

Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language

Language and Speech

1–23

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830916648720

las.sagepub.com



Leona Polyanskaya*

Università degli Studi di Padova, Dipartimento di Studi Linguistici e Letterari (DiSL), Italy

Mikhail Ordin*

Basque Center on Cognition, Brain and Language (BCBL), IKERBASQUE – Basque Foundation for Science, Spain

Maria Grazia Busa

Università degli Studi di Padova, Dipartimento di Studi Linguistici e Letterari (DiSL), Italy

Abstract

We investigated the independent contribution of speech rate and speech rhythm to perceived foreign accent. To address this issue we used a resynthesis technique that allows neutralizing segmental and tonal idiosyncrasies between identical sentences produced by French learners of English at different proficiency levels and maintaining the idiosyncrasies pertaining to prosodic timing patterns. We created stimuli that (1) preserved the idiosyncrasies in speech rhythm while controlling for the differences in speech rate between the utterances; (2) preserved the idiosyncrasies in speech rate while controlling for the differences in speech rhythm between the utterances; and (3) preserved the idiosyncrasies both in speech rate and speech rhythm. All the stimuli were created in intoned (with imposed intonational contour) and flat (with monotonized, constant F0) conditions. The original and the resynthesized sentences were rated by native speakers of English for degree of foreign accent. We found that both speech rate and speech rhythm influence the degree of perceived foreign accent, but the effect of speech rhythm is larger than that of speech rate. We also found that intonation enhances the perception of fine differences in rhythmic patterns but reduces the perceptual salience of fine differences in speech rate.

Keywords

Foreign accent, second language acquisition, prosodic timing, rhythm, speech rate

*The first and the second authors acted as principle investigators and made equal contribution to this work.

Corresponding author:

Mikhail Ordin, Basque Center on Cognition, Brain and Language (BCBL), Paseo Mikeletegi 69, Donostia, 20009, Spain.

Email: m.ordin@bcbl.eu

Introduction

Speech produced by individuals who have acquired a certain language from birth (first language (L1) speakers) differs from the speech of those who have been exposed to and learnt the same language after puberty (second language (L2) speakers). L2 speech usually exhibits a certain degree of foreign accent that is generally easily perceived by L1 speakers. Perceived foreign accent results from the production of specific segmental and prosodic characteristics that deviate from those produced by L1 speakers. Research in the production and perception of accentedness has focused on two questions: (1) Do segmental and prosodic characteristics independently influence perceived foreign accent, or do segmental and prosodic deviations from the L1 norms interact in the perception of L2 speech and, in the latter case, what is the nature of this interaction? (2) What is the relative contribution of segments and prosody, and of separate prosodic systems (i.e., stress, intonation, timing, phrasing, etc.) to perceived accentedness (e.g., Boula de Mareüil & Vieru-Dimulescu, 2006; Brahim, Boula de Mareüil, & Gendrot, 2004; Jilka, 2000; Magen, 1998; Munro, 1995; Sereno, Lammers, & Jongman, 2014; Ulbrich & Mennen, 2015, among many others). However, these issues lack many answers and remain in need of further research. In our study we wanted to concentrate on the influence of prosodic timing into perceived foreign accent. In particular, we are interested in whether differences in speech rhythm between the utterances of L2 learners at different proficiency levels affect the perception of accentedness, when speech rate is controlled for. Provided that speech rhythm and speech rate interact both in speech production and perception (Dellwo, 2006; Dellwo & Wagner, 2003), controlling for speech rate is necessary. As far as we are aware, no study has ever tried to disentangle the effects of speech rate and speech rhythm and to estimate their relative contribution to perceived accentedness. Yet, this issue is of interest not only for basic science, but also has added value in applied research related to language pedagogy, training and testing pronunciation mastery (Adams, 1979; Kang, 2010; Kang, Rubin, & Pickering, 2010; Munro & Derwing, 1998; Tajima, Port, & Dalby, 1997; Taylor, 1981), forensic aspects related to speaker and language identification (e.g., Kolly & Dellwo, 2014; Vieru, Boula de Mareüil, & Adda-Decker, 2011), speech technologies for automatic L2 speech recognition and automatic identification of language and dialect (Roach, Sergeant, & Miller, 1992; Rouas, Farina, Pellegrino, & Andre-Obrecht, 2005). We also wanted to investigate the interaction of prosodic timing patterns with other prosodic systems and, in particular, with intonation. Earlier research showed that the effect of segments and different prosodic systems in L2 speech perception is not additive, but interactive (e.g., Ulbrich & Mennen, 2015), and we aimed to explore a possible interplay between rhythmic patterns and intonation in the perception of accented speech.

To address these questions, we synthesized sentences with native British English segments (and intonation) and with rhythmic patterns of L2 English produced by French learners at different proficiency levels. Then we asked native British English listeners to assess the degree of foreign accent in the original sentences produced by French learners and in the resynthesized sentences with non-native prosodic timing patterns. We used the MBROLA speech synthesis software because it allows synthesizing utterances using the diphone database of native British English speakers and controlling for the parameters of duration and F0 fluctuations with a high level of precision. We could manipulate speech rate, speech rhythm and intonational contours independently.

The rest of this introduction has the following structure. We first provide an overview of different factors that affect L2 speech and create the auditory impression of foreign accent, and review discussions on what makes a greater contribution into foreign accent: prosody or segments, as well as what prosodic systems affect the degree of accentedness in L2 speech. Then we discuss the contribution of prosodic timing patterns to accentedness. The final part of the introduction is focused on the cross-linguistic differences in speech rate and speech rhythm, and on how these

timing patterns develop in the course of L2 acquisition. Then we report experiments aimed to investigate how these rhythm and rate characteristics of L2 speech affect perceived accentedness.

1.1 Contribution of segments and prosody to foreign accent

Most early studies on perception of accented speech focused on deviations in L2 speech from the native norms at the segmental level. For instance, Caramazza, Yeni-Komshian, Zurif, and Carbone (1973), Flege and Eefting (1987) and Port and Mitleb (1983), among others, found voicing contrasts and voice onset time (VOT) to be different in L1 and L2 speech. Flege, Schirru, and MacKay (2003), Kehoe (2002) and Flege, Bohn, and Jang (1997) investigated the differences between L1 and L2 vowel production. Flege, Takagi, and Mann (1995) investigated the differences in production and perception of liquid sounds in L1 and L2 speech. The deviations in L2 speech from native norms leading to foreign accent were found in each study, with the degree of accentedness proportional to the magnitude of the acoustic difference between the realizations of L1 and L2 segments. These studies led to the development and corroboration of two main L2 speech models: the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995). Both models are based on the assumption that listeners map L2 phonemes onto L1 phonemes, and thus fail to perceive the minute distinctions, or fine phonetic details, to which L1 users attend to. The failure to perceive the fine phonetic details that are important to L1 users results in production difficulties, which create the foreign accent effect. Neither of the L2 learning models, however, accounts for prosodic aspects in the perception and production of accented speech. The same reservations also hold true for important L1 acquisition models. For example, Kuhl's Native Magnet Model of language acquisition (1991) is based on phoneme-sized vocalic linguistic units and ignores development of language-specific prosodic characteristics in the course of L1 acquisition.

Nevertheless, Derwing and Munro (1997), Derwing and Rossiter (2003), Field (2005) and others showed that improvement in prosodic proficiency rather than correction of segmental errors will result in reduced perceived accentedness. The existing body of research showed that prosody makes a unique contribution to perceived foreign accent, independently of the contribution made by segmental errors and by the interaction of segmental and prosodic features in L2 speech. Boula de Mareüil and Vieru-Dimulescu (2006) used prosody transplantation between Italian and Spanish L1 and L2 speech to create stimuli with L1 prosody and L2 segments, and L2 prosody and L1 segments. They concluded that prosody is important in identifying foreign accent in Spanish and Italian. Moreover, they tentatively suggested a higher ranking of prosody over segmental realization in creating the effect of foreign-accented speech. Anderson-Hsieh, Johnson, and Koehler (1992) used correlational methods and pronunciation ratings given to 60 non-native speakers, and also concluded that prosody makes a greater contribution to accentedness than segments. Tajima, Port, and Dalby (1997), Maassen and Povel (1985), Ulbrich and Mennen (2015) and Sereno, Lammers, and Jongmann (2014), on the other hand, carried out a series of experiments and determined a much larger contribution of segmental features to foreign accent compared to suprasegmentals.

