Contribution of prosodic timing patterns into perceived Foreign Accent

Leona Polyanskaya

Dissertation zur Erlangung des akademischen Grades Doctor philosophiae (Dr. phil.) vorgelegt an der Fakultät für Linguistik und Literaturwissenschaft der Universität Bielefeld am 30. Juni 2014.

Prüfungskommission:
Prof. Dr. Petra Wagner (Betreuerin und Gutachterin)
Prof. Dr. Christoph Gabriel (Gutachter)
Prof. Dr. Stavros Skopeteas
Prof. Dr. Joana Cholin

Datum der mündlichen Prüfung: 8. Januar 2015

Gedruckt auf säurefreiem, alterungsbeständigen Papier (nach ISO 9706)

моему дорогому М.Ю. посвящаю....

CONTENT

LIST OF FIGURES

Kurzzusammenfassung

Wenn Menschen eine Fremdsprache lernen, verbessert sich mit den allgemeinen Fortschritten in deren Beherrschung auch die Kontrolle der *Timing*mechanismen beim Sprechen in der Fremdsprache. Ziel dieser Arbeit ist es herauszufinden, ob diese Veränderungen perzeptiv relevant sind, und ob die verbesserte Kontrolle der *Timing*mechanismen bei fortgeschrittenen Sprachlernern deren wahrgenommenen Akzent reduziert.

Sprachspezifische Unterschiede in prosodischen *Timing*mustern sind gut dokumentiert. So weisen etwa die Dauern von vokalischen und konsonantischen Intervallen in den Sprachen, die traditionell als betonungszählend klassifiziert werden, eine höhere Variabilität auf als in Sprachen, die traditionell als silbenzählend klassifiziert werden. Silbenzählende Sprachen weisen außerdem eine höhere Sprechrate auf als betonungszählende Sprachen. Darüber hinaus zeigen Untersuchungen zu verschiedenen Sprachen, dass Nichtmuttersprachler eine geringere Sprechrate und niedrigere Variabilität im *Timing* gesprochener Äußerungen aufweisen als Muttersprachler. Diese Unterschiede beeinflussen die Verständlichkeit von gesprochenen Äußerungen von Nichtmuttersprachlern sowie deren wahrgenommenen fremdsprachlichen Akzent (FA). Allerdings sind die Geschwindigkeit – gemessen in sprachlichen Intervallen pro Zeiteinheit – und die Variabilität der Dauern dieser Intervalle in gesprochenen Äußerungen miteinander korreliert: Je höher die Sprechgeschwindigkeit ist, desto geringer ist die Variabilität der Intervalldauern. Dies wirkt sich auch in der Wahrnehmung aus. Daraus ergibt sich die Frage, in welchem Maß beide Faktoren zur Wahrnehmung eines FA bei Nichtmuttersprachlern beitragen. Um diese Frage zu beantworten, müssen beide Faktoren isoliert betrachtet werden.

Tempo und *Timing*variabilität beim Sprechen einer Fremdsprache erhöhen sich im Verlauf von deren Erwerb, unabhängig davon, ob sich Mutter- und Fremdsprache (im Folgenden: L1 und L2) hinsichtlich ihrer *Timing*charakteristika unterscheiden. Der Grad dieser Veränderung sollte folglich auch die Stärke des

wahrgenommenen FA reflektieren. Wenn die *Timing*unterschiede perzeptiv relevant sind, sollten Äußerungen, deren *Timing*muster der eines fortgeschrittenen L2-Lerners entsprechen, als schwächer akzentuiert wahrgenommen werden als solche, deren *Timing*muster denen eines Anfängers entsprechen, auch wenn spektrale und intonatorische Unterschiede eliminiert werden. Dabei wird die Frage zu klären sein, in welchem Maß die beiden Faktoren Tempo und *Timing*variabilität den wahrgenommenen FA beeinflussen. Grundannahme dieser Arbeit ist, dass der Einfluss der Variabilität geringer ist, wenn L1 und L2 ähnliche *Timing*charakteristika haben.

In dieser Arbeit werden die *Timing*muster von deutschen und französischen Lernern des Englischen hinsichtlich ihres Einflusses auf den wahrgenommenen FA untersucht, wobei zusätzlich jeweils Anfänger und fortgeschrittene Lerner getestet werden. Die *Timing*charakteristika des Deutschen ähneln denen des Englischen, während sich das Französische in dieser Hinsicht deutlich vom Englischen unterscheidet. Daraus ergeben sich zwei Hypothesen: (1) Im Englischen fortgeschrittener deutscher Lerner (gegenüber Anfängern) äußert sich die Reduktion des wahrgenommenen FA stärker in einer Erhöhung der Sprechrate; (2) Bei französischen Englischlernern spielt die *Timing*variabilität eine größere Rolle als die Veränderung der Sprechrate im Verlauf des Spracherwerbs.

Diese Hypothesen wurden anhand von vier Forschungsfragen überprüft:

1. Nehmen Muttersprachler der Zielsprache (Englisch) Unterschiede im *Timing* gesprochener Äußerungen zwischen Anfängern und fortgeschrittenen Englischlernern wahr?
2. Korreliert die Reduktion des wahrgenommenen FA mit den Veränderungen der *Timing*muster im Verlauf des L2-Erwerbs?
3. Welche Anteile haben die einzelnen Faktoren Sprechtempo und *Timing*variabilität am wahrgenommenen FA?
4. Zeigen sich hinsichtlich der separaten Anteile von Sprechtempo und *Timing*variabilität am wahrgenommenen FA Unterschiede zwischen Lernern mit typologisch unterschiedlichen Muttersprachen?

In dieser Arbeit wird über die Ergebnisse von drei

Wahrnehmungsexperimenten berichtet, die zur Beantwortung der Forschungsfragen durchgeführt wurden. Die Arbeit ist wie folgt strukturiert: Im ersten Kapitel werden der theoretische Hintergrund vorgestellt und die Arbeitshypothesen erläutert. Das Kapitel beginnt mit einer Definition des Begriffs „FA“ und einer Diskussion der wichtigsten Faktoren, die zur Wahrnehmung des FA beitragen. Dabei wird auch ein kurzer Überblick über Modelle des Zweitspracherwerbs gegeben. Weiterhin werden segmentale und prosodische Unterschiede zwischen L1 und L2 und deren Einfluss auf den wahrgenommenen FA diskutiert, und es wird erörtert, wie diese Unterschiede in verschiedenen Modellen des Zweitspracherwerbs erklärt werden. Zudem wird es auch auf die Frage eingegangen, ob segmentale oder prosodische Faktoren einen größeren Einfluss auf den wahrgenommenen FA haben. Schließlich folgt eine Diskussion des Prosodiebegriffs, unter Einbeziehung der Subsysteme Betonung, Intonation und *Timing*. Im Besonderen wird auf *Timing*muster eingegangen, wobei der Fokus darauf liegt, wie *Timing* in gesprochener Sprache gemessen werden kann, wie Sprechrate und *Timing*variabilität zusammenhängen, und wie *Timing*unterschiede zwischen Muttersprachlern und Sprachlernern die Wahrnehmung von FA beeinflussen. Anschließend an diese Diskussion wird die oben eingeführten Arbeitshypothesen formuliert und motiviert.

In den Kapiteln 3, 4 und 5 werden die einzelnen Wahrnehmungsexperimente beschrieben, im letzten Kapitel zusammengefasst und diskutiert. Die Ergebnisse der Experimente stützen die oben eingeführte Hypothese und können wie folgt zusammengefasst werden:

1. Muttersprachler nehmen die Unterschiede zwischen den *Timing*mustern in den L2-Produktionen fortgeschrittener und weniger fortgeschrittener Sprachlerner wahr. Sie tendieren jedoch dazu, Unterschiede in der Variabilität in Klassifikationsaufgaben und in nichtsprachlichen Stimuli zu ignorieren. Je natürlicher und sprachähnlicher die Stimuli sind, desto stärker werden Unterschiede in der *Timing*variabilität wahrgenommen.

2. Die Stärke des wahrgenommenen FA korreliert, wie vorhergesagt, mit Veränderungen in Sprechtempo und *Timing*variabilität, die mit steigender

Kompetenz in der L2 einhergehen. Fortgeschrittenere Sprecher sprechen schneller und mit höherer Variabilität von sowohl Vokal- als auch Silbendauern. Äußerungen mit höherem Sprechtempo und höherer *Timing*variabilität werden von Muttersprachlern des Englischen als weniger stark akzentuiert wahrgenommen.

3. Der kombinierte Beitrag von Sprechtempo und *Timing*variabilität zum wahrgenommenen FA ist größer als die Summe der Effekte beider Faktoren in Isolation. Experimente, in denen jeweils einer der beiden Faktoren kontrolliert wird, zeigen, dass beide zum wahrgenommenen FA beitragen.

4. Die relative Gewichtung beider Faktoren hängt davon ab, ob L1 und L2 hinsichtlich ihrer *Timing*charakteristika ähnlich oder verschieden sind. Wenn sich L1 und L2, wie im Fall von Französich und Englisch, stark unterscheiden, ist der Beitrag der Variabilität größer; wenn sich L1 und L2 hinsichtlich ihrer *Timing*charakteristika ähneln – wie im Fall von Deutsch und Englisch – spielt das Sprechtempo für die Wahrnehmung des FA die wichtigere Rolle.

Acknowledgement

I am thankful to Prof. Petra Wagner for her supervision and advice, and for the personal attitude to me that encouraged me to efficiently accomplish this work, and for the assistance in overcoming multiple hurdles on the way to the defence, and for emotional support. I am also thankful to Prof. Christoph Gabriel for his thorough and sympathetic review of my thesis with useful suggestions. Special thanks go to the members of the dissertation committee panel, Prof. Stavros Skopeteas and Dr. Joana Cholin, for their stimulating questions and especially for personal and professional support before as well as after the defence.

I am indebted to Dr. Mikhail Ordin for the inspiration, support and understanding during my work on this long-term endeavour, and for having me as the research assistant on his projects. His project on development of rhythmic patterns in language acquisition was the greatest inspiration for me to find out whether the developmental differences between proficiency levels of L2 learners are perceived and contribute to the foreign accent. The gentle introduction into genuine research work was the greatest incentive to become an independent researcher myself, and equipped me with the skills I used in my project.

Many thanks go to the members of Phonetics and Phonology Research Group at Bielefeld University. Special thanks to Andreas Windmann for his sympathy and for his help and permanent readiness to help. I am also thankful to Laura, Joanna and Ola for nice talks we had.

I want to express my gratitude to Prof. Barbara Kühnert form Université Sorbonne Nouvelle Paris III for hosting me during my research stay in Laboratoire de Phonetique et Phonologie and for providing access to the technical facilities of the laboratory needed for the data collection.

I am thankful to Dr. Christiane Ulbrich for her collaboration and for allowing me to work at Ulster University, UK, Phonetics Laboratory, which made it possible to carry out the second experiment reported in my thesis.

I am indebted to my family for their support, love and understanding.

1. Introduction

The factors that contribute to the percept of foreign accent and their relative weighting has been a matter of scientific debate in second language acquisition, phonetic, phonology and psycholinguistics. Earlier research emphasized either prosodic or segmental characteristics of the non-native speech that make it sound more or less accented or intelligible. However, the issue of relative importance of segmental or prosodic factors in reducing or enhancing the degree of accentedness is not yet resolved. Moreover, some researchers concentrated on the contribution of separate prosodic systems – stress patterns (location, distribution and phonetic realization), pitch range or level, intonation, speech rate and rhythm – into foreign accent. The present study is aimed at detecting to what extent prosodic timing patterns influence perception of accentedness.

It is well established that prosodic timing patterns differ between languages. For example, variation in duration of vocalic and consonantal intervals and of syllables is higher in languages that are traditionally classified as stress-timed compared to the languages that are traditionally classified as syllable-timed. Besides, speech in prototypically stress-timed languages is delivered at slower rate than in prototypically syllable-timed languages. It has also been firmly established that durational variability and speech rate differ between utterances produced by native and non-native speakers of the same language. Speech of non-native speakers is characterized by lower durational variability and slower tempo, all other factors being equal. These differences in prosodic timing between native and non-native speech influences the intelligibility of the non-native speech and the degree of the perceived foreign accent. Speech tempo and durational variability of speech intervals have been shown to be interrelated and reversely correlated. The faster the speech, the less variable are the durations of the speech intervals. This is also reflected in perception. Therefore, it is impossible to say whether the differences in speech rate or the differences in durational variability contribute to the perceived foreign accent. If the differences between native and non-native speech both in terms of speech rate and durational ratios affect the degree of the perceived foreign accent, then we will need to address the issue of

the relative contribution of speech tempo and durational ratios – which prosodic timing patterns have larger weight in assessing accentedness.

Previous research also revealed that speech tempo and durational variability increase as L2 acquisition progresses. This increase is detected regardless of whether the native and the target languages of the learners are similar or different in terms of durational variability and speech tempo. In a series of experiments, we tried to figure out whether the developmental changes in speech rate and durational variability are perceptually relevant, and whether these changes affect the degree of the perceived foreign accent.

As 1) durational variability and speech tempo change as L2 acquisition progresses, and 2) durational variability and speech tempo influence the degree of the perceived foreign accent, that the developmental changes in timing patterns will supposedly also affect the degree of the perceived foreign accent. When individual differences in sounds and in intonation are eliminated and only the differences in timing patterns are preserved, we expect the perceived accentedness to be lower for the sentences which preserve durational variability of more advanced L2 speakers compared to those of lower-level L2 learners. However, it is hard to say what will have more weight in perceived accentedness – variability or rate at which speech is delivered. Based on the literature review and logical inferences, a working hypothesis was formulated. The initial hypothesis that the differences in relative contribution of speech tempo and durational variability determined by the native language of the L2 speaker. If the timing patterns of the source and the target language is similar, durational variability is assumed to have lesser effect on the perceived foreign accent than when the source and the target language exhibit distinctly different timing patterns.

English was chosen as the target language for further analysis how timing patterns in speech of German and French learners of English at different proficiency levels affect the perceived foreign accent. German and French were chosen as native languages of the learners of English because German is similar to and French is different from English in terms of durational variability and speech rate. That said, we expect that the developmental differences in durational variability in L2 English produced by German learners will have little effect on the foreign accent ratings, but the increase in speech tempo with proficiency will contribute a lot into perceived accentedness. As for French learners of English, it

is anticipated to detect a significant and very substantial effect of developmental changes in durational variability on the perceived accentedness.

Four research questions were set up to test this hypothesis:

1. Do native speakers of the target language indeed hear the differences in prosodic timing patterns (durational variability and speech rate) between proficiency levels in L2 speech?
2. Does the degree of the perceived foreign accent change as the timing patterns pertaining to the speech tempo and durational variability develop?
3. What is the shared contribution of timing patterns into degree of the perceived foreign accent, and what is the separate contribution of speech tempo and durational variability patterns into the perceived foreign accent?
4. Is the contribution of speech tempo and durational variability into the perceived foreign accent different for speakers from typologically different L1 backgrounds, in which different timing patterns are displayed (e.g., German and French learners of English).

Within the framework of the project, three perception experiments were set up to answer these questions and to confirm the hypothesis. These experiments are described in this thesis. The thesis has the following structure. In the first chapter, the detailed theoretical background is provided. This theoretical background led to the predictions and allowed to formulate the tested hypothesis. First, the notion of the foreign accent is defined. Further, the main factors that lead to the percept of the foreign accent in non-native speech are mentioned. Here, a brief overview of language learning models is also provided. The chapter follows with a discussion of segmental and prosodic differences between native and non-native speech, their contribution into the percept of accentedness, and how the emergence of these differences can be explained within Language Learning Models. The chapter continues with the discussion of an open issue of what makes a bigger contribution into foreign accent: Segmental deviations or Prosody. Further, prosody is tackled in more details by considering separate prosodic systems (e.g., stress, intonation, timing) rather than prosody as a whole. One of the prosodic systems that can be discussed is the system of prosodic timing patterns, including durational variability of speech intervals (i.e., speech rhythm) and mean durations of speech intervals (i.e., speech rate). This is discussed in the third section of the chapter. The section describes how prosodic timing patterns

can be objectively measured, how speech rate and durational ratios of speech intervals are related, and how exactly the differences in timing patterns between native and non-native speech affect the degree of the perceived foreign accent. This discussion leads to formulation of the hypothesis that the development of timing control in non-native speech with proficiency growth might reduce the degree of the perceived foreign accent. It is also justified (1) why speech rate is predicted to have a bigger contribution into the perceived foreign accent, if the native and the target language of the learner are rhythmically similar, and (2) why the developmental differences in durational ratios contribute to the accent reduction more than changes in speech rate, if the target and the native languages of the learner are rhythmically different. Second, third and fourth chapters describe the perception experiments set up to test the hypothesis. The formulated expectations are confirmed. In the last chapter, the major conclusions are presented. The conclusions could be summarized to the following:

1. Native speakers of the target language hear the differences in the examined timing patterns between the sentences produced by L2 learners at different proficiency levels. However, they tend to ignore the differences in durational variability in classification task and in non-linguistic stimuli. The more natural and the more speech-like the stimuli are, the more perceptually prominent the differences in durational variability become.

2. The degree of the perceived foreign accent is indeed affected by the developmental changes in speech rate and speech rhythm. More advanced learners tend to speak with more rapidly and with higher degree of durational variability at the timescale of vowels and concomitants and at the timescale of syllables. Sentences with higher degree of durational variability and faster sentences are perceived as less accented by the native speakers of English, all other factors being equal.

3. Combined contribution of durational variability and speech rate into the perceived foreign accent is bigger than the unique contribution of either durational variability or speech rate. However, we have also found the contribution of durational variability when controlling for speech rate and the contribution of speech rate when controlling for speech rhythm. Thus, there is not only combined but also unique contribution of the interrelated

characteristics of tempo and durational variability into the perceived foreign accent.

4. The relative contribution of durational variability and speech tempo into the perceived foreign accent depends on whether the native language of the learner is rhythmically similar to or different from the target language. The contribution of durational variability into foreign accent is bigger when the native language and the target language are rhythmically contrastive (e.g., English as the target language and French as the native language of the learner), than when the languages are rhythmically similar (e.g., English and German). If the languages are rhythmically similar, then the native speakers pay much more attention to the speech rate (faster sentences are assessed as less accented) when evaluating the degree of accentedness.

Some results of the thesis have been published in:

Polyanskaya, L., Ordin, M., Ulbrich, C. (2013). Contribution of timing patterns into perceived foreign accent. In P. Wagner (Ed.). *Elektronische Sprachsignalverarbeitung 2013, 71-79.* Dresden: TUDpress.

We have also presented the results at the international research workshop Multilinguality in Speech Research: Data, Methods and Models (Dagstuhl, Saarbruecken University (9. – 11. April, 2014), at the conference P&P9, Trends in Phonetics & Phonology in German-speaking Europe, Zuerich, 11. – 12. October 2013) and at the conference ESSV 2013 (Konferenz zur Elektronischen Sprachsignalverarbeitung, Bielefeld, 26. – 28. March 2013).

Previous research in which I have participated and which have been used as a theoretical basis for the hypothesis of this thesis include:

1) Ordin, M., Polyanskaya, L., Ulbrich, C. (2011). Acquisition of Timing Patterns in Second Language*. Interspeech 2011*, 1129-1132.

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

The speech material that was used for this project had been collected within a different project in which I collaborated. Part of the corpus is described in Ordin, M., Polyanskaya, L., Ulbrich, C. (2011). Acquisition of Timing Patterns in Second Language*. Interspeech 2011*, 1129-1132. However, the description of the full

corpus has not yet been published and therefore the description of the collected speech material is contained in the detail in this thesis.

2. Chapter I: Theoretical Background

This chapter introduces the theoretical background for the study. At first, the notion of the Perceived *Foreign Accent* (FA) is introduced. Then the factors that contribute to creating the percept of FA are discussed. The open and actively debated issue of whether the specific pronunciation features at prosodic or segmental levels make a bigger contribution into creating the percept of FA is discussed. The chapter further continues by considering the relative contribution of separate prosodic systems including stress, pitch range, intonation, pausation (duration, distribution, frequency and type of pauses) and timing into the FA. As the literature overview highlights the relatively substantial contribution of timing into the FA, the chapter further focuses on timing patterns in native and in non-native speech, and on influence of non-native timing patterns on perceived accentedness and intelligibility of non-native speech by native listeners. Language-specific patterns of durational variability at the timescale of syllables, vowels and consonantal clusters contribute to the impressions of language-specific rhythm, and language-specific phonological characteristics like phonotactic constraints on the consonantal clusters influence the mean duration of syllables, thus making some syllable perceptually faster than others. Timing patterns pertaining to the speech rhythm and speech rate as well as the interaction of perception of the speech rate and speech rhythm are discussed in more details. At the end, the hypothesis regarding the contribution of speech rate and speech rhythm is formulated based on the provided literature overview, and the research questions are set up in order to address the hypothesis.

2.1. The notion of “Foreign Accent”

Speech produced by individuals who have acquired a language from birth (L1 speakers) differs from speech of those individuals who have been exposed to and acquired the same language later in life (L2 speakers). L2 speech usually exhibits a certain degree of foreign accent (FA) which generally can be easily perceived by L1 speakers (Mennen, 2011; Scovel, 2000). Scovel (2000) in his review seminal paper states that the general consensus among specialists in language acquisition is that it is nearly impossible for second language learners to achieve native-like pronunciation, especially when learning started after the so-

called Critical period, which is often associated with puberty. Scovel (2000) and Piske, McKay and Flege (2001) found that the age of acquisition is the most informative predictor regarding the degree of foreign accent in speech of a language learner.

However, even when exposure to the second language started as early as 2 years in age, people sometimes still exhibit differences in pronunciation from monolingual speakers who have not been exposed to any second language (Kehoe, 2002; Flege, 1995). Scovel (2000) goes as far as to claim that acquiring accentless pronunciaition in two languages or in a second language is impossible, even in case of a very early immersion (2-5 years) or even in the case of simultaneous bilingual acquisition. Long (1990) suggested that this happens due to maturational constraints that occur after puberty in the central nervous system. With age the brains gradually loses a huge deal of flexibility and plasticity (when different brain areas are able to take different roles) and the ability to rapid development of new neural connections with increasing myelination of the already formed connections. Myelination promotes connectivity and signal transfer in the already existing connection, but at the same time it impedes making new ones, which is the case in more mature brains. Brain maturation starts as early as 6 years in age (Long, 1990), although there is evidence saying that the process starts even earlier, almost from the first months of age. Later in life, when the brain matures, acquiring sound patterns of another language becomes more problematic.

