The effect of speech rhythm and speaking rate on assessment of pronunciation in second language

RUNNING TITLE: Rhythm and rate in L2 pronunciation

Polyanskaya Leona[1*]
Mikhail Ordin[1,2]

[1] BCBL – Basque Centre on Cognition, Brain and Language, Donostia, Spain
[2] IKERBASQUE – Basque Foundation for Science, Bilbao, Spain

[*] Corresponding author (l.polyanskaya@bcbl.eu; +34 943 309 300 Ext.: 202)
BCBL, Mikeletegi 69, Donostia, Gipuzkoa, Spain

ABSTRACT

The study explores the effect of deviations from native speech rhythm and rate norms on the assessement of pronunciation mastery of a second language (L2) when the native language of the learner is either rhythmically similar to or different from the target language. Using the concatenative speech synthesis technique, different versions of the same sentence were created in order to produce segmentally and intonationally identical utterances that differed only in rhythmic patterns and/or speaking rate. Speech rhythm and tempo patterns modelled those from the speech of French or German native learners of English at different proficiency levels. Native British English speakers rated the original sentences and the synthesized utterances for accentedness. The analysis shows that 1) differences in speech rhythm and speaking tempo influence the perception of accentedness, 2) idiosyncratic differences in speech rhythm and speech rate are sufficient to differentiate between the proficiency levels of L2 learners; 3) the relative salience of rhythm and rate on perceived accentedness in L2 speech is modulated by the native language of the learners; 4) intonation facilitates the

perception of finer differences in speech rhythm between otherwise identical utterances. These results emphasize the importance of prosodic timing patterns for the perception of speech delivered by L2 learners.

Keywords: foreign accent, speech rhythm, speech rate, timing patterns, pronunciation assessment

INTRODUCTION

It is easier to discriminate between some languages than others. Discriminating between utterances from different languages is based on segmental and suprasegmental differences related to phonetic and phonological structures. Suprasegmental differences related to the temporal organisation of utterances (articulation rate, number, distribution and duration of pauses, segment-to-segment and syllable-to-syllable durational variability, etc.) are particularly perceptually salient. These patterns of prosodic timing contribute to perceived cross-language differences in speaking rate and rhythm, and have an effect on the speech processing strategies adopted by speakers of languages with different temporal organisation (Cutler & Norris, 1988; Segui, Dupoux & Mehler, 1990; Ramus et al., 1999).

The term *speech rhythm* presents a serious terminology challenge because it is used differently by different authors (Nolan & Jeon, 2014; Turk & Shattuck-Hufnagel, 2014). In particular, there is an ongoing debate as to whether this term refers to a clearly measurable linguistic construct (Arvaniti, 2012). In order to avoid referring to rhythm as a linguistic construct, some researchers referred to the notions of *temporal organization* (Clopper & Smiljanic, 2015) or *prosodic timing* (Henriksen & Fafulas, 2017), to refer both to the timing patterns that result from the durational ratios of vowels and consonants and from characteristics of speech tempo. According to Dauer (1983), rhythm can be more exactly described by a set of phonological properties related to vowel reduction in unstressed syllables, presence of phonologically short and long vowels in the phonemic inventory, phonotactic constraints on syllabic complexity, etc. Adams (1979) and Tyler (1981) suggested that speech rhythm is the product of perception. They propose that listeners reconstruct rhythm based on clues provided by the speaker (e.g., by maintaining the ratios of phonemic durations during speech production). In the present article, for convenience, we will use the term *rhythm* despite these debates, mainly as a short cut, to refer to *patterns of durational variability in speech intervals*. Speech intervals such as syllables, vocalic or consonantal

intervals, interstress intervals, and feet are not necessarily isochronous, but exhibit systematic patterns in durational ratios, or durational variability. This understanding of speech rhythm as surface timing patterns is in line with Dasher and Bolinger (1982), Dauer (1983), Ramus et al., (1999), and Low, Grabe, and Nolan (2000), among others.

Patterns of durational variability, or durational ratios of vowels and consonants within utterances, are influenced by the mean duration of the corresponding speech intervals, i.e., by articulation rate (measured in the number of syllables produced per second). Thus, it is necessary to normalize the values of durational variability for the rate at which the utterance is delivered (Dellwo, 2006). For example, dividing the standard deviation of vowel durations by the mean duration of the vowels in an utterance will remove (or reduce) the influence of the articulation rate on the durational ratios of vowels, and will allow us to compare the durational variability of speech intervals in utterances produced at different rates.

Durational ratios – when normalized for articulation rate – differ between languages. Languages that are traditionally classified as ”stress-timed”, exhibit greater durational variability and a lower proportion of vocalic speech material compared to languages that are traditionally classified as “syllable-“ or “mora-timed” (Bunta & Ingram, 2007; Gervain, Nespor, Mazuka, Horie & Mehler, 2008; Grabe & Low, 2002; Nazzi, Bertoncini & Mehler, 1998; Payne, Post, Astruc, Prieto, & Vanrell, 2012; Ramus et al., 1999; White & Mattys, 2007). These differences persist even when phonological properties and phonotactic factors, which have an effect on durational ratios, are controlled for (Prieto, Vanrell, Astruc, Payne, & Post, 2012). Articulation rate also differs between languages. Germanic languages, including English and German, are spoken at a slower rate (≈4.2–5.2 syl/sec) than Romance languages (over 6 syl/sec), including French (Anderson-Hsieh & Venkatagiri, 1994; Clopper & Smiljanik, 2011; Dauer, 1983; Pellegrino, Coupe, & Marciso, 2011; Quene, 2005; 2007)[1].

[1] There is some evidence suggesting no significant differences in speech rate between languages, when speech rate is measured in segments per second (Roach, 1998 for an overview). However, in speech production the syllable is regarded as the basic unit of speech planning. Speakers plan and produce syllables or syllable parts (onsets and codas), not separate segments (Mooshammer et al., 2012). Articulatory gestures also correspond to syllable parts encompassing several segments rather than to separate individual segments (Goldstein, 1988). Therefore, measuring speech rate in syllables per second is ecologically more valid than in segments per second.

When articulation rate and rhythmic patterns in the native (L1) and target (L2) languages of a language learner differ, timing organisation in the L2 is often non-target like but tends to improve as L2 proficiency increases (Ordin & Polyanskaya, 2014; 2015; White & Mattys, 2007). L2 speech has frequently been shown to be generally slower than L1 speech (Guion et al., 2000; Lennon, 1990; Munro & Derwing, 1998), and articulatiin and speech rates positively correlate with the proficiency level of L2 speakers (Anderson-Hsieh & Venkatagiri, 1994; Ordin & Polyanskaya, 2014; 2015b).

A number of studies have also shown that the durational ratios in L1 and L2 speech are different. The general finding is that word, syllable, vocalic and consonantal durations in L2 English speech are less variable than those in English produced by native speakers (Baker et al, 2011; Bond & Fokes, 1985; Grenon & White, 2008; Ordin & Polyanskaya, 2015b; Tortel & Hirst, 2010; White & Mattys, 2007). Ordin and Polyanskaya (2014; 2015b) showed that rhythmic patterns change in the course of language acquisition as a function of proficiency.

Adams (1979) and Tyler (1981) suggested that deviations from the target temporal organisation influence the degree of the perceived foreign accent (FA) in L2 speech. A failure to provide a sufficient number of clues to enable the native listener to extract and recognize specific patterns of prosodic timing reduces the intelligibility of L2 speech and increases FA (Taylor, 1981: 224-225). A number of subsequent empirical studies confirmed that deviations in durational ratios in L2 speech affect the comprehensibility and intelligibility of speech (Baker et al., 2011; Munro & Derwing, 2001; Quene & van Delft, 2010; Tajima, Port & Dalby, 1997) and lead to the perception of L2 speech as more accented (Derwing, Munro, & Wiebe, 1998 Polyanskaya, Ordin, & Busa, 2017).

A slower articulation rate in L2 also influences the degree of perceived FA. Kang (2010) and Kang, Rubin, Pickering (2010) showed that speech delivered at a faster rate is perceived as less accented, although the contribution of rate is smaller than that of other

prosodic features, e.g. pitch range, mean length of silent pauses, etc. It should be noted that the researchers used natural accented speech in which rate varied with other prosodic parameters and segmental phonetic peculiarities, such that faster utterances may have also exhibited more severe deviations from the expected norms at the segmental level, while slower sentences might have included fewer pronunciation errors. Munro and Derwing (1998; 2001) studied the impact of rate on the strength of the perceived accent under experimental conditions. They confirmed that speech delivered at a faster rate (up to a certain threshold) is indeed judged to be less accented. The authors concluded that this effect was due to the rate differences themselves, and independent of differences in proficiency (although proficiency does vary with articulation rate, as shown by Anderson-Hsieh and Venkatagiri, 1994). These studies tell us that the degree of the perceived FA is influenced by differences in articulation rate[2] and rhythm both between L1 and L2 speech and between L2 speech produced by learners at different proficiency levels.