As the issue of the relative effect of prosody and segments on perceived foreign accent remains open, researchers have decided to concentrate on the interaction of suprasegmental and segmental features (Ulbrich & Mennen, 2014); on the contribution of suprasegmental deviances in L2 speech from native norms related to separate prosodic systems; and on the interplay between different prosodic systems in creating the percept of foreign accent (e.g., Kang, 2010). Suprasegmental deviances can be related to stress assignment and stress realization, intonation, alignment of F0 peaks with metrically strong syllables, timing patterns pertaining to speech rhythm and speech rate, pause frequency and duration, etc. (Anderson-Hsieh et al., 1992; Koster & Koet, 1993; Trofimovich & Baker, 2006, among many others).

Bond (1999), Hahn (2004) and Field (2005) studied the realization of primary stress in L2 speech and its influence on degree of perceived accentedness and on intelligibility of accented speech. L2 speech was evaluated more favourably when primary stress was correctly placed, while phonetic details of stress realization contributed to accentedness less than stress assignment. Bond (1999) even claimed that the contribution of stress to foreign accent in English is greater than that of segmental errors.

Mennen (2004) studied intonation in Dutch-Greek speakers and detected noticeable differences in the realization of intonational patterns in L1 Greek and L2 Greek produced by Dutch learners. Lepetit (1989) detected realizational differences in L2 French produced by Japanese and English learners, and suggested that these differences may affect accentedness. Jilka (2000) examined L2 English and German speech produced by either L1 English or L1 German learners. He aimed to disentangle the contribution of intonation, prosody (excluding melody but including rhythmic, fluency and timing, intensity factors) and segmental characteristics into degree of accentedness. He found that the major contribution to accentedness was the improper distribution of pitch accents in L2 speech. Native listeners were exposed to (1) low-pass filtered, (2) monotonized and (3) non-manipulated original L2 speech, and were asked to identify whether it was L1 or L2 speech, and to give foreign accent ratings to the utterances. The results showed that listeners were able to identify the speaker's language on the basis of purely prosodic information, and that intonation was the major factor contributing to the degree of perceived accentedness.

Kang et al. (2010) investigated a range of suprasegmental features of different prosodic systems, and their relative contribution to accentedness and judgements of the learner's proficiency in L2 English. Sixty-second recordings of 26 male learners from various L1 backgrounds (Arabic, Korean, Spanish, Chinese) were judged for oral proficiency by 188 English-speaking undergraduates. Twenty-nine acoustic variables, including measures of speech rate, pause, stress, melody, etc., were selected for analysis. Multiple regression was used to assess the contribution of each variable and of different clusters of variables to accentedness and perceived proficiency in L2 English. The authors found that prosody in general accounts for 50% of accentedness ratings, and the most influential predictor of accentedness degree was suprasegmental fluency (the joint factor of interacting prosodic features of speech rate, articulation rate, speech-to-pause ratio, length of speech units between boundaries), followed by the right choice of pitch accents (especially rising tones) and boundary markers (low termination tones and number of silent pauses). A similar study by Kang (2010) yielded similar results: the overall contribution of prosodic characteristics accounted for 41% of variance in accentedness rating in L2 speech. The most substantial contribution to accentedness ratings was made by pitch span (L2 speech with a wider span was perceived as less accented by L1 English speakers), followed by the ratio of stressed words to non-stressed words, mean length of silent pauses, ratio of atypical boundary pauses and articulation rate. It should be noted that in neither study (Kang, 2010; Kang et al., 2010) did the authors include durational cues in their analysis (except for fluency measures, e.g., speech rate measures). Fewer studies have been done to estimate the contribution of prosodic timing patterns to perceived foreign accent, and we are providing the overview of these studies in the following section.

1.2 Contribution of prosodic timing patterns into foreign accent

Very few studies have paid attention to the independent contribution – not overlapping with the contribution of other prosodic or segmental features of speech – of timing to perceived foreign accent. Speech rate (measured in syllables per second) and articulation rate (measured in syllables per second excluding pauses) are among the most widely studied temporal aspects of L2 speech. L2 speech was found to be slower than L1 speech (Guion, Flege, Liu, & Yeni-Komshian, 2000;

Lennon, 1990; Munro & Derwing, 1998). For example, the mean articulation rate for L1 users of English is 5 syllables per second, the mean rate for highly proficient L2 English speakers is 4.4 syllables per second and for intermediate-proficient group is 3.3 syllables per second (Anderson-Hsieh and Venkatagiri, 1994). Some studies, however, did not reveal significant correlations between speech rate and perceived foreign accent (Anderson-Hsieh & Koehler, 1988), while others showed a significant contribution of tempo characteristics to accentedness (Kang, 2010). This inconsistency in results obtained by different researchers can be attributed to the correlational nature of the studies. The researchers investigated natural speech that contained not only the differences in speech rate, but also differences in other prosodic and segmental characteristics, and deviances in speech rate in certain speech samples could have been overshadowed by the deviances in segmental and suprasegmental characteristics – these may have given a greater contribution to foreign accent and thus may have diminished the effect and the perceptibility of rate differences. To overcome this disadvantage of correlational studies, Munro and Derwing (2001) examined the effect of speaking rate on accentedness in experimental conditions. They found that, by and large, speech delivered at a faster rate was judged as less accented, up to a certain threshold. Very fast (over 5 syl/s) as well as slow (2 syl/s) speech was rated as more accented than speech delivered at “optimal” rate for L2 learners (4.76 syl/s).

Differences in speech rate between L2 and L1 speakers cannot be attributed to perceptual assimilation, or to transfer and interference between L1 and L2, and thus they do not fit the existing speech learning and acquisition models, which are based on phoneme-sized units (Best, 1995; Flege, 1995; Kuhl, 1991). L2 speakers deliver speech at a slower rate, probably due to slower lexical access or less established articulatory movements involved in the production of non-native sounds and clusters. These factors are not accounted for in either of the two L2 speech production models. Besides, it is not fully understood to what extent tempo differences between utterances produced by L2 learners at different proficiency levels and L1 speakers influence accentedness. Munro and Derwing (2001) clearly showed a significant effect of speech rate on perceived foreign accent; however, the magnitude of this effect relative to the effect of other segmental and suprasegmental factors remains unknown.

Durational ratios of vocalic, consonantal, syllabic, word and foot durations also differ between languages and between L1 and L2 speech (e.g., Low, Grabe, & Nolan, 2000; Ramus, Nespors, & Mehler, 1999; White and Mattys, 2007, among others) and thus may contribute to the degree of perceived accentedness. Several studies have addressed this issue. Tajima et al. (1997) manipulated segmental durations in L1 English and in L2 English produced by Mandarin speakers. They concluded that intelligibility of Chinese-accented speech improved by 15–25% when phonemic durations were warped to match native temporal patterns. When phonemic durations were corrected according to the non-native patterns, the intelligibility of L1 English deteriorated by 15%. The influence of durational ratios on the degree of perceived accentedness was studied by Quene and Delft (2010). They manipulated the segmental durations of L1 and L2 Dutch sentences to create four types of stimuli: native segments with native durations; native segments with non-native durations; non-native segments with native durations; and non-native segments with non-native durations. The effects of speaking rate differences between L1 and L2 were neutralized by equalizing the durations using the Praat-embedded PSOLA algorithm, and intonation differences were neutralized by pitch stylization in L1 sentences and further transplantation of the pitch contour onto L2 sentences. Despite normalization for the speech rate, the authors found that non-native durational ratios add to perceived foreign accent.

Baker et al. (2011) analysed word durations in L1 and L2 English produced by Korean, Chinese and American English speakers. They found that L1 speakers produced shorter words (which can be related to faster speech rate in L1), had greater within-speaker variance for word durations

(which can relate to greater reduction of function words) and higher between-speaker variance. They also found that these durational differences are correlated with the degree of perceived foreign accent in L2 speech. The L2 speakers with higher within-speaker variance in duration, greater reduction of function words and shorter words overall were perceived as less accented.

These studies revealed the effect of rhythm (defined as durational variability of speech intervals such as vocalic sequences, consonantal clusters, syllables, feet) on perceived foreign accent, but they did not separate the effect of speech rate and durational ratios, and did not estimate the effect of prosodic timing patterns (pertaining to speech rhythm and rate) compared to the effect of other prosodic and segmental differences between L1 and L2 speech and between L2 speech produced by learners at different proficiency levels.