Kehoe (2002) also showed that the speech of simultaneous bilinguals (people exposed to two languages from birth and raised in bilingual environment where both languages are actively exploited on a daily basis) differs from speech of monolingual speakers of these two languages. As Grosjean (1989) pointed out, this is usually due to the fact that the languages in bilingual environment are used for different purposes and to talk to different people, therefore they have different uses. Monolingual speakers of these languages can easily hear these differences in speech productions of bilingual speakers or later language learners and interpret these differences as some kind of accent, be it social accent (Hughes, Trudgill, & Watt) or foreign accent (Gut, 2007).

When people in everyday life talk about L2 speech, they mean speech with foreign accent. The word “foreign accent” is easily understood in non-professional

communication, but for linguistics experts this term can be ambiguous, if a clear definition is not provided beforehand. A few examples of ambiguity for the word “foreign accent” are provided below.

Cook (2009; 2010) and Grosjean (1989) firmly stated that the speech of bilinguals or multilinguals growing up with two or more languages simultaneously cannot be compared with the speech of monolinguals. Even if bilinguals have been exposed to and used both languages from birth, their speech productions in both languages differ from those of monolingual speakers of these languages. Monolingual speakers easily pick up on the differences between speech of monolinguals and bilinguals and often interpret these differences in pronunciation between monolinguals and bilinguals as deviations from the native pronunciation norms. Should a researcher define the speech of simultaneous bilinguals “accented” despite the speaker has equally developed linguistic competence in both languages? Should a researcher say that speakers from remote areas of the English-speaking world have some pronunciation features that are interpreted as non-native and even stigmatized (Gluszek & Dovidio, 2010) by speakers belonging to the so-called “inner circle countries” (Kachru, 1992). Therefore the distinction between native and non-native speakers is not at all clear-cut, and amongst linguists there is up to date no commonly accepted agreement of what constitutes a foreign accent.

As the divergence between native and non-native speakers is not at all straightforward, the division between L1 and L2 speakers instead of making the distinction between native and non-native speech and accents will be made. FA is the characteristics of speech produced by second language speakers, i.e., those who have started learning the language after puberty. As Scovel (2000) shows, a large number of studies give an affirmative answer to the question whether the speakers who have acquired a certain language from birth can distinguish their peers from the speakers who have started learning this language after puberty, i.e., L1 and L2 speakers of the same language.

Following Gut (2007, referring back to Scovel 1969: 38) and Scovel (2000), foreign accent is defined as the set of deviations in pronunciation in L2 speech from what is perceived as the norm by L1 speakers. The norms are the expectations of what native speech is supposed to sound like held by monolingual native speakers of the target language who are making the judgment and evaluate

the degree of accentedness. It should also be admitted that in this case the degree of accentedness and the judgment on native vs. accented can be influenced by the individual expectations of people who are asked to express their opinion on nativeness and accentedness of the speech productions (Gluszek & Dovidio, 2010).

2.2. Main contributors into perceived Foreign Accent

Further in this chapter, the pronunciation peculiarities that make L2 speech sound accented to the ear of the L1 monolingual speaker of the same language are discussed. The foreign accent is defined as a set of deviations in speech of second language learners from expectations regarding pronunciation norms held by and shared by monolingual speakers of the target language who are making the judgment on accentedness.

The deviations that naïve listeners most frequently point to are those regarding the pronunciation of individual sounds, i.e., on the segmental level. These deviation might be due to phonemic and phonetic errors (we follow Hughes, Trudgill and Watt (2012: 19-37) and Wells (1982) in defining phonetic and phonological differences between linguistic varieties). The former arise as a consequence of discrepancies between the phonemic inventory of the interlanguage – the idiosyncratic linguistic system of an L2 learner – and of the target language. These discrepancies will cause so-called systemic differences between L1 and L2 pronunciation, as they are defined by Wells (1982) and Hughes, Trudgill and Watt (2012). Such errors can result not only in enhanced accentedness of L2 speech, but also impede intelligibility and comprehensibility of speech.

Besides systemic differences between phonemic inventories, foreign accent can arise due to the differences in realization of the same phoneme by L1 and L2 speakers. The same phoneme can exist in the inventory both of the target language and in the *interlanguage* of the learner, yet at the same time can be realized differently by L1 speakers and L2 learners. Such differences are most common and they are most easily picked up by the native speakers of the target language. Such differences can be identified as the “carriers” of the foreign accent. However, after a short adaptation exposure to the L2 speech exhibiting realizational deviations from the expected norms, intelligibility and

comprehensibility of the L2 speech by the native speakers of the target language is not compromised.

Distributional differences or differences in free variation of the phonemes or realizations usually go unnoticed and have no impact on the degree of the perceived foreign accent; therefore these are not discussed any further.

2.2.1. Models of Language Learning

Studies that focused on segmental characteristics of L2 speech and systematic and realizational differences between L2 and L1 speech led to development and corroboration of two main L2 speech models: Perceptual Assimilation Model (Best, 1995) and Speech Learning Model (Flege, 1995). The Speech Learning Model postulates that language learners form new phonetic categories in L2 if the tokens of this category in L2 are perceptually different from the tokens of any existing category in L1 of the learner. Therefore the learner finds it easier to acquire the segments of the target language if they are perceptually sufficiently different from the sounds in his native language to warrant a new category formation.

The Perceptual Assimilation Model states that the non-native sounds are perceived in terms of similarities and dissimilarities to native segments and phonological contrasts. The listeners try to map the non-native sounds into the system already existing for categorizing the linguistic segments. This mapping mechanism should supposedly be the same regardless of whether the listeners are monolingual speakers exposed to accented speech in their native language, or L2 learners who should discover how a new language divides the acoustic continuum into categories. The categories of a target language might be different from those in their native language, and a new mapping scheme must be developed and acquired by the L2 learners.

If necessary, the non-native sounds are perceptually assimilated to the closest category of the native phonemic inventory. The sounds that can be mapped onto any of the native category with adequate goodness of fit are processed as exemplars of this category. If the goodness of fit is not adequate, then the segment cannot be fitted into already existing category and a new category is to be formed. Alternatively, a non-native segment can be assimilated equally well (or badly) to two native categories. When a monolingual listener is processing L2 speech in his native language, he has to assimilate the non-native

segments (segments that deviate from the norms or his expectations) to the categories he has formed. If the goodness of fit is high, the degree of the perceived FA is low. The lower the goodness of fit is, the higher degree of accentedness is attached to the L2 speech.

Both models are based on the assumption that the listeners map L2 phonemes onto L1 phonemes, and thus fail to perceive the minute distinctions, or fine phonetic details, to which the L1 users attend to. Kuhl's Native magnet model of language acquisition (1991) is also based on phoneme-sized linguistic units. The failure to perceive the fine phonetic details which are important to the L1 users results in production difficulties, which create the foreign accent effect. Even though prosodic characteristics have been shown to play an important role in the production and perception of FA, none of the models accounts for suprasegmental aspects of speech.

Very few studies tried to adapt the well-established approaches and model of phonetic learning to suprasegmentals. There are possible four reasons why suprasegmentals are not often handled in L2 learning research. The first one is complexity and multidimensionality of suprasegmentals. Besides, the technical advances and especially reliable algorithms that allowed studying intonation, tone and many other prosodic systems are more recent than those that allowed acoustic analysis of segments. Many of the technical requirements that allow studying segments had been available for the researcher for decades before the accessibility of techniques to study intonation, for example, became general. Moreover, our understanding of prosodic primitives is very limited. It is not known, for example, what the building blocks of intonation are. In some approaches, these are the level tones (high and low) and the boundary tones (e.g., Pierrehumbert & Steele, 1990; Pierrehumbert, 1990; Pierrehumbert & Hirschberg, 1990). In other studies, intonation is considered as consisting of primitives of quite a different nature – pitch transitions with defined starting and ending points (t' Hart and Cohen, 1973). Fujisaki and Hirose (1984) combine both high and low targets and transitional specifications in their model. It is not even known whether the perception of the primitives is categorical or gradient, whether the contours are discreet or gradient (see a good review in Pierrehumbert, 1990 and Gussenhoven, 2004: 62-70). Finally, suprasegmentals are highly variable. Between-speaker variability is much higher on suprasegmental level than on segmental level.

Another source of variation is paralinguaitic, for example, Frequency, Effort, and Production biological codes for F0 fluctuations (Gussenhoven, 2004: 71-96). Therefore the researchers comparing the prosody in L2 and in L1 have problems with figuring out whether the differences come from phonological representations, phonetic implementations of phonological primitives, between-speakers differences (including anatomical and cultural differences), or are bound in other paralinguistic uses of prosody. These problems make it difficult to adopt and adapt the existing models of language learning and acquisition to study acquisition of prosody in the second language. The only difference is LILT (Longitudinal intonation learning transcript) approach suggested by Mennen (2004; 2007) and based on Ladd (2004; 2008) to study L2 intonation. She suggested that intonation is to be studied in four domains: systemic (single out the inventory of intonational primitives in the target language and in the interlanguage of the L2 learner); realizational (describe phonetic implementation of these primitives); distributional (analyze possible combinations of these primitives and their relative frequency); functional (align each primitive with a certain use, i.e., a certain function). Therefore the first step for the researcher studying L2 intonation is to carry out contrastive analysis comparing L1 and L2, or comparing groups of L2s (for example, L2 produced by learners with different native languages), or comparing L2 on different levels of acquisition. Analysis should be performed in all four dimensions, i.e., in systemic, realizational, distributional and functional domains. The second step is to find out whether the detected differences will tell on perception. Only those differences that can be perceived are relevant for language learning or language acquisition models (although perceptually undetected differences can be important in studies of L2 speech production). Although such approach is very effort-consuming, it seems to be the only reasonable way to systematically study L2 intonation, and therefore this approach is used in the presented study, if it turns out to be adaptable and adoptable to study the prosodic systems other than intonation in the L2.

Studies of prosody and suprasegmentals in L2 do indeed present considerable difficulties for the researchers, yet the contribution of prosody into perception of foreign accent has been widely attested (e.g., Jilka, 2000; Boula de Mareuil & Vieru-Dimulescu, 2006; Munro, 1995 and many others). In the following part of the thesis, the state of the art in research is presented, on what deviations

from the speech of monolingual speakers of the target language on segmental and on suprasegmental levels contribute to the perceived FA, and discuss the issue of relative contribution of segmental and prosodic deviations into foreign accent. This issue is far from settled, yet it potentially has an important impact on speech technologies (to make speech synthesis systems sound more natural), on foreign teaching pedagogy (pronunciation training and improving listening comprehension), on forensics (to identify an L2 speaker, his linguistic background and potentially his native language), on speech therapy (to treat speech disorders like Foreign Accent Syndrome – medical condition in which patients develop what appears to be a foreign accent in their native language) and in other areas where the question of relative contribution of prosody and segments into foreign accent can have practical applications, in addition to providing academic insights in the field of phonetics, phonology and second language acquisition. The debate regarding prosody, segments and their interaction in creating the effect of accentedness in L2 speech is overviewed in the following chapter.

Although the most widely accepted models fail to recognize the contribution of prosody to foreign accent, many studies have emphasized and even brought forward the deviation from the native norms at prosodic level into creating the effect of accentedness in L2 speech. The deviations in L2 speech can happen not only on segmental level, but also on suprasegmental level, including intonation, stress, tones, intensity fluctuations, timing patterns, syllabic structures and other phenomena.

Besides these most widely used models of (second) language acquisition (Best, 1995; Flege, 1995; Kuhn, 1991), a more general Ontogeny and Phylogeny model of second language phonological development (Major, 2001). Major says that the acquisition of L2 phonology is governed by the phonological structure of the L1, phonological structure of the L2, the universal factors like cross-linguistic markedness[*] (Eckman, 1977; 1991), and also by the interaction of these three

[*] The presence of the cross-linguistically marked feature implies also the presence of a corresponding unmarked feature, while the presence of an unmarked feature does not always imply the presence of the marked feature. For example, all languages with ergative case also have nominative case, so the presence of the ergative implies the presence of nominative. But the reverse is not always true. Thus ergative is marked. The presence of complex syllable codas implies the presence of simple codaless syllables, but the reverse is not true. Thus, complex clusters in codas are marked compared (Eckman, 1977). Eckman (1999) proposed the Markedness differential hypothesis that says that the marked features of the target language are more difficult to acquire in the course of L2 acquisition, if the native language of the learner does not have the corresponding marked structures or features.

components. Thus, Interlanguage includes the parts of L1 of the learners, the parts of L2 of the learner, and universals that are neither the parts of L1, nor the parts of L2. This model means that interlanguage is not necessarily somewhere in the middle between L1 and L2, and consequently those segmental and suprasegmental deviations from the target language norms in speech of L2 learners cannot always be explained by the transfer or interference between two (or more, in case of third language acquisition or in case of a bilinguals speakers learning a foreign language) linguistic systems.

Major (2001) built the model that predicts that the influence of L1 on the interlanguage decreases as L2 acquisition progresses, the influence of the phonological structure on the target language increases thoughout the acquisition progress, and the influence of universal factors in the course of L2 acquisition first increase, and then decrease. The proportion in which the elements of L1, L2 and universals affect the rate of acquisition is determined by whether the acquired phenomena in the native and the target languages of the speakers are similar or different, and whether the acquired feature of the L2 is cross-linguistically marked (see also The similarity differential rate hypothesis with similar ideas, proposed by Major and Kim, 1996).

Major (2001) says that those phenomena which are similar in L1 and L2 are acquired more slowly, the influence of the native language on interlanguage persists, and thesubstitution of L1 patterns for similar L2 patterns in learner's interlanguage undergoes very slowly. The influence of universals is also minimal. Thus, the accetnedness isdominated by the differences in fine phonetic details between similar L1 and L2 phenomena.

When the learner is acquiring marked phonological features of the target language, the influence of L1 on interlanguage decreases rapidly, the acquisition of marked L2 features undergoes slowly, and the acquisitional dynamics is mostly controlled by universals. The L1 phenomena are quickly substituted by the unmarked and cross-linguistically frequent universal phenomena, and then universals are very slowly substituted for the acquired phenomena of the target language.

Phonological phenomena that are neither marked nor similar between the target and native language of the learner are acquired at a faster rate, and the course of acquisition is affected by the universals more at earlier stages, while the

influence of the universlas is decreasing rapidly at later stages of L2 learning, and the influence of the native language is diminishing throughout the acquisition process.

These predictions have been made and tested again and again based on the dynamics of acquisition of L2 segmental structure, although this models lends itself to be easily tested based on acquisition of prosody and further generalized over other language layers like vocabulary, syntax, etc (Major & Kim, 1996)

2.2.2. Deviations in prosody and segments in perception of accentedness

As it has been mentioned at the beginning of the chapter, the perceived FA results from the production of specific segmental and prosodic characteristics that deviate from those produced by L1 speakers. Literature abounds in studies on whether prosody or segments make a greater contribution in perceived FA and how prosodic and segmental features interact to create the percept of accentedness. This issue is acutely debated in overflowing publications, however it is far from being resolved. Research in the production and perception of FA has focused on two questions: (I) do segmental and prosodic characteristics independently influence FA production and perception and (II) what is their individual relative contribution (e.g., see Boula de Mareüil and Vieru-Dimulescu, 2006; Brahimi, Boula de Mareuil, & Gendrot, 2004; Jilka, 2000; Magen, 1998; Munro, 1995 among multiple other publications). In the following part, overview of the studies on the contribution of segmental deviations into the FA when the influence of prosody is controlled for is given. This will allow discussing the unique, combined, and shared contribution of prosodic and segmental pronunciation features into FA perception.

Initially most studies on FA focused on segmental characteristics. For instance, Caramazza, Yeni-Komshian, Zurif, and Carbone (1973), Flege and Eefting (1987a; b), Port and Mitleb (1983) among others found voicing contrasts and voice onset time (VOT) – the duration interval between the release of the plosive consonant and the beginning of the vocal folds vibration indicating the onset of the vowel - to be different in L1 and L2 speech and between monolingual and bilingual speech.

Caramazza, Yeni-Komshian, Zurif, and Carbone (1973) studied VOT in speech of French monolingual speakers, English monolingual speakers and English-French bilinguals. They found the VOT to be the longest in speech of English monolinguals, shortest VOT was displayed in speech of French monolinguals, while bilinguals displayed intermediate VOT values. Bilinguals did differentiate between VOT in two languages – when they spoke French, VOT was shorter in bilingual speech than when they spoke English, yet English bilingual speech was characterized by shorter VOT than English monolingual speech, and French bilingual speech was characterized by longer VOT than monolingual French speech.

On the perception task, the two groups of monolinguals showed different crossovers in perception of voiced vs. voiceless plosives contrast – known to be manifested be differences in VOT between voiced and voiceless consonants – and these crossovers for bilinguals were also different from those exhibited by monolinguals. The authors concluded that language switching in bilinguals is well controlled for production but poorly controlled for perception at the phonological level.

Flege and Eefting (1987a; b) studied VOT in English spoken by Dutch (1987a), Spanish (1987b) speakers, Port and Mitleb (1983) investigated VOT produced by Arabic speakers of English and came to the conclusion that L2 speech is similar to bilingual speech in respect that the VOT values in L2 speech were intermediate between those typical of monolingual speech produced by native speakers of the target language. Monolingual speaker were also able to detect these deviations in L2 speech and reported them as non-native, i.e., contributing to the perceived FA.

Flege, Schirru and MacKay (2003), Kehoe (2002), Flege, Bohn and Jang (1997) investigated the differences between L1 and L2 production of vowels or compared monolingual and bilingual vowel productions. Flege, Schirru and MacKay (2003) investigated vowels in L2 English produced by Italians who were emerged into English-speaking environment early in life (around 7 years in age) or later in life (around 20 years in age) and those who used native language frequently (around 50% of speaking time Italian is used) or rarely (less than 10% of time Italian is used). The authors focused on realization of /ei/ diphthongs in monolingual English speech as compared to L2 English as produced by Italians

belonging to one of the above-mentioned four groups. They found L2 vowels are different from L1 vowels, but where the differences lie depends on the amount of L1 use and the age of arrival. Italians who arrived to Canada early and used primarily English from the time of arrival exaggerated the diphthongal glide and produced a significantly larger movement of the tongue than monolingual English speakers. The authors take it as an indicator of the fact that in early age the participants managed to form a new category, non existent in their native language, for the diphthong /ei/. Perceptual difference between Italian /e/ and English /ei/ lead to establishing a new category, and L2 learners deemed it important to underline this dissimilation by a greater tongue movement, greater than in monolingual English speech, and this is perceived as exaggerated by the native speakers of English. Late bilinguals failed to establish a new category and merged Italian /e/ and English /ei/ into one, which resulted in significantly less movement compared to monolingual English speech. Both merger of the phonetic categories and exaggerated dissimilation are deviations from the expectations of the native English speakers and contribute to the perceived FA.

Flege, Bohn and Jang (1997) compared vowel productions of monolingual English speakers with those in L2 English speech produced by German, Spanish, Korean and Mandarin speakers and found differences between L1 and L2 English in vowel productions for all groups of English learners. These segmental differences were diminished with experience in L2 and also depended on the native language of English learners, but never disappeared, and consequently these segmental differences from native norms or expectations also contributed to the perception of accentedness.

Kehoe (2002) studied length vowel contrasts in German vowels produced by German monolingual children and by German-Spanish bilingual children. She found that even when the children were brought up from birth in bilingual environment and had more exposure to German than to Spanish, they acquired German vowel length contrasts later than their monolingual peers and the length contrasts in monolingual children speech were more substantial than those in bilingual children speech. Again, bilingual speech differs from the expected norms of monolingual speech on the segmental level.

Flege, Takagi and Mann (1995) investigated the differences in liquid production and perception in L1 and L2 speech and found that the segmental

differences between monolingual English and L2 English produced by Japanese speakers in production of liquids do contribute not only to the degree of the perceived FA, but also to the comprehensibility of L2 speech.

Although the research into the influence of segmental deviations in L2 speech from the expected pronunciations by L1 speakers of the target language has been started very early and resulted in widely accepted models of language learning (see previous chapter), many more recent studies have emphasized the role of prosody and suprasegmentals in creating the percept of foreign accent. The researchers also tried to address the question whether deviations at the segmental or at the suprasegmental level play a major role in perception of foreign accent, what is the relative contribution of prosody and segments into assessment of the accentedness, and how the deviations at different levels interplay in L2 speech, enhancing or impeding the degree of the perceived foreign accent.

Derwing and Munro (1997), Derwing and Rossiter (2003), Field (2005) and others, for example, have shown that an improvement in the prosodic rather than in the segmental level of speech, results in increased comprehensibility and reduced perceived FA. Field (2005) found that the errors in lexical stress location affect intelligibility of English speech by both native and L2 listeners. Bond (1999) showed that the contribution of stress to FA is greater than that of segmental errors, at least for English listeners. Derwing and Munro (1997) found that prosody has a greater contribution than segmental errors on intelligibility and accentedness of English speakers whose first language is Cantonese, Japanese, Polish and Spanish. Derwing and Rossiter (2003) split 48 learners of English as second language into three groups. One group received pronunciation training to improve prosody, the second group received pronunciation instruction to improve pronunciation at the segmental level, and the third group did not receive any pronunciation training. The L2 learners were evaluated by expert phoneticians and native speakers of the target language before training and after 12 weeks of training. The authors found that improvement at the level of prosody resulted in a significantly and substantially greater reduction of accentedness and improvement in fluency and intelligibility.