While both rhythmic patterns and articulation rate matter for the perception of accentedness in L2, we still do not know which matter most, and how non-target productions are modulated by the native language of the speaker and by his proficiency in the L2. Another open question is whether idiosyncratic segmental properties and various prosodic features (e.g., intonation) make differences in rate and rhythm perceptually more prominent, or, on the contrary, divert the listener's attention from non target-like prosodic timing in L2 utterances. So far, the evidence regarding the role of intonation in timing perception and language discrimination is inconsistent. Polyanskaya et al. (2017) showed that intonation enhances the perception of fine differences in prosodic timing. Vicenik and Sundara (2013) demonstrated that intonation might be used by listeners to discriminate between languages and even between dialects. However, results reported by Ramus and Mehler (1999) showed that the presence of intonation does not provide any advantage for perceiving differences in prosodic

[2] However, see Flege (1988) and Anderson-Hsieh and Koehler (1988), who did not detect any effect of articulation rate on the comprehensibility or perceived accentedness of L2 speech, despite significant differences in articulation rate between learners at different proficiency levels.

timing, and cannot be used to discriminate between degraded utterances from different languages. These conflicting results point to the need for further investigation of the potential interplay between intonation and the perception of speech rhythm.

A number of methodological challenges complicate an exploration of the relative contributions of rhythm and rate to perceived accentedness. In real speech, articulation rate and rhythmic patterns interact: at a faster rate, speech intervals are perceived to be less variable in duration (Dellwo, 2006; Wiget et al., 2010). Therefore, real utterances allow us to investigate the influence of temporal organisation on perceived accentedness, yet do not allow us to disentangle the perception of rate and rhythm. Only few attempts have been made to separately study the perceptual salience of rhythm and rate, and their unique roles in speech processing and the perceived accentedness of L2 speech. Further, these studies have provided contradictory results. Ordin and Polyanskaya (2015a) used L2 English utterances produced by German learners of English at different proficiency levels, transforming all vowels in these utterances into *'a'* and all consonants into *'s'* with monotonized F0 (following the '*sasasa'* transform method suggested by Ramus & Mehler, 1999). Native English listeners were trained to classify the stimuli into three categories, corresponding to those derived from utterances produced by German learners of English at elementary, intermediate and advanced proficiency levels. Since the derived '*sasasa*' stimuli preserved prosodic timing patterns, while the segmental and melodic differences between the utterances were eliminated, the only cues for classification pertained to tempo (the number of syllables per second) and rhythm (variability in the duration of vowels and consonants). After training, the listeners performed the actual test, splitting the stimuli into two categories – fast and slow – based on the number of '*sa*' syllables per second. Both categories contained stimuli with high and low durational variability of vowels and consonants (see Arvaniti and Rodriques, 2013, for similar findings in a discrimination task). Ordin and Polyanskaya (2015a) concluded that listeners relied entirely on speech rate and ignored differences in rhythm between the stimuli, both when

classification (i.e., identification) and discrimination tasks were performed. In a later study, Polyanskaya et al. (2017) resynthesized utterances produced by French learners of English at different proficiency levels so that the new utterances mimicked the rhythm and rate characteristics of the corresponding sentences spoken by French learners, and substituted pitch contours and segments with those of native English. Native English speakers were recruited to listen to the resynthesized sentences and rate them on a foreign accent scale. The results showed that both rate and rhythm affect the degree of FA, but the effect of rhythm was larger than that of rate.

The results of the studies by Ordin and Polyanskaya (2015a) and Polyanskaya et al. (2017) might seem contradictory in regard to whether speech rhythm or speech tempo has more perceptual salience. Two possible explanations can be proposed: the difference in the nature of the stimuli, and the difference between the prevailing rhythmic patterns in the native languages of the participants in these two studies. It is possible, that the '*sasasa*' stimuli in Ordin and Polyanskaya (2015a) are not perceived as speech, and thus fail to engage the full set of speech processing mechanisms and phonological filters shaped by native language. Non-speech stimuli may be processed at a lower, psychoacoustic level, using a direct physiological response to sharp increases in intensity that correspond to vowel onsets. In this scenario, a physiological response to the rate of alternation between '*s*' and '*a*' in '*sasasa'* stimuli would determine the rate at which neurons fire. This direct physiological response would make it easy to discriminate between and assign labels to fast and slow stimuli, but would reduce the perception of differences in durational variability. Thus the rate of intensity peaks in the acoustic stimuli would outweigh any difference in rhythmic patterns. By contrast, resynthesized sentences are processed as speech, and rhythm plays a very important role in speech processing (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996; Nazzi & Ramus, 2003; Kim, Davis, & Cutler, 2008; Murty, Otake, & Cutler, 2007). Rhythm should have a

bigger influence on perceived accentedness, when listeners have to process meaningful sentences.

Alternatively, it is possible that native English listeners were able to detect differences in speech rhythm between English utterances produced by French learners of English at different proficiency levels because French and English are rhythmically different languages. The deviations in rhythm in L2 English spoken by French learners are highly salient. By contrast, German and English are rhythmically closer, and the relatively small deviations in rhythmic patterns in L2 English spoken by German learners might not be sufficient to be perceptually relevant. In this study, we aimed to evaluate these two possible explanations by estimating the relative effects of rate and rhythm in L2 speech on perceived accentedness when native and target languages of the learners are rhythmically more or less distant. We also wanted to clarify the role of intonation in the perception of differences in timing organization between utterances.

METHOD

To address these research questions, we used the modified approach first implemented in Polyanskaya et al. (2017) of the re-synthesis technique to manipulate rhythmic patterns while controlling articulation rate. We used natural utterances produced by learners of English at different *proficiency* levels (beginners, intermediate and advanced). The original utterances differed both at the segmental and prosodic levels, i.e., in the realization of phonemes, intonation, stress, speech rhythm, speech rate, etc. We used the native British English diphone database (en1) in MBROLA (Dutoit et al., 1996) for re-synthesis to prepare stimuli in three *conditions*: those that differed either 1) only in durational ratios, 2) only in articulation rate, 3) both in durational ratios and articulation rate. Durational ratios and articulation rates of the resynthesized stimuli were the same as those in the original utterances spoken by L2 learners at different proficiency levels. All stimuli were made in flat (monotonized) and intoned versions. Native English listeners were recruited to listen to the stimuli and original

utterances, and to rate the degree of FA of each utterance, on a 6-point scale. We explored the unique and combined contributions of articulation rate and speech rhythm to perceived accentedness by using *condition* (rate only, rhythm only, rate and rhythm), *proficiency* (beginner, intermediate and advanced) as factors, and *accent ratings* on each stimulus as dependent variables in a repeated-measures MANOVA in SPSS (v. 18.0.1). The analysis was performed separately on flat and intoned stimuli.

We compared the effect sizes for *proficiency* and *condition* factors to explore the effect of F0 contour on the perception of timing patterns pertaining to rhythm and to rate. Further in this section, we provide more details on the speech material, stimuli preparation, and experiment procedure.

Speech material

Because we were interested in the effects of rate and rhythm on the perceived accentedness of the speech of L2 learners whose native languages were either rhythmically close to or more distant from the target language, we decided to work with L2 English produced by French and German learners. Ordin and Polyanskaya (2015) showed that the durational variability of syllables, consonantal clusters and vocalic sequences increases both in L2 and L1 English as acquisition progresses, regardless of whether the native language of the learners is rhythmically similar to (German) or rhythmically different from (French) the target language (English). The authors also showed that, although the direction of the development of speech rhythm was the same in both groups of English learners despite their rhythmically different native languages, French learners exhibited less target-like temporal organisation in L2 English compared to German learners.

