuous; the division of languages into stress-versus syllable-timed is therefore not dichotomous. Rather, languages can be positioned on a continuum between the opposing rhythm extremes. Moreover, estimating cross-linguistic rhythmic differences requires calculating the durational variability of speech intervals within individual utterances. Utterances even within the same language can also vary in the extent of stress- or syllable-timing. Therefore, we will define rhythmic characteristics in terms of the degree of syllable- versus stress-timing, referring to the *degree* of regularity (isochrony) in the duration of V, C and S intervals. The higher the degree of durational variability of speech intervals, the more stress-timed the utterance is.

1.2 GENERAL MECHANISMS OF RHYTHM PERCEPTION

Abundant empirical evidence shows that adults, pre-linguistic babies and even animals are able to differentiate between languages that are perceived to be rhythmically contrastive (e.g., Japanese or French vs. Dutch or English), while they have difficulty discriminating rhythmically similar languages (e.g., Japanese vs. French, or Dutch vs. English) (Nazzi, et al., 1998; Nazzi & Ramus, 2003; Ramus et al., 1999; Ramus & Mehler, 1999; Ramus et al., 2000; Tincoff et al., 2005; Toro et al., 2003). This widespread ability suggests that discrimination of rhythmic patterns could be governed by general properties of the auditory system and cognitive mechanisms shared by all humans, irrespective of their native language.

In fact, several recent studies have shown the existence of a neural basis for the ability to detect rhythmic changes (Hickok et al., 2015), and for the ability of acoustic rhythm to modulate the excitability of the auditory system (Ghitza et al., 2013; Greenberg & Ainsworth, 2004; Hickok et al., 2015). These abilities are presumably based on entrainment of neural oscillations to the acoustic rhythms (Giraus & Poeppel, 2012; Gitza, 2011; Ghitza et al., 2013; Howard & Poeppel, 2012). Regular rhythms couple with neural oscillations and facilitate attention better than irregular rhythms (Barnes & Jones, 2000; Howard & Poeppel, 2010; Jones, et al., 2002). Regularity also allows predicting the onset of the following vowels, and a rhythm change is detected when this prediction is not met. Rhythms characterized by a higher degree of

stress-timing make it difficult to build expectations of when the next beat should happen, and this uncertainty leads to poorer preparation and slower responding (Ellis & Jones, 2010). McAuley and Fromboluti (2015) showed that variability in the timing of tones weakens the onset timing effect, leads to distortions in perception of tone durations, and undermines attentional entrainment. They proposed that the greater variability in timing impairs the entrainment of acoustic rhythms and neural oscillations and disrupts perception of interval durations.

The existing evidence suggests that syllable-timed, regular rhythm should lead to a better coupling between speech and attentional rhythms and thus facilitate attending compared to stress-timed, irregular rhythm. If so, it might be easier to detect a change of rhythm when the listener is first tuned into an acoustic stream with regular vowel onsets, which are salient auditory events in speech-like signals. This faculty, being purely physiological in nature, should not be affected by linguistic experience and the range of speech rhythms in the native language of a listener.

1.3 LANGUAGE-BASED HYPOTHESIS OF RHYTHMIC DISCRIMINATION

However, there is also some prior evidence in favor of the hypothesis that rhythmic discrimination can be modulated by linguistic experience. Speakers of rhythmically contrastive languages employ different weightings of speech constituents (morae, syllables, feet, inter-stress intervals) and durational cues when segmenting continuous speech into discrete words and phrases (Cutler & Butterfield, 1992; Kim et al., 2008; Murty et al., 2007; Polka & Sundara, 2011; Smith et al., 1989)[1]. Erra and Gervain (2016) showed that the statistical structure of a particular language is shaped, among other things, by rhythmic patterns, which, in turn, can fine-tune the auditory codes to the statistics of the native language for efficiency of neural coding of speech. Non-native rhythmic patterns in utterances produced by second language (L2) speakers contribute to perceived accentedness and impair intelligibility of L2 speech (Polyanskaya et al., 2017; Tajima et al., 1997). Rhythm plays a very important role in acquisition of linguistic features of particular languages (Langus et al., 2018) and dialects (Clopper & Smiljanic, 2015; Polka & Sundara, 2011). These studies confirm the psychological and linguistic reality of rhythm and the fundamental role of

[1] The consistent use of distinct segmentation units in different languages is challenged by a range of studies e.g., in Content et al. (2001).

rhythm in language acquisition and in speech processing, suggesting that processing of speech rhythm can be language-based.

Native speakers of more stress-timed languages have an advantage of being familiar with a wider range of linguistic rhythms because utterances in such languages vary in how stress-timed they are. Children acquiring a stress-timed language start with syllable-timed rhythm (Bunta & Ingram, 2007; Ordin & Polyanskaya, 2014; Payne et al., 2012; Polyanskaya & Ordin, 2015). Child-directed speech also exhibits a higher degree of isochrony compared to adult-directed speech, and the stylistic differences are more extreme in more stressed-timed languages. Thus, children, in the course of L1 acquisition, are exposed to more syllable-timed utterances via child-directed speech and other children's speech. At the same time, they get exposure to more stress-timed utterances via adult-directed speech, and their own rhythm patterns also become increasingly more stress-timed as acquisition progresses (van Maastricht et al., in press; Polyanskaya & Ordin 2015; Prieto et al., 2012). Importantly, languages that are generally more stress-timed allow utterances with low durational variability of the speech constituents. In contrast, languages that are generally more syllable-timed rarely include stress-timed utterances (Ordin & Polyanskaya, 2015a). This is in part due to the strict phonotactic constraints which prohibit complex consonantal clusters and lead to utterances consisting of predominantly CV syllables with more isochronous speech intervals in syllable-timed languages. In stress-timed languages, with lax phonotactic constraints, some utterances may be more syllable-timed, consisting of predominantly CV syllables and monosyllabic words, and other utterances may contain words with complex consonantal clusters and multiple cases of reduced vowels (Prieto et al., 2012). Consequently, native speakers of stress-timed languages have experience with a wider range of possible rhythmic patterns, more and less isochronous, including those occurring in more syllable-timed languages. A language-based hypothesis would predict that native speakers of more stress-timed languages might benefit from experience with a wider spectrum of rhythms and perform better when asked to discriminate between contrastive rhythms. Indeed, Lidji, Palmer, Peretz & Morningstar (2011) showed that native speakers of English (a stress-timed language) entrain their tapping performance to the rhythm in stress-timed utterances better than native French

speakers, while native French speakers did not reveal a better entrainment of their motor output to the syllable-timed utterances compared to French.

1.4 PREDICTIONS

If perception of speech rhythm depends fundamentally on the aforementioned features of the neural and auditory system, then listeners, irrespective of their native language, should perform better detecting the switch from syllable-timed to stress-timed patterns. If instead (or in addition) linguistic experience matters, then native speakers of less syllable-timed languages (e.g., German/English) should outperform the speakers of more syllable-timed languages (e.g., Spanish/French) due to their exposure to a wider range of rhythms in their native languages, irrespective of the direction of the rhythmic change.

To choose between these alternatives, we conducted two AX discrimination experiments. In these experiments, a trial included two successive streams of syllables that could either match or mismatch. For the first experiment, we recruited native speakers of French and German (French is more syllable-timed and German is more stress-timed) and measured the effect of (a) their native language and (b) the rhythm in the paired stimuli on decision reaction time and accuracy. For the second experiment, we recruited native speakers of Spanish and English (Spanish is more syllable-timed and English is more stress-timed), and modified the difficulty level by making the rhythmic differences between the stimuli more subtle. We compared performance in rhythmic discrimination on linguistic and non-linguistic stimuli, providing a wider range of languages and stimuli. In both experiments, the first member of an AX pair was either stress-timed or syllable-timed. A significant effect of this factor (from syllable-timed or from stress-timed, to either a rhythmically similar or a rhythmically contrastive stimulus) would provide evidence in favor of the hypothesis that rhythmic discrimination is governed by the design of the auditory system. A significant effect of native language (syllable-timed vs. stress-timed) would provide evidence that rhythm discrimination is affected by linguistic experience.

2 EXPERIMENT I

2.1 METHOD

2.1.1 PARTICIPANTS

We recruited monolingual native speakers of two rhythmically different languages – French and German – French being more syllable-timed than German (Ordin & Polyanskaya, 2015a;b). We performed a power analysis to decide on the number of participants necessary to achieve significance, assuming at least a medium effect size (for $p<0.05$, partial eta square used for the effect size measure). Based on the power analysis, we recruited 25 French (Paris 3: Université Sorbonne Nouvelle, age: 18-30 y.o., median age=24 y.o., 18 females) and 25 German (Bielefeld University, age: 18-35 y.o., median age=24 y.o., 14 females) undergraduates. None of the participants reported any speech or hearing problem or proficiency in a foreign language.

2.1.2 STIMULI

We used an inventory of five consonants ([s,z,f,v,ʃ]) and five vowels ([a,u,i,e,o]). We decided to limit the consonants to fricatives because, unlike plosives, they allow stretching and compressing without losing naturalness, and unlike sonorants, they are difficult to confuse with vocalic intervals even when their durations are short. In pilot testing, synthesized stimuli with only fricative consonants received the highest ratings for naturalness.

Concatenations of the five consonants and five vowels produce 25 possible CV syllables. We created two random sequences, each with 3000 syllables, using these 25 syllables (see below for the synthesis procedure). Each syllable occurred 120 times in a sequence, with at least two different syllables between repetitions of the same syllable. One sequence was used to create stimuli with a relatively high degree of vocalic durational variability and lower %V (typical of stress-timed languages), and the other sequence was used to create stimuli with lower vocalic durational variability and higher %V (typical of syllable-timed languages). Each sequence was split into 120 25-syllable passages. In the stimuli with stress-timed rhythm, vowel durations were between 10 ms. and 150 ms. with zero skewness from a normal distribution centered around an 80 ms mean. In the stimuli with syllable-timed rhythm, vowel durations were between 80 ms. and 120 ms. with zero skewness from a normal distribution centered around a 100 ms mean. Durations of consonants were calculated by subtracting the duration of each subsequent vowel from 200 ms., thus producing streams of 25 CV syllables, 200 ms. each. Different values of median vowel durations,

with constant syllable durations, lead to variations in the %V measure, producing rhythmic differences between syllable- and stress-timed patterns. The metric scores, capturing the rhythmic distinctions we implemented in our syllable sequences, are presented in Table 2. The implemented differences were designed to be more extreme than those typical of natural languages (Grabe & Low, 2002; Ramus & Mehler, 1999; White & Mattys, 2007) to ensure that the rhythmic differences would be easily perceivable. These durations were fed into the MBROLA speech synthesizer (Dutoit et al., 1996) to synthesize 120 stimuli (25-syllable sequences) of each rhythm, using the IT4 voice.