Boula de Mareüil and Vieru-Dimulescu (2006) used prosodic transplantation between Italian and Spanish L1 and L2 speech to the stimuli with L1 prosody but L2 segments and vice versa. The stimuli were evaluated for the degree of foreign

accent in order to find the relative contribution of segmental and prosodic deviations from the native norms into perceived FA. The authors concluded that prosody is important in the identification of FA 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), based on 60 non native speakers from eleven different L1 backgrounds, explicitly stated that prosodic characteristics make a larger contribution to the perception of FA. They investigated the relationship between judgments of nonnative pronunciation and actual deviations from the native pronunciation norms in segments, prosody, and syllable structure. It was found that although all three components – segmental and prosodic deviations and deviations in syllable structure – were important in identifying the degree of the perceived FA, the contribution of prosody was the largest overall and for each of the eleven L1 backgrounds (although the relative contribution of segmental deviations and errors in syllable structure differed between L1 backgrounds).

Tajima, Port and Dalby (1997) as well as Maassen and Povel (1985), on the other hand, claimed that segmental features provide more cues to accentedness and influence intelligibility.

Maassen and Povel (1985) recorded 10 deaf Dutch children pronouncing each 30 Dutch sentences. The sentences were further digitally manipulated to transform separate speech sounds, temporal structures and durational relationships between speech constituents, intonational contours. Segmental correction of sentences produced by deaf children had a substantially greater effect on intelligibility of these sentences (from 24% to 72%) – which was mostly due to correction of the vowel formants – than correction of temporal relationships or intonational contours (24% to about 34%). Combination of segmental and prosodic corrections has an additive effect and improves intelligibility to almost 100%. The authors conclude that segmentals have a greater influence on intelligibility than suprasgmentals. As the degree of accentedness is closely related to intelligibility – the higher the degree of the perceived FA, the lower the intelligibility of speech is (James, 1998) – segmentals might have a more substantial effect on the perceived FA than prosody.

Tajima, Port and Dalby (1997) made temporal correction of Cantonese-accented English and found that intelligibility of foreign-accented speech can be improved if explicit training is provided on temporal properties of their speech. However, the temporal correction cannot override the impeding effect of deviations at the segmental level on intelligibility of accented speech.

Magen (1998) investigated the contribution of initial epenthetic schwa, non-initial epenthetic schwa, vowel reduction, keeping lax-tense vowel opposition in speech, final /s/ deletion, manner (/t͡ʃ–ʃ/), fricative voicing (/z-s/), stop voicing, lexical stress and phrasal stress position and phonetic realization into perceived FA in English produced by Spanish learners. English-speaking listeners rated the extent of foreign accent of the phrases by Spanish learners before and after acoustic corrections were performed to the accented productions with those of the the English native pronunciation norms. She found that the degree of the perceived FA was affected by the factors pertaining to the syllable structure and stress at the prosodic level as well as by the manner of consonant realization at the segmental level (but not affected by voicing differences). She did not conclude that either prosody or segmentals make a greater contribution into the degree of the perceived FA.

Based on this short literature review it cab be concluded that the issue of the relative importance of prosody and segments in perception of FA has not yet been resolved, and needs further examination. To shed more light on the combined, shared and unique contribution of prosody and segments into the percept of FA, some researchers decided to focus on separate prosodic systems. They concentrated on how the degree of FA is affected by intonation as opposed to other prosodic systems and segments (Jilka, 2000), stress location and realization (Bond, 2000), pitch span (Eckert & Laver, 1994), prosodic timing (Tajima, Port, & Dalby, 1997) and other prosodic systems. The next section gives the review of the studies that focussed on separate prosodic systems rather than on prosody as a whole, in an attempt to estimate the relative contribution of separate prosodic subsystems into the percept of the FA, and to analyse how separate prosodic systems (e.g., stress or intonation or tone) interplay with segments in perception of L2 speech.

2.2.3. Unique contribution of separate prosodic systems into foreign accent

Rather than trying to account for the contribution of prosody vs. segments, a few studies have tried to tackle the separate and unique impact of different prosodic systems (intonation, stress patterns and pitch accents, temporal characteristics, etc.) into the FA.

Hahn (2004) investigated the influence of location and presence of lexical stress on intelligibility of L2 speech by English speaking international teaching assistants for L1 English speaking undergraduates. Her results show that correct stress placement allows undergraduates to better recall the material later, makes L2 speech more intelligible and leads to more positive evaluation of the assistant by the students, and evaluation of their speech as less accented and more comprehensible. Field (2005) also revealed that non-native stress patterns handicap L2 speech intelligibility for L1 listeners of English and result in a higher degree of perceived FA. The result pattern was very similar to that of Hahn (2004) – intelligibility and accentedness was highly affected when stress patterns were deviant and that L2 speech with correctly placed primary stress was evaluated more favourably than L2 speech with missing or incorrectly placed stress. Bond (1999) investigated the contribution of stress location and stress distribution into the perceived FA and compared that with the contribution made by deviations at segmental level and deviations in other prosodic systems. He concluded that the contribution of stress to the FA is greater than that of segmental errors, at least for English listeners, but he did not make any inference regarding the contribution of stress in relation to intonation or other prosodic systems.

Another dimension of prosody in which deviations from the native norms are common – albeit less easy to pinpoint where exactly the deviance lies with naked ear, yet very easy to pick up that there is a deviance– is the pitch range. Pitch range can be analysed in two domains: pitch span – the range of frequencies covered by the pitch fluctuations – and pitch level – the overall height of speaker's voice (Ladd, 2008: 192-210).

Pitch range can be a source of cross-linguistic differences in addition to other sources of tonal differences between languages including the form of salient events in the F0 contour, the timing of these events and alignment of tonal events with the segmental string, as well as the relations between the form and the

function and meaning of different contours. For example, Germans are sometimes reported to sound as unfriendly and dull by the British English speakers, and British English sound as overexcited to German speakers (Eckert & Laver, 1994; Gibbon, 1998).

Pitch range in L1 and L2 both in the domains of the pitch span and pitch level was tackled by Mennen (2007) who empirically confirmed that 1) speakers of different languages indeed might reveal significant differences between in pitch, and 2) there are differences in pitch span and level between L1 and L2 in German and in English (speakers use different pitch range when they speak different languages, but the pitch range of the L2 speakers also differed from that of L1 speakers). These findings were confirmed by Scharff–Rethfeldt, Miller and Mennen (2008). The differences in pitch range can be attributed to the difference in distribution of the pitch accents in L1 German and L1 English (Mennen, Schaeffler, Docherty, 2012), and to the transfer of distributional differences in pitch accents from L1 into L2. Binghadeer (2008) carried out comparisons of pitch range in speech produced by Arabic L2 speakers of English and L1 speakers of English and found that the pitch range produced by the L2 speakers was narrower in sentences with falling intonation. This can be interpreted as realizational differences in pitch range between L1 and L2 produced by Arabic learners.

Mennen, Schaffler and Docherty (2012) showed that the differences in pitch range between English and German are related to the cross-linguistic differences in the intonational structure, not in the pitch range per se. In particular, the first initial peak H* is higher in English than in German, while non-initial H* tones are, on the contrary, higher in German. Although the study did not reveal any differences in the inventory of the tones, the authors found cross-linguistic differences in the phonetic realization of intonational primitives depending on their position in a sentence. Another difference is the low number of L* tones in English and higher number of such tones in German, which is a distributional difference between intonational systems of these two languages. As the differences in pitch range between languages are easily perceived and even give rise to anecdotal evidence, yet these differences are related to the intonational structure, it is not surprising to also have evidence of intonation making a contribution into perception of foreign accent. Some researchers focussed exclusively on the contribution of intonation into the perceived FA. Jilka (2000) examined the relative

contribution of intonation, prosody (excluding intonation but including rhythmic and intensity factors) and the segmental characteristics to perceived FA. He did this research on the basis of data obtained from native German learners of American English and American English learners of German. He found differences in the distribution and types of pitch accents in L1 and L2 in German and in English, and carried out a series of perception tests to prove that these acoustic manifestations in L2 speech are indeed heard by the L1 speakers of German and English and influence the degree of perceived FA. Native listeners of each language were presented with three types of stimuli: a low-pass filtered, monotonized and non-manipulated speech and asked to identify the language and accent of the speakers. Results showed that listeners were able to identify the speakers' language on the basis of purely prosodic information, and that prosody was relevant in the recognition of FA. Furthermore, low-pass filtered stimuli with monotonous intonation attracted higher FA ratings than those with preserved intonation suggesting that intonation is the most important prosodic cue in perceptions of accentedness in L2 speech. 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 actual differences in intonation in L2 French produced by Japanese and English learners.

Kang, Rubin and Pickering (2010) investigated a whole bunch of suprasegmental features and their relative contribution to the strength of the accent and the judgments of the learner's proficiency in English. 60-second recordings of 26 male learners of English from various L1 backgrounds (Arabic, Korean, Spanish and Chinese) were judged for oral proficiency and comprehensibility by 188 English speaking undergraduates. 29 different acoustic variables including measures of speech rate, stress, pitch, the frequency, type and duration of pauses, etc. were selected for analysis. Multiple regression analysis was used to assess the contribution of each individual measure and different clusters of acoustic variables to the strength of the FA accent, comprehensibility and oral proficiency. The authors found that prosody, in general, accounts for 50% of the variance in accent ratings, and among the most influential prosodic factors are suprasegmental fluency (the construct composed of speech rate, articulation rate, speech-to-pause ratio, length of speech units between boundaries), followed

by the right choice of the pitch accents (esp. 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 the prosodic characteristics accounted for 41% of variance in the strength of the accent rating in L2 speech. The most substantial contribution to an accent rating was made by pitch span (L2 speech with 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 the authors did not include durational cues in their analysis (except for fluency measures, e.g. speech tempo measures).

Gut (2007) analyzed L2 English (produced by learners with 17 different L1s) and L2 German (produced by learners with 24 different L1s) and found that the only acoustic features which correlated with the FA ratings obtained from L1 speaker of the target languages were temporal features, i.e. the length of stressed syllables, the length of reduced syllables, the articulation rate. Pitch range, vowel reduction and vowel reduction measured in the ratio between unreduced and reduced syllables did not correlate with the perceived accent. Other prosodic features, neither temporal nor pitch, correlated with the FA ratings received by L2 learners of English.

The reviewed studies show that there is no consensus regarding the relative contribution of prosodic and segmental characteristics, or regarding the unique contribution of separate prosodic systems (like stress, intonation, pitch range, pausation, i.e., pause durations, types, distribution and location, frequency) into perceived FA. Even less is exactly known regarding the interaction of prosodic features with the segmental material in the perception of the strength FA. Therefore, we intend to further expand our knowledge in this domain and study the separate contribution of prosodic timing patterns into the perceived FA. The aim is to determine the relative strength of contribution of different aspects of prosodic timing patterns and how the contribution of timing patterns into perceived FA is interrelated with the contribution of prosody in general and also with that of the segments.

The contribution of prosodic timing patterns into the FA is investigated in this study. Timing patterns is a general term that embraces, in addition to phonemic durational contrasts, speech rhythm (when defined as durational

variability of speech intervals), phrase-final lengthening, tempo characteristics, polysyllabic shortening and lengthening and other phenomena. Timing patterns can be considered as a prosodic system in its own right or as a set of phenomena related to separate prosodic systems including stress patterns, rhythm and tempo, system of boundary signals marking phrase, utterance and word edges. What will be in the focus of our investigation are the durational variability of syllables, vocalic sequences and consonantal clusters, and the mean duration of the above-mentioned speech intervals. Mean durations are tempo-related characteristics. Durational variability creates the auditory impression of language-specific rhythm. Consequently, the contribution of *speech rate* and *speech rhythm* into the FA is the focus of this research.

2.3. Issues of Rhythm and Timing in Foreign Accent

Very few studies aimed to estimate the unique contribution of timing into the perceived FA. The word 'unique' refers to the contribution that is not overlapping with the contribution made by tonal and segmental characteristics of speech, to the perception of foreign accent. Amongst the most widely studied temporal aspects of L2 speech and their influence on the accentedness are speech rate (measured in syllables per second) and articulation rate (measured in syllables per second excluding pauses). However, other timing patterns might also contribute to the degree of foreign accent in L2 speech, and they, unfortunately, remain underresearched.

The differences in timing patterns between L1 and L2 add to the perception of foreign accent (Polyanskaya, Ordin, & Ulbrich, 2013) and impair intelligibility of L2 speech (Tajima, Port, & Dalby, 1997). As it has been said above, in the focus of this study are the prosodic timing patterns pertaining to the speech rate and speech rhythm (more details on the interrelations of speech rhythm and timing are provided further in this section). The hypothesis that deviations in speech rhythm add to the perceived foreign accent has initially been proposed by Adams (1979) and Taylor (1981). The authors suggest that speech rhythm is a fundamental organizing principle in speech. Rhythm is reconstructed by the listeners based on clues provided by the speaker. The failure to provide a sufficient number of clues to enable the listener to extract and recognize specific rhythmic patterns reduces the intelligibility of L2 speech and increases accentedness (Taylor, 1981: 224-

225). Taylor (1981) and Adams (1979) emphasize the importance of adequate production of speech rhythm for accent reduction and intelligibility.

Although acquisition of native-like rhythmic patterns is one of the most difficult aspects of mastering English pronunciation (Adams, 1979), very few studies have directly focused on rhythm development in second language acquisition, and we are not aware of any study that has directly compared the development of the speech rhythm in L2 and in L1 acquisition (Ordin & Polyanskaya, 2014). This comparison is deemed useful because it is still not known to what degree the development of rhythmic patterns in L2 is determined by the native language of the learner. In particular, it is not clear which aspects of rhythm development are universal and thus shared by L1 and L2 learners, and which aspects of rhythm development are due to L1-transfer and language-specific rhythmic patterns of the learner's native language. We are not aware, to the best of our knowledge, of any study so far that compares the development of L2 rhythm in speech of learners with rhythmically contrastive native languages by investigating the changes between different proficiency levels in the target language in different groups of learners.

Considering theoretical and practical importance of understanding the impact of timing on foreign accent, my empirical studies, shedding light on the contribution of timing patterns into perception of foreign accent in L2 speech, seems timely and functional.

2.3.1. Notion of Timing Patterns

The word *timing patterns* further refers to the range of durational cues that might differ between languages and consequently between L1 and L2 speech in the same language. Durational cues mark segmental distinctions (e.g. voicing or difference between phonologically long and short vowels), location of word boundaries in continuous speech, phrase boundaries, lexical stress, focus/topic distinctions, etc (e.g., Klatt, 1976; Quene, 1992; Bion, Benavides-Varela & Nespor, 2011).

Duration patterns convey much of linguistically-relevant information in a language-specific way, therefore duration patterns may express phonological and phonetic distinction between languages. For example, duration patterns may discriminate between short and long vowels within one language (e.g., German,

English), while this distinction might not be present in another language (Italian, Spanish). Cross-linguistic difference are referred to as phonological, i.e. relating to the language structure. There might also be differences in how a certain phonological durational contrast is realized in speech. For example, final lengthening – the increase in duration of speech constituents like syllables or words in the vicinity of the right edge boundary with lengthening proportional to the boundary strength (Turk & Shattuck-Hufnagel, 2007) – is a universal phenomenon, yet its phonetic realization differs between languages. The presence of final lengthening has been attested in all languages in which it was sought after, but at the same time the domain of phonetic implementation, peculiarities of lengthening, units to which the lengthening is applied (syllables, segments, feet) and the degree of lengthening is language-specific. See, for example, Nakai, Turk, Kari, Granlund, Ylitalo, and Kunnari (2012) for Finnish; White & Turk, (2010) for English; Cambier-Langeveld, Nespor, and van Heuven (1997) for Dutch; Frota (2000) for Portuguese; Gorka, Frota, and Vigário (2005) for Spanish and Portuguese; D'Imperio, Elordieta, Frota, Prieto and Vigário (2005) for several other Romance languages. The differences in how the same phenomenon is realized are referred to as phonetic differences between languages further in this dissertation. As the speakers might transfer phonetic peculiarities of realizations of durational patterns from their native language into L2, and as the L1 and L2 of the speaker might feature different phonological durational patterns that L2 speakers need to master when learning the language, it is not surprising to see the differences between L1 and L2 in timing patterns. It is also expected that these timing differences will be detected by the L1 speakers of the target language and contribute to the degree of the perceived FA in L2 speech.

Such language-specific timing patterns led many researchers to the assumption that the perception of speech rhythm, for example, arises from a combination of purely phonetic and phonological factors that enhance or inhibit variation in the duration of syllables, vowels and consonantal clusters. For example, the lack of vowel reduction in unstressed syllables and the relatively less significant role of increase in duration to manifest stress cause smaller differences between stressed and unstressed syllables. For example, if vowels are not reduced in unstressed syllables, this will result in smaller differences between stressed and unstressed syllables. In some languages (e.g., Japanese), strict

phonotactic constrains lead to predominantly CV (consonant-vowel) syllables, and/or only few and shorter consonantal clusters, which lowers the difference in duration of adjacent syllables compared to the languages that allow more complex syllables (e.g., CCVCC). If a language has both short and long vowels, it will also enhance durational variability of syllables and vocalic speech intervals. Current studies have shown that these patterns of durational variability contribute to the perception of rhythm differences between languages. We have investigated how such timing patterns influence the perception of FA in L2 speech, how these patterns interplay with speech rate characteristics in perception of rhythm and in perception of accentedness, and what is their relative contribution into FA compared to the overall contribution of prosody and of the deviations from the expected norms at the segmental level.

2.3.2. From Timing to Durational Variability and Rhythmic Patterns

Timing patterns are perceptually very salient to the listeners. Empirical evidence suggests that not only adults (Ramus & Mehler, 1999), but also infants might discriminate languages with different patterns of durational variability of speech intervals (Nazzi & Ramus, 2003). It has been shown that adults (Ramus & Mehler, 1999) and even infants (Ramus, Nespor, & Mehler, 1999; Nazzi & Ramus, 2003) can differentiate rhythmic patterns of the languages that are traditionally labeled as stress-timed (e.g., Russian, English, German, Dutch) and syllable-timed (e.g., French, Spanish, Italian).

The words stress-timed and syllable-timed refer to the rhythmic classes. The word “rhythm” is one of the most ambiguous terms in linguistics and it causes a lot of intense debates, often because the researchers mean different things when they talk about speech rhythm. In a review article, Turk and Shattuck-Hufnagel (2013) illustrate the range of topics the term rhythm covers and conclude that researchers in speech sciences can understand rhythm differently. We will briefly provide different ideas researchers specify when they use the word *rhythm*, and then we will define what we will understand by *speech rhythm*.

The word rhythm bears an implicit assumption of periodicity and isochrony (such as in music, where certain patterns re-occur at regular intervals). This assumption led James (1940), Pike (1945), Abercrombie (1967) and Ladefoged

(1993) to dividing the languages into stress-timed (in which intervals between stressed fragments are thought to be of equal duration, e.g., German, English, Dutch, Russian), syllable-timed (in which syllables are thought to be of equal duration, e.g., French, Italian, Spanish) and mora-timed (Japanese, Austronesia languages like Gilbertese and Hawaiian, there is evidence that Finnish and West Greenlandic can also exhibit mora-timed rhythm). This distinction was made only on auditory impressionistic analysis, and acoustic measurements reported in later studies failed to support to this claim (Dauer, 1983; Roach, 1982; Pamies Bertran, 1999). Nevertheless, psychological reality of rhythm and the capacity of humans to discriminate rhythmic patterns of languages that are traditionally considered to be stress-timed and syllable-timed has been experimentally confirmed (Bosch & Sebastian-Galles, 1997; Bertoncini, Floccia, Nazzi, & Mehler, 1995; Ramus & Mehler, 1999; Nazzi & Ramus, 2003; Nazzi, Bertoncini, & Mehler, 1998). Adults, children and even neonates can discriminate between the rhythm patterns of rhythmically different languages (English vs. French or German vs. Italian) and cannot distinguish between those of similar languages (even if both languages are not native to them). As languages with distinctly different rhythmic patterns also possess different morphological and syntactic properties and syllable structures, it was hypothesized that rhythm might be something that helps infants to extract and acquire such linguistic properties like word order, phonetic features, word structure and morphological type of language (Ramus & Mehler, 1999; Nazzi & Ramus, 2003; Nazzi, Bertoncini, & Mehler, 1998). Speech rhythm might also play a role in segmentation strategies i.e., extracting discreet units like words and phrases from continuous speech stream that does not always have clear pauses or other unambiguous cues to the word edges (Cutler & Butterfield, 1992; Smith, Cutler, Butterfeld, & Nimmo-Smith, 1989). Speech segmentation strategies differ between speakers with syllable-timed and stress-timed native languages (Murty, Otake, Cutler, 2007; Kim, Davis, & Cutler, 2008). These results support the Rhythm Class Hypothesis stating that languages can be categorized into distinct rhythmic classes based on their rhythmic patterns, and these classes also correspond to a number of linguistic properties. The exposure to languages with certain properties determines how speakers handle the complexities of their native and also foreign languages. As psychological reality of rhythm and the fundamental role of rhythm

in language acquisition and in speech processing have been confirmed, the search for concrete acoustic correlates of rhythm has never been abandoned.

Recent studies have demonstrated that the acoustic correlates of speech rhythm can be found in systematicity of timing in the speech signal, which defines speech rhythm as a purely phonetic rather than phonological property. Systematicity in phonetic surface timing is described by patterns of variability in duration of syllables, vowel sequences and consonantal clusters. The variation in duration corresponds to patterns of alternation of more salient segments in speech stream against the background of less salient ones because duration is one of the major correlates of stress and prominence.

Durational variability of speech intervals is an interplay of several other durational cues which emerge from language properties like syllable complexity, phonotactic constraints, reduction and assimilation processes, phonological opposition between long and short vowels and between geminate and non-geminate consonants, vowel harmony, final lengthening and others (Dauer, 1983; 1987). Dauer (1987; 1983) linked these language-specific properties to the auditory impression of speech rhythm which differs between the so-called “stress-timed” languages and “syllable-timed” languages.