1.3 Cross-linguistic and developmental differences in speech rate and speech rhythm

Speech rate is usually expressed in syllables per second. Speech rate has been reported to differ between languages.¹ Germanic languages usually have a slower speech rate compared to Romance languages (Anderson-Hsieh & Venkatagiri, 1994; Clopper & Smiljanik, 2011; Ordin and Polyanskaya, 2015b; Pellegrino, Coupé, & Marciso, 2011; Quene, 2005, 2007). Spontaneous speech in Dutch, German and English is produced at an average rate of 5.2 syl/s (the reported mean syllable durations vary between 0.194 and 0.239 msec). Speech in Romance languages is delivered at the faster rate of approximately 6 syl/s (Dauer, 1983; Pellegrino et al., 2011).

More proficient L2 learners of English tend to deliver L2 speech at faster rates compared to less proficient speakers (Anderson-Hsieh and Venkatagiri, 1994; Ordin and Polyanskaya, 2015a, 2015b). Ordin and Polyanskaya (2015b) compared speech rate in the same set of sentences produced by French and German learners of English at different proficiency levels. In both studies the differences in speech rate between proficiency levels and between L1 and L2 speech produced even by advanced learners exceeds the 5–8% just-noticeable threshold for tempo differences (Quene, 2007), and is therefore perceivable.

Durational ratios – timing patterns pertaining to speech rhythm – are captured with the help of the so-called rhythm metrics. Table 1 presents the overview of the most popular metrics, which we also used in this study.

These metrics capture the durational variability of speech intervals that corresponds to the auditory impression of stress- and syllable-timing. Romance languages were shown to have an overall lower degree of durational variability and higher proportion of vocalic material compared to Germanic languages (Bunta & Ingram, 2007; Payne, Post, Astruc, Prieto, & del Mar Vanrell, 2012; Prieto, del Mar Vanrell, Astruc, Payne, & Post, 2012; Ramus et al., 1999; White and Mattys, 2007, among others), which corresponds to a lower degree of stress-timing.

Differences in rhythmic patterns between L1 and L2 speech have also been confirmed in several studies (Bond & Fokes, 1985; Grenon & White, 2008; Ordin & Polyanskaya, 2015a; White & Mattys, 2007). L2 English exhibits a lesser degree of stress-timing compared to L1 English, irrespective of the L1 of the learner.

Ordin and Polyanskaya (2014, 2015a, 2015b) and Tortel and Hirst (2010) investigated changes in durational ratios in the course of acquisition of L1 and L2 English. They clearly showed that durational variability increases as acquisition progresses. As the differences in rhythm between L1 and L2 utterances were reported to impede intelligibility and to affect the degree of perceived foreign accent in English produced by L2 learners, we assume that the developmental differences in speech rhythm between the utterances produced by learners of English at different proficiency levels could also be perceived and contribute to the accentedness of L2 speech. As L2 English

Table 1. Overview of the most widely used rhythm metrics.

Global metrics			Local metrics		
Metrics name	Description	Reference	Metrics name	Description	Reference
Raw metrics					
Δ (delta)	Standard deviation in duration of a certain speech interval, e.g., ΔS (for syllables), ΔV (for vocalic sequences), ΔC (for consonantal clusters).	Ramus, Nespor, & Mehler, 1999	rPVI	Raw pairwise variability index for a certain type of speech intervals (e.g., rPVI-S – for syllables, rPVI-V – for vocalic sequences, rPVI-C – for consonantal clusters).	Grabe & Low, 2002
Normalized metrics					
Varco	Coefficient of variability in duration of a certain speech intervals (VarcoS = $\Delta S/\text{meanS}$; VarcoV = $\Delta V/\text{meanV}$, VarcoC = $\Delta C/\text{meanC}$)	Dellwo, 2006	nPVI-v and nPVI-c	Normalized pairwise variability index for a certain type of speech intervals (e.g., nPVI-S – for syllables, nPVI-V – for vocalic sequences, nPVI-C – for consonantal clusters). Normalization is done by the mean duration of the corresponding interval type, which reduces the influence of rate on the measure of durational variability.	Grabe & Low, 2002
%V	Amount of vocalic material in an utterance in percent	Ramus, Nespor, & Mehler, 1999			

becomes more stress-timed in the course of acquisition and as durational variability in L2 English approaches that in L1 English as acquisition progresses, we assume that the non-native rhythmic patterns of beginners will contribute more to their perceived accentedness level than the L2 rhythmic patterns of advanced learners of English. To test this hypothesis, we examined the independent and combined effects of speech rhythm (durational variability of sequential speech units captured by the rhythm metrics) and speech rate (the number of units produced per second) on perceived accentedness of L2 speech. English utterances were resynthesized so that they mimic the speech rate and/or rhythm characteristics of the corresponding sentences spoken by French learners of English at three different proficiency levels. Native British English participants listened to the stimuli and rated them on a foreign accent scale.

Intonation has been reported to interplay with segmental characteristics of L2 speech in the perception of L2 speech and in contributing to perceived accentedness (Kolly & Dellwo, 2014; Ulbrich & Mennen, 2015). Also, Vicenik and Sundara (2013) and Kolly and Dellwo (2014) showed a complex interaction of F0 contours and timing patterns in language and dialect discrimination and identification. In our study, we wanted to study whether and to what extent intonation interacts with rhythmic and temporal patterns in the perception of accented L2 speech. Therefore, we created stimuli with intonational contour, and with flat F0. If intonation indeed interacts with prosodic timing, we should expect differences in the perception of intoned and flat stimuli with non-native timing patterns. In a series of perception experiments, we tested how speech rate and speech rhythm indeed influence the perception of foreign accent, to estimate the strength of both of these sources of influence, and explored the interaction of melodic and timing prosodic systems in judging the degree of foreign accent in L2 speech.

2 Method

2.1 Speech material

For the perception experiments, we used a corpus of speech samples representing L2 English produced by monolingual French learners of English as a foreign language. This is a small subset of a larger corpus of L2 speech collected and described by Ordin and Polyanskaya (2015a, 2015b). The speech corpus contains the recordings of 48 monolingual French learners of English from the Parisian area. The recordings were done in the Laboratory of Phonetics and Phonology at Sorbonne Nouvelle Paris III University at 48 kHz, 32 bit mono. Recordings were carried out with each speaker individually, and each session lasted approximately 30 minutes. Each recording consisted of two parts: interview and sentence elicitation task. During the first part, the participants were asked 10 questions – the same set of questions for every speaker – related to personal biography, music and reading preferences, educational choices, etc. This interview lasted 10–12 minutes per participant. These interviews were anonymized and subsequently given to three native speakers of English, who were also teachers of English as a foreign language and had experience in language testing.

We asked the native English speakers to listen and evaluate each interview 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 linguistic performance. We calculated the agreement between raters for each parameter using Cronbach's alpha. The values of Cronbach's alpha (<.88 for each parameter) show high consistency between the ratings given to each learner by the three teachers. A high value of Cronbach's alpha does not show absolute agreement, thus we use the word *consistency*. A high alpha shows agreement of a relative kind. The teachers are supposed to agree on which of the learners is more fluent, while the absolute fluency assessments of

individual learners may differ between teachers. Our reliability test shows that the scoring patterns are highly consistent between raters, while the raw ratings from teachers may differ, for example, due to different expectations that may affect the initial reference for comparison. The Cronbach's alpha $< .75$ between parameters for each teacher was also high. This means that the ratings given to each learner are correlated across different parameters, that is, a higher rating assigned to fluency of a certain learner also means higher ratings assigned to grammatical accuracy and vocabulary.

The ratings for each learner were averaged across teachers and across parameters to obtain the overall mean assessment of learners' proficiency. These overall assessments were used to split the learners into three proficiency groups: beginners (with an overall mean rating between 4 and 6), intermediate learners (with an overall mean rating between 6 and 8) and advanced learners (with an overall mean rating above 8).