To construct AX discrimination pairs, we presented the stimuli with a 1-second pause between the two members of a pair. There were 30 pairs in which the rhythm before and after the pause was syllable-timed (syl-syl), 30 pairs in which the rhythm before and after the pause was stress-timed (str-str), 30 pairs in which the stimuli with syllable-timed rhythm were followed by those with stress-timed rhythm (syl-str), and 30 pairs in which stimuli with stress-timed rhythm were followed by stimuli with syllable-timed rhythm (str-syl). Note that all pairs differed in their phonetic content due to the random concatenation procedure for sequences. Each stimulus was paired only once, in one of the 120 pairs.

2.1.3 PROCEDURE

The experiment was carried out in sound-treated booths. The participants were told that they would listen to utterances in an “extraterrestrial language” followed by a pause and by utterances either in the same or a different language. Their task was to identify whether the utterances before and after the pause were in the same or a different language. The answer was given by pressing the button “1” or “2” on the keyboard. The buttons “1” and “2” to respond “same” or “different” were counterbalanced between listeners. The participants were instructed not to wait until the utterance after the pause finished, and to respond immediately when they recognized whether the languages before and after the pause were different or the same. The order of pairs was randomized. To familiarize the participants with the procedure, the experiment was preceded by a training session with 12 additionally prepared pairs, with feedback as to the accuracy of the response given on each trial. Accuracy of responses and RTs (onsets of the recorded RTs locked to the onsets of the X stimuli in the AX discrimination pairs) were subjected to statistical analysis.

2.2 RESULTS

2.2.1 DEALING WITH OUTLIERS

As the listeners need at least three syllables to evaluate the durational variability, we excluded 38 responses with RTs less than 600ms. An additional 33 responses were excluded because the RT exceeded the mean RT for the participant +2SE. In total, 1.18% of all responses were excluded.

2.2.2 RESPONSE ACCURACY

As the AX discrimination task can be seen as a change detection task, we decided to adopt a signal detection theory (SDT) approach. SDT accuracy (or "sensitivity") measures are designed to be independent of any response bias (e.g., if a participant generally tends to respond "X is different from A"). The statistical procedures separate decision factors from sensory ones, which was our primary rationale for this methodological choice. In addition, sensitivity measures are comparable between experiments because they are expressed in the same units.

Sensitivity and bias measures of participants' responses were computed both for the full set of experimental items and for the subsets in which the first element of the pair was syllable- or stress-timed. We used the measures A' and B_D'' as presented by Donaldson (1992) and implemented in R by Pallier (2002). These measures represent nonparametric alternatives to the classical measures d' and β of signal detection theory, and were preferred because A' can be computed even in cases where d' is infinite or undefined. A' ranges between 1.0 and 0.0: a value of 1.0 means that detection performance is perfect, whereas a value of 0.5 means that performance is at chance. A' values that are significantly above 0.5 indicate sensitivity in detecting rhythmic changes. The bias index B_D'' ranges between -1.0 and 1.0. A value of 0.0 means that a listener shows no bias in reporting whether A and X differ or not, a value of 1.0 means that the listener always responds that X has the same rhythm as A, and a value of -1.0 means that the listener always responds that X has a different rhythm than A. Thus, biases in detection can be assessed by checking whether B_D'' differs significantly from 0.0. B_D'' is undefined when classification performance is fully correct or fully incorrect (i.e., when the hit rate equals 1.0 and the false alarm rate equals 0.0, or vice versa; this shortcoming also applies to β). Undefined values of B_D'' were obtained for three participants

(two French and one German) when analyzing the items in which A was syllable-timed, and for one German participant when analyzing the items in which A was stress-timed. The degrees of freedom of the statistical tests for B_D'' were adjusted accordingly, reflecting the reduced number of subjects in those cases.

The analysis of all items together showed that detection of rhythmic changes was good in both language groups, A' = .858 for German listeners, significantly above 0.5, $t(24) = 15.36$, $p < .0005$; and A' = .832 for French listeners, also significantly above 0.5, $t(24) = 11.22$, $p < .0005$ (see **Figure 1**; error bars stand for ±2SE in all figures). These A' values were statistically indistinguishable from one another, $t(48) = 0.68$, $p = .5$. Response bias was low for both language groups: B_D'' = 0.143 for Germans, indistinguishable from 0, $t(24) = 1.29$, $p = .21$, and B_D'' = 0.155 for French, again indistinguishable from 0, $t(24) = 1.58$, $p = .13$. The B_D'' values for German and French listeners were not different from each other, $t(48) = -0.077$, $p = .94$.

When we analyzed the responses only to the stimuli in which A was stress-timed (i.e., str-str and str-syl stimuli pairs), detection sensitivity remained high: A' = .825 for Germans, significantly above 0.5, $t(24) = 10.08$, $p < .0005$, and A' = .776 for French, significantly above 0.5, $t(24) = 7.21$, $p = .0005$. A' values for German and French listeners did not differ from each other, $t(48) = 0.99$, $p = .33$. Response bias was low for both groups: B_D''=0.138 for Germans, indistinguishable from 0, $t(23) = 1.03$, $p = .32$, and B_D'' = 0.044 for French, indistinguishable from 0 as well, $t(24) = 0.39$, $p = .7$. B_D'' values for German and French listeners did not differ from each other, $t(47) = 0.54$, $p = .59$.

The analysis of responses to the stimuli in which A was syllable-timed (i.e., syl-syl and syl-str stimuli pairs) yielded a similar pattern of results. Detection of rhythmic changes was good, A' = .881 for Germans, significantly above 0.5, $t(24) = 17.90$, $p < .0005$, and A' = .877 for French listeners, significantly above 0.5, $t(24) = 14.54$, $p = .0005$. A' values for French and German participants did not differ from each other, $t(48) = 0.13$, $p = .90$. Response bias was low for Germans, B_D'' = 0.056, indistinguishable from 0, $t(23) = 0.42$, $p = .68$. It was somewhat higher for French listeners, B_D'' = 0.222, but this value was not significantly different from 0 either, $t(22) = 1.84$, $p = .088$. The difference between B_D'' values for the two language groups was not significant, $t(45) = -0.93$, $p = .36$.

Contrasting performance in items in which A was syllable-timed vs. items in which A was stress-timed, we confirmed that both French and German listeners showed higher sensitivity in detecting rhythmic changes when A was syllable-timed than when A was stress-timed, $t(24) = 2.30$, $p = .03$ for Germans and $t(24) = 4.27$, $p = .0003$ for French participants.

These sensitivity results provide preliminary answers to the central questions of the current study. First, language background did not affect performance, indicating that rhythm processing here was dominated by language-independent factors. Second, listeners were more accurate when the vowel onsets in the first stimulus in a pair were relatively regular, with little variability in vowel durations.

2.2.3 REACTION TIME

The error rates were 22.0% and 20.2% for the French and the German listeners, respectively. An ANOVA on the reaction times, with *L1* of the listener and *Accuracy* (correct vs. incorrect) as factors, revealed a significant effect of *Accuracy,* (1, 48) = 36.896, $p < .0005$, $\eta_p^2 = .435$. There was no effect of *L1*, $F(1,48) = .437$, $p = .512$ and no significant interaction, $F(1,48) = 1.004$, $p = .321$. Both German and French listeners responded more slowly when they gave an incorrect answer (see **Figure 2**). Therefore, the following analyses of reaction times were performed only using correct responses, to avoid contaminating the results with longer RTs on incorrect responses.

To explore the impact of linguistic experience on the RT for discriminating between rhythm types, we performed a 2-way ANOVA with *L1* and *Pair-type* as factors. The four pair-types consisted of the 2x2 crossing of stress-timed and syllable-timed sequences with the two positions in an AX trial. **Figure 3** shows RTs split by L1 and Pair-type. The effect of native language was not significant, $F(1,48) = .112$, $p = .739$; German and French listeners did not differ in their time to decide whether the rhythm of the two speech stimuli was the same or different. The effect of *Pair-type* was significant, $F(3,46) = 5.033$, $p = .004$, $\eta_p^2 = .247$: Pairs with a syllable-timed stimulus in the first position were responded to more quickly than those with a stress-timed stimulus in that position. There was no interaction between *L1* and *Pair-type*, $F(3,46) = .092$, $p = .964$, $\eta_p^2 = .006$, reflecting a similar advantage for the initial syllable-timed item regardless of the listeners' linguistic backgrounds. Pairwise comparisons (Bonferroni corrected) confirmed that an initial

syllable-timed item produced faster responses both when the second item in a pair was syllable-timed, p = .014; and when the second item was stress-timed, p = .022. Listeners responded significantly more quickly when the first stimulus in the pair had a regular temporal structure. The facilitatory effect of a regular temporal structure was not modulated by native language of the listener. Thus, the conclusions from the reaction time data are exactly those that the sensitivity analyses supported.

To make sure that the null effect of linguistic background does not stem from a lack of statistical power, we computed Bayes factors using the BayesFactor package in R (Morey & Rouder, 2012). Bayes Factor (BF) is the ratio between the likelihood of a set of observed data given two statistical models. If these two models are assumed to be equally probable a priori, then the BF corresponds to the ratio of probabilities of these models, given the observed data. This means that BFs may be useful not only to quantify the degree of support that some data provide for the alternative hypothesis, but also to quantify whether they support the null hypothesis[2] (Morey & Rouder, 2011).

We conducted Bayesian analyses on accuracy rates and response times considering each of the four item types (syl-syl, syl-str, str-syl, str-str) separately. We analyzed accuracies and response times (for correct answers only) using 4×2 mixed models, with item type as a within-subject factor, native language as a between-subject factor, and subjects as a random factor. We computed BFs contrasting each full regression model (both factors plus interaction) against the corresponding model with neither the interaction nor the main effect of native language. The BFs obtained were 0.0287 for accuracy and 0.0248 for response times. These numbers indicate that outcomes of both measures strongly support the null model, a model lacking any effect of native language. That is to say, the null models are 34.9 times and 40.3 times (for accuracy and response time, respectively) more likely than the full regression models, given the data. Moreover, if the full models and the models with no effect of native language are assumed to be equally probable (50%) a priori, these BFs indicate that the probability of the model lacking native language grows to 97.2% a posteriori for accuracy, and to 97.6% a posteriori for response times.