We used the corpus of L2 speech originally collected to investigate the developmental changes in speech rhythm associated with progress in L2 acquisition for the current study. This corpus represents the speech of L2 learners at different proficiency levels. The corpus itself is publicly available on IRIS, and the researchers who collected the corpus provided a

detailed description in Ordin and Polyanskaya (2015). The speech corpus contains recordings of 48 French learners of English from the Parisian area and 51 German learners of English who grew up in or near the city of Bielefeld in NRW (the variety of German spoken in the area is close to the Northern Standard Variety of German). The recordings were made either in a sound-treated booth at an audio-visual studio at Bielefeld University (at 44 kHz, 16 bit, mono), or in a sound-treated booth at the laboratory of Phonetics and Phonology at Sorbonne Nouvelle Paris III University (at 48kHz, 32 bit, mono). Each speaker was individually recorded. All participants were asked the same ten questions related to reading and music preferences, career choices, lifestyle, biography and childhood. This informal interview lasted 10-12 minutes per participant. Immediately after the interview, a sentence elicitation task was performed, using 33 picture prompts. The pictures were presented one by one to participants, with a simple descriptive sentence written under each picture. The participants were instructed to look at the pictures and to remember the sentence. The participants could look through the pictures at their own pace and go backwards and forwards as they wished. When they said they were ready, the same pictures, without accompanying sentences, were shown again, and the participants were asked to say aloud the sentence that described the picture. These utterances were recorded for the corpus. The interviews were given to three native English teachers of English as a foreign language for evaluation on three parameters: grammatical accuracy, fluency and vocabulary. Each parameter was evaluated on a 10-point scale by each teacher, with 10 points indicating native-like performance. Consistency in evaluations between raters on each parameter separately, and consistency in evaluations within raters between parameters were assessed using Cronbach's alpha. High alpha values (<.88 for each parameter between teachers and < .75 between parameters within teacher) indicated that the ratings were consistent. These ratings, obtained based on evaluation of the learners' performance during the interview, were used to estimate the L2 proficiency of the learners (see Polyanskaya et al., 2017 or Ordin & Polyanskaya, 2015a for details).

From the 33 sentences produced by each speaker during the sentence elicitation task, a subset of 15 random sentences was chosen. Each sentence was produced by German and French learners at advanced, intermediate and beginning proficiency levels, yielding six utterances per sentence. Below, we use the word *utterance* to refer to the actual production of a specific sentence by a learner, thus the raw material includes 90 utterances of 15 sentences.

Stimuli preparation

The selected 90 utterances were used to create further stimuli. The original utterances differed both at segmental and prosodic levels, i.e., in the realization of phonemes, intonation, stress, speech rhythm, speech rate, etc. We measured the durations of segments in Praat (Boersma, 2001). The boundaries of the phonemic realizations were detected based on the criteria outlined in Stevens (2002). Then, we created three sets of synthetic stimuli, with the segmental durations of French and German learners of English. In the first set, the three utterances of the same sentence differed only in rate, in the second set the utterances differed in rhythm and rate, and in the third set the utterances differed only in rhythm (see figures 1-3 for an overview). Since the same diphone database was used for resynthesis, all segmental differences between utterances produced by learners with different L1s and proficiency levels in the L2 were neutralized and did not have an effect on accentedness ratings. We used the en1 diphone database to ensure all phonemic realizations were native-like.

In order to create stimuli that were different only in rate (*rate only* condition), we fed the phonemic durations measured on the utterances spoken by intermediate learners into MBROLA, and synthesized the utterances with these durations. Next, the fast and slow versions of these utterances were synthesized by increasing and decreasing the vowel durations by 15% (see figure 1)[3]. Polyanskaya et al. (2017), following the methodology of Munro and Dewing (1998; 2001), manipulated speech tempo by stretching and compressing whole sentences. It should be noted that in natural speech vowels and consonants are not

[3] Based on the just-noticeable differences in speech rate measured by Quené (2007), we assume that our manipulation of vowel durations produces perceptually detectable changes in speech rate.

stretched and compressed to the same degree. In the current study, the methodology was improved by accounting for the fact that vowels allow greater modifications of duration compared to consonants, a modification which gives us more confidence in the results. The implemented manipulations do not change the durational ratios of vowels and consonants, thus do not alter tempo-normalized durational variability (i.e., rhythmic patterns). The only differences between the fast, normal and slow versions of the same utterance pertain to the articulation rate.

Next, we synthesized the stimuli using the segmental durations measured in the utterances produced by beginning and advanced learners of English. The resynthesized utterances with the durations of advanced learners were then stretched or compressed to make them equal to the duration of the fast utterances in the *rate only* condition. The resynthesized utterances with the durations of beginning learners were stretched or compressed to make them equal to the duration of the slow utterances in the *rate only* condition. Finally, the resynthesized utterances with the durations of intermediate learners were made equal to the utterances with a normal rate in the *rate only* condition. The resulting versions of these sentences differed in patterns of temporal organisation pertaining both to articulation rate and the durational ratio of vowels to consonants, characteristic of the different proficiency levels of language learners. Below, we refer to these stimuli as the *rhythm+rate* condition. Figure 2 illustrates the procedure for creating these stimuli.

Finally, we created the third set of stimuli (*rhythm only* condition) by making the *rhythm+rate* versions of the utterances of each sentence equal in duration, as illustrated in figure 3. When the differences in rate (number of syllables per second) between the utterances representing different re-synthesized versions of the same sentence in the *rhythm+rate* condition are neutralized, the only difference that remains between the versions pertains to rhythm.

To create the intoned variants of the resynthesized utterances, we recorded a male native British English speaker producing the chosen 15 sentences. The configurations of his F0 contours were imitated in the resynthesized utterances of each sentence, thus creating similar intonation contours for each intoned synthetic utterance per sentence, and neutralizing intonational differences between utterances of the same sentence, caused by L2 *proficiency* or the native language of the learners.

Finally, we prepared:

I) The original utterances: 15 sentences, each produced by six L2 learners of English (three German and three French leaners);

II) Resynthesized utterances for the *rate only* condition: 3 intoned and 3 flat versions of each of the 15 sentences produced by German learners, and 3 intoned and 3 flat versions of each of the 15 sentences produced by French learners.

III) Resynthesized utterances in the *rhyhm+rate* condition: 3 intoned and 3 flat versions of each of the 15 sentences produced by German learners, and 3 intoned and 3 flat versions of each of the 15 sentences produced by French learners.

IV) Resynthesized utterances in the *rhythm only* condition: 3 intoned and 3 flat versions of each of the 15 sentences produced by German learners, and 3 intoned and 3 flat versions of each of the 15 sentences produced by French learners.

INSERT FIGURES 1-3 SOMEWHERE HERE

Stimuli verification

Given that creating the stimuli involved the synthesis technique, it is possible that the stimuli in one condition were more natural-sounding than those in another condition, and that the differences in accentedness ratings could be potentially explained not only by differences in timing organisation, but also by differences in naturalness. To factor out this explanation, we performed a verification test on the stimuli.

We recruited 20 native English speakers (different from those who participated in the main experiment), and asked them to listen to the synthetic stimuli with segmental durations of French or German learners of English. The stimuli were played to the participants one by one, upon each stimulus the listeners had to answer the questions "Was this utterance produced by a human or by a machine?" and then "Is this utterance comprehensible?" The responses were registered on an 8-point scale, from 1 – "it was definitely produced by a human" to 8 – "it was definitely produced by a machine" for the naturalness scale, and from 1 – "the utterance is totally comprehensible" to 8 – "the utterance is absolutely incomprehensible". For each listener, we averaged the naturalness and comprehensibility ratings across the stimuli in *rhythm only*, *rate only*, and *rhythm+rate* conditions. The stimuli were rated as moderately natural (likely produced by a machine than by a human), and as highly comprehensible (figure 4).

INSERT FIGURE 4 SOMEWHERE HERE

Importantly, we wanted to show that different types of manipulation, used to prepare the stimuli, did not produce differences in the perceived naturalness or comprehensibility of the synthetic stimuli, we had to test the null hypothesis (namely, no difference in naturalness and comprehensibility ratings between the conditions). In order to estimate the support for the null hypothesis, we calculated the Bayes factors (JASP v.0.8.0.2) for *condition* as a within-subject factor, potentially affecting the naturalness (BF_{10}=.3) and comprehensibility (BF_{10}=.1) ratings (uninformative default priors: the probability of the null and the alternative hypotheses are taken as equal in the priors). The Bayes factors provide a very strong support for the null hypothesis for comprehensibility ratings (the scores of the Bayes factors are interpreted based on Jarosz & Wiley, 2014), and a positive support for the null hypothesis for naturalness rating: the probability that the implemented manipulations produced differences in perceived naturalness and comprehensibility of the stimuli from different conditions is very low.