The second part of the recording included a sentence elicitation task. For this purpose, 33 picture prompts were used. The pictures were presented as separate slides to the participants, together with an accompanying descriptive sentence. The participants were instructed to look at the pictures and to remember the sentences in a self-paced manner, and were allowed to go backward and forward at their own pace. When they said they were ready, the pictures were presented to them without the accompanying sentences, each picture on a separate slide, and the participant was asked to retrieve the sentence from memory and to say it aloud. Ninety-three per cent of the sentences were produced correctly and without hesitations. In 7% of the sentences there was a deviation between the expected and produced sentence (e.g., *The dog is running after the cat* instead of *The dog is chasing the cat*) or the participant could not recall the sentence. In these cases, verbal prompts were provided to elicit the sentence that corresponds in the lexical material and in syntax to the sentence presented during familiarization. For example, the interviewer said: "Yes, that's perfect. You can also use the word *chase* for *run after*. *Chasing* and *Running after* mean the same. Could you please rephrase the sentence with the word *to chase* and say it again?" One verbal prompt was sufficient to elicit the expected sentence in all cases. This procedure allowed one to record lexically and grammatically identical sentences from all the learners who were at different proficiency levels in L2, and at the same time to avoid a reading mode.

A subset of 15 randomly chosen sentences was used as speech material for the perception study reported here. In the material, each of the 15 chosen sentences was produced by a learner at advanced, intermediate and beginning levels of proficiency. This resulted in 45 utterances in total, serving as raw material for the perception experiments. We will use the word *utterance* to refer to the actual production of a certain sentence by a learner, thus the raw material includes 45 utterances of 15 sentences.

The selected utterances were annotated in Praat (Boersma, 2001). Each utterance was divided into consonantal (C) and vocalic (V) intervals and into syllables (S). Segmentation was carried out manually, based on the criteria outlined by Peterson and Lehiste (1960) and Stevens (2002) for C and V intervals. Syllabification was done following Wells (2008). Adjacent vowels and consonants were combined into the same V and C intervals, even when the consonantal cluster or a sequence of vocalic segments straddles the syllabic boundary. Pauses and hesitations within utterances were excluded when calculating rhythm measures, but vowels and consonants on either side of the pause were not combined into the same interval.

Traditional rhythm metrics that capture durational variability of speech intervals were calculated on each utterance. We used rate-normalized rhythm metrics (nPVI, Varco) because they capture durational variability of speech intervals (i.e., rhythmic patterns) independently of the mean duration of these intervals (i.e., independently of speech rate). Thus, if we modify speech rate in the process of stimuli preparation, the rhythmic characteristics captured by these rhythm metrics will remain intact. The metrics were calculated on V, C and S intervals, and further compared to

confirm that the rhythmic patterns indeed differ between sentences produced by L2 learners at different proficiency levels.

2.2 Stimuli preparation

To estimate the contribution of the differences in speech rate to the degree of accentedness while controlling for speech rhythm, and to estimate the contribution of speech rhythm to perceived foreign accent while controlling for differences in speech rate, it was necessary to disentangle these closely related timing phenomena. Thus we prepared four sets of stimuli:

The first set of stimuli comprised the 45 original utterances, that is, 15 sentences, each produced by a beginner, an intermediate and an advanced learner of English. The three versions of these sentences differed both at segmental as well as prosodic levels, that is, in the realization of phonemes, intonation, stress, speech rhythm, speech rate, etc.

The selected 45 utterances were used to create further stimuli for the perception experiment. Praat was used to measure the durations of phoneme realizations (consonants and vowels). These durations were then used to resynthesize the utterances using the MBROLA speech synthesis system (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) with the native British English diphone database (en1). In order to test the expected interplay between rhythmic patterns and intonation in the perception of accented speech, the utterances were resynthesized in two versions. Firstly, 45 utterances were resynthesized with a flat F0 set to 115 Hz (average F0 for male European voices). Secondly, we created 45 utterances with intonation (intoned stimuli). For this, a male native British English speaker pronounced the chosen 15 sentences. The configurations of these F0 contours were imitated in three resynthesized versions of the same sentence that were based on the segmental durations of a beginner, an intermediate and an advanced learner.

In addition, we took the resynthesized utterances based on the segmental durations of intermediate learners, and increased the overall duration of each utterance by 10% to create slow utterances, and decreased the overall utterance duration by 10% to create fast utterances in a linear fashion (Figure 1, second set of stimuli). As a result, three versions of the same sentence differed only in speech rate (i.e., mean duration of syllables). Differences in phonemic realizations were neutralized because the same diphone database was used for resynthesis. Differences in intonation did not exist because either the set F0 contour was imposed on each version of the sentence, or the utterances were resynthesized with flat intonation. The rhythmic patterns were identical because rate-normalized durational variability measures did not differ between the versions of the same sentence. The 10% manipulation was chosen because it slightly exceeds the 5% just noticeable difference in tempo for linguistic stimuli (Quene, 2007), and we can be sure that the 10% acceleration or deceleration are perceptually salient for the human auditory system. Thus, if we indeed observe the differences in accent ratings between faster and slower versions of the same sentence, these differences in the degree of perceived accentedness can be attributed to differences in speech rate.

To create the third set of stimuli, we modified the overall duration of the resynthesized utterances with segmental durations from the advanced and intermediate learners and from the beginners. We made the overall duration of the resynthesized utterances with the segmental durations from the advanced learners equal to the overall duration of the corresponding fast utterances. The overall duration of the resynthesized utterances with segmental durations from the beginners was made equal to the overall duration of slow utterances. The overall duration of the resynthesized utterances with segmental durations from the intermediate learners was not modified and was left equal to the overall duration of the utterances at normal rate (Figure 2). The manipulation of the overall duration did not influence the rhythmic patterns of the resynthesized sentences because when the whole utterance is stretched or compressed, all segments, syllables, C and V intervals are

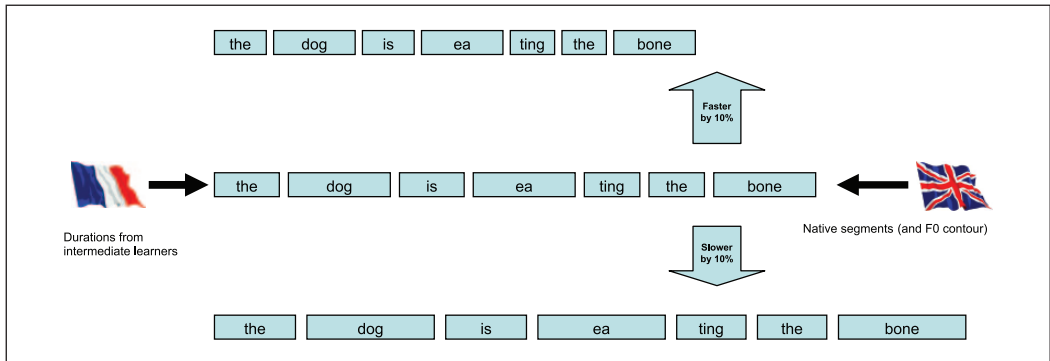


Figure 1. Creating the utterances that differed only in speech rate – second set of stimuli.

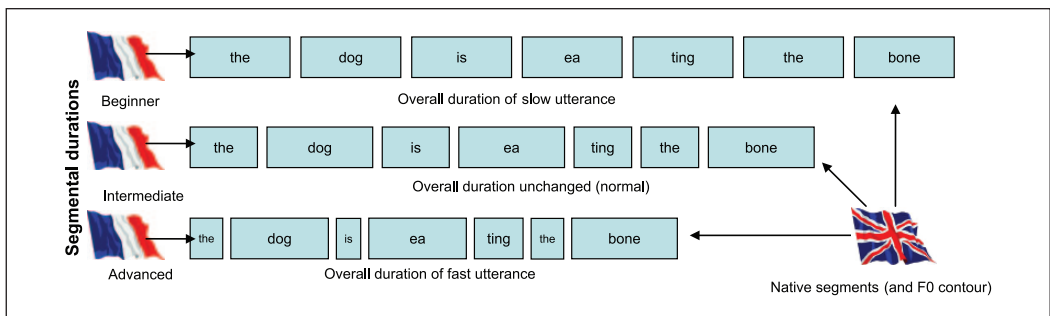


Figure 2. Creating the utterances that differed only in speech rate and speech rhythm – third set of stimuli.

stretched or compressed proportionally, and the durational variability of speech intervals captured by the rate-normalized rhythm metrics is not affected. As a result of such manipulations, we obtained three versions of the same sentence that were different in speech rate and in speech rhythm. Again, we created flat and intoned utterances.

To create the fourth set of stimuli (again, both intoned and flat versions), we equalized the overall utterance durations of three versions of the same sentence (Figure 3), and the resulting utterances differed only in rhythmic patterns (i.e., durational variability characteristic of L2 English produced by French learners at beginning, intermediate and advanced levels), but not in segmental realizations, intonation or speech rate.