2.3 DISCUSSION

[2]For a simple statistical test (e.g., a t-test), a BF of 1.0 means that the data provide evidence in favor of neither the alternative nor the null hypothesis, whereas a BF smaller (resp. greater) than 1.0 means that the data support the null (resp. alternative) hypothesis.

The results showed that people were quite good in general at discriminating rhythms in all conditions, but detecting the rhythm change in one direction (from more regular to irregular) was faster and more reliable than in the opposite direction. When the A stimulus was syllable-timed, participants' sensitivity was higher, and RT was lower – for both French and German listeners – compared to the conditions when the A stimulus was stress-timed. This asymmetry, similar to one reported by McAuley and Fromboluti (2015) for non-linguistic tone sequences, is shared by listeners with rhythmically different native languages, supporting the view that it is grounded in a general auditory mechanism. Although Germans are familiar with a wider spectrum of rhythmic patterns than French listeners, there was no influence of the native language on sensitivity or RT. Therefore, we tentatively conclude that the ability to discriminate utterances with different linguistic rhythms is not biased by linguistic experience.

Experiment 1 only tested the efficiency of rhythm discrimination by native speakers of one pair of rhythmically different languages: German and French. If we wish to draw broad conclusions, it is necessary to show that the observed pattern is not limited to a particular pair of languages. To this end, a second AX discrimination experiment was conducted using a different pair of languages that differ in terms of being more syllable-timed (Spanish) or more stress-timed (English). Another objective was to include two different sets of stimuli: one that is perceived as being made up of utterances in a natural language, and the other, which is perceived as being less like natural language. This manipulation is designed to investigate whether the performance in the rhythm discrimination task is modulated by whether the stimuli are linguistic or non-linguistic. Finally, we made the stimuli more ecological by modelling the durational values based on the values of a real language and by allowing a wider variety of syllable types, which is typical of real languages.

3. EXPERIMENT II

3.1 METHOD

3.1.1 PARTICIPANTS

We recruited native speakers of two rhythmically different languages – Spanish and English – Spanish being more syllable-timed than English (Payne et al., 2012; Prieto et al., 2012; Ramus et al., 1999). We recruited

29 participants per group, but we had to discard two Spanish participants for not performing the task as instructed. None of the participants reported any speech or hearing problem. Based on self-report, Spanish participants were fluent in Basque (another syllable-timed language), and English participants were fluent in Spanish, with proficiency varying from intermediate to high, which increases their linguistic experience with both syllable- and stress-timed rhythmic patterns.

3.1.2 STIMULI

For this experiment, we created two different types of stimuli, which we will refer to as *linguistic* stimuli and *non-linguistic* stimuli. Linguistic stimuli were intended to encourage listeners to engage the processing mechanisms that are brought to bear while listening to real language. Linguistic stimuli should be perceived as plausible utterances of a natural language, manifesting natural prosody and properties of a real human language. Non-linguistic stimuli are intended to be speech-like sequences of syllables, which, nevertheless, are processed at more of a psychoacoustic level, without engaging additional mechanisms involved in listening to real language. For this, we used random syllabic sequences with a monotonous pitch. These were devoid of prosody, recognizable hierarchical linguistic structures, or segmentable discrete constituents, thus making the stimuli sound less like real language.

To prepare non-linguistic stimuli, we used an inventory of five consonants [s,m,n,l,f] and five vowels [a,u,i,e,o]) to create 25 possible CV syllables. We created 240 sequences, in which the 25 possible syllables were randomized, with each syllable occurring only once per sequence. In each sequence, five random syllables were marked as stressed, with the only restriction being that two consecutive syllables could not be stressed.

Half of the sequences were used to prepare stimuli with high durational variability (more stress-timed), and the other half to prepare stimuli with low durational variability (more syllable-timed). In stimuli with higher durational variability, the vowel durations for the stressed positions varied between 80ms and 240 ms with 40-ms steps (one value per syllable per sequence), and vowel durations in unstressed syllables varied between 40ms and 80 ms with 10-ms steps (four values per syllable per sequence). In stimuli with lower durational variability, the vowel durations for the stressed positions varied between 160ms and 200

ms with 10-ms steps (one value per syllable per sequence), and vowel durations in unstressed syllables varied between 80ms and 100 ms with 5-ms steps (four values per syllable per sequence). The total syllable duration was set to 280 ms for stressed syllables and 180 ms for unstressed syllables. Durations of consonants were calculated by subtracting the duration of each subsequent vowel from 280 ms (stressed syllables) or 180 ms (for unstressed syllables). In this way, we ensured that the stimuli differed in rhythmic characteristics (durational variability), but not in speech rate (number of syllables or phonemes per second). The sequences were synthesized with these phoneme durational parameters in MBROLA, with the ES2 voice, F0 set to 200Hz.

To prepare linguistic stimuli, we took the recordings of a native Welsh speaker reading 38 sentences (this was a Welsh-English bilingual recorded for a different experiment reported in Ordin & Mennen, 2017). The sentences were annotated in Praat, and durations of vowels and consonants were measured. We used a Spanish set of phonemes to re-synthesize the Welsh sentences with the durations of vowels and consonants of a Welsh speaker. In case a Welsh phoneme did not exist in Spanish, it was substituted with the closest Spanish phoneme. Then we imposed an intonational contour of a Welsh sentence on resynthesized versions of the sentences. The values of the rhythm metrics are similar to those of Spanish and contrastive to those of English, which leads to Welsh being positioned closer to the syllable-timed end of the rhythm spectrum (Grabe & Low., 2002). This was confirmed by the scores of vowel duration variability calculated on the individual's recording used in our study.

We created a second set of sentences in which we multiplied the durations of stressed vowels by 2 and unstressed vowels by 0.6 in order to enhance the durational contrasts between stressed and unstressed vowels. The original re-synthesized sentences represent more syllable-timed rhythms, and the re-synthesized sentences with enhanced durational ratios represent more stress-timed rhythms. The sentences were concatenated into 2-sentence pairs (each pair making one stimulus), 5.5 seconds in duration (±500ms), thus matching the duration of non-linguistic and linguistic stimuli. Each sentence was used 5-7 times for the stimuli of each rhythm type. We created 120 linguistic stimuli of each rhythm type. A preliminary pilot experiment was performed to make sure that the linguistic stimuli in our experiment

received higher naturalness ratings and were more likely to be perceived as real language compared to the non-linguistic stimuli (see Appendix I for the results of the pilot study).

We constructed 120 pairs of stimuli for each stimulus type (linguistic and non-linguistic) with a 1.5-second pause between the two members of a pair. For each stimulus type, there were 30 pairs in which the rhythm before and after the pause was syllable-timed (syl-syl), 30 pairs in which the rhythm before and after the pause was stress-timed (str-str), 30 pairs in which the stimuli with syllable-timed rhythm were followed by those with stress-timed rhythm (syl-str), and 30 pairs in which stimuli with stress-timed rhythm were followed by stimuli with syllable-timed rhythm (str-syl). Each stimulus was paired only once, in one of the 120 pairs.

In the first experiment the differences in durational variability between rhythms were pushed to the extremes, whereas in the second experiment, the scores of the rhythm metrics were based on the values typical of natural languages. This made the differences in durational variability more natural and less salient, thus making the discrimination task more challenging.

3.1.3 PROCEDURE

Experiment 2 consisted of two sessions held with an interval of 10 days (±3 days) between the sessions. In one of the sessions, participants listened to linguistic stimuli, and in the other session the stimuli were non-linguistic. The order of sessions was counterbalanced across participants. The experiment was carried out in sound-attenuated booths. The participants were told that they would listen to pairs of utterances in an "extraterrestrial language". Their task was to identify whether the rhythm in the utterances before and after the pause was the same or different. The answer was given by pressing the button "1" or "2" on the keyboard. The buttons "1" and "2" to respond "same" or "different" were counterbalanced between listeners. The order of pairs was randomized. To familiarize the participants with the procedure, the experiment was preceded by a training session with 12 additionally prepared pairs, with feedback as to the accuracy of the response given on each trial during training.

Unlike the first experiment, when we asked participants not to wait until the end of the second utterance and to make the response as soon as they thought they knew the answer, in this experiment the

response could only be given after the participant had finished listening to the second utterance in the trial. This was deemed necessary because the task was subjectively much more challenging, and we wanted to prevent a possible speed-accuracy trade-off by giving the participants the same time on each trial to make their decision. Note that this approach prevented us from using RT as a dependent measure this time.

At the end of the second session, participants performed an additional test. Ten randomly chosen stimuli of each type (linguistic and non-linguistic stimuli), half with high variability durations and half with low, were played to the participants in a random order. Upon listening to each stimulus, the participant was asked to mark on an 8-point scale how likely it seemed that the stimulus represents an utterance of a real language. This was done to ensure that individual participants indeed perceived the Welsh-based stimuli as more linguistic than random syllabic concatenations. The analysis of this additional test confirmed that our linguistic stimuli were indeed perceived as more like a real language; the rhythmic type had no effect on the perceived naturalness of the stimuli (see Appendix II for the test results).

3.2 RESULTS

We kept all the responses and participants; no outliers were removed. As in Experiment 1, we used the non-parametric measures A' and B_D'' to index sensitivity and bias.

3.2.1 LINGUISTIC STIMULI RESULTS

We computed the sensitivity and bias measures of participants' responses for the full set of experimental items and for the subsets in which the first element of the pair was syllable- or stress-timed. **Figure 4a** shows the sensitivity results. The analysis on the full set showed that detection of rhythmic changes was significantly above chance for both language groups, A' = .642 for English listeners, $t(28) = 10.62$, $p < .0001$; and A' = .590 for Spanish listeners, $t(26) = 4.79$, $p < .0001$. A' values were significantly larger for English listeners, $t(54) = 2.30$, $p = .03$. **Figure 4b** shows that the response bias was low for both languages: B_D'' = 0.017 for English, indistinguishable from 0, $t(28) = 0.28$, $p = .78$ and B_D'' = 0.052 for Spanish, again indistinguishable from 0, $t(26) = 0.82$, $p = .42$. Moreover, B_D'' values for English and Spanish listeners were not different from each other, $t(54) = -0.41$, $p = .68$.