Procedure

We recruited 48 native speakers of British English who were not fluent in any foreign language. Each participant came to the experiment four times, with an interval of at least 2 weeks between sessions. In each session, participants listened to the stimuli in one of the conditions (*original*, *rhythm only*, *rate only*, *rhythm+rate*). The stimuli within each session were presented three times, in blocks. The order of utterances was randomized within blocks. Thus, within each session the listeners had to evaluate either 135 original utterances, or 270 synthetic utterances.

The listeners were asked to use the mouse and rate the degree of perceived FA in each utterance on a 6-point scale by clicking one of the six buttons on the screen: 6 (native-like), 5 (mild accent), 4 (moderate accent), 3 (rather strong accent), 2 (strong accent), and 1 (strongest accent). The stimuli were played via headphones connected to the computer. A new stimulus was played automatically after the response was given. The listeners could replay each stimulus a maximum of three times before evaluating its accentedness.

Given the number of utterances to be rated, half of the participants listened to the French-based and the other half listened to the German-based stimuli. The order of stimuli within each session was randomized. The order of sessions was completely counterbalanced (thus 24 participants heard the German-based and 24 participants heard the French-based set of stimuli).

The responses given to the first block were discarded from further analysis. We assumed that during the first block listeners were familiarized with the range of degrees of accentedness, and constructed an internal scale and reference to use in the evaluation of subsequent incoming stimuli. As participants have familiarized themselves with the spread in the degrees of accentedness by the end of the first block, their ratings within the session become more consistent (Guttman split-half coefficients >.82, using a split-half test with the second and third blocks as the two halves). This consistency resulted in an absence of outlier data points (ratings exceeding 2SE) for any participant in any condition (also, we verified that

there were no outlier data points across participants within any condition and verified the assumption of normality). The ratings assigned to the same utterance in the second and third blocks were subjected to a RM AMOVA analysis as a dependent variable.

Predictions

As differences in timing patterns between L1 and L2 utterances have been reported to impede intelligibility and to strengthen perceived accentedness in English produced by L2 learners, we assumed that developmental differences in speech rhythm between the utterances produced by learners of English at different proficiency levels could also be detected and contribute to the perceived accentedness of L2 speech. Timing organisation in L2 English becomes more native-like as L2 proficiency grows. Therefore, we predicted that *proficiency* should be a significant factor in all conditions: utterances with the timing patterns of more proficient learners would be rated as less accented. We also expected that rhythm and rate would make unique (i.e., independent) contributions to perceived accentedness, and that the combined effect of rhythm and rate should be greater than the unique effects of rhythm and rate. Therefore, we expected that *condition* would be a significant factor. Polyanskaya et al. (2017) demonstrated that rhythm makes a bigger contribution than tempo to the perceived accentedness of utterances produced by French learners. We did not know if this would still be the case in utterances produced by German learners. The results of Ordin and Polyanskaya (2015a) suggest that this might not necessarily be the case, since differences in the rhythmic patterns between utterances produced by German learners of English are less perceptually relevant than the rhythmic differences between utterances of French learners at different proficiency levels. Therefore, we expected a significant interaction between *condition* and *proficiency*, with the effect size of the *proficiency* factor for the *rhythm only* condition to be larger for the group listening to French-based stimuli than for the group listening to German-based stimuli. The opposite relation between effect sizes should be observed in the *rate only* condition. Finally, we expected that the presence of intonation should enhance the perception

of subtle differences in timing patterns, therefore we expected the effect size of *proficiency* to be larger for *intoned* than for *flat* stimuli.

RESULTS

First, we wanted to make sure that the original utterances produced by learners at different proficiency levels did indeed differ in accentedness ratings. For this purpose, we performed a repeated measures analyses with *proficiency* as a factor on the accent ratings assigned to the original utterances produced by beginning, intermediate and advanced learners of English. The main analyses were followed with planned comparisons, in order to test whether the differences in perceived accentedness between beginners and intermediate learners and between intermediate and advanced learners were significant. The descriptive data is given in table 1. This analysis revealed significant differences in accent ratings assigned to the utterances delivered by German learners of English at different proficiency levels, $F(2,718)=1064.776$, $p<.0005$, $\eta_p^2=.748$. Planned comparisons revealed significant differences in ratings assigned to the utterances delivered by advanced and intermediate learners, $p<.0005$, $\eta_p^2=.301$, and by intermediate and beginning learners, $p<.0005$, $\eta_p^2=.581$. The pattern of results for ratings assigned to the utterances by French learners was similar, $F(2,718)=1304.469$, $p<.0005$, $\eta_p^2=.791$, with planned comparisons revealing significant differences in ratings assigned to the utterances delivered by advanced and intermediate learners, $p<.0005$, $\eta_p^2=.570$, and by intermediate and beginning learners, $p<.0005$, $\eta_p^2=.530$. The analyses revealed that the utterances produced by advanced learners were rated as least accented, and utterances by beginners were rated as most accented. This pattern of results is presented in figure 5.

INSERT FIGURE 5 SOMEWHERE HERE

To explore if and to what extent differences in patterns of timing organization between proficiency levels contributed to pronunciation assessment of L2 utterances, multivariate analyses were performed on *accent ratings* with *proficiency* (*beginning, intermediate,*

advanced) and *condition* (*rhythm-only*, *rate-only*, *rhythm and rate*) as factors. The results showed that the effect of *proficiency* was significant and substantial, explaining over 40% of the variance in ratings of the utterances with French timing patterns and around 30% of the variance in ratings of the utterances with French timing patterns (see table 2 for the exact statistics). This reveals the major finding of our study: the differences in timing patterns between proficiency levels of L2 learners are not only perceivable, but also sufficient to reliably assign pronunciation ratings that vary in the same direction as the ratings assigned to the original utterances, which exhibited segmental and intonational differences in addition to differences in prosodic timing (figures 6-7). The main analyses also showed significant interactions of *condition* and *proficiency* factors (table 2), indicating that the contribution of *proficiency* to accentedness varies across the different *conditions*, i.e., according to whether participants are listening to stimuli with different rhythms or at different rates.

INSERT FIGURES 6-7 SOMEWHERE HERE

To explore the effect of proficiency, the ratings assigned to the stimuli with the timing organization of beginning learners were compared to the ratings assigned to the stimuli with the timing organization of intermediate learners, and the ratings assigned to the stimuli with the timing organization of intermediate learners were compared with the ratings assigned to the stimuli with the timing organization of advanced learners. These planned comparisons clearly showed significant differences for *ratings* of utterances with timing patterns of beginning and intermediate learners, as well as between utterances with timing patterns of intermediate and advanced learners.

For the *condition* factor, planned comparisons were constructed to compare the effects of *rhythm* and *rate* with the overall experimental effect. An insignificant contrast would show that the unique contribution of either rhythm or rate differences between proficiency levels is not different from the overall experimental effect, while a significant contrast would

demonstrate that the unique contribution of either rhythm or rate differences between proficiency levels is different from the overall experimental effect.

INSERT TABLE 2 SOMEWHERE HERE

The results (table 2) show that the overall experimental effect on the French-based stimuli is the same as the effect of rhythmic differences (p=.493 for intoned stimuli and p=.896 for flat stimuli), and is different from the effect of rate differences (p<.0005 for intoned and p<.0005 for flat stimuli). This means that the difference in accent ratings for French-based stimuli between *conditions* was due to differences in ratings assigned to *rhythm-only* stimuli. This confirms the conclusion of Polyanskaya et al. (2017) that rhythm makes a bigger contribution to perceived FA than rate, if the native and the target languages of the learners are rhythmically different. However, the overall experimental effect on the German-based flat stimuli is the same as the effect of rate differences (p=.397) and is different from the effect of rhythm differences (p<.0005), which means that rate differences make a bigger contribution to perceived FA than rhythm for flat stimuli, provided that the native and the target languages of the learners are rhythmically similar. For the intoned German-based stimuli, the overall experimental effect is different from the effect of rhythm differences (p<.0005) and from the effect of rate differences (p=.007), representing the contribution of both rhythm and rate to perceived accentedness. This indicates that rhythmic differences, at least in German-based stimuli, become more prominent in intoned stimuli.