In the end, we had four types of stimuli (cf. supplementary material):

- I) 45 original utterances, 15 sentences, each produced by a French learner of English at beginner, intermediate or advanced proficiency level (*Originals*);
- II) 45 intoned and 45 flat resynthesized utterances, three intoned and three flat versions of each of the 15 sentences that differed only in speech rate (*Rate Only*);
- III) 45 intoned and 45 flat resynthesized utterances, three intoned and three flat versions of each of the 15 sentences that differed only in speech rhythm and speech rate (*Rhythm and Rate*);
- IV) 45 intoned and 45 flat resynthesized utterances, three intoned and three flat versions of each of the 15 sentences that differed only in speech rhythm (*Rhythm Only*).

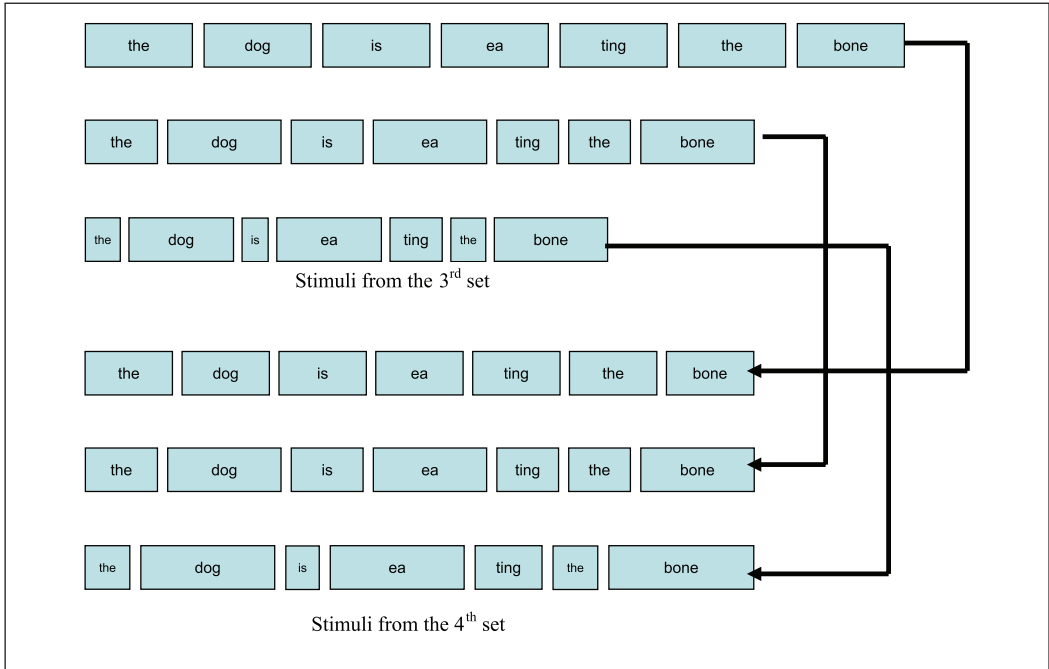


Figure 3. Creating the utterances that differed only in speech rhythm – fourth set of stimuli.

These stimuli were used for the perception experiments to verify the effects of differences in speech rate and speech rhythm on the degree of perceived foreign accent, and to estimate the relative contribution of speech rate and speech rhythm into the degree of accentedness.

2.3 Procedure

We recruited 14 monolingual British English native speakers to participate in the experiment. The participants did not speak French and were not fluent in any other foreign language. Each participant took part in four different sessions. The minimum interval between each session was 14 days. In each session, the participants listened to stimuli of a certain type, first flat, then intoned (except for the original utterances, we did not have monotonized versions of the original utterances). The order in which the different types of stimuli were presented in different sessions was randomized. The listener had to listen to the stimuli, one by one, and evaluate the degree of the perceived foreign accent on a six-point scale, from 1 (strongest accent) to 6 (native-like). The degrees in-between were verbalized as 5 (mild accent), 4 (moderate accent), 3 (rather strong accent) and 2 (strong accent). The stimuli were presented to the participants via headphones connected to the computer. The participants had to listen to the stimulus and press one of the six buttons on the screen with the mouse to rate the degree of foreign accent. The participants could replay the stimulus two more times after the initial presentation by clicking a *replay* button. After the degree of accentedness was evaluated by clicking one of the buttons, a new stimulus was played. The average duration of one session was either 20 minutes (for the original sentences) or 40 minutes (for the resynthesized stimuli in flat and intoned conditions).

Forty-five utterances of each type were presented three times in blocks. The order of utterances was randomized within each block. The participants were not informed that the same stimulus

Table 2. Assessment of reliability and consistency in foreign accent rating between and within listeners.

Stimulus type	Between-listeners consistency		Within-listeners agreement
	Cr. alpha (2nd block), N = 14	Cr. alpha (3rd block), N = 14	Guttman Split-Half coeff.
Type I: Originals	.951	.922	.968
Type II: Rate Only	.805	.805	.916
Type III: Rhythm and Rate	.78	.833	.881
Type IV: Rhythm Only	.794	.801	.889

will be presented several times, and there were no pauses or experiment interruptions between the blocks. Thus, the total number of each type of stimuli the listeners had to evaluate was 135 (for the originals) and 270 (for the resynthesized utterances, 135 flat stimuli followed by 135 intoned stimuli).

2.4 Assessment of reliability and consistency of ratings

The first block of 45 responses presented for the first time within each session was considered the familiarization phase and the responses were not included in further analysis. We assumed that during familiarization, as the listeners gain experience with the stimuli, they build up the reference line to rate the degree of foreign accent in the subsequent sentences.

The ratings given to the same stimulus when it was presented in the second and in the third blocks were used to evaluate within-listener agreement to make sure that the same stimulus will receive the same (or comparable) rating when it is evaluated several times. For this, we are more concerned about absolute agreement. We were also interested in agreement between listeners, that is, that the utterance that is rated as more accented by one listener will also be rated as more accented by the other listeners. We were more concerned about the relative consistency of rating patterns between listeners than about their absolute agreement. One of the listeners might be a stricter or a more tolerant listener than the others, which should affect the mean ratings, but should not affect the consistency of the rating pattern (that is, if utterance *A* is rated lower than utterance *B* by one listener, it will also receive the lower ratings from other listeners, although a more tolerant listener might give a higher rating to utterance *A* than a stricter listener).

To estimate the within-listeners consistency, we used a Guttman Split-Half test. We considered the second and the third blocks as the same test administered to the same participants twice, which means we can estimate true test–retest reliability. The test reveals very high within-listener reliability (Table 2), that is, each listener tends to give the same or very similar ratings to the same stimulus, when the stimulus is presented several times for evaluation. To estimate the between-listeners consistency, we calculated the Cronbach's alpha separately for the second and the third block. High alpha values (Table 2) indicate high consistence in rating between participants. High between- and within-listeners reliability justified using the obtained ratings for further analysis on how speech rhythm and speech rate affect the degree of perceived foreign accent.

3 Results

Firstly, we decided to compare rhythm measures in the selected sentences in order to confirm that rhythmic patterns indeed differ between productions of L2 learners at different proficiency levels.

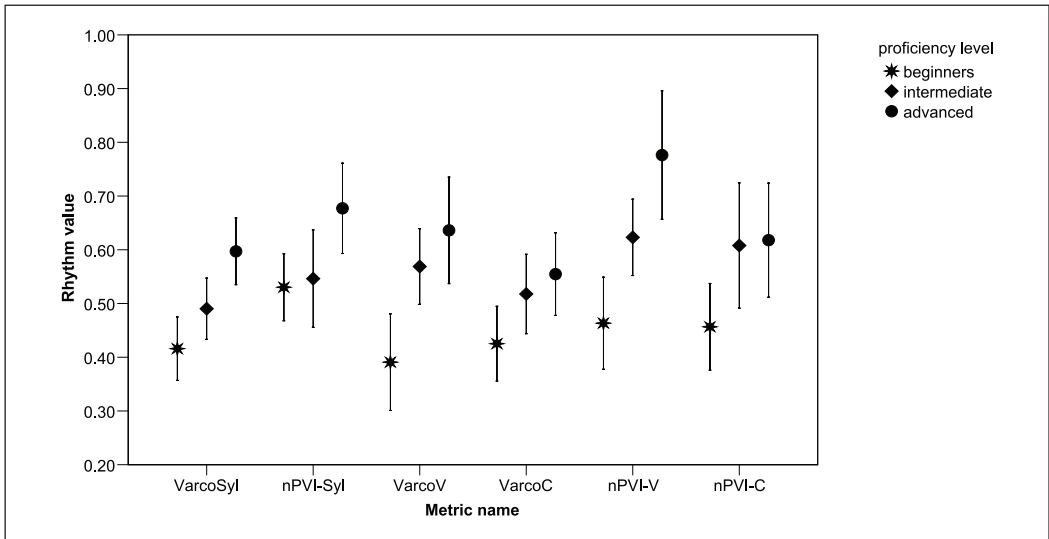


Figure 4. Mean scores (± 2 SE) of the rate-normalized metrics calculated on the utterances produced by French learners of English at different proficiency levels.