The analysis of responses only to the stimuli in which A was syllable-timed revealed good detection of rhythmic changes, A' = .708 for English, significantly above 0.5, $t(28) = 10.93$, $p < .0001$, and A' = .652 for Spanish listeners, significantly above 0.5 as well, $t(26) = 6.37$, $p < .0001$. A' values for English and Spanish participants marginally differed from each other, $t(54) = 1.86$, $p = .07$. Response bias was low for English, B_D'' = 0.082, indistinguishable from 0, $t(28) = 1.06$, $p = .30$, as well as for Spanish listeners, B_D'' = 0.160. This value was marginally different from 0, $t(26) = 1.97$, $p = .06$. The difference between B_D'' values for the two languages was not significant, $t(54) = -0.70$, $p = .49$.

When we analyzed the responses only to the stimuli in which A was stress-timed, detection sensitivity dropped: A'=.555 for English, significantly above 0.5, $t(28) = 2.55$, $p = .02$, and A' = .516 for Spanish, statistically indistinguishable from 0.5, $t(26) = 0.73$, $p = .47$. A' values for English and Spanish listeners did not differ from each other, $t(54) = 1.28$, $p = .21$. Response bias was low for both groups: B_D'' = -0.050 for English, indistinguishable from 0, $t(28) = -0.82$, $p = .42$, and B_D'' = -0.061 for Spanish, indistinguishable from 0 as well, $t(26) = -0.95$, $p = .35$. B_D'' values for English and Spanish listeners did not differ from each other, $t(54) = 0.12$, $p = .91$.

Contrasting performance on items in which A was syllable-timed vs. items in which A was stress-timed, we confirmed that both English and Spanish listeners showed higher sensitivity in detecting rhythmic changes when A was syllable-timed than when A was stress-timed, $t(28) = 5.04$, $p < .0001$ for English and $t(26) = 4.99$, $p < .0001$ for Spanish participants.

Altogether, the analysis of sensitivity shows that both English and Spanish listeners were good at discriminating the two rhythms only when the first stimulus in the test pairs was syllable-timed (i.e., rhythmically regular). Performance was substantially worse (at the chance level for Spanish participants and slightly above chance for the English participants, with no significant difference between the groups) when the first stimulus was stress-timed. This shows that listeners were more accurate when the first stimulus in the pair is characterized by a regular distribution of vowel onsets, with relatively little variability in vowel durations.

3.2.2 NON-LINGUISTIC STIMULI RESULTS

The analysis of all items together showed that detection of rhythmic changes was above chance in both language groups, A' = .572 for English listeners, significantly above 0.5, $t(28) = 4.47$, $p = .0001$; and A' = .563 for Spanish listeners, also significantly above 0.5, $t(26) = 3.24$, $p = .003$ (see **Figure 4a**). The two A' values were statistically indistinguishable from one another, $t(54) = 0.36$, $p = .72$. Response bias was low for both languages: B_D'' = 0.048 for English, indistinguishable from 0, $t(28)= 0.78$, $p = .44$ and B_D'' = 0.032 for Spanish, again indistinguishable from 0, $t(26) = 0.51$, $p = .62$. Moreover, B_D'' values for English and Spanish listeners were not different from each other, $t(54) = 0.18$, $p = .86$.

The analysis of responses only to the stimuli in which A was syllable-timed yielded a similar pattern of results. Detection of rhythmic changes was good, A' = .631 for English, significantly above 0.5, $t(28) = 6.71$, $p < .0001$, and A' = .652 for Spanish listeners, significantly above 0.5, $t(26) = 5.68$, $p < .0001$. A' values for English and Spanish participants did not differ from each other, $t(54)= -0.66$, $p = .51$. Response bias was low for English, B_D'' = 0.081, indistinguishable from 0, $t(28) = 1.13$, $p = .27$, as well as for Spanish listeners, B_D'' = 0.065, $t(26) = 0.73$, $p = .47$. The difference between B_D'' values for the two languages was not significant, $t(54) = 0.14$, $p = .89$.

When we analyzed the responses only to the stimuli in which A was stress-timed, detection sensitivity vanished: A' = .505 for English, not significantly above 0.5, $t(28) = 0.18$, $p = .86$, and A' = .463 for Spanish, not significantly above 0.5, $t(26) = -1.37$, $p = .18$. A' values for English and Spanish listeners did not differ from each other, $t(54) = 1.09$, $p = .28$. Response bias was low for both groups: B_D'' = 0.022 for English, indistinguishable from 0, $t(28) = 0.29$, $p = .78$, and B_D'' = 0.022 for Spanish, indistinguishable from 0 as well, $t(26)= 0.27$, $p = .79$). B_D'' values for English and Spanish listeners did not differ from each other, $t(54) = 0.0004$, $p = .99$.

Contrasting performance on items in which A was syllable-timed vs. items in which A was stress-timed, we confirmed that both English and Spanish listeners showed higher sensitivity in detecting rhythmic changes when A was syllable-timed than when A was stress-timed, $t(28) = 3.75$, $p = .0008$ for English and $t(26) = 4.82$, $p < .0001$ for Spanish participants.

Altogether, the analysis revealed the same result pattern that we observed with linguistic stimuli: both Spanish and English listeners performed better on the task when the first stimulus in a test pair was characterized by regular rhythm.

3.2.3 COMPARISON OF SENSITIVITY ON LINGUISTIC AND NON-LINGUISTIC STIMULI

We performed repeated-measures ANOVAs on A' scores with the *rhythm in the first stimulus* in the AX pair (stress- vs. syllable-timed) and the *type of the stimuli* (linguistic vs. non-linguistic) as within-subject factors, and with *native language of the listener* (Spanish vs. English) as a between-subject factor. The analysis showed that sensitivity was significantly higher both for linguistic and non-linguistic stimuli when the first stimulus was syllable-timed, $F(1,54) = 63.495$, $p < .0005$, $\eta_p^2 = .54$ (η_p^2 is partial eta squared). Overall, participants performed slightly better on linguistic stimuli than on non-linguistic stimuli, $F(1,54) = 7.361$, $p = .009$, $\eta_p^2 = .12$. The effect of the native language, on the other hand, was not significant, $F(1,54) = 3.05$, $p=.086$, $\eta_p^2 = .05$. Importantly, native language modulated neither the effect of the stimulus type (the language*stimulus type interaction was not significant, $p = .257$), nor the effect of the first stimulus in AX pairs (language*rhythm in the first stimulus of a pair was not significant, $p = .551$). The three-way interaction of *language*preceding stimulus rhythm*stimulus type* was not significant either, $p = .142$.

This analysis shows that the results are stable for both stimulus types and language groups. Overall, the effects of the stimulus type (linguistic vs. non-linguistic) and of the rhythmic characteristics of the first stimulus in the test pairs significantly affected sensitivity, whereas the effect of the native language produced a null result. Overall, both language groups yielded similar results, but the native English listeners showed some sensitivity on linguistic stimuli, provided that the first stimulus in the discrimination pair was not stress-timed. To estimate the support for the null hypothesis regarding the effect of the native language on performance in Experiment 2, we analyzed the accuracy rates for this experiment using Bayes factors in a similar manner to the previous experiment. The BFs obtained were 0.0444 for linguistic stimuli and 0.0157 for non-linguistic stimuli, yielding strong evidence for the null models. These are 22.5 times and 63.7 times more likely than the full regression models, given the data. If the two models are assumed to be

equally probable a priori, then the probability of the model lacking native language grows to 95.7% *a posteriori* for linguistic stimuli, and to 98.5% *a posteriori* for non-linguistic stimuli.

3.3 DISCUSSION

The result pattern observed in the second experiment confirmed and strengthened the preliminary conclusions offered in the first experiment. The ability to discriminate rhythmic patterns in speech is not affected by linguistic experience stemming from the native language of the speaker. Rhythm discrimination is facilitated if the first stimulus manifests regular rhythm. If the first stimulus has an irregular rhythm, the task becomes more challenging and performance drops, even if the native language of the participants includes utterances with irregular, stress-timed rhythm. The same pattern was observed for non-linguistic and linguistic stimuli. The former may not engage the full set of speech processing mechanisms, while the latter presumably do because they are perceived as the utterances of a real language. These results suggest that rhythm discrimination in speech is controlled to a large extent by cognitive mechanisms shared by all humans, irrespective of their native language, based on the properties and the general design of the auditory system.

4. GENERAL DISCUSSION

Although people indeed find it easier to detect slight rhythmic differences in linguistic than in non-linguistic stimuli, our results showed that discrimination of rhythms even in linguistic stimuli is not affected by linguistic experience. Instead, performance is dependent on the rhythmic characteristics of the first stimulus in an AX pair. Our results revealed that utterances characterized by higher regularity of vowel onsets and lower variability in vowel durations enhanced attention to the rhythmic details of the acoustic signal. As a result, performance on the AX rhythm discrimination task was better when the X stimulus was preceded by a syllable-timed A stimulus.

This pattern shows that the native language is not a factor in rhythm discrimination performance, which suggests that the stimuli are not filtered via the phonology of the native language, even when these stimuli are perceived as speech in a natural language. Given that rhythm discrimination is so important for speech processing and language acquisition (Hickok et al., 2015; Mehler et al., 1996; Nazzi & Ramus, 2003),

that strategies of speech segmentation and acquisition are so different in rhythmically different languages (e.g., Cutler & Butterfield, 1992; Kim et al., 2008; Langus, Mehler & Nespor, 2018), and that deviations from the native norms in rhythmic structure affect speech intelligibility and increase accentedness of foreign speech (Polyanskaya et al., 2017; Tajima et al., 1997), it is surprising that linguistic experience plays a very minor role in performance on the rhythm discrimination task. As performance is similar on linguistic and non-linguistic stimuli, it appears that rhythm discrimination is primarily based on a domain-general mechanism defined by the general design of the auditory system. Presumably, rhythm discrimination happens at early stages of auditory processing, irrespective of whether the stimuli are linguistic or not, before language processing skills are engaged. The auditory system is designed so that acoustic rhythms modulate the firing pattern of the auditory nerves (Ghitza et al., 2013; Hickok et al., 2015), which leads to coupling of neural cortical oscillations and the acoustic rhythms (Howard & Poeppel, 2012), resulting in a facilitatory effect of a regular rhythm in the first stimulus. Rhythms characterized by regularity couple with internal oscillators and facilitate attention better than irregular rhythms (Jones, Moynihan, MacKenzie, & Puente, 2002).