Variation in duration of speech intervals is influenced by many factors that may be summarized under three headings. The most evident factor is phonology, including phonotactic constraints on allowed consonantal clusters, assimilation processes that allow for cluster simplification, syllable complexity, presence or absence of phonological opposition between short and long vowels and single and double consonants (non-geminates and geminates respectively), vowel reduction in unstressed syllables, free or fixed location of lexical stress in a word, etc. Variation in duration can also be influenced by purely phonetic properties like final lengthening, polysyllabic and polysegmental shortening, magnitude of durational increase of stressed syllable, the relative roles of duration and f0 in manifesting prominence at different levels (phonological word and phrase), tempo fluctuations, etc. For example, final lengthening is assumed to be universal because it has been attested in every language that has been studied for it, but phonetic implementation of final lengthening and the domain over which final lengthening operates differ across languages (Nakai, Turk, Kari, Granlund, Ylitalo, & Kunnari, 2012 for Finnish; White & Turk, 2010; Turk & Shattuck-Hufnagel, 2007; Wightman,

Shattuck-Hufnagel, Ostendorf, & Price, 1992 for English; Cambier-Langevelt, Nespor, & van Heuven, 1997 for Dutch; Frota, 2000 for Portuguese and Spanish; D'Imperio, Elordieta, Frota, Prieto, & Vigario, 2005 for French and other Romance languages). A detailed discussion of these and other phonetic and phonological factors that enhance or inhibit variation in duration, or that correlate with stress- and syllable-timing (e.g., presence of vowel harmony is correlated with lower duration variability and therefore with syllable-timing) can be found in Dauer (1983) and Schiering (2007).

There are also non-linguistic factors that have an impact on the rhythm measures. Loukina et al. (2013) found that speaker-specific timing strategies influence the metrics. Ordin and Polyanskaya (2014) showed that durational variability changes in child speech as first language acquisition progresses and partially relate these changes to the development of motor control. A complex interplay of phonetic, phonological and non-linguistic (e.g., cognitive) factors gives rise to the emergence of language-specific timing patterns that can be captured by the rhythm metrics, and also to the perception of speech rhythm that is based on systematicity in durational variation of speech intervals and to perception of speech rhythm.

A number of the so-called rhythm metrics were suggested in order to capture the durational variability of speech intervals. Languages which allow higher variability in segmental durations exhibit vowel reduction, more complex syllabic clusters and more versatile syllable types, opposition of long and short vowels, germinate and non-germinate consonants, and are less likely to exhibit vowel harmony and fixed stress. Such languages also create the impression of "stress-timing" (Dauer, 1983; 1987; Schiering, 2007).

Most of these metrics describe rhythm with durational measurements. Among the most widely used rhythm measures are the overall proportion of vocalic intervals in the utterance (%V, in percent to the duration of the whole utterance), standard deviation of vocalic and consonantal intervals (ΔV and ΔC accordingly), coefficient of variability of vocalic and consonantal durations (VarcoV and VarcoC), normalized and raw pairwise variability index in duration of vocalic (nPVI-v and rPVI-v) and consonantal (nPVI-c and rPVI-c) intervals. Below I will review these metrics and provide interpretation of what these metrics might stand for.

Ramus, Nespor and Mehler (1999) and Ramus and Mehler (1999) suggested that adults and infants can differentiate languages belonging to different rhythmic classes based on standard deviation of vocalic and consonantal intervals and the vocalic proportion of the total utterance duration (ΔV, ΔC, %V accordingly). Dellwo and Wagner (2003) and then Dellwo (2006) suggested VarcoV and VarcoC metrics, which are calculated by dividing the standard deviation of consonantal or vocalic interval durations by the mean consonantal or vocalic duration in the sentence. These metrics capture variability in the entire sentence.

Grabe and Low (2002) proposed pairwise variability indices that capture the difference between successive intervals pair by pair. The normalized PVI (1) is calculated by dividing the difference in duration within each pair of successive intervals by the mean duration of this pair, summing these differences, and dividing the sum by the number of pairs. Calculating raw PVI (2) does not involve the normalization factor. PVI metrics better capture variability in duration of successive intervals, and thus, arguably, better capture auditory impression of durational variability.

$$nPVI = 100 \times \left[\sum_{k=2}^{n} \left| \frac{d_k - d_{k-1}}{(d_k + d_{k-1})/2} \right| /(n-1) \right] \quad (1).$$

$$rPVI = 100 \times \left[\frac{\sum_{k=2}^{n} \left| d_k - d_{k-1} \right|}{n-1} \right] \quad (2),$$

where

n – number of interval in an utterance for which PVI is calculated,

d – duration of k-th interval.

The Control and Compensation index (CCI) suggested by Bertinetto and Bertini (2008) measures a language-specific degree of segmental lengthening and shortening according to the context. The authors argued that languages, which are traditionally classified as syllable-timed, allow a lower degree of compression and maintain segmental durations in the unstressed position with higher precision compared to languages which are traditionally classified as stress-timed. Thus they called the latter controlling languages (because these languages control segmental durations regardless of the stressed/unstressed position), and the former – compensating languages (because they allow shortening in unstressed positions to compensate for lengthening in stressed syllables). The CCI index is calculated by the following formula (3):

$$CCI = \frac{100}{m-1}\sum_{k-1}^{m-1}\left|\frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}}\right|, \quad (3)$$

where *m* is the number of intervals in an analysed utterance, *d* is the (in ms) of k^{th} interval (vocalic or consonantal), and *n* is the number of segments within the relevant interval.

All these indices can be applied to various intervals (Nolan & Asu, 2009; Mok, 2011), e.g. syllables, vocalic and consonantal intervals within the utterance, successive feet, etc. It captures the average variability in the duration of the analysed intervals.

Higher values of %V and lower values of other metrics correspond to the auditory impression of "syllable-timing" (White & Mattys, 2007; Low, Grabe, & Nolan 2000; Ramus, Nespor & Mehler, 1999; Grabe & Low, 2002; Dellwo & Wagner, 2003).

Dauer (1983; 1987; Shiering, 2007) analysed the phonological structure of languages which give the impression of stress-timing or syllable-timing. Those languages which produce the effect of stress-timing display vowel reduction, more complex clusters, have more different syllable types, opposition between phonologically long and short vowels, between germinate and non-germinate consonants, are less likely to exhibit vowel harmony and fixed stress. Rhythm metrics can reflect some of these language-specific phonological properties.

To name a few examples, ΔC is thought to be indicative of the syllabic structure, syllable complexity and consonantal phonotactic constraints. ΔV is supposed to be indicative of the degree of vowel reduction. VarcoV and VarcoC reflect the same properties ΔC and ΔV do, but Varco measures are supposed to neutralize the effect of the tempo differences, and thus to reduce the effect of idiosyncrasies in speech production (see below in the next chapter). %V indicates the syllabic structure and inventory. Languages with more restricted syllabic inventory operate less complex syllables, usually of CV structure. The more types of syllables there are in the language inventory, the more consonants are added to the onset or coda of the syllables. This reduces the proportion of vocalic intervals to the overall duration of the utterance, and %V decreases.

A lot of other rhythm metrics have been suggested as well, a good review can be found in Loukina et al. (2013). However, only those metrics that are

discussed above have been used in this study because usefulness of these metrics has been empirically verified, the changes in these metrics will tell on perception, and their power to differentiate between languages with different rhythms has been attested. Besides, these metrics are among the most widely used ones, and therefore applying the same metrics in my study to new speech material will make the results of my experiments more comparable with the results of previous studies and will better fit into linguistic debate.

Although rhythm metrics have been successfully applied to study cross-linguistic differences, differences in speech rhythm between monolinguals and bilinguals, between adults and children, between L2 learners on different proficiency levels, between regional and stylistic varieties within the same language, the metrics have recently also have received much criticism. There has been a lot of debate regarding the usefulness of the rhythm metrics. For example, Arvaniti (2009, 2012) and Wiget et al. (2012) showed that metrics for the same language may be different depending on the speech material used to calculate the metrics and depending on the individual peculiarities of the speaker's speech production. Therefore it was claimed by Arvaniti (2012) that the same language may fall into different rhythm class depending on the speaker and the analysed utterance.

However, it makes sense to separate the language-specific (determined by the phonological regularities) and the speaker- or utterance-specific rhythmic patterns. For example, it is possible to construct sentences with an equal number of CV (consonant-vowel) syllables in English and in French, and the rhythm metrics calculated for these sentences will be more similar than one would expect from these rhythmically different languages. This will only mean that the utterances are rhythmically similar, but not that the languages are rhythmically similar. The frequency of CV syllables in French is much higher than in English. Sentences which include only CV syllables are unlikely in English but in French most syllables have a CV structure. Therefore, it is possible to fine or construct English sentences which exhibit more syllable-timed rhythmic characteristics, but such sentences are less frequent or less likely in natural speech compared to those exhibiting more stress-timed characteristics.

There were empirical studies supporting the usefulness of the rhythm metrics despite Arvaniti's criticism (2012). Prieto et al. (2012) controlled for

phonological differences in speech material representing English, Spanish and Catalan and found that rhythm metrics discriminate between rhythmically contrastive languages even when phonological properties and syllable complexity are controlled for. White, Payne and Mattys (2009) found differences in the rhythmic patterns of Venetian and Sicilian dialects of Italian, which they did not relate to phonotactics. White and Mattys (2007) and Gabriel and Kireva (2014) investigated the differences in speech rhythm between utterances delivered by L1 and L2 and revealed systematic differences in rhythmic patterns between L1 and L2 speech. These results reveal collective preferences shared within a certain linguistic community for suppressing or reinforcing durational variability by phonetic means. Therefore, despite the criticism the rhythm metrics have recently received, they still remain widely applied in linguistic research and, provided that the methodological considerations are accounted for, yield valuable insight into cross-linguistic or cross-interlanguage rhythmic differences.

2.3.3. Interaction of Speech Rate and Durational Variability

The patterns of durational variability that are captured by the so-called rhythm metrics differ between languages and are closely related to speech tempo characteristics, including speech rate (measured in syllables per second) and articulation rate (measured in syllables per second excluding pauses and hesitations). For example, some languages are reported to be faster than others. Dutch, for example, has the average speaking rate of 4.5 – 5.5 syl/sec, as reported by Quene (2007). Other Germanic languages including English (Clopper & Smiljanik, 2011; Anderson-Hsieh & Venkatagiri, 1994), German (Pellegrino, Coupe, & Marciso, 2011) share this relatively slow speaking rate. Speech in Romance languages is delivered at a faster rate exceeding the average tempo of 6 syl/sec, and this tempo is confirmed for French, Italian, Spanish (Dauer, 1983; Pellegrino, Coupe, & Marciso, 2011). Japanese is spoken at a rate of almost 8 syl/sec (Pellegrino, Coupe, & Marciso, 2011). Malisz (2011) report Polish to be delivered at almost 7 syl/sec.

This is also reflected in the changes of some rhythm metrics. Standard deviation of speech intervals, for example, decreases as the overall duration of these intervals decreases. That is, the faster the person talks, the shorter the

vowels, consonants and syllables are. The shorter these units are, the smaller the standard deviation of their durations is.

To compensate for the effect of speech rate, a normalization factor has been introduced into some of the metrics. Dividing standard deviation of syllable duration by the mean duration of the syllable in the sentence, for example, diminishes the effect of the speech rate. Thus Varco coefficients are more robust to the tempo variations compared to the Δ coefficients, while reflecting the same properties, namely, the syllabic structure, syllable complexity and consonantal phonotactic constraints, the degree of vowel reduction. Further studies have confirmed that Varco coefficients and %V are robust to the variations in tempo (White & Mattys, 2007; Wiget et al., 2010), and stay stable across many productions of the same speaker when the changes in the speech tempo happen. These metrics are also more stable across similar productions from different speakers in the same language than Δ coefficients.

Calculating the normalized PVI also involves normalization factor (see formula 1), which is the averages duration of the pair of successive units. This factor is introduced to compensate the idiosyncratic speech rate differences, thus this index is thought to be more robust and less influenced by tempo differences between speakers and utterances. The raw PVI (rPVI) does not involve the normalization factor (2), which makes rPVI sensitive to tempo variability.

The increase in tempo results in a significantly greater contraction of stressed vowels, while the duration of unstressed vowels and consonants is relatively unchanged. This makes the durational difference of stressed and unstressed vowels smaller, thus yielding a lower nPVI for fast speech compared to slow speech. On the contrary, in fast speech consonantal clusters in stressed and unstressed syllables are contracted to the same degree. This keeps the durational variability in consonantal intervals relatively stable. Therefore, in natural speech acceleration, as a rule, results to bigger modifications in durational variability for vocalic intervals than for consonantal intervals. That is why normalization for tempo is necessary for the vocalic intervals, and is less important for consonants. Besides, normalization of PVI on consonantal intervals also neutralizes the effect of the language-specific constraints on syllables and consonantal clusters, eliminating the differences of the rhythm metric between languages. That is why the normalized version of PVI is usually applied to the vocalic intervals (to

neutralize the effect of tempo on the value of the rhythm metric), while the raw version is applied to consonantal intervals (Grabe & Low, 2002; Low, Grabe & Nolan, 2000).

Interaction of speech tempo and durational variability also influences perception. Dellwo (2006) showed that the perception of stress-timing or syllable-timing is closely connected to the speech tempo. The faster the speech, the more syllable-timed it is perceived. It might be one of the reasons why Romance languages (delivered at faster rate) are perceived as more syllable-timed compared to Germanic languages (delivered at slower rate).

2.3.4. Development of Timing Patterns as function of proficiency growth

A lot of effort has been devoted to understanding how rhythmical characteristics are acquired in L2 and L1 in bilingual and monolingual acquisition (Grabe, Post & Watson 1999; Grabe, Gut, Post & Watson 2000; Low, Grabe & Nolan 2000; Whitworth, 2002; Work, Andruski, Casielles & Kim 2005; White & Mattys, 2007; Bunta & Ingram, 2007; Diez, Dellwo & Gavalda, 2008; Grenon & White, 2008; Shport, 2008; etc).

Generally it was found that although simultaneous bilingual children are able to differentiate rhythm patterns of two languages in speech production, their rhythm patterns are different from those of monolingual speakers (Bunta & Ingram, 2007). It should be noted that this result was not supported in the study by Mok (2011) who compared rhythm metrics in Cantonese and English in the speech of English-Cantonese bilinguals. Rhythm patterns were not different in bilinguals' speech when they switched from one language to another (except for rPVI on consonantal intervals). The discrepancies between the results in these two studies can be explained by the fact that Mok (2011) analyzed speech of children at three years of age, while Bunta and Ingram (2007) worked with 4- and 5-year old children. Besides, Bunta and Ingram (2007) tried to work with balanced bilinguals, while English-Cantonese bilinguals in Mok's study (2011) were Cantonese-dominant.

Both of these studies report that rhythmic patterns in the speech of monolingual children differs from early on, and Payne et al. (2012) even found that language-specific rhythmic patterns are already observable in the speech of 2-

year old children. They analyzed the speech of Catalan, Spanish and English children and showed that the rhythmic patterns in English were clearly distinguished from those of Catalan and Spanish already at the age of 2 years. As babies at the age of two almost exclusively use CV syllables, the differences in rhythm metrics between languages can be mostly attributed to phonetic language-specific timing patterns, which are evident in very early speech productions.

Based on the analysis of the previous research, it can be concluded that there appear to be similarities in how rhythm metrics develop as the first language acquisition progresses, regardless of the target language. Grabe, Post and Watson (1999) showed that French and English monolingual children have similar values of the Pairwise Variability Index closer to the values typical of syllable-timed languages, although French and English adults have drastically different rhythmic patterns. Later, English-speaking babies develop distinctive rhythm patterns different from their French speaking peers. That in turn tells on the values of PVI metrics of English and French speaking kids, which start to diverge. Payne et al. (2012) found that vocalic variability in duration increases from two to six years of age in the speech of monolingual Catalan, Spanish and English children. The values of the rhythm metrics in vocalic intervals grow with age, and %V decreases, which indicates the development towards more stress-timed speech in all three languages, although Catalan and Spanish speakers were clearly distinct from English in timing patterns from the age of two, exhibiting more syllable-timed patterns. But the values of the consonantal metrics develop contrary to what was predicted. They significantly and substantially decrease with age in all three groups of children.

Even when the languages which the babies acquire belong to the same end of the rhythm spectrum (e.g. German and English), babies first go through the stage of syllable-timed rhythm before developing their language-specific rhythmic patterns (Grabe, Gut, Post & Watson, 2000).

Whitworth (2002) investigated rhythm acquisition by English and German bilingual children. He showed that bilingual children also develop rhythm metrics from those typical of more syllable-timed languages to more stress-timed languages when both languages they acquire are stress-timed. Bunta and Ingram (2007) investigated rhythm acquisition by Spanish and English bilinguals and

found the same trend when the languages differ drastically in their rhythmic characteristics.

Lots of efforts have been focused on detecting the differences between L1 and L2 speech, and between L2 speech delivered by learners with the native languages either rhythmically similar to, or contrastive to the target language (White and Mattys, 2007; Gabriel, Stahnke, & Thulke, 2015a; Gabriel, Stahnke, & Thulke, 2015b; Ordin et al, 2015; Li & Post, 2014). The general finding is that the rhythm in L2 is determined not only by the transfer from the native into the target language, but also by the universals, thus L2 rhythm is not always in between the target and the native languages of the learner. Although some researchers place more emphasis on positive transfer (Gabriel et al., 2015a; b), others emphasize the influence of universal tendencies on rhythm development in acquisition (Li & Post, 2014; Ordin et al., 2015). Yet all researchers admit the influence of different factors, including the transfer and that of the universal tendencies, although the evaluation of contribution of these factors into the developmental trajectory of the L2 speech rhythm differs between studies. I believe that what we must account for in the future is the influence of language use, and this may explain why either transfer or universal factors are emphasized in different studies. For example, Gabriel et al. (2015 a; b) studies rhythm in French and English as a foreign language spoken by either monolingual German pupils, or pupils from German schools who use either Chinese or Turkish at home (trilinguals), and they detected the additive influence of the heritage language to the rhythmic patterns in the target foreign language. More multilingual exposure also means broader variety of rhythmic patterns the learners have been exposed to. Thus trilingual learners turned out to be more efficient compared to monolingual German learners of French (and English) in acquisition of speech rhythm of the target language. The study was carried out in multilingual classroom. In contrast, Ordin et al. (2015) concentrated solely on monolingual German and French learners of English, and the study was done in strictly controlled laboratory conditions, which could have inhibited the influence of positive transfer. The material analysed in Gabriel et al. (2015) could have inhibited the influence of universal tendencies. This could have led to different proportions of transfer and universals (e.g., markedness) effects on rhythmic patterns in a foreign language in these studies. In the future, I will account for such factors as language use and environment.

Ordin and Polyanskaya (2014) found that speech rhythm, as captured by the rhythm metrics, changes as a function of age as L1 acquisition progresses and as a function of length of residence in the target language environment in L2 acquisition. The direction of the rhythm development is the same in both groups of the acquirers, namely, from more syllable-timed towards more stress-timed, i.e. durational variability increases with age in L1 acquisition and with proficiency in L2 acquisition. Tempo (measured in syllables per second) also increases in both groups of acquirers as learning progresses.

Ordin and Polyanskaya (2014) compared development of durational variability in speech of children in L1 acquistion and in L2 acquisition by speakers with L1 that exhibits very different timing patterns of durational variability. They compared acquisition of English by monolingual English children and speakers from the so-called syllable-timed languages. Therefore the trend in L2 acquisition could potentially be explained by the transfer from the first language of L2 learners of English. However, Ordin, Polyanskaya, and Ulbrich (2011) and Ordin and Polyanskaya (accepted) show the same developmental trend, i.e., an increase in durational variability as a function of acquisition progress, in the speech of German adult learners of English, which means that L2 learners whose L1 is stressed-timed also undergo a similar developmental path in L2 rhythm acquisition. Ordin, Polyanskaya and Wagner (2015) used carefully controlled speech of French and German learners of English to compare acquistional paths of durational variability and tempo features by people from distinctly different L1 background, and again, the same developmental path was found in controlled speech elicited in laboratory settings from French and German learners of English. Therefore the authors suggest that this acquisition pattern is derived to a certain degree from rhythm universals in the process of language acquisition.

Ordin, Polyanskaya and Wagner (2014) seek additional support to this idea in parallels between rhythm development in language evolution and in language acquisition, i.e., between ontogenesis and phylogenesis of language development. Schiering (2007) gave an interesting review of language diachronic changes mirroring synchronic changes in speech rhythm development. Venneman (1988) shows how syllable complexity increases in Germanic languages via the loss of unstressed vowels. Vowel syncope increases syllable complexity during the transition from Old French (Horne, 1989). Blevins (2004) provides examples of

Australian languages in which syllable complexity is increased via vowel syncope and evolution of geminates. Syllabic complexity is a phonological correlate of speech rhythm (Dauer, 1983, 1987; Schiering, 2007), thus we can say that many Indo-European languages develop towards a higher degree of stress-timing in diachrony. Bybee, Chakraborti, Jung, and Scheibman (1998) analyze changes in stress patterns in languages and conclude that these changes within separate languages can be modeled as a shift from syllable-timing towards stress-timing. Based on this evidence, Ordin and Polyanskaya and Wagner (2014) suggest that rhythm development in interlanguage of an L2 learner and in L1 acquisition recapitulate the rhythm development in language evolution. As diachronic changes in speech rhythm in many regards mirror the rhythm development in L1 and L2 acquisition, there might be some universal principles which govern rhythm development from easier-to-acquire syllable-timed patterns towards more complex stress-timed patterns, in addition to those above-mentioned factors which are specific for L1 and L2 language acquisition.

Speech tempo increases with language mastery in L2 speech. At the same time duration variability also increases. Considering the interrelation of speech tempo and durational variability, durational variability is expected to decrease with the increase in speech tempo. The findings that durational variability and speech tempo evolve in the same direction, from lower to higher values. It can be concluded that the timing patterns in these two dimensions develop independently, and variability changes as a function of proficiency rather than as a function of speech tempo development. This is also confirmed that the development of durational variability is much better captured by the so-called rhythm metrics that are normalized for tempo. Durational variability in L2 speech increases as proficiency growth independently of the tempo changes and L1 of the learner.