The metrics show that the durational variability is higher in the utterances produced by advanced learners of English and lower in the utterances produced by beginners (Figure 4).

To check whether the differences between speech rhythm in utterances from learners at different proficiency levels actually exist, we performed a one-way multivariate analysis of variance (MANOVA) for each rhythm metric calculated on utterances produced by beginners, intermediate and advanced learners. The model was significant, $\Lambda = .398$, $F(12, 74) = 3.608$, $p < .0005$, $\eta^2 = .369$, which means that the proficiency level indeed has a significant effect on the values of the rhythm metrics. To test which metrics do differ between proficiency levels, the MANOVA was followed by a series of ANOVAs, separate for each metric, and with controlled repeated contrasts (the Bonferroni correction applied to the contrasts), to find out whether each metric differs between utterances produced by advanced and intermediate learners, and by intermediate learners and beginners (Table 3).

The analysis confirmed that the differences are significant. The effect sizes are large, which indicates that durational ratios of syllables, consonantal and vocalic intervals differ substantially between sentences produced by learners with different degrees of L2 mastery, and provide sufficiently reliable cues for the listeners and for us to investigate whether and to what degree these rhythmic differences affect perceived foreign accent.

To test whether listeners give different foreign accent ratings to sentences produced by L2 learners at different proficiency levels, we performed a one-way repeated-measures ANOVA with *level* (beginning, intermediate and advanced) as the factor and *accent rating* (rating scores the original sentences received) as the dependent variable. As the Mauchly's test was significant, $\chi^2(2) = 9.738$, $p = .008$, which indicates that the assumption of sphericity is probably violated, we decided to report multivariate tests (the correction for df in the main test, $\epsilon = .956$, would not change the result pattern and hardly influence the effect size, which is why the choice of the statistics to report makes it simply a matter of preference here). The results show that *level* has a significant and substantial effect on *accent rating*, $\Lambda = .18$, $F(2, 208) = 472.9$, $p < .0005$, $\eta^2 = .82$. The utterances produced by advanced learners received significantly higher scores than those produced by intermediate

Table 3. Statistic data for the effects of the speaker's proficiency levels on the values of the rhythm metrics. One asterisk indicates significant effect, while two asterisks indicate significant effects after the Bonferroni–Holm correction is applied to the model. All pairwise contrasts are corrected by the classical Bonferroni method.

Metric	$F(2,42)$	p	η^2	Significance of contrasts	
				Beginner–Intermediate	Intermediate–Advanced
Varco-S	10.816	<.0005**	.34	.064	.009**
nPVI-S	4.677	.015**	.182	.764	.017**
Varco-V	9.698	<.0005**	.316	.004**	.249
Varco-C	3.785	.031*	.153	.063	.45
nPVI-V	12.666	<.0005**	.367	.014**	.018**
nPVI-C	3.595	.036*	.146	.03**	.882

learners, $F(1, 209) = 325.6, p < .0005, \eta^2 = .609$, and the utterances produced by beginners received significantly lower ratings compared to those of intermediate learners, $F(1, 209) = 317.1, p < .0005, \eta^2 = .603$.

For the analysis of the rating scores given to the synthetic stimuli, we applied multilevel modelling. Our design involved multiple sessions per rater, each session yielding a set of accentedness scores for the stimuli, thus the ratings are nested within listeners. Individual differences in assessment of accentedness degree may potentially produce correlated residuals within raters across sessions, yet not necessarily correlated residuals across raters within conditions. Considering this, we had to apply multi-level modelling, which allows interpreting the effect of *level* (beginners versus intermediate versus advanced), *type* (flat versus intoned) and *stimuli* (*Rhythm and Rate*, *Rhythm Only* and *Rate Only*) on *foreign accent rating* as a dependent variable despite potential correlated residuals. The analysis was performed using SPSS 15.0 Mixed Models, with the restricted maximum likelihood criteria. Raters were level 2 units of analysis; the accent scores (nested within listeners) were introduced on level 1. Since we expected the ratings to vary over the participants, with some raters being stricter than others, or weighting different features of L2 pronunciation in their assessments of degree of accentedness, we factored the potential between-listeners variability in by fitting a random intercept model for the raters to the data.

The analysis revealed a significant main effect of *level*, $F(2, 3749) = 115, p < .0005$, *stimuli*, $F(2, 3749) = 36.5, p < .0005$ and *type*, $F(1, 3749), p = .006$, moderated by a significant interaction between *stimuli* and *level*, $F(4, 3749) = 5.324, p < .0005$. The interaction between *stimuli* and *level* indicates that the effect of *level* on accent rating depends on whether the synthetic stimuli preserve only rhythmic, only rate or rhythmic and rate between-level differences. This means that different types of timing patterns have a different impact on foreign accent rating, and the presence of both types of tempo differences influences the perception of rhythmic differences (although we still cannot say whether the effect is enhancing, impeding or a more complex one).

The significant interaction makes the straightforward interpretation of the main effects problematic. To analyse the influence of different types of prosodic timing patterns on perceived accentedness, we performed a planned analysis of simple effects. We performed a repeated-measures ANOVA with *accent rating* as a dependent variable and with *level* as a factor, separately for each type of stimuli (*Rhythm and Rate*, *Rhythm Only* and *Rate Only*). The analysis was performed separately on intoned and on flat stimuli. The ANOVAs were followed by controlled comparisons (with the Bonferroni correction). We compared the ratings for the stimuli with the timing patterns of the

Table 4. Statistical data for the effect of *level* on the *accent rating* for different types of flat stimuli.

Stimuli type	Repeated-measures ANOVA				Contrasts					
					Advanced–Intermediate			Intermediate–Beginner		
	Λ	$F(2,208)$	p	η^2	$F(1,209)$	p	η^2	$F(1,209)$	p	η^2
Rhythm and Rate	.595	70.781	<.0005	.405	24.08	<.0005	.103	73.872	<.0005	.261
Rhythm Only	.68	49.049	<.0005	.32	.352	.554	.002	84.212	<.0005	.287
Rate Only	.751	34.543	<.0005	.249	5.64	.018	.026	42.891	<.0005	.17

ANOVA: analysis of variance.

Table 5. Statistical data for the effect of *level* on the *accent rating* for different types of intoned stimuli.

Stimuli type	Repeated-measures ANOVA				Contrasts					
					Advanced–Intermediate			Intermediate–Beginner		
	Λ	$F(2,208)$	p	η^2	$F(1,209)$	p	η^2	$F(1,209)$	p	η^2
Rhythm and Rate	.597	70.226	<.0005	.403	27.783	<.0005	.117	73.588	<.0005	.26
Rhythm Only	.648	56.404	<.0005	.352	12.885	<.0005	.058	57.645	<.0005	.216
Rate Only	.807	24.847	<.0005	.193	6.826	.01	.032	23.852	<.0005	.102

ANOVA: analysis of variance.

advanced and intermediate learners, and then the ratings for the stimuli with the timing patterns of the beginners and intermediate learners.

The analysis shows that accent ratings differ between proficiency levels for all types of stimuli, both in the flat (Table 4) and the intoned (Table 5) conditions, but the magnitude effect of these differences and significance of pairwise comparisons between levels is modulated by which timing patterns are preserved in the synthetic stimuli. When listeners had to evaluate the resynthesized utterances that maintained only the rhythmic patterns of English learners at different L2 proficiency levels, there was a clear difference in the result patterns depending on whether the listeners evaluated intoned or flat utterances. In the flat condition, listeners gave similar accentedness ratings to the utterances that preserved the rhythmic patterns of advanced and intermediate learners, and the accent ratings differed significantly only between the stimuli with the rhythmic patterns of intermediate learners and beginners. However, when intonation was added, listeners could perceive the difference in rhythmic patterns between the utterances of advanced and intermediate learners, as well as between intermediate learners and beginners. Therefore, we conclude that the presence of intonation enhances the perception of rhythmic differences and strengthens the effect of rhythm on perceived foreign accent.