This interpretation is in line with the Dynamical Attending Theory (Barnes & Jones, 2000; Howard & Poeppel, 2010; Jones, et al., 2002), which predicts that the presentation of a regular rhythm entrains attentional oscillations, which in turn creates stronger expectations for upcoming events and thus leads to better discrimination performance. Regular rhythm, or the regular occurrence of vowel onsets, supports anticipatory mechanisms for temporal prediction of when the following onset is expected to happen; deviations from these expectations are therefore detected more reliably and faster (Barnes & Jones, 2000; Lakatos et al., 2008; ten Oever, Schroeder, Poeppel, et al., 2014). Rhythm has a direct influence on the perception of discrete acoustic events, including salient vowel onsets (Hickok et al., 2015). Selective active attending to a stimulus generates a dynamically evolving neural model of the acoustic stream to which the listener is attending, in the form of neuronal oscillations (Lakatos et al., 2013). Listeners are constantly updating a reference pattern as the stimulus develops in time, based on expectations built on rhythmic regularity (Lakatos et al., 2013), and building the reference is more difficult when the reference stimulus is

more stress-timed, i.e., exhibits irregular vowel onsets. The detection of rhythmic changes is then more difficult because the reference is still being updated when the rhythmic change happens. The listener, having no clear expectations, can only react to the change that has already happened. Regular rhythm allows future-oriented attending, while irregular rhythm requires direct comparison of the rhythmic patterns, using past-oriented attention, in order to compare the current rhythm with the reference rhythm (ten Oever, Schroeder, Poeppel, et al., 2014).

The stimuli for the first experiment were based on extreme values of variability parameters to model rhythmic differences, which reduce the challenge of the discrimination task but also potentially reduce the ecological validity of the stimuli. The stimuli in the second experiment were based on the less extreme values observed in real languages. This made the rhythmic differences subtler, and substantially increased the difficulty of the task; hence the overall performance was lower in the second experiment.

Note that the non-linguistic stimuli were composed only of CV syllables. Rhythmically contrastive languages also exhibit differences in syllabic complexity: Languages featuring a higher degree of stress-timing allow more complex syllables due to loose phonotactic constraints (Prieto et al., 2012). In regard to phonotactic complexity, all non-linguistic sequences were more syllable-timed than stress-timed. In principle, this asymmetry could explain the observed facilitation effect for syllable- over stress-timed stimuli. The linguistic stimuli, in contrast, have the properties characteristic of natural languages (e.g., prosody, statistical cues, and extractable constituents), thus engaging the additional mechanisms that are involved in listening to real speech.

Although linguistic stimuli were indeed perceived as though they could have been utterances of natural languages, the pattern of results for linguistic stimuli was not different from that for non-linguistic ones. This equivalence illustrates the crucial role played by the general design of the auditory system and low-level physiological mechanisms that are not affected by individual experience. The mammalian auditory system has not evolved specifically for speech processing, and linguistic experience is unlikely to shape the general design of the auditory system. On the contrary, it is more likely that the structure of speech is shaped by the design of the auditory system, which ensures that the speech stream is

processable by a general auditory mechanism that allows entrainment of the neural oscillations to the environmental rhythms. However, linguistic experience can modulate the *output (neuronal firing rate)* of early auditory mechanisms, when the *output* is passed forward for processing at a higher level, e.g., via the phonological filter of the listener's native language.

In the current study, we observed the outcome of such general perceptual mechanisms in adults processing speech-like stimuli without access to higher-level linguistic information. The auditory mechanism underlying rhythm discrimination, and probably rhythmic cognition in general, is not specific to language processing. It is shared by humans irrespective of their native language. Together with other rhythm-based general perceptual mechanisms (e.g., iambic-trochaic grouping regularities), it manifests itself in the behavior of pre-linguistic babies (Mehler et al., 1996) and non-human species (de la Mora et al., 2013; Tincoff, et al., 2005; Toro & Nespor, 2015). Universal rhythm processing mechanisms have even been mentioned as the precursors to speech emergence and development in phylogenesis (Ghazanfar & Takahashi, 2014 a; b; MacNeilage, 1998; Merker, Madison, & Eckerdal, 2009). The existence of general, rather than language-specific mechanisms for extracting and classifying rhythmic patterns may be a pre-requisite of language acquisition in ontogenesis (Mehler & Nespor, 2004). Psychoacoustic, low-level processing of rhythmic patterns provides pre-linguistic infants recourse to language-independent cues for differentiating and classifying utterances and languages (Nazzi, et al., 1998), for extracting linguistic structure from the ambient language (Mehler, et al., 1996), and for segmentation of the continuous acoustic stream into discrete constituents (Nazzi & Ramus, 2003).

The distinction we are making between general auditory mechanisms and language-specific processing is consistent with the assumptions of Werker and Curtin's (2005) PRIMIR model (Processing Rich Information from Multidimensional Interactive Representations). The model posits that the information extracted from speech is processed through filters based on (1) biological biases, (2) the competence level of the listener in the language that is being processed, and (3) the requirements of the specific task. These filters attract the listener's attention to one of three representational dimensions: general perceptual dimensions, a wordform dimension, or a phonemic dimension (extendable to further

linguistic dimensions). Mapping of the signal simultaneously to both a general perceptual dimension and to the wordform dimension is possible via the mechanism of statistical learning. To engage this mechanism, statistical regularities in the input are necessary. The material in Experiment 1 and the non-linguistic stimuli of Experiment 2 were designed to be devoid of these statistical regularities. As a result, the conditions needed to activate the mechanism of statistical learning were not present. Thus, the results only reflect processing at the general perceptual plane through general auditory mechanisms that are based on biological biases, present in all humans irrespective of their native language.

If rhythm change detection is indeed controlled by general properties of the human auditory system, and if the ability to discriminate rhythmic patterns of different utterances is a prerequisite for successful language acquisition, then performance in such a task can potentially serve as one of the indices of healthy phonological development. This suggestion is consistent with recent neurophysiological evidence that the disruption of synchronization between acoustic and neural oscillations can lead to deficits in both phonological and reading skills (Molinaro et al., 2016). Dyslexic readers exhibit impaired neural entrainment to speech, with impaired coupling between neural oscillations in the auditory cortex and the left inferior frontal region. This pattern is often accompanied by delays in phonological development (Lallier et al., 2017). Goswami (2011) has proposed an integrated theoretical framework postulating that phonological auditory deficits, stemming from atypical neural oscillations, lead to developmental dyslexia. These theoretical and empirical developments suggest that a behavioral test of rhythm detection on speech-like stimuli, which could be designed as a game for young children, could provide a non-invasive diagnostic for phonological deficits. The procedures in the current study clearly could be modified to develop such a test, though of course considerable further research would be needed to establish clinically-relevant norms.

ACKNOWLEDGMENT:

The authors acknowledge support from the Spanish Ministry of Economy and Competitiveness (MINECO) Grant # PSI2017-82563-P (to AGS), from the 'Severo Ochoa' Programme for Centres/Units of Excellence in R&D (SEV-2015-490), and from the Basque Foundation for Science (IKERBASQUE). DMG was supported by Grant PIA/Basal FB0003 from the Chilean Research Council (CONICYT). LP was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) via Juan de la Cierva fellowship. We thank two anonymous reviewers for their constructive suggestions on an earlier version of this paper.

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, pp. 97-99.

Aslin, R. N., Saffran, J. R. and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science 9*,321-324.

Barnes, R., & Jones, M. (2000). Expectancy attention and time. *Cognitive Psychology 41*, 254-311.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking four-and five-year-old children. *Journal of Speech, Language, and Hearing Research 50*, 999-1014.

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics 49*, 1-15.

Content, A., Meunier, C., Kearns, R.K., & Frauenfelder, U. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes 16*, 609-636.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language 31*, 218-236.

Dauer, R. (1983) Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics 11*, 51-62.

De la Mora, D., Nespor, M. & Toro, J.M. (2013). Do Humans and non-human animals share the grouping principles of the Iambic - Trochaic Law? *Attention, Perception, & Psychophysics 75(1)*, 92-100.

Dellwo,V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.). *Language and language-processing* (pp.231-241). Frankfurt am Main: Peter Lang.

Donaldson, W. (1992). Measuring recognition memory. Journal of Experimental Psychology: General, 121(3), 275-277.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. Philadelphia: ICSLP.

Ellis, R.J., & Jones, M. (*2010*). Rhythmic context modulates foreperiod effects. *Attention*, *Perception, & Psychophysics 72*, 2274–2288.

Erra, R., Gervain, J. (2016). The Efficient Coding of Speech: Cross-Linguistic Differences. *PLoS ONE 11(2)*, https://doi.org/10.1371/journal.pone.0148861.

Gervain, J. Nespor, M., Mazuka, R., Horie, R., & Mehler J. (2008) Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cognitive Psychology 57(1)*, 56-74.

Ghazanfar, A., & Takahashi, D. (2014a). Facial expressions and the evolution of the speech rhythm. Journal of Cognitive Neuroscience 26(6), 1196-1207.

Ghazanfar, A., & Takahashi, D. (2014b). The evolution of speech: vision, rhythm, cooperation. Trends in Cognitive Sciences 18(10), 543-553.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. https://doi.org/10.3389/fpsyg.2011.00130

Ghitza, O., Giraud, A.-L., & Poeppel, D. (2013) Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Front. Hum. Neurosci*. **6**:340. https://doi.org/10.3389/fnhum.2012.00340

Goswami, U. (2011). A temporal sampling framework for dyslexia. *Trends in cognitive sciences*, *15*(1), 3-10.

Grabe, E. & Low, L. (2002). Durational variability in speech and the rhythm class hypothesis. Laboratory Phonology 7, 515-546.

Greenberg, S. & Ainsworth, W. (2004). Speech processing in the auditory system: An Overview. In S. Greenberg, W. Ainsworth, A. Popper, & R.Fay (Eds.). Speech Processing in the Auditory System (pp. 1-62). New York: Springer Verlag.