2.3.5. Contribution of Timing Patterns into perceived Foreign Accent

Many studies have revealed the cross-linguistic differences in timing patterns (see above the differences in speech rate between Japanese, Polish, Germanic and Romance languages, as reported in Quene, 2007; Clopper & Smiljanik, 2011; Anderson-Hsieh & Venkatagiri, 1994; Pellegrino, Coupe, & Marciso, 2011; Dauer, 1983; Malitz, 2011). L2 speakers might use the timing

patterns of their native language when they speak L2. Besides, speaking in L2 demands more cognitive resources, which might lead to a slower speaking rate. L2 speech has frequently been shown to be generally slower than L1 speech (Guion, Flege, Liu & Yeni-Komshian, 2000; Lennon, 1990; Munro & Derwing, 1998), and to correlate with the proficiency level of the L2 speakers. Anderson-Hsieh and Venkatagiri (1994) found the mean articulation rate for L1 speakers of English is 5 syllables per second, the mean rate for highly proficient L2 English speakers is 4.4 and for the intermediate-proficient group is 3.3 syllables per second. These tempo differences are exceed the just-noticeable differences for speech tempo (Quene, 2007), and therefore the logical question to answer is whether the differences in speech rate affect intelligibility of L2 speech, whether they are perceived attentively, and if so, whether they influence the degree of perceived FA, or the perceived proficiency of the L2 learner. These questions were addressed in a number of studies that are reported below.

Anderson-Hsieh and Koehler (1988) investigated the effect of speech rate in L2 on comprehension and FA perception. They stated a close correspondence and correlation between the degree of the perceived FA and comprehensibility of L2 speech, i.e. more accented speech was perceived as less intelligible. However, they did not find a significant effect of speech rate on both the degree of the perceived FA and the degree of comprehensibility of L2 speech. The authors concluded that the speech rate of Chinese learners of English, although it increases with the proficiency growth, does not result in improvement of intelligibility or reduction of the degree of perceived FA. More proficient Chinese learners of English could sound faster and at the same time less intelligible and more accented than low-proficiency Chinese learners. This conclusion was also supported by the empirical evidence in Flege (1988), who did not find any influence of speech tempo on the perceived FA either.

On the contrary, Kang (2010) and Kang, Rubin, Pickering (2010) showed a significant contribution by the articulation rate to FA. However, this contribution was smaller compared to that made by other prosodic features, e.g. pitch range, mean length of silent pauses, etc. It should be noted that in these studies speech rate varied with other prosodic and spectral phonetic features because the authors used natural speech for analysis, and the obtained results are based on purely correlational research.

Munro and Derwing (1998; 2001) studied the effect of the speaking rate on the strength of the accent perception in experimental conditions. They found that in general speech delivered at a faster rate is judged less accented (up to a certain threshold), and the authors concluded that the effect was due to the rate differences themselves, and not to the differences in proficiency (although proficiency does vary with speech rate, as shown by Anderson-Hsieh and Venkatagiri, 1994). Munro and Derwing (2001) estimated that the optimal rate for L2 speakers to sound less accented is to speak faster than L2 leaners usually tend to do, but not necessarily as fast as L1 speakers. Very slow (2 syl\sec) and very fast (over 5 syl/sec) speech was rated more accented than speech delivered at the optimal rate for L2 learners (4.76 syl/sec). The authors manipulated speech tempo by compressing and expanding the duration of whole sentences. Changes in phonetics of segments, although inevitably induced by such manipulations, were not considered.

Differences in speech rate between L2 and L1 speakers cannot be attributed to perceptual assimilation, or to transfer / 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 production of non-native sounds and clusters. These factors are not accounted in neither of the L2 speech production models. The previous studies clearly showed that speech rate might influence the perceived FA, but its exact contribution, relative to the contribution of other prosodic and segmental aspects, is not very well understood.

Durational variability has also been reported to vary between languages, especially between languages that are percieived as prototypical representatives of distinct rhythm classes (Ramus, Nespor, & Mehler, 1999). Payne, Post, Astruc, Prieto, and del Mar Vanrell (2012) found that language-specific rhythmic patterns are already observable in the speech of two-year-old children. They analyzed the speech of Catalan, Spanish and English children and showed that the rhythmic patterns in English were clearly distinguished from those of Catalan and Spanish already at the age of 2 years. As babies at the age of two almost exclusively use CV syllables, the differences in rhythm metrics between languages can be mostly

attributed to language-specific timing patterns, which are evident in very early speech productions.

Empirical evidence also suggests that individuals are sensitive to durational variability, i.e., timing patterns captured by the rhythm metrics. The ability to attend to these duration cues is of utmost importance for language acquisition. Previous research has shown that babies use durational cues to discriminate languages of different classes (Ramus, Nespor & Mehler, 1999) and evidence has been provided indicating that these durational cues are used to bootstrap the syntactic properties of a language (Nespor, Guasti & Christophe, 1996; Mazuka, 1996). Furthermore speech rhythm has been shown to aid segmentation purposes (Christophe, Gout, Peperkamp & Morgan, 2003; Cutler & Butterfield, 1992; Morgan, 1996) as well as word extraction and learning (Thiessen & Saffran, 2007). Similar results were found for adults who were also able to discriminate languages belonging to different rhythmic classes (Ramus & Mehler, 1999).

Several studies also tackled whether durational variability in L1 and L2 speech differs. Bond and Fokes (1985) found, for example, that syllable durations in L2 English speech was less variable than in L1 English speech, so it is concluded that durational variability of syllabic duration, captured by the so-called rhythm metrics applied to the syllabic durations, differs between L1 and L2 speech. Baker, Baese-Berk, Bonnasse-Gahot, Kim, Van Engen and Bradlow (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 a greater reduction in function words), and higher between-speaker variance. They also found that these durational differences correlate with the degree of the perceived FA 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 by native speakers of English. These differences undoubtedly follow from the differences in durational variability of vowels and syllables between L1 and L2 speech, or variability in segmental durations between L1 and L2 speech. White and Mattys (2007) showed that there are differences between L1 and L2 speech in the same language in durational variability. A considerable number of other more recent studies have been carried out to investigate the differences

between L1 and L2 timing patterns that can be captured by rhythm metrics (e.g. Diez, Dellwo, & Gavalda, 2008; Grenon & White, 2008; Shport, 2008; Mok, 2011). The general finding is that although individuals can evaluate cross-linguistic timing differences in speech perception with relative ease, they cannot imitate language specific timing patterns in L2 speech production.

These patterns of durational variability are pertaining to the impression of speech rhythm. The hypothesis that the deviations in speech rhythm adds to the perceived FA has already been proposed by Adams (1979) and Taylor (1981). The results of both studies suggested that speech rhythm is a fundamental organizing principle in speech. Rhythm is reconstructed by the listeners based on clues provided by the speaker. The failure to provide a sufficient number of clues to enable the listener to extract and recognize specific rhythmic patterns reduces the intelligibility of L2 speech and increases perceived FA (Taylor, 1981: 224-225). These studies therefore emphasize the importance of adequate production of speech rhythm for accent reduction and intelligibility. However, as far as we are aware, no studies were carried out to date to estimate the contribution of rhythmic patterns to the perceived FA. Despite these very logical contemplation, very few perception studies have been conducted in order to see if the deviations from L1 rhythmic patterns in L2 speech are detected by the L1 speakers of the target language, and if they cause or contribute to the FA.

Tajima, Port and Dalby (1997) manipulated segmental durations in L1 English and in L2 English produced by Mandarin speakers. They concluded that the intelligibility of Chinese-accented speech improved by 15-25% when phonemic durations were warped to match native temporal patterns. The Intelligibility of L1 English utterances deteriorated by 15% when phonemic durations were corrected according to non-native patterns. These variations in duration of native and non-native segments could actually be captured by the so-called rhythm metrics, and therefore the differences in the values of the rhythm metrics between L1 and L2 speech could contribute to the perception of the degree of accentedness in L2 speech.

Quene and Delft (2010) 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 effect of

speaking rate differences between L1 and L2 were neutralized using 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 patterns reduced intelligibility and added to perceived FA.

Therefore non-native durational patterns and durational variability can influence the degree of the perceived FA and hinder intelligibility of L2 speech. Consequently, there are language-specific patterns of variability in the duration of segments, and the ear of the L1 speaker is tuned to this variability and probably can detect the deviations from the expected timing patterns.

2.4. Hypotheses, Predictions and Research Questions

The literature review shows that timing patterns referring to durational variability of speech intervals and to speech tempo (measured in syllables per second) differ between languages. See Ramus, Nespor, and Mehler (1999); Payne, Post, Astruc, Prieto, and del Mar Vanrell (2012); Bond and Fokes (1985); Baker, Baese-Berk, Bonnasse-Gahot, Kim, Van Engen and Bradlow (2011) for differences in durational variability between languages and Quene (2007); Clopper and Smiljanik (2011); Anderson-Hsieh and Venkatagiri (1994); Pellegrino, Coupe, and Marciso (2011); Dauer (1983); Malisz (2011) for tempo.

White and Mattys (2007) and Mok (2011) showed that durational variability differ between L1 and L2 speech in the same language. Speech tempo is also different in L2 speech from L1 speech in the same language, L2 learners exhibit slower speech rate compared to native speakers of the target language (Guion, Flege, Liu and Yeni-Komshian (2000); Lennon (1990); Munro and Derwing (1998).

Anderson-Hsieh and Venkatagiri (1994), Guion, Flege, Liu and Yeni-Komshian (2000); Lennon (1990), Munro and Derwing (1998) showed that more advanced L2 learners speak more rapidly than beginners and intermediate learners. Ordin, Polyanskaya, and Ulbrich (2011) and Ordin, Polyanskaya (2014) for durational variability proved that durational variability increases with proficiency growth in L2 speech, regardless of the native language of the learner and the target language, and regardless of whether the native and the target language of the learners are similar or different in regard to how timing patterns are displayed.

Deviations in speech tempo in L2 speech from native norms influence the degree of the perceived foreign accent (Munro & Derwing, 1998; 2001; Kang, 2010; Kang, Rubin, & Pickering, 2010). Deviations in durational variability in L2 speech from native norms also influence the degree of the perceived foreign accent (Quene & Delft, 2010; Tajima, Port, & Dalby, 1997).

Speech tempo and durational variability of speech intervals are interrelated. They are reversely correlated. The faster the speech, the less variable are the durations of the speech intervals. This is also reflected in perception (Dellwo, 2006; Dellwo & Wagner, 2003; Wiget, White, Schuppler, Grenon, Rauch, and Mattys, 2010). Therefore speech tempo and durational variability make not only unique but also shared contribution into perceived FA.

However, to the best of our knowledge, no perception study has been done up to now to estimate the separate contribution of durational variability into perceived FA and to estimate the relative contribution of durational variability compared to the contribution of the whole set of timing patterns. It is also not known to what extent timing patterns influence the degree of perceived accentedness compared to that of prosody on the whole and segmental deviations from the norms. These issues need to be resolved, and this provides the rational for my research.

As (1) durational variability and speech tempo change with L2 acquisition progress, and (2) durational variability and speech tempo influence the degree of the perceived FA, the developmental changes in timing patterns will supposedly be perceived by the native speakers of the target language and will affect the degree of the acentedness in L2 speech. When individual differences in sounds and in intonation are eliminated and only the differences in timing patterns are preserved, the perceived accentedness is expected to be lower for the sentences which preserve durational variability of more advanced L2 speakers compared to those of lower-level L2 learners.

We hypothesize that speech tempo and durational variability have a combined effect on the accentedness ratings, but they also make unique, separate contribution into perceived FA. Hence the aim of my study is to estimate the unique contribution of speech tempo and rhythmic patterns in the perceived FA. Since we were interested in the independent contribution of these two types of timing patterns in the FA, we set out to find a way to separate speech rate and

other durational cues which are captured by the rhythm metrics. The effect of other prosodic cues (stress, acoustic correlates of stress, pitch range, intonational patterns) and segmental realizations that are known to influence perceived FA had to be eliminated.

In order to test these hypotheses, the following research questions were set.

1. Do native speakers of the target language indeed hear the differences in prosodic timing patterns (durational variability and speech rate) between proficiency levels in L2 speech?
2. Does the degree of the perceived FA changes as the timing patterns pertaining to the speech tempo and durational variability develop?
3. What is the shared contribution of timing patterns into degree of the perceived FA, and what is the separate contribution of speech tempo and durational variability patterns into the perceived FA?
4. Is the contribution of speech tempo and durational variability into the perceived FA differs for speakers from typologically different L1 backgrounds, in which different timing patterns are displayed (e.g., German and French learners of English).

Based on the literature review, it is expected to find the changes in accentedness ratings for L2 speech produced by speakers on different proficiency levels. The more advanced speakers are expected to sound less accented. This result pattern is anticipated on original sentences spoken by L2 learners, and also on modified sentences in which idiosyncratic differences in segmentals and in intonation will be neutralized. The combined effect of speech tempo and durational variability will be bigger than unique effect of these two types of timing patterns. However, it is hard to say what will have more weight in perceived accentedness – variability or rate at which speech is delivered. We expect to find the differences in relative contribution of speech tempo and durational variability determined by the native language of the L2 speaker. If the timing patterns of the source and the target language will be similar, the durational variability is assumed to have lesser effect on the perceived FA than when the source and the target language exhibit distinctly different timing patterns.

That said, the developmental differences in durational variability in L2 English produced by German learners are expected to have little effect on the FA

ratings, but the increase in speech tempo with proficiency is expected to contribute a lot into perceived accentedness. As for French learners of English, a significant and very substantial effect of developmental changes in durational variability on the perceived accentedness is anticipated. German and French learners of English were chosen because their native languages exhibit contrastively distinct timing patterns, and the patterns of German are similar to those of English, while French displays substantially lower durational variability, faster speech rate and different phonetic implementation of final lengthening compared to the Germanic languages. Thus French learners of English will have to change their timing, while German learners will not have to modify their timing in L2 English to the degree required from French learners.

3. Chapter II: Experiment 1 (based on delexicalized stimuli)

The literature review shows that durational variability and speech rate increases in L2 speech as L2 acquisition progresses. More advanced learners deliver speech at higher rate and with higher durational variation of speech intervals. Indirect evidence allowed us to suppose that such developmental changes in L2 speech can be perceived by the native speakers of the target language and contribute to the perceived FA. The first perception experiment was set up to test this prediction and to address the first research question – whether native speakers of English can indeed hear the differences in speech rhythm and tempo between sentences spoken by learners on different proficiency levels in the L2.

We used the sentences produced by German learners of English with different degrees of L2 mastery for this experiment. German was chosen as the source language because German and English are supposedly similar in timing patterns: Both languages belong to more stress-timed end of the rhythm continuum and are produced at similar speech rates. German and English share phonological characteristics that impact durational variability: Both languages exhibit phonological characteristics typical of stress-timed languages (Dauer, 1987; 1983; Schiering, 2007). Therefore, German learners of English do not have to acquire phonological characteristics like production of complex syllables and complex consonantal clusters, opposition of long and short vowels, acoustic correlates of lexical stress and the location of lexical stress that is relatively free. As both English and German are similar in timing patterns and phonological characteristics that influence the timing patterns, the differences in rhythm and tempo in L2 English produced by German learners on different proficiency levels are expected to be subtle. If native English speakers are able to hear such subtle differences, it can be concluded that the developmental changes in L2 timing patterns are indeed perceived.

3.1. Speech material

The speech material was selected from a speech corpus previously collected by Ordin, Polyanskaya and Ulbrich (2011). The corpus consists of recordings of L2 English speech produced by 51 native German speakers. It is well known that factors such as age of arrival, length of residence, exposure to L1, gender, formal instruction, motivation, language learning aptitude, etc. (see Piske, McKay &

Flege, 2001 for a review) might influence the pronunciation of speakers and consequently potentially affect the degree of perceived FA. Therefore we collected data from a homogeneous group of speakers. The speakers were selected based on questionnaires that were distributed amongst potential participants. The speakers were native Germans from families with monolingual German-speaking parents. All speakers grew up in or near the city of Bielefeld, which is a region in North-Rhine Westfalia in Germany, where inhabitants speak Westphalian standard variety that is close to what is perceived as German standard pronunciation. None of them had resided abroad for reasons other than tourism and only for short holidays. All of them started learning English as a compulsory subject at school and received most of their training through formal classroom instructions. None of them learnt English by immersion. None of them studied English as a major subject at University.

During the recording session we asked participants general questions about their preferences in reading and music, lifestyle, career choice. Questions regarding biography and childhood were asked in order to verify the information extracted from the language background questionnaire. The individual interviews lasted for approximately 10 minutes.

After the interview, we ran a sentence elicitation task following Bunta and Ingram (2007). The participants viewed thirty three picture slides on a computer screen. Each picture was accompanied by a written descriptive sentence. The participants were instructed to memorise the sentence that accompanied each picture. The participants were allowed to move on to the next picture or to return to the previous slide at their own pace. After the participants indicated that they would be able to recall all the sentences, they were asked to look at the pictures again. This time there was no text on the screen and they had to produce the sentence which they had previously found to accompany the individual pictures. This procedure helped us to avoid reading mode and made the analyzed speech material more comparable to natural spontaneous speech.

The recordings were made individually with every participant in a sound-treated booth of the audio-visual studio at the University of Bielefeld. The recordings were made in WAV PCM format at 44 kHz, 16 bit in mono.

Three experienced teachers of English as a foreign language (over four years of teaching and testing experience, certified TEFL teachers), native English

speakers, listened to the recorded interviews and evaluated the learners' performance on three parameters: grammatical accuracy (number of morphological and syntactic mistakes), fluency (distribution and length of pauses, pronunciation) and vocabulary resources (synonyms, avoiding lexical repetitions, selecting exact words to express meaning, selecting words that best fit the conversation style) . They used a 10-point scale for the evaluation on each parameter. To estimate the consistency between the assessors on each parameter we used Cronbach alpha, which scored .92 for accuracy, .89 for fluency, and .90 for vocabulary. The scores show a high consistency between the assessors and the reliability of their assessments. The scores given by the teachers were used to assess L2 mastery of the speakers. The speakers were divided into three groups: lower-intermediate, intermediate and advanced learners.

Eighteen out of 33 sentences per speaker were selected for stimuli preparation. All the sentences had either three beats (e.g. *the ‘dog is ‘eating the ‘bone*), two beats (e.g. *the ‘book is on the ‘table*) or one beat (e.g. *it's ‘raining outside*). The beat corresponded to the number of phrasal accents. The sentences for stimuli were selected so that we had an equal number of sentences with three beats, two beats and one beat (6 sentences per speaker per category). It is possible to produce the same sentence with different number of accents (phrasal stresses). For example, one can produce the sentence *it's ‘raining out‘side* with two beats. Therefore, the selected sentences produced by the selected speakers were listened to in order to make sure that the sentences were indeed pronounced with the expected number of beats. The productions of seven speakers per proficiency group were selected for perceptual experiment. The speakers from the advanced group had the highest averaged grades for fluency, accuracy and vocabulary depth, the speakers from the lower-intermediate group had the lowest average grades.

3.2. Participants

For the perception experiment twenty five native English speakers were recruited to act as listeners in the perception study (Age range – 21-24 years, M=22; 13 females). As the listener's background may influence the way he or she perceives or evaluates the degree of foreign accent in L2 speech (e.g. Bent & Bradlow, 2003), a homogeneous group of monolingual listeners was formed from the

student community from the same geographical area. Participation in the experiment was voluntary; participants did not receive any monetary or other compensation. All participants were students of Ulster University, monolingual English speakers from monolingual families, all of them grew up in or around Belfast.

3.3. Procedure

The speech resynthesis technique was used to degrade the segmental and most of the prosodic information (e.g. tonal contours, phonotactic cues, F0 peaks and valleys, spectral cues, etc.) by replacing all consonantal intervals with '*s*' and all vocalic intervals with "*a*" and synthesizing sentences in MBROLA with a constant fundamental frequency of 133 Hz, thus leaving differences in timing patterns as the only cue for discrimination between the varieties. The technique was formerly used and evaluated by Ramus and Mehler (1999). Although this resynthesis technique shifts the syllabic boundaries, it still preserves the patterns of durational variability of vocalic sequences and consonantal intervals, thus preserving sufficient details to discriminate the utterances with contrastively different rhythms. Speech rate is not affected because this transformation does not change the overall duration of the utterance and the number of syllables (in my speech material). Thus, the number of syllables per second remains intact both in the original sentence and in the resynthesized stimulus. I wanted to find out whether the differences in durational variability of vowels and consonants and in speech rate are sufficient for the listeners to distinguish between utterances in the same language but produced by learners at different proficiency levels.

The stimuli were presented in two blocks, (training and testing). The listeners were not informed that the stimuli were derived from L2 English speech because we did not want the listeners to be biased and to use linguistic knowledge and expectations. Instead, the listeners were told that the stimuli are derived from three less known African languages which were coined as Mahutu (productions of the lower-intermediate L2 speakers converted into "*sasasa*" stimuli), Losto (converted productions of the upper-intermediate speakers) and Burabah (converted productions of the advanced speakers).

At the beginning of the training session, the listener was exposed to nine stimuli (3 stimuli per proficiency group, i.e. per "African language") and had one

minute to listen to them and to get familiar with what the stimuli from different “languages” would sound like. After that, another stimulus was presented, and the listener had to identify from which language (Mahutu, Losto or Burabah) it originates. On the response, the listener was provided with a feedback (which “language” it really was), and the next stimulus was played.

108 stimuli were prepared for the training session (18 stimuli per speaker, 2 speakers per proficiency group). When all 108 stimuli were presented, the participant had a 2-minute break before the stimuli were played again. The training procedure was repeated three times. After the third round, the testing session began. Supposedly, during the training session the participants formed new perception categories for further discrimination between linguistic varieties.

For the testing session 270 stimuli were prepared (different from those used in the training session, 5 speakers per proficiency group, 18 sentences per speaker). During the testing session, the listeners had no feedback.

The whole experiment ran 90-110 minutes. The participants could take a short break and have a rest pause during the training session and between the training and the testing session. During the experiment the participants were offered hot and cold drinks and sweet snacks to help them cope with possible fatigue. The participants could have their drinks and snacks during the rest pauses as well as during the training session (but not during the testing session). The order of stimuli presentation was randomized using the internal Praat algorithm in attempt to counterbalance for possible fatigue effect.