To explore the interaction of *proficiency* and *condition* factors, we performed repeated measures analyses of variance for *ratings* on *condition* separately with *proficiency* as a factor, and confirmed that the differences of accent ratings between proficiency levels are significant in all conditions, both for flat and intoned stimuli, for French-based and German-based stimuli. The results are shown in table 3. Interestingly, the effect size of the proficiency factor for the accentedness ratings assigned to intoned French-based stimuli in the *rhythm only* condition (η_p^2=.211) is larger than that for flat stimuli (η_p^2=.179). At the same time, the effect

size for the French-based flat *rate-only* stimuli (η_p^2=.19) is larger than that for intoned stimuli (η_p^2=.15). The pattern of results for German-based stimuli is similar: the effect size of *proficiency* on *accent ratings* in the *rhythm only* condition is larger for intoned stimuli (η_p^2=.075) than for flat stimuli (η_p^2=.057), confirming the earlier interpretation of the planned comparisons for different conditions, and in the *rate only* condition the effect is larger for flat stimuli (η_p^2=.134) than for intoned stimuli (η_p^2=.081). This perfectly replicates the findings reported by Polyanskaya et al. (2017) showing that intonation impedes the perception of small differences in rate and enhances the perception of small differences in rhythm in L2 English produced by French learners, and allows us to generalize this finding to L2 English produced by German learners, despite the differences in timing organization in learners' native languages. It is also important to note that the effect sizes both for flat and intoned stimuli and for both French-based and German-based stimuli are larger in the *rhythm+rate* condition than in the *rhythm only* or *rate only* conditions, which means that the combined contribution of rhythm and rate is larger than the unique effects of rhythm or rate, taken separately, and for German-based stimuli even the summed effect size for *rhythm only* and *rate only* is smaller than the effect size in the *rhythm+rate* condition, suggesting that the patterns of rhythm and rate interact in perception in a complex multidimensional manner, and the contribution of deviations in the temporal organization of L2 speech from native norms is larger than the sum of the unique contributions of patterns pertaining to rate and to rhythm.

Finally, the effect sizes reported in table 3 indicate that for intoned French-based stimuli, the effect of *proficiency* on accentedness is stronger for the *rhythm only* condition (η_p^2=.211) than for the *rate only* condition (η_p^2=.158), and for German-based stimuli, the effect of *proficiency* is stronger for the *rate only* condition (η_p^2=.081) than for the *rhythm only* condition (η_p^2=.057). This indicates that in L2 English produced by French learners, the contribution of durational ratios (i.e., rhythm) to perceived accentedness is larger than the

contribution of rate. By contrast, in L2 English produced by German learners, the contribution of rate is larger than the contribution of rhythm.

INSERT TABLE 3 SOMEWHERE HERE

DISCUSSION

The results showed that differences in prosodic timing patterns between proficiency levels of L2 learners are perceived by native speakers of the target language and contribute to their assessment of pronunciation. Moreover, these differences may be sufficient to differentiate between the proficiency levels of L2 learners. The results also revealed that although variation in durations of speech constituents (i.e., rhythmic patterns) and the number of syllables per second (speech tempo) are closely related in speech production and perception, they nevertheless make separate, unique contributions to the perceived accentedness of L2 speech. However, speech rhythm and tempo contribute to foreign accent in different proportions, depending on the temporal organization of the native and target languages of the learners. If the target and native languages are rhythmically similar, then the effect of rhythm on accent ratings will be smaller than the effect of rate. If native and target languages are rhythmically different, then rhythm makes a bigger contribution to perceived accentedness than rate.

The ratings assigned in the *rhythm only* condition with rhythmic patterns of German learners of English at advanced and intermediate levels do not differ significantly, which indicates that German L2 learners of English achieve the ceiling effect in L2 rhythm acquisition by the intermediate proficiency level. Supposedly, the contribution of speech tempo to pronunciation assessment increases if the variation in rhythmic patterns is within the native norms of the target language. This assumption is in line with Ordin and Polyanskaya (2015b), who showed that advanced German learners of English do not differ in rhythmic patterns from native speakers of English. We conclude that rhythmic deviations from native norms are already not perceptually relevant for L2 English produced by German learners, and thus do not contribute to perceived accentedness by the intermediate stage. French

intermediate learners of English, by contrast, exhibit rhythmic deviations from native norms that affect pronunciation assessments and contribute to the degree of perceived FA. The perceptual salience of rhythmic deviations from the norms in L2 English spoken by French learners outweighs the prominence of rate differences, thus we found that only rhythmic differences contributes to perceived accentedness in French-based stimuli. In German-based stimuli, by contrast, we observed the contribution of both rhythm and rate deviations from native norms, because rhythmic deviations do not overshadow the perceptual salience of rate deviations.

We have also confirmed the hypothesis that intonation enhances the perception of finer differences in rhythmic patterns, and slightly inhibits the perception of small differences in rate. The impeding effect of F0 contour on perception of fine differences in articulation rate can explain the results reported by Ramus and Mehler (1999) on re-synthesized "sasasa" stimuli, showing lower discrimination performance for stimuli with F0 contours than for flat stimuli. If "sasasa" stimuli are processed at the psychoacoustic level, without engaging higher-order language-processing mechanisms, than the rate of "s" and "a" alternations can be more perceptually relevant for discrimination of stimuli (this interpretation is also in line with Ordin & Polyanskaya, 2015b), and the rate differences can be more perceptually salient in flat stimuli. The facilitatory effect of the presence of intonation on rhythm perception is probably explained by the fact that durational ratios interact with F0 fluctuations in prominent syllables that manifest lexical and phrasal stress. The perception of linguistic prominence is enhanced when several cues work in synergy, and perception of one cue leads to better processing of the other (see Jun, 2014 for similar ideas). Jun (2014) proposed that languages with lower durational variability of speech intervals (i.e., French) exhibit stronger tonal rhythm (i.e., F0 peaks and valleys that are larger in magnitude and are more regularly distributed in time) than languages with bigger durational ratios of speech intervals (i.e., German and English). Thus,

intonational contour enhances perception of fine rhythmic differences in French-based stimuli and has no effect on the perception of rhythmic differences in German-based stimuli.

Our study clearly demonstrates that attainment of gradient properties referring to durational ratios of speech intervals and a typical articulation rate can lead to significant improvement in L2 pronunciation by reducing perceived accentedness, especially if the target and the native languages of the learners are rhythmically different. However, as we noted in the Introduction, speech rhythm is determined by a set of properties related to durational vowel contrasts due to the lengthening degree in stressed syllables, phrase-final lengthening (Byrd, 2000; White & Mattys, 2007), distribution of prominence and acoustic correlates of phrasal prominence and lexical stress (Jun, 2014), as well as phonotactics (Arvanitti, 2012; Prieto et al., 2012). Consequently, production of target-like rhythmic patterns also requires development of control in multiple segmental and prosodic domains, and acquisition of segmental characteristics at earlier stages of L2 learning is a pre-requisite for successful acquisition of prosodic timing at later stages, when the relative contribution of timing patterns to the assessment of accentedness becomes more salient (Polyanskaya et al., 2017).

Many theories of L2 learning suggest that similarity between the target and native languages makes L2 acquisition easier (e.g., Flege, 1995; Best, 1995), which is, to some extent, also true for acquisition of L2 rhythm (Ordin & Polyanskaya, 2015; White & Mattys, 2007). However, recent empirical studies suggest that similarity alone cannot explain the acquisition of rhythm. Van Maasticht et al. (2018) investigated the development of rhythmic patterns in Spanish by Dutch learners and in Dutch by Spanish learners. The authors reported that Dutch learners of Spanish (a more syllable-timed language with relatively low durational variability of speech intervals) were in general more successful in acquiring rhythm than Spanish learners of Dutch (a more stress-timed language with relatively high durational variability of speech intervals). Consequently, the rhythmic difference between languages cannot account for all of the variance in rhythm attainment, and the direction of acquisition

matters as well (from more syllable-timed to more stress-timed, or from more stress-timed to more syllable-timed). These results suggest that rhythm in languages that exhibit higher durational variability (i.e., more stress-timed languages) is more challenging. As rhythmic characteristics in such languages are universally more marked (Ordin & Polyanskaya, 2015), the results corroborate the Markedness Differential Hypothesis, stating that features that are more marked are also more difficult to attain and are acquired later in L2 and L1 acquisition. Our study leaves open the question of whether the relative contributions of rhythm and rate differ depending on the learning direction. Is the relative contribution of rhythm higher than that of rate in L2 French produced by English learners because the languages are rhythmically different, or must the learning direction also be taken into account? This research question can only be answered when new empirical data becomes available, and may have important pedagogical implications in the domain of pronunciation training, both for the order in which pronunciation features are targeted, and at which proficiency levels specific features should be targeted over the course of learning and teaching L2 languages other than English.