Hay, J., & Diehl, R. (2007). Perception of rhythmic grouping: Testing the Iambic/Trochaic law. *Perception & Psychophysics 69*, 113–122.

Hickok, G., Farahbod, H., & Saberi, K. (2015). The Rhythm of Perception: Acoustic Rhythmic Entrainment Induces Subsequent Perceptual Oscillation. *Psychological Science 26(7)*, 1006–1013.

Howard, M., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustic but not comprehension. *Journal of Neurophysiology 104*, 2500-2511.

Howard, M., & Poeppel, D. (2012). The neuromagnetic response to spoken sentences: Co-modulation of theta band amplitude and phase. *NeuroImage 60*, 2118-2127.

James, L. (1940). *Speech Signals in Telephony*. London: Sir Isaac Pitman & Sons.

Jones, M., Moynihan, H. M., MacKenzie, N., & Puente, J. K. (*2002*). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science 13*, 313–319.

Kent, R.D., Weismer, G., Kent, J.F., & Rosenbek, J.C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders 54*, 482–499.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech, 51(4)*, 343-359.

Ladefoged, P. (1993). *A course in phonetics.* 3rd edition. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. Science, 320, 110–113.

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. Neuron 77, 750–761.

Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., & Carreiras, M. (2017) Amodal atypical neural oscillatory activity in dyslexia: A cross-linguistic perspective. Clinical Psychological Science 5(2), 379-401.

Langus, A., Mehler, J., & Nespor, M. (2018). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews 81*, 158-166.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: entrainment to English and French speech rhythms. *Psychonomic Bulletin and Review 18*, 1035-1041.

Liss, J., White, L., Mattys, S., Lansford, K., Lotto, A., Spitzer, S. & Caviness, J. (2009). Quantifying Speech Rhythm Abnormalities in the Dysarthrias. *Journal of Speech, Language and Hearing Research 52*, 1334-1352.

Low, E.L., Grabe, E., & Nolan, F. (2000). Quantitative characterization of speech rhythm: syllable-timing in Singapore English. *Language and Speech 43(4)*, 377-401.

MacNeilage, P. (1998). The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 21(4), 499-511.

McAuley, J.D., & Fromboluti, E.K. (2014). Attentional entrainment and perceived event duration. *Philosophical Transactions of the Royal Society B,* 369: 20130401.

Mehler, J. & Nespor, M. (2004) Linguistic rhythm and the development of language. In A. Belletti & L. Rizzi (Eds.) *Structures and Beyond: The Cartography of Syntactic Structures*. Oxford: Oxford University Press. Pages 213-221.

Mehler, J., Dehaene-Lambertz, G., Dupoux, E. & Nazzi, T. (1996) Coping with Linguistic Diversity: The Infant's Viewpoint. In J.Morgan and K. Demuth (Eds.) *Signal to Syntax*. Hillsdale, NJ:LEA, 101-116..

Merker, B., Madison, G., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. Cortex 45, 4–17.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016) Out-of-synchrony speech entrainment in developmental dyslexia. Human Brain Mapping, 37, 2767–2783.

Morey, R. D., & Rouder, J. N. (2012). BayesFactor: An R package for computing Bayes factor for a variety of psychological research designs (available in the Comprehensive R Archive Network).

Muneaux, M., Ziegler, J.C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: evidence from French. *Neuroreport 15*, 1255-1259.

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora Rhythm. *Language and Speech, 50(1)*, 77-99.

Nazzi, T., & Ramus, F. (*2003*). Perception and acquisition of linguistic rhythm by infants. *Speech Communication 41*, 233-243.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance 24(3)*, 756-766.

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System 42*, 244-257

Ordin, M., & Polyanskaya, L. (2015). Acquisition of English speech rhythm by monolingual children. *Proceedings of Interspeech 2015*, 3120-3124.

Ordin, M., & Polyanskaya, L. (2015a). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America 138(2)*, 533-544.

Ordin, M., & Polyanskaya, L. (2015b). Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology 6: 316*, 1-15.

Pallier, C. (2002). Computing discriminability and bias with the R software. Available at http://www.pallier.org/pdfs/aprime.pdf. Last accessed on November 4, 2018.

Pamies Bertran, A., (1999). Prosodic Typology: On the Dichotomy between Stress-Timed and Syllable-Timed Languages. *Language Design*, *2*, 103-130.

Payne, E., Post, B., Astruc, L., Prieto, P., & del Mar Varnell, M. (2012). Measuring child rhythm. *Language and Speech, 55(2)*, 203-229.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phrase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex 23*, 1378–1387.

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

Polka, L., & Sundara, M. (2011). Word Segmentation in Monolingual Infants Acquiring Canadian English and Canadian French: Native Language, Cross-Dialect, and Cross-Language Comparisons. *Infancy 17*, 198-232.

Polyanskaya, L., & Ordin, M. (2015). Acquisition of speech rhythm in first language. *The Journal of the Acoustical Society of America 138(3)*, 199-204.

Polyanskaya, L., Ordin, M., & Busa, M. (2017). Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language. Language and Speech 60(3), 333-355.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication 54*, 681-702.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America 105 (1)*, 512-521.

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

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition 73(3)*, 265-292.

Roach P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.). *Linguistic Controversies.* (Edward Arnold, London), 73-79.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: The role of distributional cues. Journal of Memory and Language, 35,606-621.

Schiering, R. (2007). The phonological basis of linguistic rhythm. Cross-linguistic data and diachronic interpretation. *Sprachtypologie und Universalienforschung* 60, 337-359.

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research 32*, 912-920.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics 25*, 1-24.

ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. Neuropsychologia, 63, 43–50.

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

Toro, J.M., & Nespor, M. (2015). Experience-dependent emergence of a grouping bias. *Biology Letters 11*.

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

Van Maastricht, L., Krahmer, E., Swerts, M., & Prieto, P. (in press). Learning direction matters. *Studies in Second Language Acquisition* 1–35.

Werker, J., & Curtin, S. (2005) PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development 1*, 197-234.

White, L., & Mattys, S. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics 35(4)*, 501-522.

White, L., Mattys, S., & Wiget, L. (*2012*). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language 66*(4), 665-679.

FIGURE CAPTIONS

Figure 1. Sensitivity (a) and response bias (b) of German and French listeners for different stimuli pairs from Experiment I. For example, „str-syl“ refers to performance in AX items, in which A was stress-timed and X was syllable-timed. Horizontal line indicates the chance level (50%). Error bars ±2SE around the mean.

Figure 2. RT for correct and incorrect answers given by German and French listeners. Error bars ±2SE around the mean.

Figure 3. RT for correct responses by German and French listeners for different pair-types. Error bars ±2SE around the mean.

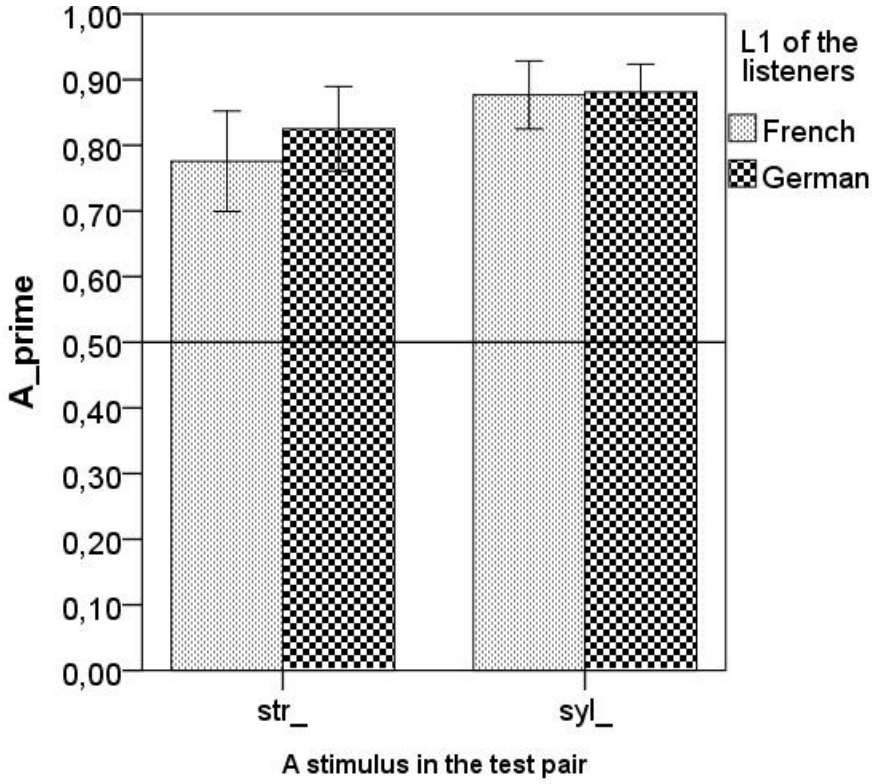
Figure 4. Sensitivity (a) and response bias (b) of German and French listeners for different stimuli pairs from Experiment II. For example, „str-syl“ refers to performance in AX items, in which A was stress-timed and X was syllable-timed. Horizontal line indicates the chance level (50%). Error bars ±2SE around the mean.

Table 1. Overview of the rhythm measures.