3.4. Results and Discussion

Only the answers given during the testing session were used for the analysis. For the initial analysis, the mode of the listeners’ response was extracted (i.e., the most frequent answer from three possible options: Burabah, Losto, Mahutu) for each of 270 stimuli. Each stimulus was evaluated by 25 listeners, and only the stimuli with unimodal distributions of answer frequencies (i.e., when one of the three possible responses was undoubtedly more frequent than the others) were included into further analysis. After the mode extraction, 79 stimuli which were classified as Burabah by the majority of listeners, 67 stimuli classified as Losto, and 53 stimuli classified as Mahutu. 71 stimuli were excluded from analysis either

for bi-modal distribution of the frequency of answers, or because the difference between the mode and the next most frequent answer was less than two.

The Chi-square test ($\chi 2=12.333$, df=4, p=.015) shows a significant association between the proficiency level of the speaker whose production was used to synthesize the stimulus, and the most frequent response given by the listeners. However, the strength of association, as shown by Cramer's V=.176, is not high. This indicates that there are much more powerful predictors of the most frequent response than merely the proficiency level of the speaker whose speech was converted into the 'sasasa' stimuli. Cross-tabulation details are in table 3-1.

	Proficiency level		
	lower-intermediate	**upper-intermediate**	**advanced**
Mahutu	24	17	**12**
Losto	17	31	**19**
Burabah	23	22	**34**

Table 3-1. Agreement between the groups into which the listeners divided the stimuli and the proficiency levels.

Figures from 3-1 to 3-6 show the differences of the rhythm metrics between the stimuli classified as Burabah, Losto or Mahutu.

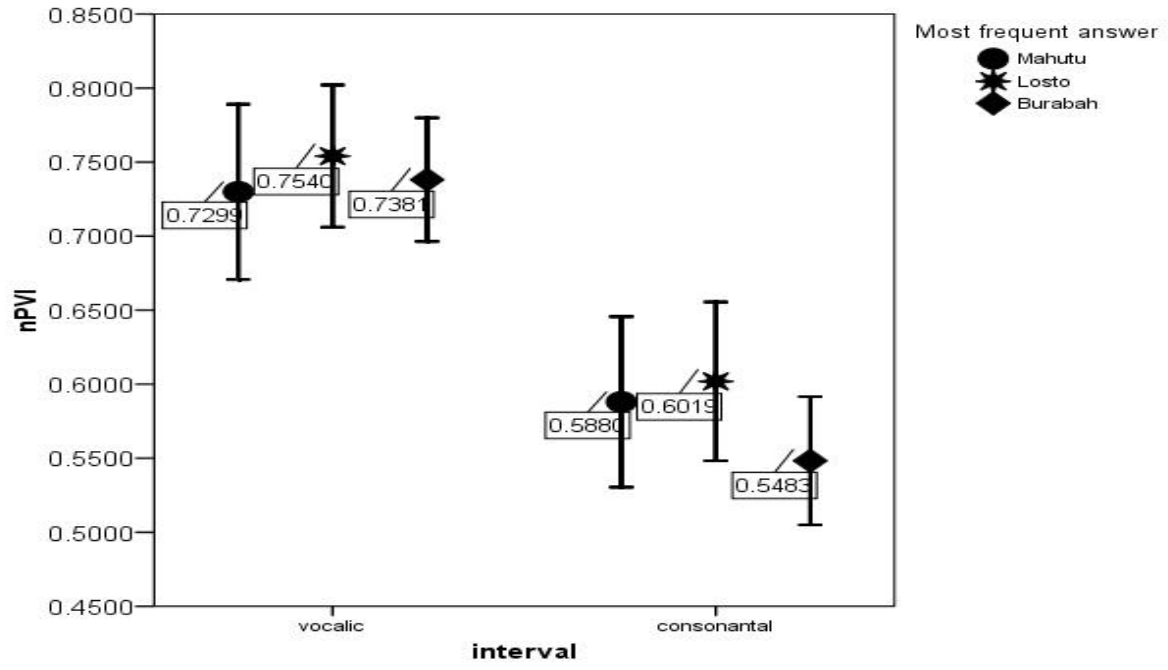


Figure 1. 3-1: Means of nPVI-v, nPVI-c for Mahutu, Losto and Burabah. Error bar shows ±2 S.E.

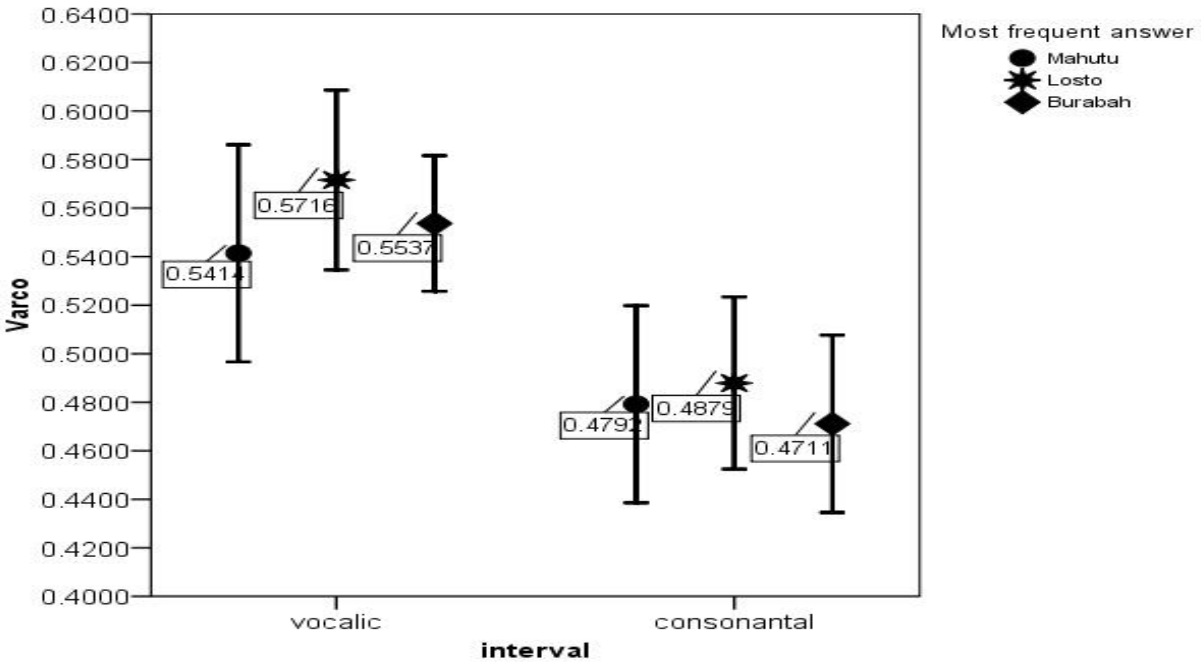


Figure 2. 3-2: Means of VarcoV and VarcoC for Mahutu, Losto and Burabah. Error bar shows ±2 S.E.

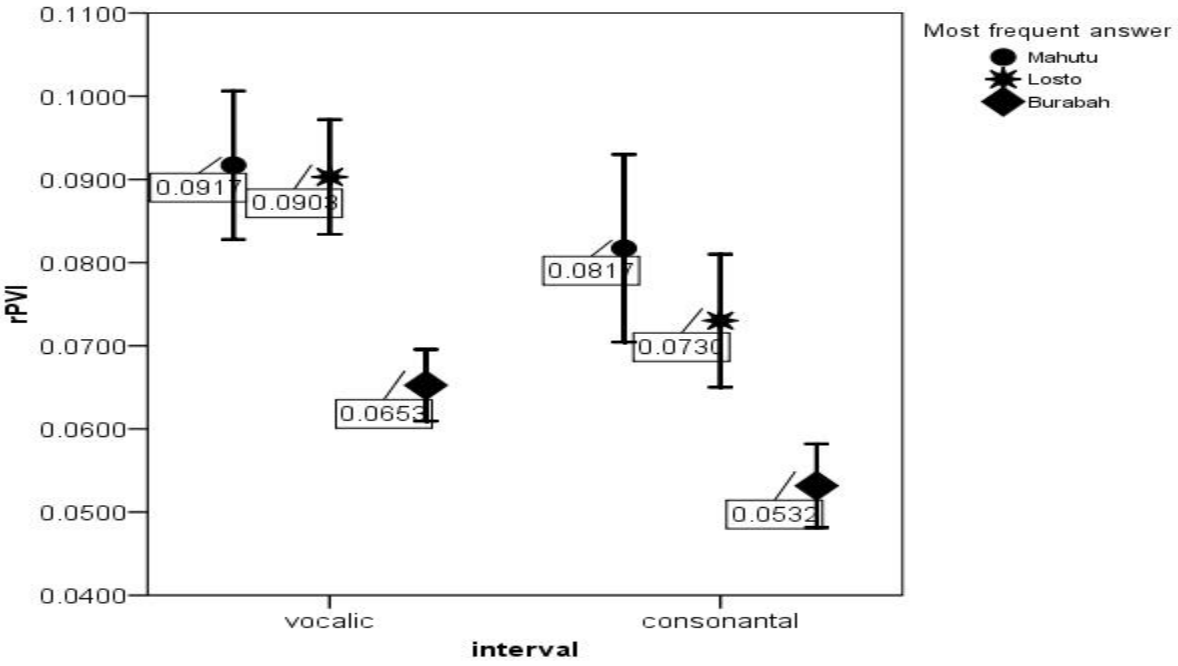


Figure 3. 3-3: Means of rPVI-v and rPVI-c for Mahutu, Losto and Burabah. Error bar shows ±2 S.E.

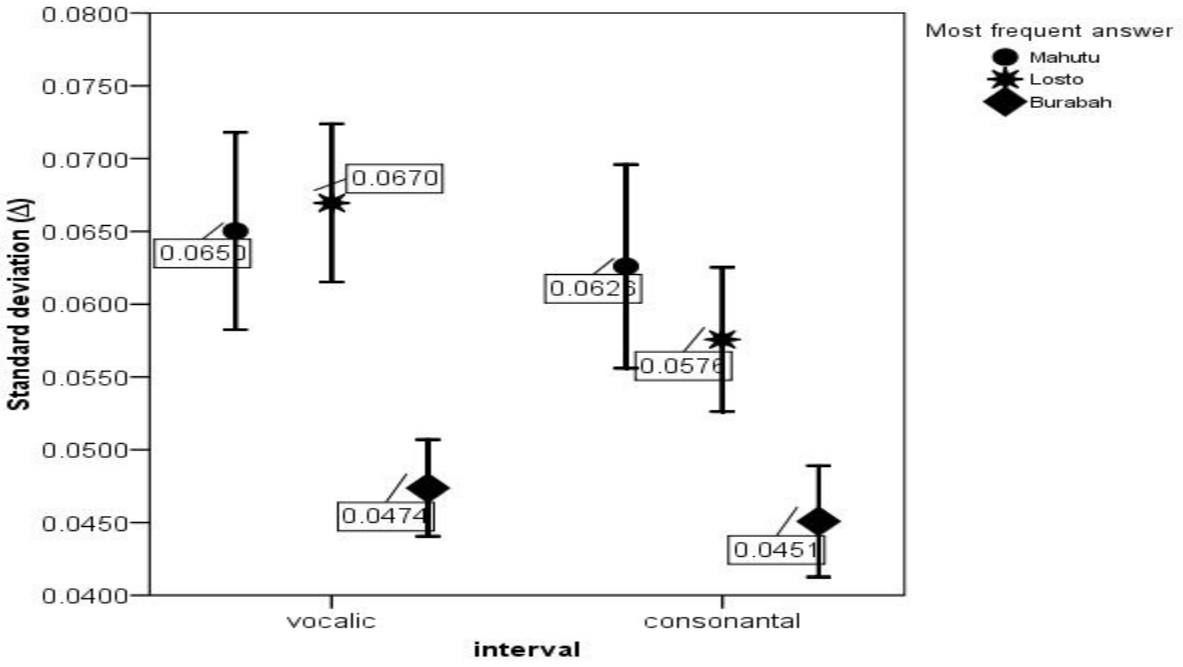
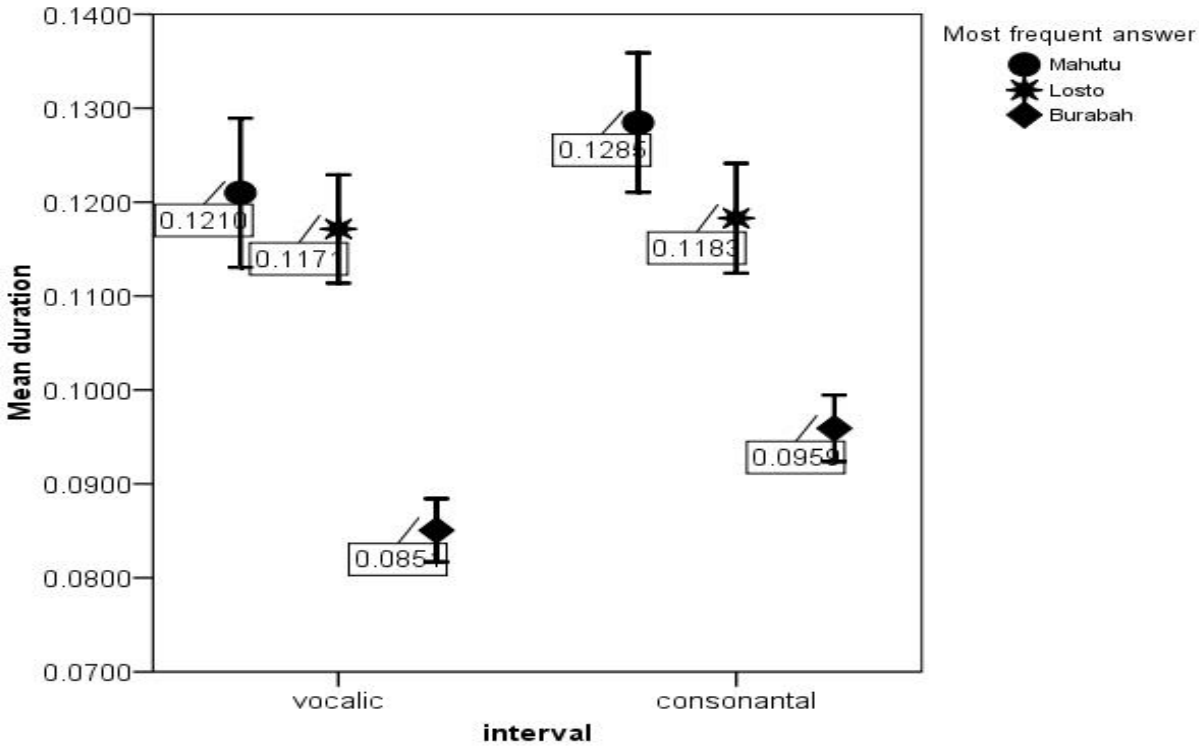


Figure 4. 3-4: Means of ΔV and ΔC for Mahutu, Losto and Burabah. Error bar shows ±2 S.E.



5. Figure 3-5: Means of meanV and meanC for Mahutu, Losto and Burabah. Error bar shows ±2 S.E.

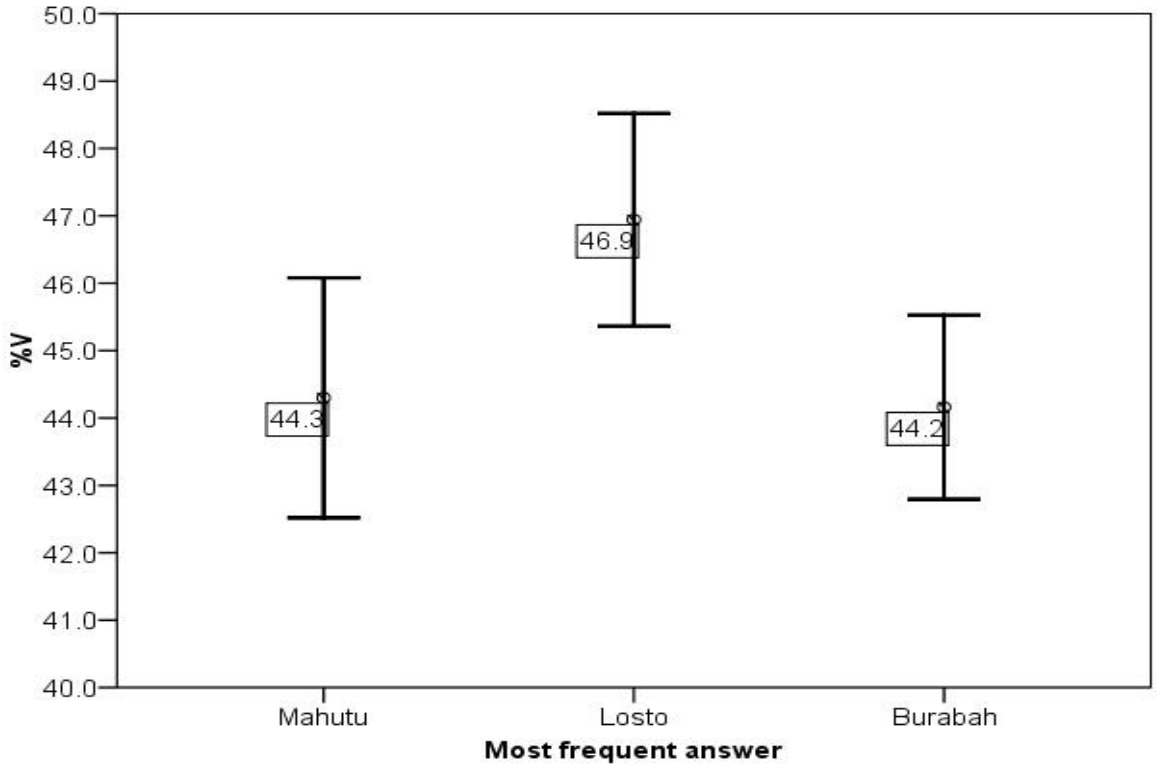


Figure 6. 3-6: Means of %V for all for Mahutu, Losto and Burabah. Error bar shows ±2 S.E.

The figures clearly show that only mean durations of the vocalic and consonantal intervals and non-normalized metrics (rPVI-n, rPVI-c, ∆V and ∆C) differ significantly between the stimuli identified as Burabah, Mahutu and Losto. Normalized rhythm metrics and %V that capture durational variability independent of tempo characteristics do not differ between the stimuli identified as Burabah, Mahutu and Losto. Mean durations of the syllables, vocalic and consonantal intervals are tempo-related measures (the faster the tempo, the shorter are the mean durations). Raw metrics are non-normalized for tempo, thus tempo influences the values of the raw PVI and delta coefficients (the faster the tempo, the lower the variability). Considering this, it is not possible to answer the question as to what people actually were listening for. They might have been listening for tempo (rate of '*s*' and '*a*' changes in the stimuli, or the number of '*s*' and '*a*' in a given period of time), or patterns of durational variability captured by the raw rhythm metrics, or both durational variability and tempo.

To answer this question, the frequency for Burabah, Losto and Mahutu response for each stimulus was calculated (i.e., how many listeners out of 25 in totals identified each stimulus as Burabah, Losto or Mahutu). After that, the

stepwise multiple regression with *Frequency_Burabah* (how many people identified the given stimulus as Burabah) as dependent variable and *means* and *raw metrics* as predictors was performed.

The constructed model included only three steps and only tempo-related measures (mean durations). See table 3-2 for the details. R^2 changes after adding raw metrics into the model were petty and insignificant.

Step	**Metrics**	**β**	**t**	**B**	**p**	**R^2**	**R^2 change**	**Significance of R^2 change**
1	meanV	-.638	-13.57	-102.9	<.005	.407	.407	**<.005**
2	**meanV**	**-.489**	**-10.50**	**-78.82**	**<.005**	**.519**	**.111**	**<.005**
	meanC	**-.365**	**-7.85**	**-64.49**	**<.005**			

Table 3-2. *Coefficients and parameters of the regression model with Frequency_Burabah as the dependent variable.*

The results show that the most important predictors are mean durations of vocalic and consonantal intervals. Mean durations are negatively correlated with the frequency of Burabah-response, which means that the shorter the means, i.e. the faster the tempo, the more likely the listener to classify the stimulus as Burabah.

Then the stepwise multiple regression with *Frequency_Mahutu* (how many people identified the given stimulus as Mahutu) as dependent variable and *means* and *raw metrics* as predictors was performed. See table 3-3 for the details.

Step	**Metrics**	**B**	**T**	**B**	**p**	**R^2**	**R^2 change**	**Significance of R^2 change**
1	meanV	.525	10.1	61.108	<.0005	.276	.276	**<.0005**
2	**meanV**	**.377**	**7.163**	**43.814**	**<.0005**	**.386**	**.110**	**<.0005**
	meanC	**.363**	**6.914**	**46.28**	**<.0005**			

Table 3-3. *Coefficients and parameters of the regression model with Frequency_Mahutu as the dependent variable.*

The regression model for the *Frequency_Losto* as dependent variable and *means* and *raw metrics* as predictors is presented in table 3-4

Step	**Metrics**	**β**	**t**	**B**	**p**	**R^2**	**R^2 change**	**Significance of R^2 change**
1	meanV	.427	7.741	41.821	<.0005	.183	.183	**<.0005**
2	**meanV**	**.358**	**5.993**	**35.018**	**<.0005**	**.207**	**.024**	**=.005**
	meanC	**.170**	**2.847**	**18.207**	**=.005**			

Table 3-4. *Coefficients and parameters of the regression model with Frequency_Losto as the dependent variable.*

The analyses show that the most influential predictors for the Losto and Mahutu response are means, the increase of the means results in a higher likelihood that the stimulus will be classified as Losto or Mahutu. The frequency of Mahutu response is influenced by the means to a greater degree than the frequency of Losto, but the direction of this influence is the same.