ACKNOWLEDGEMENT

Support provided by the IKERBASQUE – Basque Foundation for Science and by Ministerio de Ciencia E Innovacion through Ayuda Centro de Excelencia Severo Ochoa SEV-2015-0490.

REFERENCES

Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. Mouton.

Anderson-Hsieh, J., & Koehler, K. (1988).The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning 38*, 561-613.

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of intermediate and high proficiency Chinese ESL speakers. *TESOL Quarterly 28*, 807-812.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3), 351-373.

Arvaniti, A., & Rodriquez, T. (2013). The role of rhythm class, speaking rate and *F0* in language discrimination. *Laboratory Phonology 4*, 7-38.

Baker, R.E., Baese-Berk, M.M., Bonnasse-Gahot, L., Kim, M., Van Engen, K., & Bradlow, A. (2011). Word durations in non-native English. *Journal of Phonetics 39(1)*, 1-17.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Bond, Z. S., & Fokes, J. (1985). Non-native patterns of English syllable timing. *Journal of Phonetics 13*, 407-420

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.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica 57*, 3–16.

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics 39*, 237-245.

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

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14(1), 113-121.

Dasher, R. & Bolinger, D. (1982). On pre-accentual lengthening, *Journal of the International Phonetic Association* 12, 58-69.

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

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

Derwing, T., Munro, M. & Wiebe, G. (1998) Evidence in favour of a broad framework for pronunciation instruction. *Language Learning 48,* 393-410.

Dutoit T., & Leich H. (1993). MBR-PSOLA: Text-to-Speech Synthesis Based on an MBE Re-Synthesis of the Segments Database. *Speech Communication 13,* 435-440.

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.

Flege, J. (1988). Factors affecting the degree of perceived foreign accent in English sentences. *Journal of Acoustical Society of America 91*, 370 – 389.

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

Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica 45*, 140-155.

Grabe, E., & Low, L. (2002). Acoustic correlates of rhythm class. In C. Gussenhoven & N. Warner (Eds.). *Laboratory Phonology 7* (515-546). New York: Mouton de Gruyter.

Grenon, I., & White, L. (2008). Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese. In *Proceedings of the 32nd Boston University Conference on Language Development*, 155-166.

Guion, S., Flege, J., Liu, H., & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics 21*, 205–228.

Henriksen, N., & Fafulas, S. (2017). Prosodic timing and language contact: Spanish and Yagua in Amazonian Peru. *Studies in Hispanic and Lusophone Linguistics 10(2)*, 225–257.

Jarosz, A.F., Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *Journal of Problem Solving 7*, 2-9.

Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In: Sun-Ah Jun (Ed.). *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 520-540). Oxford: Oxford University Press.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System 38*, 301-315.

Kang, O., Rubin,

D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgements of language learner proficiency in oral English. *The Modern Language Journal 94(4)*, 554-566.

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

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning 40*, 387–417.

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

Malisz, Z. (2011). Tempo differentiated analyses of timing in Polish. In:Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 1322–1325.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant's viewpoint. In J. Morgan and K. D. Demuth (Eds.). *From Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 101–116.

Mooshammer, Ch., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridging planning and execution: Temporal planning of syllables. *Journal of Phonetics 40(3)*, 374-389.

Munro, M., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning 48*, 159–182.

Munro, M.J., & Derwing, T.M. (2001). Modeling Perceptions of the accentedness and comprehensibility of L2 speech. The role of speaking rate. *Studies in Second Language Acquisition 23*, 451-468.

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

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: a metaphor? *Transactions of the Royal Society B* 369, 20130396. doi:10.1098/rstb.2013.0396

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

Ordin, M., & Polyanskaya, L. (2015a). Production and perception of speech rhythm in English as the second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology, 6: 316*. doi: 10.3389/fpsyg.2015.00316

Ordin, M., & Polyanskaya, L. (2015b) Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *Journal of Acoustical Society of America 138(2)*, 533-545.

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

Pellegrino, F., Coupé, Ch., & Marsico, E. (2011). Across-Language Perspective on Speech Information Rate. *Language 87(3)*, 539-558.

Polyanskaya, L., & Ordin, M. (2015). Acquisition of speech rhythm in first language. *Journal of Acoustical Society of America 138(3)*, EL199-204.

Polyanskaya, L., Ordin, M., & Busa, M.G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Language and Speech*. DOI: 10.1177/0023830916648720

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.

Quené, H. (2005). Modeling of between-speaker and within-speaker variation in spontaneous speech tempo. In *Proceedings of the ninth European conference on Speech Communication and Technology* (*InterSpeech*), 4–8 September (pp. 2457–2460). Lisbon, Portugal.

Quene, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics 35*, 353-362.

Quené, H., & van Delft, L. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication* 52, 911-918.

Quene, H., & van Delft, L.E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication 52*, 911-918.

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

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

Roach, P. (1998). Some languages are spoken more quickly than others. In L. Bauer & P. Trudgill (Eds.). *Language Myths* (150-159). London: Penguin.

Segui, J., Dupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification and lexical access. In G. Altmann (Ed.). *Cognitive Models of Speech Processing* (263-280), Cambridge Mass: MIT Press.

Stevens, K. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of Acoustical Society of America 111*(4), 1872-1891.

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

Taylor, D.S. (1981). Non-native speakers and the rhythm of English. *International Review of Applied Linguistics and English Teaching 19(3)*, 219-226.

Tortel, A., & Hirst, D. (2010). Rhythm metrics and the production of L1/L2. In *Proceedings of Speech Prosody 5*, Chicago, USA.

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: what is it used for, and how is it controlled? *Transactions of the Royal Society B* 369, 20130395. DOI: 10.1098/rstb.2013.0395

Van Maastricht, L., Krahmer, E., Swerts, M., & Prieto, P. (in press). Learning direction matters. *Studies in Second Language Acquisition 1–35*. doi: 10.1017/S0272263118000062.

Vicenik, C., & Sundara, M. (2013). The role of intonation in language and dialect discrimination by adults. *Journal of Phonetics 41*, 297–306.

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

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of American Acoustical Society 127(3)*, 1159-1569.

Table 1. Descriptive statistics for the accent ratings assigned to stimuli in different conditions, with implemented patterns of prosodic timing of learners at different proficiency levels.

		Proficiency	*L1 of the learners*	*Accent ratings*			
				mean	**SE**	**SD**	**95% CI**
Originals		Advanced	German learners	5.183	.034	.915	5.12-5.25
		Intermediate		4.425	.036	.975	4.35-4.50
		Beginners		2.935	.041	1.11	2.85-3.02
		Advanced	French learners	5.141	.033	.88	5.08-5.21
		Intermediate		3.833	.042	1.105	3.75-3.92
		Beginners		2.607	.044	1.166	2.52-2.69
Intoned	***Rhythm+rate***	Advanced	German-based	4.574	.042	1.14	4.49-4.66
		Intermediate		4.393	.045	1.205	4.3-4.48
		Beginners		3.999	.044	1.193	3.9-4.08
		Advanced	French-based	4.446	.048	1.298	4.35-4.54
		Intermediate		4.23	.051	1.36	4.13-4.33
		Beginners		3.678	.054	1.441	3.57-3.78
	Rhythm only	Advanced	German-based	4.446	.043	1.152	4.36-4.53
		Intermediate		4.497	.044	1.186	4.41-4.58
		Beginners		4.251	.045	1.209	4.16-4.34
		Advanced	French-based	4.478	.048	1.287	4.38-4.57
		Intermediate		4.318	.048	1.28	4.22-4.41
		Beginners		3.939	.05	1.368	3.84-4.04
	Rate only	Fast	German-based	4.418	.044	1.176	4.33-4.5
		Normal		4.297	.047	1.263	4.2-4.39
		Slow		4.15	.048	1.279	4.06-4.24
		Fast	French-based	4.494	.048	1.298	4.4-4.59
		Normal		4.374	.049	1.309	4.28-4.47
		Slow		4.117	.05	1.342	4.02-4.21
Flat	***Rhythm+rate***	Advanced	German-based	4.54	.042	1.128	4.46-4.62
		Intermediate		4.367	.042	1.138	4.28-4.45
		Beginners		3.975	.047	1.252	3.88-4.07
		Advanced	French-based	4.361	.046	1.232	4.27-4.45
		Intermediate		4.144	.048	1.284	4.05-4.24
		Beginniners		3.574	.053	1.42	3.47-3.68
	Rhythm only	Advanced	German-based	4.504	.043	1.15	4.42-4.59
		Intermediate		4.465	.045	1.21	4.38-4.55
		Beginners		4.219	.045	1.207	4.13-4.31
		Advanced	French-based	4.268	.048	1.281	4.17-4.36
		Intermediate		4.263	.048	1.284	4.17-4.36
		Beginners		3.846	.049	1.313	3.75-3.94
	Rate only	Fast	German-based	4.547	.044	1.181	4.46-4.63
		Normal		4.392	.044	1.173	4.31-4.48
		Slow		4.161	.047	1.249	4.07-4.25
		Fast	French-based	4.4	.05	1.344	4.3-4.5
		Normal		4.28	.051	1.373	4.18-4.38
		Slow		3.969	.051	1.381	3.87-4.07