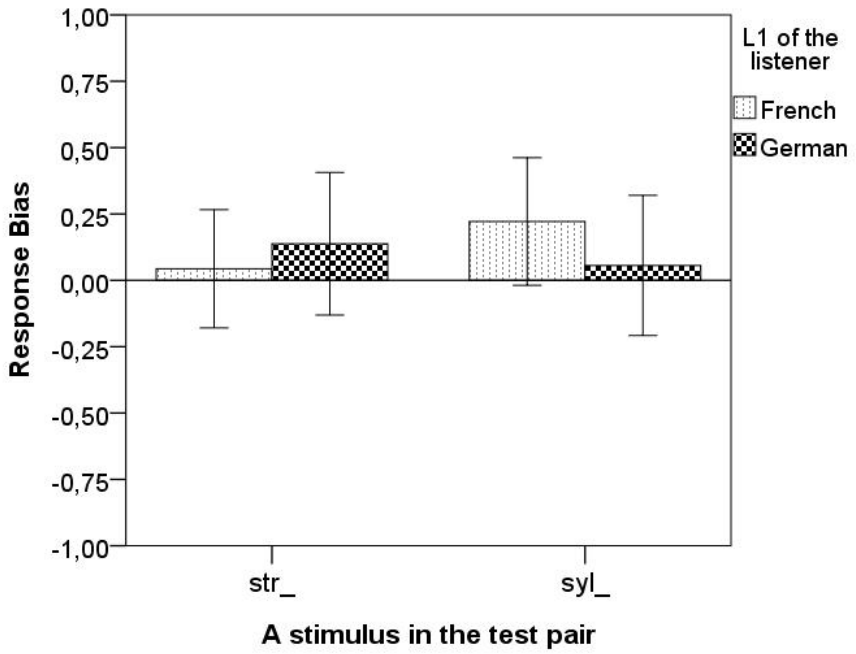
Metric name	Δ (delta)	Varco	%V	nPVI
Description	standard deviation in duration of speech intervals	Coefficient of variability in duration of speech intervals	Proportion of vocalic material in an utterance	Normalized pairwise variability index for speech intervals. $nPVI = \sum_{k=2}^{n} \left\| \frac{d_k - d_{k-1}}{(d_k + d_{k-1})/2} \right\| /(n-1)$
Reference	Ramus, et al., 1999	Dellwo, 2006	Ramus, et al., 1999	Low, Grabe, & Nolan, 2002

Table 2. Measures of durational variability of Vs and mean durations of Cs (meanC), Vs (meanV), and the proportion of vocalic material in the stimuli with two contrastive rhythms

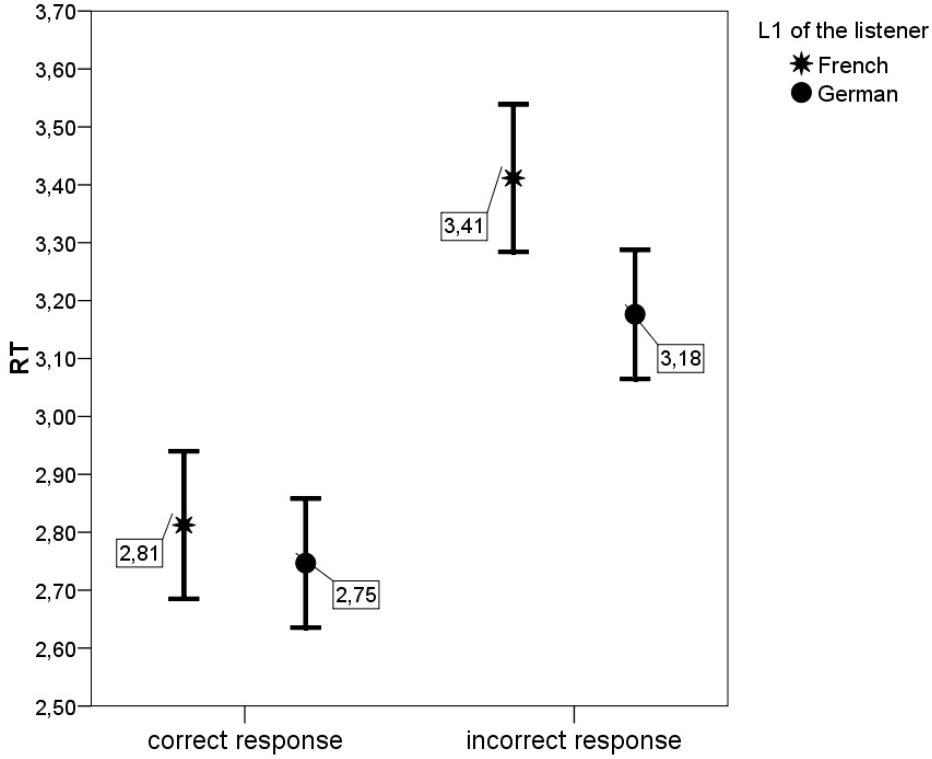
	%V	nPVI	ΔV	meanV	VarcoV	meanC
Stress-timed rhythm	40%	78	34.8	80ms	0.48	120ms
Syllable-timed rhythm	50%	7.7	10.7	100ms	0.107	100ms


1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
A_prime
str_
syl_
A stimulus in the test pair
L1 of the listeners
French
German

1a

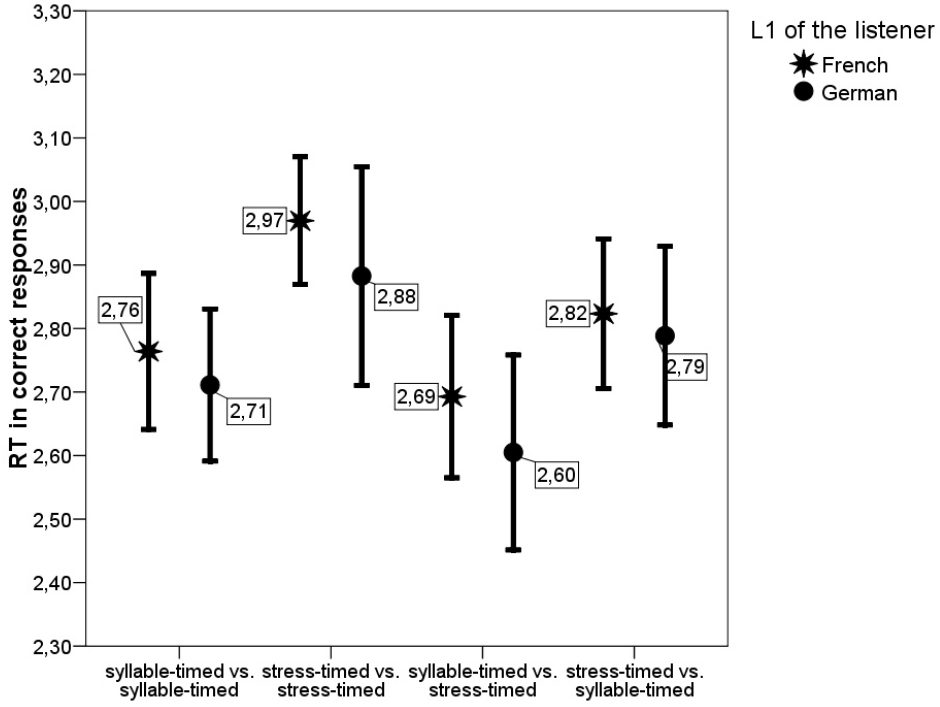

1,00
0,75
0,50
0,25
0,00
-0,25
-0,50
-0,75
-1,00
Response Bias
str_
syl_
A stimulus in the test pair
L1 of the listener
French
German

1b



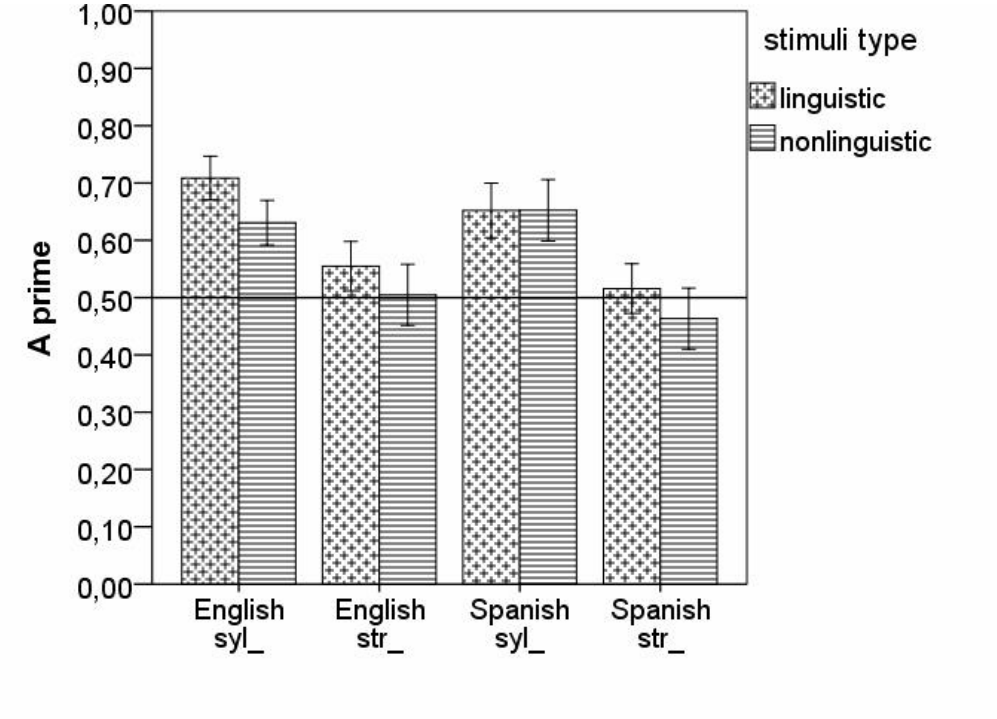
2

116x93mm (200 x 200 DPI)

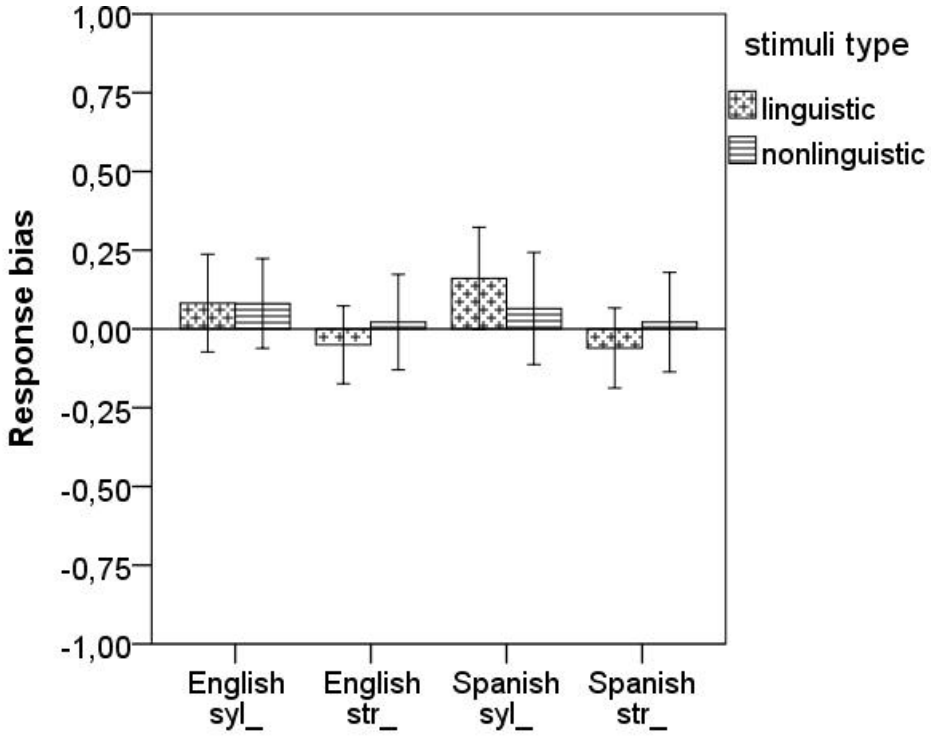


3

123x92mm (200 x 200 DPI)