A very interesting observation regarding larger effect size for *Frequency_Burabah* (R^2 = .519) and smaller effect size for *Frequency_Losto* (R^2 = .207) and for *Frequency_Mahutu* (R^2 = .386) can be explained by saying that the differences in durational variability become more salient at slower tempo, thus making the decision making more difficult for the participant. An alternative explanation, however, seems simpler and thus more probable. Faster sentences are immediately categorized as falling into Burabah category. Slower sentences should further be divided into Mahutu (very slow) and Losto (slow but faster than Mahutu), and this division seems difficult, thus reducing the effect size of the speech rate on *Frequency_Mahutu* and *Frequency_Losto*. This interpretation is also supported by the post-hoc analysis of tempo differences between the three perceptual categories (see discussion further).

The analysis shows that the listeners are sensitive to the tempo differences and use mean durations of vocalic and consonantal intervals to classify the stimuli into the three groups and to form new perceptive categories. They are less sensitive to timing patterns captured by the rhythm metrics, i.e., less sensitive to variability in the duration of vocalic and consonantal intervals.

An increase in tempo results in the decrease of the raw metrics, in other words, higher tempo removes or reduces the timing differences between stressed and unstressed syllables or vowels, and this is reflected in the values of the raw PVI and delta coefficients. Raw metrics are indistinguishable from the tempo characteristics. The tempo was controlled by normalizing the PVI by the mean duration within each pair of intervals or by normalizing delta coefficients by the mean duration of the interval in the sentence. But the normalized metrics do not discriminate between the listeners' responses. This indicates that people are probably listening not for the durational variability but for the speech tempo-related characteristics when they form new perceptive categories and classify the stimuli.

It was found that the adult listeners reliably classify the stimuli into three groups based on tempo characteristics and ignoring differences in durational

variability. This conclusion agrees with psychoacoustic data. Quene (2007) and Thomas (2007) studied just-noticeable differences in tempo, and they found that 5%-8% change in tempo (expressed as beats per minute for non-speech stimuli and syllables-per-minute for speech stimuli) is easily detected by the subjects.

The tempo differences between the stimuli which were identified as Losto, Mahutu and Burabah were analysed. The tempo was expressed in syllables per second. Each syllable in the 'sasasa' stream is of the open CV type, thus the number of 'a'-segments in the stimulus corresponds to the number of syllables, or the number of beats. Dividing the total duration of the stimulus into the number of "a"-segments is taken as a measure of tempo. It was found that tempo is the highest in the stimuli identified as Burabah (5.62 syl/sec), the lowest in the stimuli identified as Mahutu (4.1 syl/sec), and intermediate in the Losto stimuli (4.41 syl/sec). ANOVA analysis showed that the difference in tempo between the groups is significant, $F(2, 196)=64.077$, $p<.0005$. Pairwise comparisons (with the Bonferroni correction applied) reveal that the difference lies between the Losto and the Burabah stimuli, while the difference between Losto and Mahutu is not significant. The tempo differences range between 2.5 syl/sec and 7.2 syl/sec. The difference between the mean Burabah tempo and the mean Losto tempo is 1.21 syl/sec, that is, the mean tempo in the Burabah stimuli is 25.7% higher than in the Losto group, and this increase is much above the just noticeable tempo difference threshold. The difference between the mean Lost and the mean Mahutu tempo is 0.31 syl/sec, that is, the mean tempo in Losto group is only 6.6% higher than in Mahutu. Such a tempo increase is insufficient for reliable discrimination.

Figure 3-5 on the right shows that there is no statistical difference in mean durations of consonantal and vocalic intervals between stimuli identified as Losto and Mahutu, while stimuli identified as Burabah exhibit much higher speech tempo, which perfectly agrees with psycho-acoustic predictions.

A number of studies in the physiology of hearing (e.g. Greenberg, 1999; Greenberg & Ainsworth, 2004) showed that the neurons fire in response to a sharp increase in intensity, which usually coincides with the vowel onset (p-centres, as defined by Scott (1986), Marcus (1981) and Morton, Marcus and Frankish (1976). As the listeners are sensitive to the p-centres and the p-centres are not shifted by transforming the sentences into the stimuli using '*sasasa*' transformation, we once again justify the use of the chosen technique. The rate at which '*s*' and '*a*' alternate

in the stimuli, then, determines the rate at which the neurons will fire. This means, the effect of the tempo characteristics found in discriminating the stimuli into groups is based on the behaviour of human hearing.

4. Chapter III: Experiment 2 (linguistic stimuli based on utterances in L2 English produced by German learners)

The first perception experiment led us to the conclusion that native speakers of English are more sensitive to the fluctuations in speech rate and ignore the differences in durational variability in L2 speech between the sentences produced by the learners of English at different proficiency levels. However, we expected to find that the listeners would also use rhythmic patterns to perform the classification task because rhythm has been shown to be perceptually salient and important for language acquisition and speech processing. For example, it was found that adults, infants and even neonates can discriminate rhythmically distinct speech-like sound sequences, they can discriminate the languages with distinct rhythms (Nazzi, Bertoncini & Mehler, 1998; Ramus & Mehler, 1999; Ramus, Nespor & Mehler, 1999). Native speakers of rhythmically different languages use different strategies to split the continuous stream of speech into discrete units like words and phrases (Cutler & Noris, 1988; Segui, Dipoux & Mehler, 1990). Non-native rhythm in a foreign language, for example, can lead to wrong segmentation solutions by native speakers of a learnt language. Tajima, Port and Dalby (1997) found that non-native patterns of durational variability inhibit intelligibility of L2 speech. Thus native speakers of English were initially expected to be sensitive to the differences in durational variability between sentences produced by English learners on different proficiency levels, provided that the sentences are the same in regard to the number of syllables, complexity of consonantal clusters, and syntactic structure. Therefore, the obtained result that the listeners completely ignore the differences in durational variability in classification task seems somewhat strange.

The unexpected outcome of our experiment can be accounted for by the fact that the participants had to perceive the rhythmic differences within the same language in L2 speech. Moreover, the native and the target languages of the learners (whose speech productions were used to create the stimuli) were rhythmically close (German and English have similar phonological factors that are reported to influence durational variability and both language are at the more stress-timed end of the spectrum). Therefore, the differences in durational variability between the sentences produced by learners with different degrees of

L2 mastery may have been insufficient to be perceptually relevant. Besides, the chosen methods of delexicalization makes the stimuli not-linguistic, while the most results quoting that the differences in durational variability are perceivable come from participants judging linguistic stimuli. Moreover, many studies valuated not only whether the differences in speech rhythm are perceived, but whether the differences in speech rhythm between L1 and L2 make L2 speech less intelligible or more accented.

These possible explanations for the discrepancies between the results of the first experiment and the results reported in some of the previous studies could be tested by changing the nature of the stimuli and by adapting the experimental paradigm used in many studies that focussed on evaluation of the FA degree in L2 speech. Adapting a new paradigm will allow addressing the second and the third research questions of my study, namely:

- Does the degree of the perceived FA changes as the timing patterns pertaining to the speech tempo and durational variability develop?
- What is the shared contribution of timing patterns into degree of the perceived FA, and what is the separate contribution of speech tempo and durational variability patterns into the perceived FA?

In particular, the aim was to estimate the unique, independent, influence of speech rate and durational variability on perceived FA. More specifically, we intended to disentangle speech rate and speech rhythm and investigate 1) whether rhythm, on its own, when isolated from segmental and other prosodic idiosyncracies, makes a contribution to FA, and if it does, to what degree; and 2) whether and to what degree speech rate makes a contribution to FA, when other durational cues are intact.

4.1. Speech material

The same corpus of speech material was used as in the experiment 1. All speakers were monolingual native speakers of German, and learnt English though formal instructions. From the corpus, fifteen sentences have been randomly selected. Each sentence has been produced by a learner at lower-intermediate, intermediate and advanced proficiency levels in English. Below we provide the list of sentences chosen for the analysis:

1. the dog is eating a bone
2. the book is on the table
3. the girl is eating an apple
4. the ball is on the chair
5. the boy is kicking the ball
6. the knife is on the table
7. the bread is on the table
8. the cat is drinking milk
9. the baby is crying
10. the baby is sleeping
11. the cat is chasing the mouse
12. the man is walking
13. it's raining outside
14. it's snowing outside
15. the spoon is on the table

4.2. Stimuli preparation

The sentences were segmented in Praat (Boersma & Weenink, 2011), and the duration of the phoneme realizations were measured.

Then the phonemic durations were fed into MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) with an English (en1) diphone database, thus obtaining native segments quality. F0 was set to 220 Hz throughout. A new set of sentences was synthesized at 16kHz. The first set of stimuli was comprised of sentences which did not differ in segmental realizations, nor in intonation (F0 was flat throughout) or accentuation. Thus, the only differences between the produced stimuli were timing patterns: durational variability and speech tempo.

Then the obtained stimuli were manipulated using TO-PSOLA algorithm implemented in PRAAT. Three versions of each sentence were equalized in duration by either stretching or compressing the whole sentence to the average duration for the three versions, as it was done by Munro and Derwing in previous studies (1998; 2001). This manipulation affected speech tempo, i.e. removed the differences in speech rate between the lower-intermediate, upper-intermediate and advanced learners. Linear stretching or compressing the sentence keeps the relative durations between the vowels and consonants intact. That is why the

rhythmic patterns are not affected by this manipulation[*]. In the resulting second set of stimuli the sentences differed only in durational variability.

Finally, 15 stimuli from the first set – derived from the sentences produced by intermediate German learners of English – were used to prepare two more versions of each stimulus by stretching and compressing the duration by 10%, which is above the 5-8% of just-noticeable difference in tempo measured by Quene (2007). The resulting stimuli differed only in speech tempo, while durational variability of speech intervals was kept intact. Each type of stimuli included three versions of the fifteen sentences, 45 utterances in total. The versions of the same sentence differed either only in rhythm and tempo, or only in rhythm, or only in tempo, or included the whole range of segmental and prosodic cues marking the L2 speech of German leaners of English at different proficiency levels.

4.3. Procedure

Fifteen native English speakers (from the Belfast area, Northern Ireland) were asked to participate in the experiment. Participation was voluntary. Participants did not receive any monetary compensation. Effort was taken to make the group of listeners homogeneous in educational, language and regional background.

Participants had to listen to the sentences and to evaluate the degree of FA on a 6-point scale, from 6 (native or native-like pronunciation) to 1 (strongest accent). Other FA degrees on the scale were verbalized as 5(mild accent), 4 (moderate accent), 3(rather strong accent), 2 (strong accent).

Participants had to attend the experiment four times with an interval of at least two weeks between the sessions. This 2-week interval between the sessin was deemed to be necessary to avoid the effect of habituation to the task and the stimuli. During each session they were presented with one set of stimuli (original sentences, stimuli which differed only in speech rate and rhythm, stimuli which differed only in speech rhythm, and stimuli which differed only in speech rate). The order of the sessions was randomized for participants.

During each session, the stimuli were presented three times in blocks, 45 stimuli followed by the same 45 stimuli and then again by the same 45 stimuli, 135 stimuli in total. Consequently, every assessor listened to the stimuli three times.

[*] although this procedure differs from the ways that tempo alternations affect the relative durations of vowels and consonants.

The order of stimuli presentation within the blocks was randomized. The responses given to the stimuli presented for the first time (first block of 45 ratings) were not analyzed. The first block was considered as the training block used to familiarize the participants with the task and with the range of accent degrees they would have to evaluate. It was assumed that during this block the participants formed a reference to be able to say that one sentence sounds more or less accented than the other.

4.4. Results

4.4.1. Assessment of Between-Rater and Within-Rater consistency

McNamara (2000), Hughes (2002), Underhill (1987), Taylor (2011) amongst others emphasize that the individuals often disagree with each other when they have to assess L2 learners' speech productions. This disagreement between raters is measures as inter-rater variability. Instabilities and inconsistencies also occur within individual raters and leads to observable intra-rater variability, i.e. the same assessors on two different occasions may score the same utterance differently. For that reason, the obtained ratings were first analysed for the stability between and within assessors.

The ratings given to the same stimulus presented in the second and the third block were used to calculate the Spearman's correlation coefficient (ρ) for each rater. The results of the analysis revealed intra-rater consistency. Spearman correlation was preferred because it better suits the ordinal data (accent ratings). To investigate intra-speaker variability the correlations were then averaged across all raters for each type of stimuli (i.e., original sentences and the three types of acoustically manipulated sentences). The averaged ρ for the original sentences was .858 (ranging between .787 and .966); averaged ρ for the modified sentences was .722 (ranging between .558 and .904) for the sentences in which only timing patterns (both speech rate and rhythm) were preserved; averaged ρ was .575 (ranging between .345 and .893) for the sentences in which only rhythmic patterns were preserved and in the sentences where all segmental and prosodic idiosyncrasies except speech rate were neutralized averaged ρ was .683 (ranging between .378 and .911). The results show that the original sentences contain enough information for consistent assessment; timing patterns also provide

sufficient cues for consistent rating. Rhythmic patterns alone (with tempo differences between sentences equalized) or speech rate alone (without rhythmic distinctions) on the other hand seem to provide less reliable cues to the assessors for consistent test-retest marking. However, speech rate appears to be more informative for the listeners than speech rhythm, and test-retest stability is higher when the listeners can rely on speech rate compared to speech rhythm when scoring the strength of a foreign accent.

To measure inter-rater variability, Cronbach's alpha was used. Cronbach's alpha allows for an estimate of consistency between different subscales in order to see to which degree the values on these subscales measure the same underlying construct. In other words, inter-rater reliability was assessed by comparing the variability of ratings of the same sentence to the total variation across ratings and subjects. This allows estimating the consistency of ratings rather than absolute agreement in ratings from different assessors. That means, the focus is not on whether all assessors give "4" to the sentence "Y" and "2" to sentence "X". What is important is that if sentence "Y" is rated higher than sentence "X" by one participant, it will also be rated higher by the other participants. Therefore, statistical procedures that compare means (e.g., ANOVA which compares mean scores given by the assessors) are not suitable for this purpose. The mean scores given by different assessors may be distinctly different (for example, one rater being stricter than another), but highly consistent at the same time, showing that the raters have the same internal construct of the FA and that they are assessing the same construct.

The results revealed that Cronbach's alpha was .974 for the original sentences, .937 for sentences with preserved timing patterns only, .904 for sentences which preserved rhythmic patterns only, and .917 for the sentences which differed from each other only in speech rate. These values indicate extremely high inter-rater consistency (De Vellis, 2003; Kline, 1999), especially in the ratings of the original sentences. Although inter-rater consistency was lower for modified sentences, it is still well above the minimum acceptable level of .7 (DeVellis, 2003).

As intra-rater and inter-rater stability was high, it was thought possible to assign each sentence a unique FA rating score by averaging the individual ratings. The averaged ratings were statistically analysed to investigate the differences

between sentences derived from L2 German learners of English on different proficiency levels, and to probe the correlation between the perceived FA degree and the L2 fluency of the speaker.

4.4.2. Differences in Durational Variability between sentences produced by learners at different proficiency levels

First, it was analysed whether rhythm metrics indeed differ between the proficiency groups of the L2 speakers in the selected samples of sentences. The metrics are the patterns of durational variabilities to which the listeners might be sensitive to. The following rhythm metrics were calculated both for vocalic and consonantal intervals and also for the syllabic durations: nPVI, %V, CCI, Varco. These metrics were chosen because they are tempo-normalized and maintain their values when the speech rate is manipulated and when the original sentences are modified as described in the section on Stimuli preparation. To explore which rhythm metrics best predict the proficiency level of the L2 learner, a discriminant function analysis was performed with the *rhythm metrics* as independent variables and the *proficiency level* as the dependent variable. The analysis revealed two discriminant functions. Together, the discriminant functions significantly differentiated between the proficiency levels, $\lambda=.632$, $\chi^2(18)=38.117$, $p=.004$. The structure matrix (the correlations between the values of the rhythm metrics and the proficiency levels) in Table 3-1 reveals that nPVI-V, Varco-V, Varco-S, CCI-V and nPVI-S[†] loaded highly on the first discriminant function, which explained 86.6% of variance, $R^2=.321$. The second discriminant function explained only 13.4% of variance, $R^2=.069$, and without the first function it does not significantly differentiate between the proficiency levels, $\lambda=.931$, $\chi^2(8)=5.89$, $p=.66$.

To verify that the values of the nPVI-V, Varco-V, Varco-S, CCI-V and nPVI-S vary with the increase of the proficiency level, a multivariate analysis of variance with *proficiency level* as the factor was carried out. There was a significant effect of the proficiency level on the values of the metrics, $\lambda=.65$, $F(10,166)=3.993$, $p<.0005$, $\mu^2 = .194$.

Contrasts and pairwise comparisons revealed that these metrics significantly differ between sentences produced by lower-intermediate and

[†] The letter "V" after the metric stands for "vocalic intervals", "C" – for consonantal intervals and "S" – for syllabic durations.

intermediate learners, but do not differ between intermediate and advanced learners. The overall direction of the developmental change is from more syllable-timing towards more stress-timing in speech production. The values of the rhythm metrics increase with the increase in L2 mastery (Figure 4-1).

	DISCRIMINANT FUNCTION	
	1	**2**
Npvi-v	.795	**-.134**
Varco-V	.758	**.356**
Varco-S	.565	**-.358**
CCI-V	.558	**-.336**
nPVI-S	.404	**-.294**
Varco-C	.065	**-.031**
nPVI-C	.096	**-.29**
CCI-C	.171	**-.276**
%V	**.116**	**.208**

Table 4-1. Correlations between the values of the rhythm metrics and the proficiency levels (structure matrix).

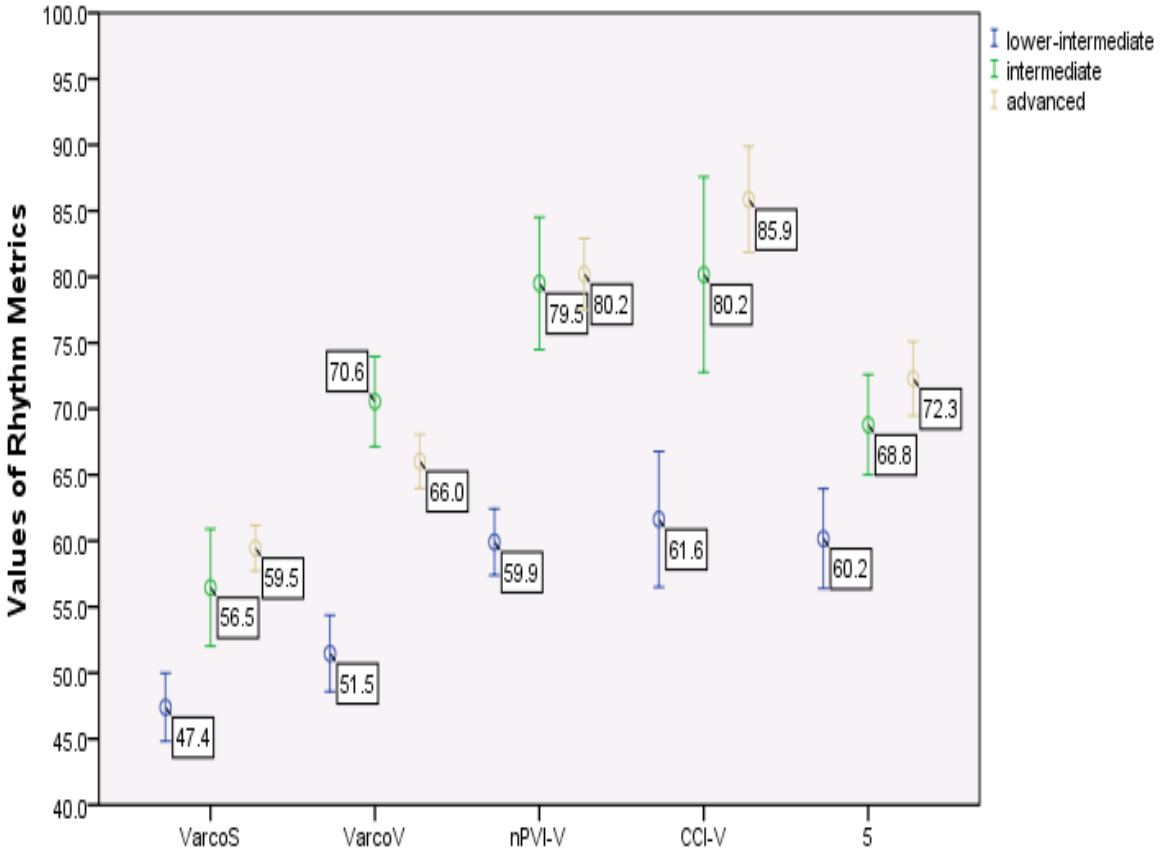


Figure 7. 4-1: Changes in the values of the rhythm metrics as a function of L2 proficiency grows.

The combination of MANOVA and Discriminant function analysis (DA) was chosen to reveal whether and / or which rhythm metrics depend on the proficiency level of the L2 learners and can be used to differentiate between the utterances produced by different L2 learners. MANOVA was deemed to be necessary to precede a series of one-way ANOVAs because separate ANOVAs would not provide us with any information on the underlying relationships between the dependent variables and would treat the rhythm metrics as parameters, which are not interconnected. DA allows us to see that nPVI-V, Varco-V, Varco-S, CCI-V and nPVI-S are interrelated, are able to differentiate between the proficiency level only in combination. Thus only these metrics were further introduced into MANOVA analysis to show that the increase in L2 fluency influences the values of these metrics.

As the MANOVA model was significant, a series of one-way ANOVAs for each metric was performed in order to find out which of the individual metrics differed between the proficiency levels. One-way ANOVAs were performed with

proficiency level as independent factor and one of the *rhythm metrics* as dependent variable. All assumptions for ANOVAs are met, including homogeneity of variance (checked by Levene's test), as well as normality and equality of samples.

Every separate ANOVA test reveals significant differences between the proficiency levels for the following rhythm metrics: nPVI-v, $F(2,42) = 6.317$, $p = .004$; Varco-V, $F(2,42) = 5.928$, $p = .005$; Varco-S, $F(2,42) = 3.38$, $p = .002$; CCI-V, $F(2,42) = 3.277$, $p = .048$; and n-PVI-S, $F(2,42) = 2.759$, $p = .03$. Pairwise comparisons (with Bonferroni correction) revealed that these metrics significantly differ between sentences produced by lower-intermediate and intermediate learners, but do not differ between intermediate and advanced learners. The other metrics did not exhibit significant differences between the proficiency levels.