When the listener rated the synthetic stimuli that differed only in speech rate, they perceived faster utterances as less accented and slower utterances as more accented than the utterances with the speech rate of an intermediate French learner of English. What is fascinating is that the difference in accent rating on *Rate Only* utterances was smaller (Figure 5) when the listeners had to rate intoned utterances, and the effect size on intoned *Rate Only* utterances ($\eta^2 = .193$) was smaller compared to that on flat utterances ($\eta^2 = .249$). This might indicate that intonation impedes the perception of small differences in speech rate. When intonation lacks, listeners have more resources to process the only parameter that differs between utterances – speech rate – and these distinctions

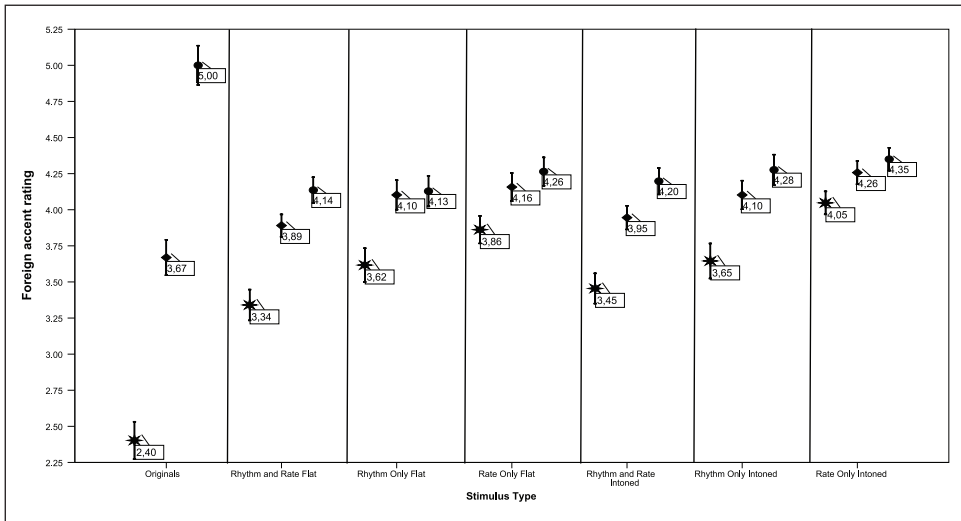


Figure 5. Accent ratings of the original utterances and of different types of synthetic stimuli. Error bars show ± 2 SE. Stars represent the ratings received by beginners, rhombi by intermediate and circles by advanced learners of English.

in speech rate become more salient in flat stimuli. However, this inference is tentative and requires further investigation.

When listeners evaluated the resynthesized utterances that differed both in rhythm and in rate, with utterances preserving the rhythmic patterns of advanced learners being faster and the utterances preserving the rhythmic patterns of beginners being slower, the differences in accentedness between the proficiency levels were significant. Thus, we conclude that the faster speech rate in the utterances with rhythmic patterns of advanced learners of English reduces the degree of perceived foreign accent, while the slower speech rate in the utterances with rhythmic patterns of beginners increases perceived accentedness. The increase in perceived accentedness can also be enhanced by unusual co-articulatory patterns and uncommon timing relations at the phonemic and subphonemic levels, because in natural speech vowels tend to be lengthened to a greater degree than consonants, while in our stimuli we compressed and lengthened the entire sentences, which led to compressing vowels and consonants in the stimuli to an equal degree. The range of accent ratings is larger in the *Rhythm and Rate* condition than in *Rhythm Only* and *Rate Only* conditions. Consequently, speech rate and speech rhythm have a cumulative as well as a unique, independent effect on perceived foreign accent.

Rhythmic patterns and speech rate together account for over 40% of variance in accent rating (both for flat and intoned stimuli), when the segmental and intonational differences between utterances produced by learners at different proficiency levels are neutralized. The differences in speech rhythm only, all other factors being equal, explain more variance (35% for intoned utterances and 32% for flat utterances) than the differences in speech rate only (19% and 25%, respectively).

This means that the participants paid more attention to rhythmic differences than to the differences in speech rate when they rate the degree of accentedness. This means that the differences in speech rhythm make a greater contribution to perceived foreign accent in English sentences produced by French L2 learners than the differences in speech rate. The effects sizes in our tests are large and the analysis is powerful, although the sample size of the listeners is relatively small ($N = 14$), which suggests that the difference in effect size between conditions is potentially

transferrable from our sample to other samples from the population of native English listeners. The independent contribution of speech rhythm to foreign accent, especially on intoned sentences, is greater than the contribution of speech rate because the range of accent ratings and the effect size of *level* on *accent rating* for *Rhythm Only* utterances are larger compared to those of *Rate Only* utterances in the intoned condition.

4 Discussion

Our results indicate that the differences in speech rhythm between utterances produced by L2 learners at different proficiency levels make a greater contribution to perceived foreign accent than differences in speech rate. This result might seem to contradict the outcome of a previous experiment series (e.g., Ordin & Polyanskaya, 2015a), where it was reported that native English speakers do not pay attention to differences in speech rhythm and rely entirely on the timing cues pertaining to the speech rate when they have to classify the stimuli into categories based on prosodic timing patterns. However, Ordin and Polyanskaya (2015a) used the “*sasasa*” transform to prepare the stimuli, that is, all the vowels in the L2 utterances were transformed into “*a*” and all the consonants into “*s*” (Ramus et al., 1999). The authors found that the stimuli derived from the L2 utterances were classified into two categories based on the number of “*sa*” syllables per second. The stimuli with the faster rate were grouped separately from the stimuli with the slower rate, both groups containing the stimuli with high and low durational variability of V and C intervals. This is in agreement with the results obtained by Arvaniti and Rodriquez (2013) who found that “*sasasa*” stimuli representing prosodic timing patterns in different languages are discriminated based on speech rate, not on durational variability.

In the present experiment we have found that native English speakers indeed attend to rhythmic cues when evaluating the degree of accentedness. The difference in the nature of the stimuli might account for the discrepancy in the result patterns between the two experiments. It is possible that “*sasasa*” stimuli are not perceived as speech, and the rate of intensity peaks outweighs the finer cues to durational variability of V or C intervals in non-speech stimuli. When dealing with speech-like stimuli, as in the present experiment, rhythmic cues become more informative. Language-specific rhythm is used to develop the segmentation strategies attuned to particular languages and to differentiate rhythmically different utterances (e.g., Kim, Davis, & Cutler, 2008; Kolly and Dellwo, 2014; Murty, Otake, & Cutler, 2007; Ramus et al., 1999). Besides, rhythmic patterns affect the intelligibility and comprehensibility of L2 speech. Thus, when dealing with more speech-like stimuli, rhythmic patterns may play a more significant role than speech rate. An alternative explanation for the dissimilarity between the results of this study and the study reported by Ordin and Polyanskaya (2015a) is that the native languages of the learners of English are different; in fact, the authors used L2 utterances produced by German learners of English to construct the “*sasasa*” stimuli. For the present experiment, the original L2 English utterances were produced by native speakers of a rhythmically contrastive language, that is, French. English and German (the native languages of the listeners and the speakers who provided the timing patterns for “*sasasa*” stimuli in the study by Ordin and Polyanskaya, 2015a) exhibit similar rhythmic patterns.

The rhythmic patterns in the English sentences produced by the German L2 learners of English at different proficiency levels might have not been sufficiently distinct to be perceptually relevant. On the other hand, the differences in rhythmic patterns in the English utterances produced by the French learners at different proficiency levels might have been more salient because the native and the target languages of the learners are substantially different rhythmically. To choose between these two explanations, it is necessary to replicate the perception experiments with the synthetic stimuli that implement the rhythm and rate of German learners of English.