1,00
0,90
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
A prime
stimuli type
linguistic
nonlinguistic
English syl_
English str_
Spanish syl_
Spanish str_

4a


1,00
0,75
0,50
0,25
0,00
-0,25
-0,50
-0,75
-1,00
Response bias
stimuli type
linguistic
nonlinguistic
English syl_
English str_
Spanish syl_
Spanish str_

4b

APPENDIX I:

To ensure that Welsh-based resynthesized sentences were perceived as linguistic stimuli for Experiment 2, we recruited 10 participants for a pilot assessment. We concatenated Welsh-based (i.e., linguistic) stimuli into two 2-minute sequences. Non-linguistic stimuli were also concatenated into two 2-minute sequences. Finally, we created an artificial language (following the approach often used in an artificial language learning paradigm, Aslin, Saffran, & Newport, 1998; Saffran, Newport, & Aslin, 1996). We used the same syllables with simple CV syllabic structure, which also comprised non-linguistic stimuli, and constructed 10 bi-syllabic nonsense words (samu, nelo, noma, namo, fenu, fale, lufe, mesu, sofu, sela). The vowel /i/ was only used in '*filler*' syllables (fi, si, mi, ni, li) that were interspersed with the nonsense words and modelled frequent structural words (articles and prepositions). Lexical stress on word-initial syllables was modelled by lengthening the vowel by a factor of 1.5. In a stream FI**MESU**MI**SELA**LI**SAMU**MI**FALE**SI**NELO**..., transitional probabilities (TPs) between syllables within words equal to 1.0, and between syllables straddling the word boundaries equal to 0.2, thus providing a reliable statistical cue for segmenting continuous syllabic streams into constituents. An intonational contour was imposed on the syllabic stream, with boundary tones aligned with some word-final syllables. This contour allowed clustering the statistical nonsense words into larger constituents, e.g., sentences. In this way, a prosodic hierarchy with smaller discrete constituents embedded into larger units (Nespor & Vogel, 2007) was modelled. Thisimplements into our artificial language a) a hierarchical structure, b) prosody (distribution of stressed and unstressed syllables, intonation); and c) statistical cues to mark discrete units and typological properties (head-prominence) of a language. In this way, we tried to make the stimuli sound similar to a real language. Such stimuli were also concatenated into two 2-minute sequences. The resulting 6 sequences of two minutes each were concatenated into a continuous stream, and we asked participants to listen to this stream. Prior to listening, participants were told that they were going to listen to some speech-like passages, some of which were a real natural language, while other passages were not.

Figure AI-1. Schematic illustration of the familiarization stream used in the pilot experiment for evaluating the stimuli.

After familiarization, participants were played 120 stimuli, 6.6-seconds each, 40 stimuli of each type. Upon listening to each stimulus, they had to indicate how likely it is that the stimulus is represents a real language. The responses were registered on an 8-point scale, from 1 – "I am sure it is a language", to 8 – "I am sure it is not a language". Figure AI-2 shows the results, error bars indicate 2 SE around the mean.

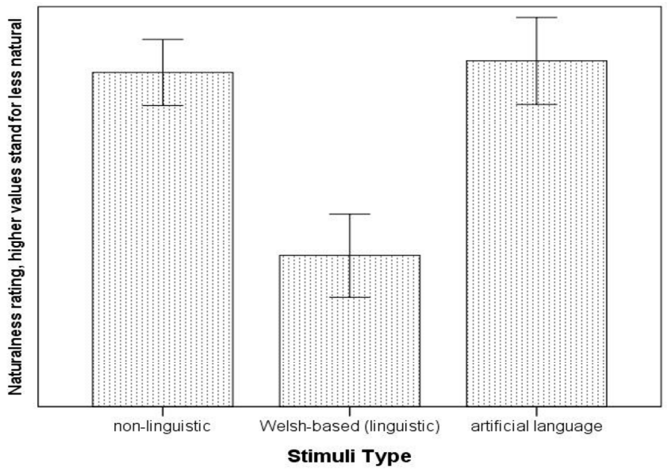


Figure AI-2. Differences in naturalness rating for different stimuli types.

The analysis revealed that stimulus type has a significant effect on the ratings, λ=.908, $F(2,398)$=20.262, p<.0005, η^2=.092. Pairwise comparisons (with Bonferroni correction applied) showed that Welsh-based stimuli were perceived to be significantly more likely to come from a real language than non-linguistic stimuli, p<.0005 for both contrasts. Ratings assigned to non-linguistic stimuli and to the artificial language did not differ, p=.6. Welsh-based stimuli were rated as significantly more natural and similar to real language utterances than the non-linguistic and artificial language stimuli. This suggests that, besides the presence of prosody and distributional cues, a larger variation in phonotactic and segmental complexity is needed for the sequence of syllables to be perceived as a real language. Based on these results, we chose Welsh-based resynthesized sentences as linguistic stimuli for Experiment 2.

APPENDIX II:

In order to ensure that linguistic stimuli indeed sound to participants as more representative of a real language, listeners were asked to perform a short test after the second session. For the test, we randomly selected 10 linguistic and 10 non-linguistic stimuli. Half of the stimuli exhibited regular rhythm (syllable-timed) and the other half had irregular rhythm (stress-timed). Participants listened to the stimuli in randomized order and responded to the question "How much does this sound like a real language". The naturalness ratings were given on an 8-point scale, from 1 – I am sure it is a language, to 8 – I am sure it is not a language. Figure AII-1 shows the results, error bars indicate 2 SE around the mean.

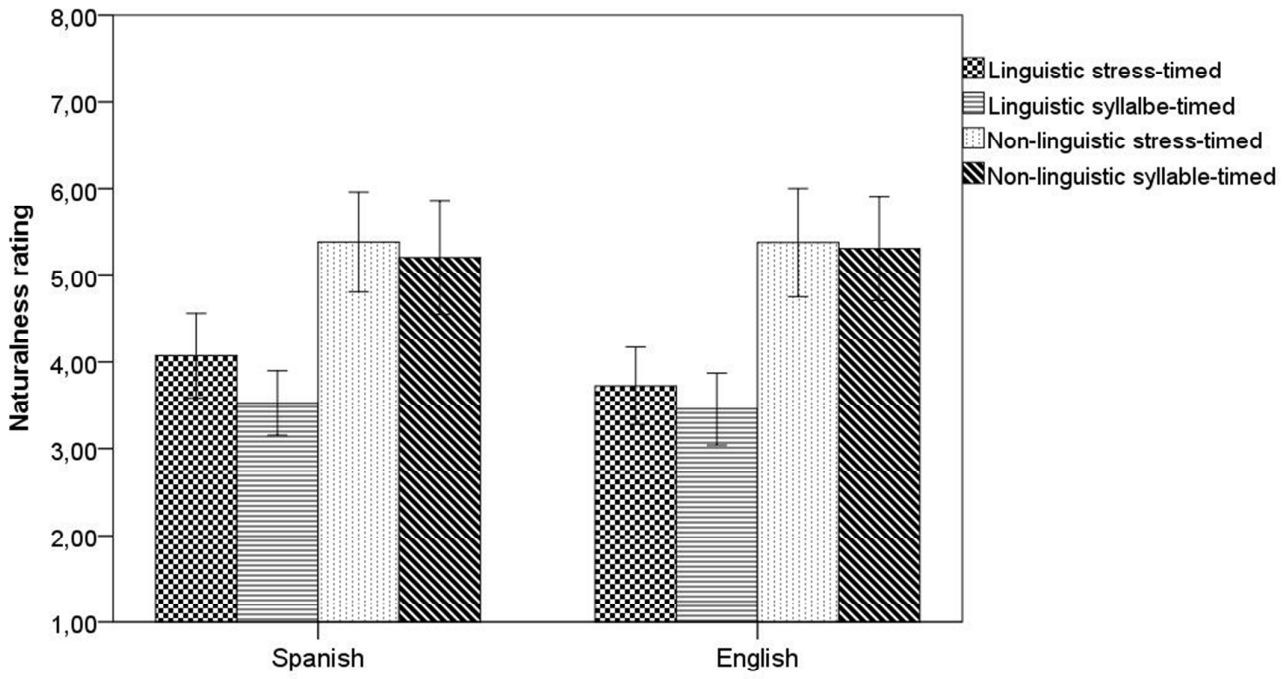


Figure AII-1. Naturalness ratings assigned by participants after the experiment to linguistic and non-linguistic stimuli with different rhythms.

An ANOVA with *L1* of the participant as a between-subject factor and *stimulus type* as a within-subject factor revealed a significant effect of stimulus type, λ=.45, F(3,52)=21.19, p<.0005, η^2=.55. There was no effect of *L1*, F(1,54)=.089, p=.767, η^2=.002 and no interaction between L1 and stimulus type, λ=.967, F(3,52)=.43, p=.73, η^2=.024. Pairwise comparisons (with Bonferroni correction) showed that the ratings assigned to linguistic stimuli were significantly lower (thus rated as more similar to real language) than the ratings assigned to non-linguistic stimuli.

The results also showed that manipulations of duration of the linguistic stimuli did not affect naturalness ratings. Within each stimulus type (linguistic or non-linguistic), the naturalness ratings assigned to the stimuli with irregular and regular rhythmic patterns (i.e., stress- and syllable-timed correspondingly) did not statistically differ. These results confirm that linguistic stimuli were indeed perceived as more representative of a real language and were more likely to engage a fuller set of speech processing mechanisms.