Table 2. Statistic data for the effect of *condition* and *proficiency* on the *ratings* (significant results remain significant after correction by the Bonferroni method).

analysis/contrast		**statistic**	**German intoned**	**French intoned**	**German flat**	**French flat**
Condition		F(2,718)	7.312	30.851	5.270	19.83
		p	.001	<.0005	.005	<.0005
		η_p^2	.02	.079	.014	.052
Proficiency		F(2,718)	130.293	250.117	.148.63	239.475
		p	<.0005	<.0005	<.0005	<.0005
		η_p^2	.266	.411	.293	.4
condition*proficiency		F(4,716)	20.599	18.409	8.131	18.884
		p	<.0005	<.0005	<.0005	<.0005
		η_p^2	.103	.093	.043	.095
proficiency **contrasts**	***adv.-inter.***	p	<.0005	<.0005	<.0005	<.0005
	inter.-begin.	p	<.0005	<.0005	<.0005	<.0005
condition **contrasts**	***Rate only***	p	.007	<.0005	.397	<.0005
	Rhythm only	p	<.0005	.493	.008	.896

Table 3. Statistical data for the effect of *proficiency* on *accent ratings* for flat and intoned stimuli in different *conditions* (significant results remain significant after correction by the Bonferroni method).

	Rhythm only flat		Rate only flat		Rhythm+rate flat		Rhythm only intoned		Rate only intoned		Rhythm+rate intoned	
	German	**French**	**German**	**French**	**German**	**French**	**German**	**French**	**German**	**French**	**German**	**French**
F(2,718)	29.06	78.19	55.58	84.43	95.34	178.91	21.61	95.93	31.63	67.24	111.49	184.54
P	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005	<.0005
η_p^2	.057	.179	.134	.19	.210	.333	.075	.211	.081	.158	.237	.340

FIGURE CAPTIONS

Figure 1. Creating the stimuli in *rate only* condition.

Figure 2. Creating the stimuli in *rhythm+rate* condition.

Figure 3. Creating the stimuli in *rhythm only* condition.

Figure 4. Naturalness and comprehensibility ratings assigned to the stimuli in different conditions. Error bars ±2SE.

Figure 5. Ratings assigned to the original utterances produced by learners of English at different proficiency levels. Error bars ±2SE.

Figure 6. Ratings assigned to the stimuli with rhythmic and temporal patterns of French learners of L2 English at different proficiency levels. Error bars ±2SE.

Figure 7. Ratings assigned to the stimuli with rhythmic and temporal patterns of German learners of L2 English at different proficiency levels. Error bars ±2SE.

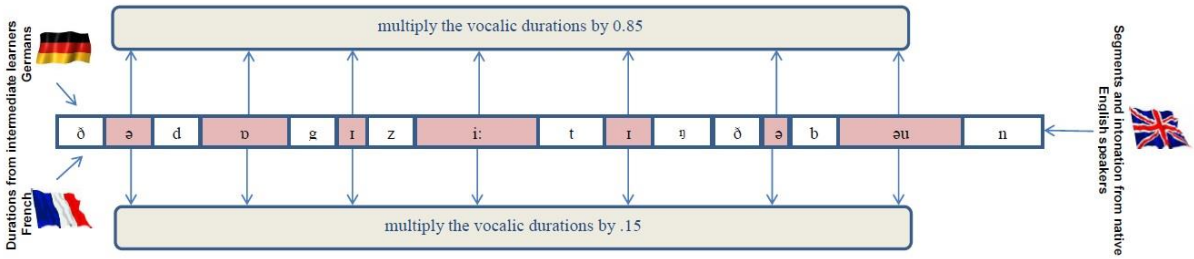

Figure 1. Creating the stimuli in *rate only* condition.

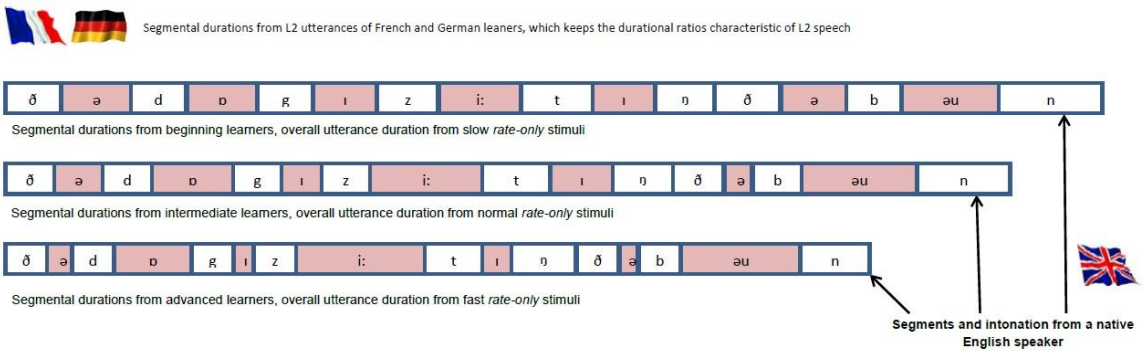

Figure 2. Creating the stimuli in *rhythm+rate* condition.

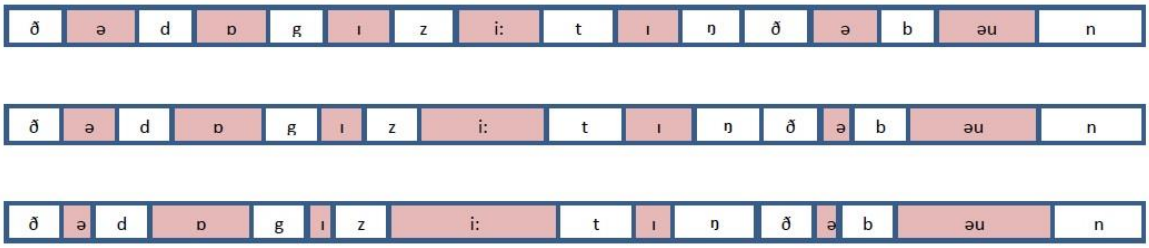

Figure 3. Creating the stimuli in *rhythm only* condition.