Our results convincingly show that the timing patterns pertaining to speech rhythm and speech rate are sufficient to reliably evaluate which utterance is more accented. It is also clear that the sentences produced by the L2 beginners received lower scores in the original version than in the resynthesized sentences (Figure 5). The opposite was found for the original sentences of the advanced learners and for the stimuli with the timing patterns of the advanced learners: the original sentences received higher scores than the synthetic stimuli. The sentences produced by the intermediate L2 learners received comparable scores both in the original and in the modified conditions. This might indicate that the relative weights of segmental characteristics and timing patterns in assessment of accentedness differ depending on the proficiency level. In other words, speech rate and rhythmic patterns are overridden by foreign accented segmental characteristics in the speech produced by L2 learners with a lower proficiency level. However, L2 timing patterns become perceptually more noticeable for native speakers of the target language when the segmental characteristics are acquired by L2 learners and exhibit less salient deviations from the native norms. Thus, we tentatively suggest that the segmental characteristics override deviations in temporality from the target patterns in L2 speech. The importance of fine-tuning rhythmic and tempo patterns grows as proficiency increases, segmental deviations diminish and more attention is diverted towards temporality during the perception of L2 speech.

The ratings given to intoned utterances are slightly but significantly higher for all types of synthetic stimuli, which might be explained by the fact that intoned stimuli sound more natural to the listeners and thus receive higher scores. We have also cautiously suggested that the presence of intonation enhances the perception of rhythmic differences between synthetic stimuli with rhythmic patterns of L2 learners at different proficiency levels. This conclusion is in need of further empirical evidence, because we need to establish that the effect of intonation is not an artefact of the difference in naturalness between flat and intoned stimuli. If the effect of the presence of intonation is genuine and linguistic, we will have to find the reason why the presence of F0 contours facilitates the perception of rhythmic differences. We tentatively suggest that prominence may play a role in the interaction of F0 contours and durational ratios in perception. Prominence at different levels of prosodic hierarchy plays a role in extracting discrete constituents from continuous speech streams (Ordin & Nespov, 2013), and thus listeners are sensitive to prominence fluctuations within utterances. Some parts of the utterance (syllables in a word or words in a phrase) are made more prominent by vowel lengthening and by local fluctuations of pitch; therefore, duration and F0 strongly interact in prominence manifestation. This explanation entails yet another question for future research, that is, what kind of F0 fluctuations enhance the perception of rhythmic patterns: local ones, that is, those associated with pitch-accented syllables, or global ones, that is, the presence of a declining F0 contour spanning over the whole utterance.

Another interesting question for further investigation is whether the presence of F0 contours per se enhances the perception of rhythmic differences, or whether language-specific details of phonetic implementation of pitch range and pitch accents facilitate discrimination of rhythmic patterns. Can non-native intonation (e.g., intonation of a Japanese learner of English) also facilitate the perception of rhythmic differences in L2 English speech by a native English listener? How native and non-native intonation contours and pitch accents interact with other prosodic cues, and how non-native intonation contours affect the perception of speech rhythm is still to be discovered.

Acknowledgements

We are thankful to Barbara Kühnert and the Laboratory of Phonetics and Phonology, Paris III Sorbonne Nouvelle, for support during our research stay and data collection.

Funding

The work was supported by the Alexander von Humboldt Foundation, Ikerbasque – Basque Foundation for Science, a dissertation grant from the Language Learning Research Club at Michigan University to the first author and by the Bielefelder Nachwuchsfond von Universität Bielefeld.

Note

1. Usually researchers distinguish between speech rate and articulation rate. Articulation rate is calculated excluding the duration of the pauses from the total duration of the utterances. Thus, a higher number and longer pauses might decrease the speech rate without affecting articulation rate. The speech material that we elicited for our study does not include pauses within utterances, and thus exhibits no difference between speech and articulation rate in the strict meaning of these terms. Therefore, we do not distinguish between speech and articulation rate, and use all terms pertaining to speech tempo interchangeably.

References

- Adams, C. (1979). *English speech rhythm and the foreign learner*. The Hague, the Netherlands: Mouton.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*, 529–555.
- 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., & 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*, 1–17.
- Best, C. (1995). A direct realist view of cross-language speech perception research with adults. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–204). Timonium, MD: York Press.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International, 5*, 341–345.
- Bond, Z. (1999). *Slip of the ear: Errors in the perception of casual conversation*. San Diego, CA: Academic Press.
- Bond, Z., & Fokes, J. (1985). The timing of English words by non-native speakers. *Journal of the Acoustic Society of America Supplement 1, 77*, S53.
- Boula de Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica, 63*, 247–267.
- Brahimi, B., Boula de Mareüil, P., & Gendrot, C. (2004). Role of segmental and suprasegmental cues in the perception of Maghrebien-accented French. In *Proceedings of Interspeech 2004, Jeju, Korea*, 341–344.
- 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.
- Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America, 54*, 421–428.
- Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 9*, 237–245.
- 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. Sziget (Eds.), *Language and language-processing* (pp. 231–241). Frankfurt am Main, Germany: Peter Lang.
- Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate. In *Proceedings of the 15th ICPHS*, 471–474.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*, 1–16.

- Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning*, 13, 1–18.
- 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, PA: ICSLP.
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39, 399–423.
- Flege, J., Bohn, O.-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Flege, J., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, 15, 67–83.
- Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40, 467–491.
- Flege, J., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, 38, 25–55.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). Timonium, MD: York Press.
- Grabe, E., & Low, L. (2002). Acoustic correlates of rhythm class. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (515–546). New York, NY: 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.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201–223.
- Jilka, M. (2000). The contribution of intonation to the perception of foreign accent. *Doctoral Dissertation, Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS)* 6(3), University of Stuttgart.
- 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, 554–566.
- Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *International Journal of Bilingualism*, 6, 315–334.
- Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech*, 51, 343–359.
- Kolly, M.-J., & Dellwo, V. (2014). Cues to linguistic origin: The contribution of speech temporal information to foreign accent recognition. *Journal of Phonetics*, 42, 12–23.
- Koster, C., & Koet, T. (1993). The evaluation of accent in the English of Dutchmen. *Language Learning*, 43, 69–92.
- Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93–107.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417.
- Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, 39, 397–413.
- Low, E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43, 377–401.
- Maassen, B., & Povel, D. J. (1985) The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America*, 78, 877–886.
- Magen, H. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, 26, 381–400.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543–563.

- 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. (1995). Non-segmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17*, 17–34.
- 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.
- Ordin, M., & Nespors, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning, 63*, 800–834.
- Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System, 42*, 244–257.
- Ordin, M., & Polyanskaya, L. (2015a). Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology, 6*, 1–15.
- Ordin, M., & Polyanskaya, L. (2015b). Acquisition of speech rhythm in second language by learners of English with rhythmically different native languages. *Journal of Acoustical Society of America, 138*, 533–545.
- Payne, E., Post, B., Astruc, L., Prieto, P., & del Mar Varnell, M. (2012). Measuring child speech. *Language and Speech, 55*, 203–229.
- Pellegrino, F., Coupé, Ch., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language, 87*, 539–558.
- Peterson, G., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of Acoustical Society of America, 32*, 693–703.
- Port, R., & Mitleb, F. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics, 11*, 219–229.
- Prieto, P., del Mar Varnell, 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.
- Quene, H., & van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication, 52*, 911–918.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265–292.
- Roach, P., Sergeant, P., & Miller, D. (1992). Syllabic consonants at different speaking rates: A problem for automatic speech recognition. *Speech Communication, 11*, 475–479.
- Rouas, J.-L., Farinas, J., Pellegrino, F., & André-Obrecht, R. (2005). Rhythmic unit extraction and modelling for automatic language identification. *Speech Communication, 47*, 436–456.
- Sereno, J., Lammers, L., & Jongman, A. (2014). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics*. DOI: <http://dx.doi.org/10.1017/S0142716414000575>.
- Stevens, K. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of Acoustical Society of America, 111*, 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*, 219–226.
- Tortel, A., & Hirst, D. (2010). Rhythm metrics and the productions of L1/L2. In *Proceedings of Speech Prosody 5, Chicago, USA*.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*, 1–30.

- Ulbrich, C., & Mennen, I. (2015). When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism*. doi:10.1177/1367006915572383
- Vicenik, Ch., & Sundara, M. (2013). The role of intonation in language and dialect discrimination by adults. *Journal of Phonetics*, 41, 297–306.
- Vieru, B., Boula de Mareüil, P., & Adda-Decker, M. (2011). Identification and characterisation of non-native French accents. *Speech Communication*, 53, 292–310.
- Wells, J. (2008). *Longman pronunciation dictionary*. 3rd edition. Longman, 922 pp.
- White, L., & Mattys, S. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501–522.