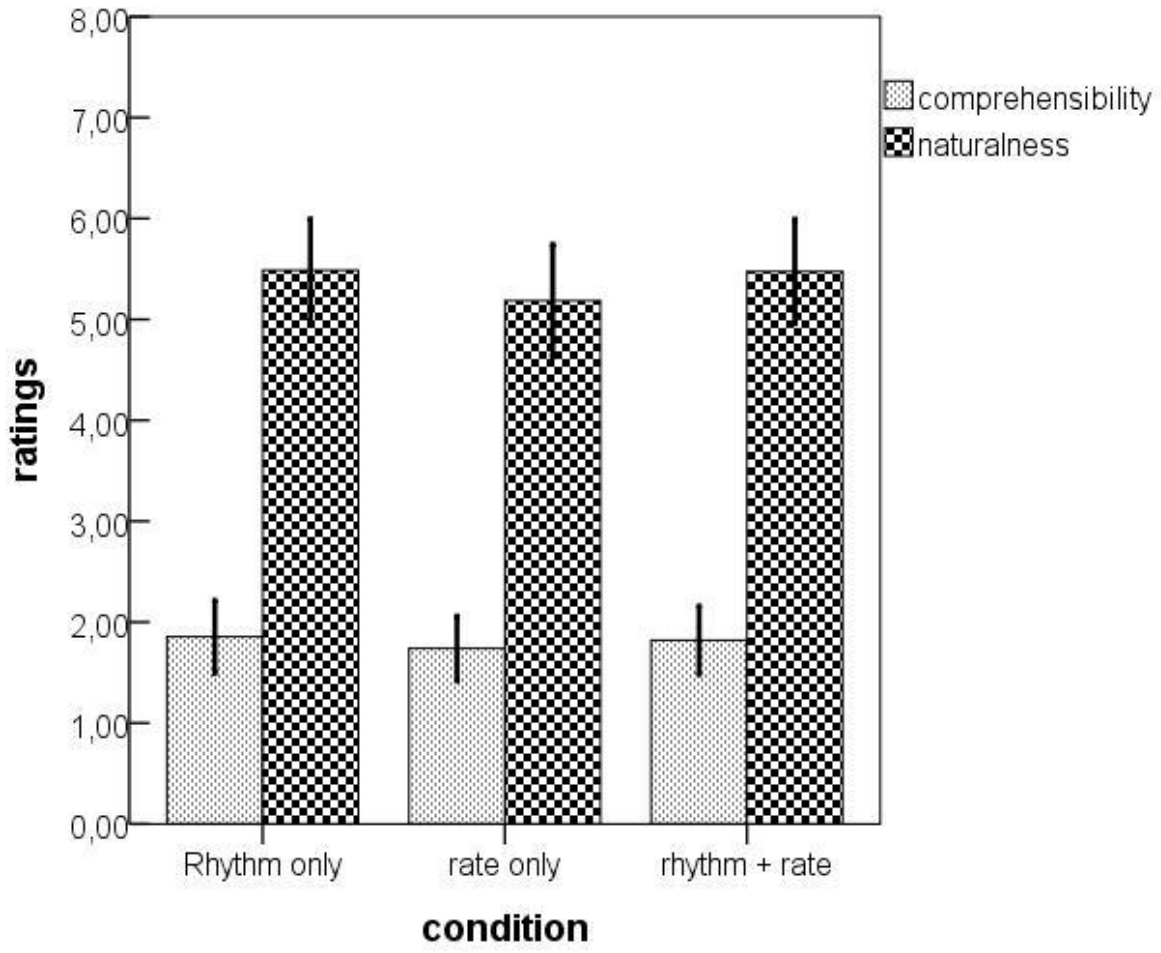

Figure 4. Naturalness and comprehensibility ratings assigned to the stimuli in different conditions. Error bars ±2SE.

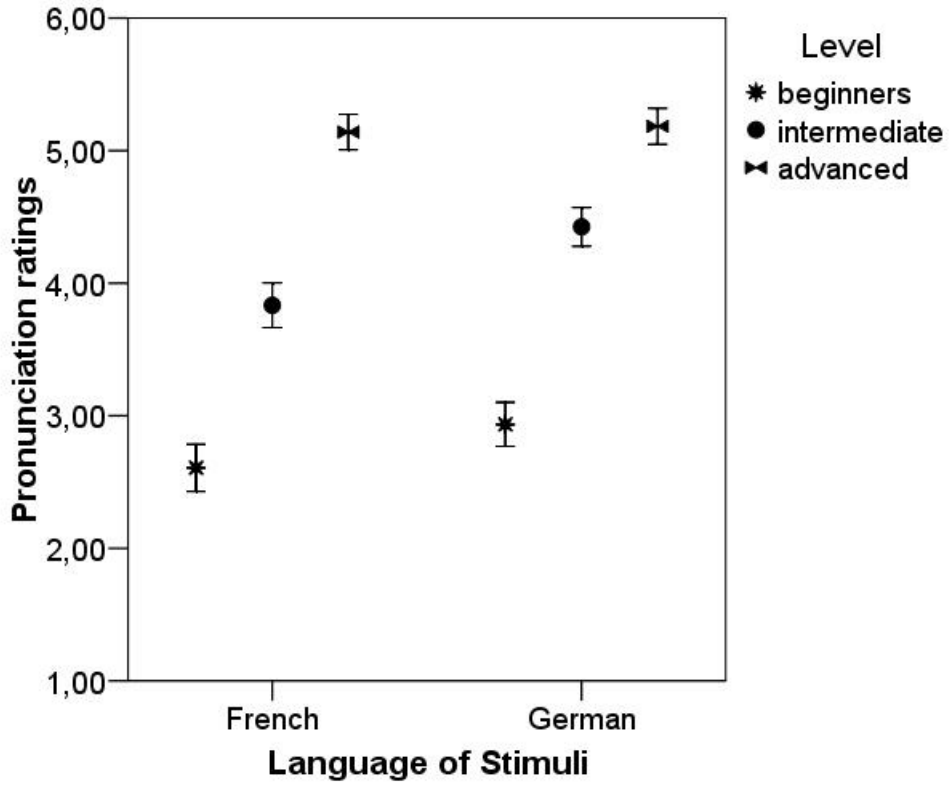

Figure 5. Ratings assigned to the original utterances produced by learners of English at different proficiency levels. Error bars ±2SE.

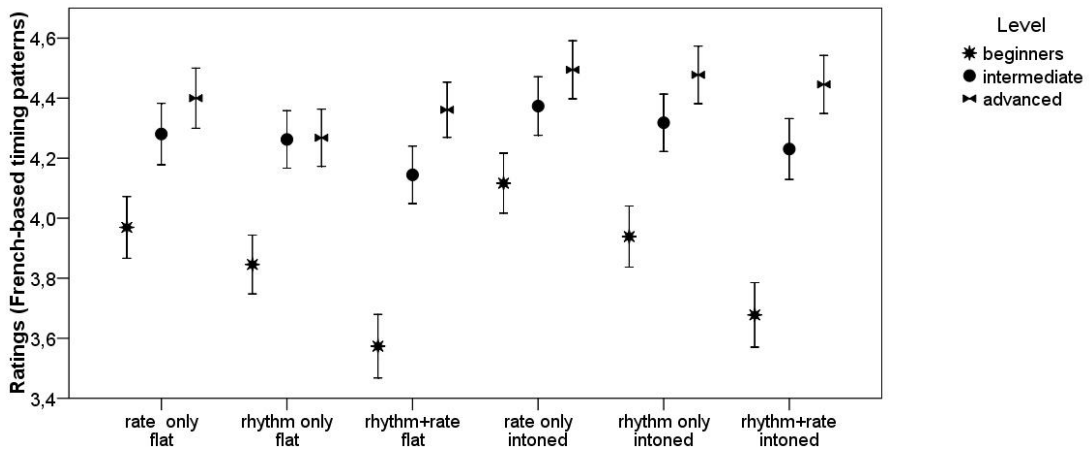

Figure 6. Ratings assigned to the stimuli with rhythmic and temporal patterns of French learners of L2 English at different proficiency levels. Error bars ±2SE.

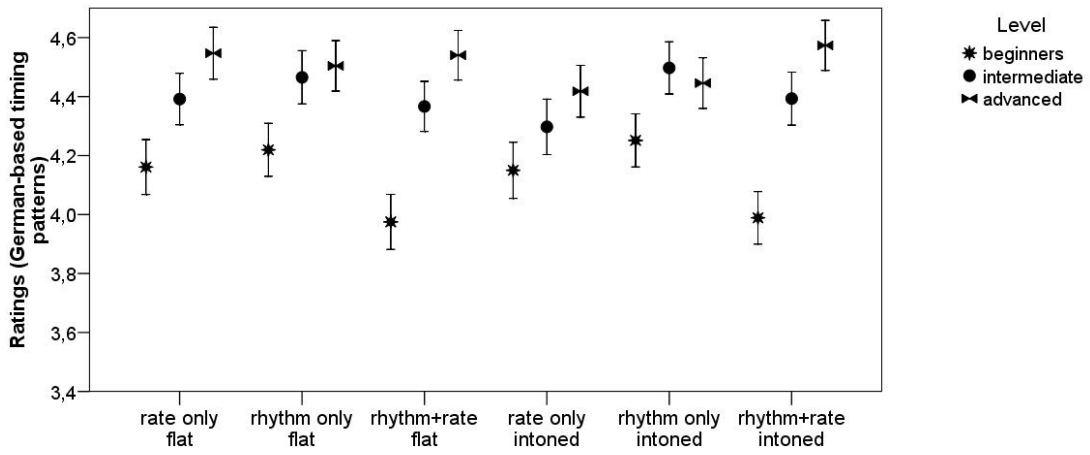

Figure 7. Ratings assigned to the stimuli with rhythmic and temporal patterns of German learners of L2 English at different proficiency levels. Error bars ±2SE.