4.4.3. Assessment of contribution of Speech Rate and Durational Variability into perceived Foreign Accent rating

In order to investigate the influence of the L2 learners' proficiency level on FA ratings analyses of variance was carried out with the *proficiency level* of the speaker as the independent variable and *FA rating score* averaged across raters as the dependent variable. The analysis was done separately for four different types of stimuli:

Type 1: the original sentences in which three versions of the same sentence produced by L2 learners on different proficiency levels differed in segmental realizations and in prosody including intonation, timing, stress, etc.

Type 2: the stimuli in which three versions of the same sentence differed only in timing including tempo differences and rhythmic patterns, while intonational idiosyncrasies and segmental / spectral differences were neutralized using the resynthesis technique;

Type 3: the stimuli in which three different versions of the same sentence differed only in durational variability of speech intervals;

Type 4: the stimuli in which three different versions of the same sentence differed only in speech rate (measured in syllables per second).

The results revealed that the proficiency level explains 91% of variance in FA ratings, λ=.086, $F(2,28)$=148.21, p<.0005, μ^2=.914 for the original sentences, The contrasts between ratings of sentences produced by lower-intermediate and intermediate L2 learners and between intermediate and advanced L2 learners were significant at p<.0005 (Figure 4-2). For the modified sentences which preserved all timing patterns, including speech rate and rhythmic features, proficiency level explains 78% of variance in rating, λ=.218, $F(2,28)$=50.202, p<.0005, μ^2=.782. The contrast between ratings given to the sentences produced by lower-intermediate and intermediate L2 learners was significant at p<.0005, and the contrast in ratings given to the sentences by intermediate and advanced L2 learners was significant at p=.007 (Figure 4-2). For the modified sentences, which preserved rhythmic patterns but neutralized differences in tempo, proficiency level explained only 46% of variance, λ=.545, $F(2,28)$=11.689, p<.0005, μ^2=.455.Here, the contrast in ratings given to the sentences produced by lower-intermediate and

intermediate L2 learners was significant at p=.001, whereas the contrast in ratings of intermediate and advanced L2 learners' productions was not significant, p=.602

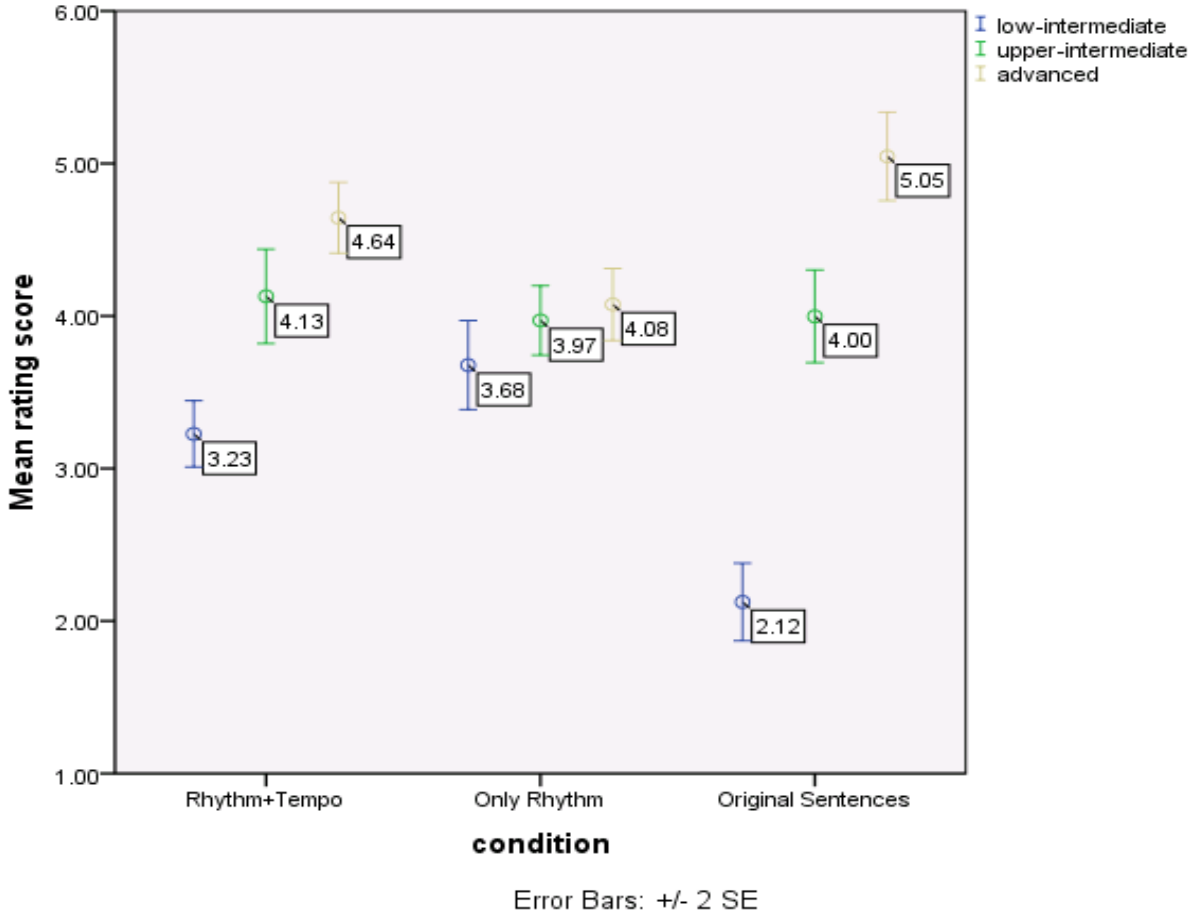


Figure 8. 4-2: Averaged FA rating scores on the original, modified sentences that preserved only timing patterns related to both speech tempo and durational variability (Rhythm+Tempo) and modified sentences that preserved only timing patterns related to durational variability (Rhythm).

Each sentence, produced by L2 speakers on different proficiency levels, was evaluated by each listener in the original and in the modified conditions, in which all segmental and prosodic idiosyncrasies, except timing patterns, were neutralized. The goal was to see if the original or modified sentences with preserved timing patterns received higher ranking. For this, a mixed analysis of variance was run with the *type of the sentence* (*original* vs. *rhythm only vs. rhythm + tempo*) as within-subject factor and the *proficiency level* of the L2 learner who produced the sentence (*lower-intermediate* vs. *intermediate* vs. *advanced*) as between-subject factor. The influence of the sentence type (i.e., manipulation type) on FA rating was significant, λ=.863, $F(2,86)$=6.819, p=.002, μ^2=.137. The

influence of the proficiency level on the FA rating was also significant, F(2, 87)=54.525, p<.0005, μ^2=.556. The interaction of *proficiency level* and *sentence type* is also significant, λ=.384, F(2,172)=26.369, p<.0005, μ^2=.38. Figure 4-2 exhibits this interesting significant interaction between *proficiency level* and *sentence type* of the FA rating.

For the modified sentences which differed only in speech rate, it was examined if the same sentence in three different tempos will incur different ratings. Tempo explained 86% of variance in rating scores, λ=.115, F(2,28)=107.324, p<.0005, μ^2=.885. The contrasts between ratings given to slow and normal and between ratings given to normal and fast sentences were significant at p<.0005 (Figure 4-3).

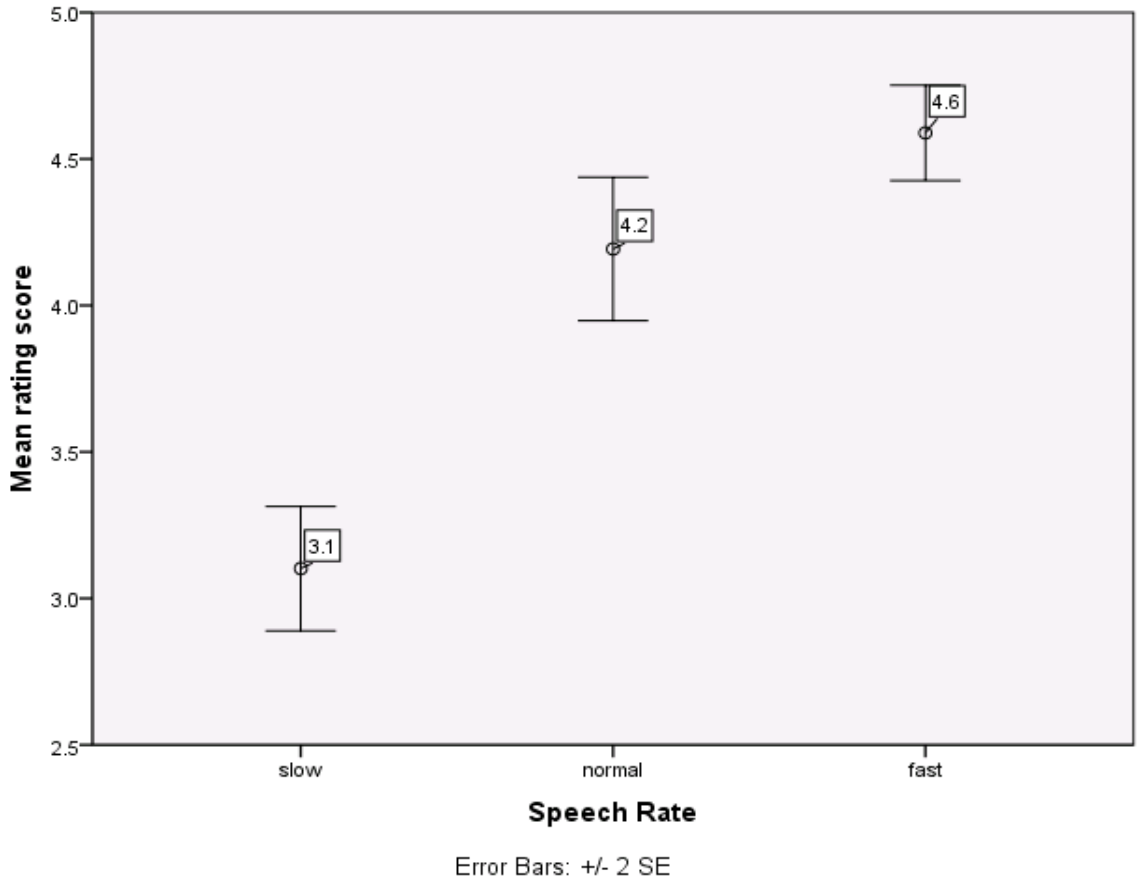


Figure 9. 4-3: Averaged rating scores obtained on the modified sentences that differed only in speech rate.

The results obtained indicate that timing patterns contribute to the perceived FA, and that speech rate and rhythm make not only a combined but also a unique contribution to the degree of the strength of the FA. Combined

contribution of speech rate and speech rhythm is greater than that of speech rhythm only, thus rhythm and tempo have additive influence on the perception of accentedness degree. It is evident that differences in speech rate have a larger effect on the perception of the degree of the FA than differences in durational variability. This confirms, to some extent, the outcome of the first experiment. However, it is not possible to say that in linguistic stimuli the differences in durational variability are ignored. Such differences are less perceptually relevant when the native and the target language of the learner are rhythmically similar.

4.5. Discussion

The presented experiment study dealt with an individual and combined contribution of speech rate and speech rhythm to perceived FA. The results show that when segmental and other prosodic idiosyncrasies are neutralized, timing patterns still contain sufficient cues for the listeners to reliably and consistently rate the degree of FA in L2 speech. Timing patterns (i.e. rhythmic patterns and speech rate) explain up to 78% of variance in the strength of the accent rating on acoustically modified sentences which neutralize all suprasegmental and segmental characteristics and preserve only timing patterns.

Rhythmic patterns significantly differ between the proficiency levels of the L2 learners and these differences can be accounted for by a number of rhythm metrics, namely nPVI-V, Varco-V, Varco-S, CCI-V and nPVI-S. If the speech rate idiosyncrasies are neutralized, speech rhythm can account for 45.5% of variance in FA rating, compared to 78% of variance accounted for by all the timing patterns, including both speech rate and rhythm. This indicated that speech rate makes a unique contribution into perceived FA, but this contribution is lower than the combined contribution of speech rate and rhythm.

In order to verify that the contribution of speech rate is unique and independent of that made by speech rhythm, the listeners were asked to rate FA in the sentences that differed only in speech tempo, while rhythmic patterns captured by the rhythm metrics nPVI-V, Varco-V, Varco-S, CCI-V and nPVI-S were neutralized. The prediction was confirmed. Faster speech rate received lower FA ratings whereas a slower speech rate was perceived as more foreign accented. The analysis revealed that speech rate influences the judgement of FA degree and

accounts for 86% of variance in ratings. Thus the contribution of speech rate to perceived FA is unique, substantial and independent of other timing patterns.

The range of scores is much narrower when listeners have to rate the modified sentences which differed only in rhythmic patterns compared to rating the sentences which provide rhythmic and tempo cues (Figure 4-2) or only tempo cues (Figure 4-3). In other words, sentences which differ only in rhythm are perceptually less distinct than those that differ in rhythm and tempo or tempo only. In addition, our results show that test-retest stability (intra-speaker variability) and between-raters consistency (inter-speaker variability) are higher when the listeners can rely on speech rate than when they rely on speech rhythm only (see the first subsection of Results). Therefore the conclusion was confirmed that the listeners rely more on tempo characteristics than on rhythmic patterns, when making judgments on the degree of FA in L2 speech. This agrees with the results of the first experiment in the dissertation.

The results can be explained by the fact that rhythm metrics, which capture the patterns of durational variability, to which the listeners are sensitive, differ depending on the proficiency level of the L2 learner, but the metrics differ between the lower-intermediate and intermediate L2 speakers only, no significant difference between rhythm metrics in speech of intermediate and advanced speakers was found. Sentences produced by lower-intermediate and intermediate L2 learners received significantly different FA scores, whilst ratings for sentences produced by intermediate and advanced L2 learners did not, which corresponds to what was established in the production analysis. Listeners in our experiment were not able to tell apart sentences produced by intermediate and advanced L2 learners relying exclusively on rhythmic cues because the rhythmic patterns did not differ (Figure 4-1). Yet when constructing the stimuli which differed in speech rate, we increased or decreased the overall duration of the whole sentence by 10%, which results in more than the minimum noticeable difference in speech tempo (Quene, 2007). Consequently, the listeners were able to discriminate fast, normal and slow sentences and make clear distinctions between fast, normal and slow utterances, as these distinctions existed in the acoustics of the speech signal.

The prevalence of speech tempo over speech rhythm can also be explained to some extent by similarities in the phonological structure of German and English, which determines the variability in duration of syllables and vowels. Both

languages are rhythmically similar, exhibit vowel reduction in unstressed positions, operate complex syllables, and feature opposition of short and tense vowels. This supposedly reduces the gross differences between languages in patterns of durational variability. Therefore German learners of English are capable of producing English-specific rhythm patterns. Presumably, rhythm patterns might have more influence on the perceived FA if the L1 and L2 of the learner are rhythmically distinct. A more salient influence of the rhythm on FA should be expected in speech of, for example, French learners of English.

As in some of the previous studies no influence of speech tempo on the accent rating has been detected (Anderson-Hsieh & Koehler, 1998; Flege, 1998), while in other studies the effect of speech tempo on accent judgements is reported (Munro and Derwing, 1998; 2001; Kang, 2010; Kang et al., 2010), the outcome of our experiments could not be predicted with certainty. In experimental conditions the influence of speech tempo on the FA rating has been revealed even when L1 and L2 of the learner are rhythmically similar and consequently the effect of differences in duration variability of segments and syllables on FA judgement is diminished. Our finding undoubtedly supports the latter standpoint.

The positive finding is attributed to the fact that we controlled for the tonal idiosyncrasies (by monotonizing F0) and for segmental idiosyncrasies when creating the stimuli. Therefore, the raters did not have to rely on contradictory cues to the degree of the strength of the accent (e.g., faster speech with more phonetic errors).

It should be noted that in the study by Munro and Derwing (2001) only 6% of the variance in accent rating was explained by the speech tempo variation. This estimate was obtained using regression analysis with speech rate (in syl/sec) as independent variable and accent rating as dependent variable. Although the direction of influence was the same, i.e. compressed stimuli (faster speech tempo) were rated as significantly less accented than normal or extended (slowed down) stimuli, the effect sizes in our study were substantially larger. This difference can also be attributed to how the stimuli were created for these experimental studies. Munro and Derwing (1998; 2001) used natural sentences and extended or compressed the duration of each sentence as a whole by 10%. Therefore, the segmental and tonal phonetic errors were also preserved, as well as the individual peculiarities of the speakers' voices. These factors also influence the degree of the

perceived FA. However, only speaking rate was included in the regression equations as an independent variable. Therefore, the set of stimuli included faster speakers exhibiting more phonetic errors and slower speakers exhibiting fewer errors. The assessors had to deal with the contradicting cues to the strength of the accent, weighted the cues and based their judgement on a variety of factors; some of these factors diminish the effect of the speech tempo. In our study the effect of phonemic realizations and tonal differences was neutralized using resynthesis technique, and the only cues the listeners could rely on for the FA judgement was speech rate. In the absence of other cues the differences in timing patterns became more perceptually salient, and this increased the effect of speech rhythm and speech tempo on FA rating in our study.

Figure 4-3 shows that the difference between FA scores given to normal and slowed sentences is larger than that given to normal and accelerated sentences. This might suggest that the listeners respond to the overall unnaturalness of decelerated sentences. When the sentence is stretched, all segments, vowels and consonants alike, are extended proportionally. However, when a speaker naturally decreases the speech tempo, vowels are stretched more than consonants, and continuant consonants are stretched more than stops. An equal degree of lengthening in extended sentences might have sounded unnatural to the listeners, and therefore slower sentences might have received a lower rating, which is explained by unnatural lengthening patterns rather than by overall speech rate. This alternative interpretation is plausible, but this is not to be forgotten that all the contrasts between slow, normal and fast sentences were significant. Besides, Munro and Derwing (2001) used the same approach to making slower and faster sentences, and in their study the ratings assigned to the natural and slow stimuli did not differ significantly at all, while the ratings assigned to normal and fast stimuli were different. This suggests that the listeners accepted 10% lengthening of the whole sentence as natural. If they had based their judgement on unnatural lengthening patterns in extended sentences, than Munro and Derwing (1998; 2001) should have observed a significant difference in ratings assigned to normal and slowed sentences. Therefore it is concluded that this alternative interpretation is possible, but less probable.

Slower speech tempo allows more time for perception and analysis of deviations from the phonetic norms of the target language, and therefore it is

perceived as more accented compared to faster L2 speech, which provides less time to concentrate on prosodic and segmental deviations and demands addressing attention and cognitive resources to processing the meaning of the utterance.

As it follows from the results, monolingual English speakers are sensitive to the deviations in timing patterns (both speech rhythm and speech tempo) in L2 speech from the expected norms of their native language, and these deviations influence the perception of the FA. Speech rate and speech rhythm, as two different aspects of timing patterns, make unique contributions to FA.

Figure 4-2 shows that the sentences produced by lower-level L2 learners received significantly lower scores in the original sentences compared to the modified sentences. The opposite was found in the comparison of original and modified sentences produced by advanced L2 learners, i.e. the original sentences received higher scores than the manipulated sentences. Sentences produced by intermediate L2 learners received comparable scores in both original and modified conditions. This might indicate that the influence of segmental characteristics differs depending on the proficiency level. In other words, tempo and rhythmic patterns are overridden by foreign accented segmental characteristics in speech produced by L2 learners with a lower proficiency level. On the other hand, timing patterns of the target language become perceptually more salient for native speakers with increased mastery of the segmental characteristics.

Thus it is tentatively suggested that from a didactic perspective, maintaining timing patterns of the target language becomes more important in higher levels of L2 mastery, while on lower proficiency levels it is worthwhile to concentrate on the segmental realizations, and probably on such prosodic aspects as stress and intonation.

Referring back to the research questions, the following answers can be given:

- Does the degree of the perceived FA changes as the timing patterns pertaining to the speech tempo and durational variability develop?

Yes, as timing patterns change between the proficiency levels, the accent rating also changes. The sentences with the timing patterns of lower-intermediate

learners are assessed as significantly more accented compared to the sentences with the timing patterns of advanced and upper-intermediate learners.

- What is the shared contribution of timing patterns into degree of the perceived FA, and what is the separate contribution of speech tempo and durational variability patterns into the perceived FA?

Durational variability and speech tempo characteristics in L2 speech make combined, overlapping, as well as unique contribution into the degree of accentedness of L2 speech. However, the relative contribution of rhythm and tempo is not possible to estimate based on the obtained results, and it is not possible to say whether rhythm, when the influence of tempo of durational variability is controlled for, or tempo, when the rhythmic patterns are not different between the sentences, make bigger contribution into the perceived FA.

5. Chapter IV: Experiment 3 (linguistic stimuli based on utterances in L2 English produced by French learners)

The third perception experiment was set up in order to address the fourth research question, namely, if the contribution of speech tempo and durational variability into the perceived foreign accent differs for learners of English with rhythmically contrastive native languages. It was also deemed necessary to refine, as far as it is possible, the answer to the third research question, namely, what are the unique contributions and what is a shared contribution of speech rate and durational variability into the perceived foreign accent, the third experiment was set up. In the previous experiments it was established that native speakers of English can hear the differences in duration variability only in linguistic stimuli. It was also established that the differences in durational variability between the utterances produced by German learners at different proficiency levels indeed contribute to the perceived foreign accent. Patterns of durational variability that are representative of L2 English produced by learners at lower proficiency levels enhance the degree of the perceived foreign accent. However, the differences in accent rating on the stimuli that differed only in durational variability were smaller compared to the differences in foreign accent rating on the stimuli that differed only in tempo, which indicates that native British English listeners are more sensitive to tempo changes than to changes in durational variability.

There is a possibility that the differences in durational variability between proficiency levels of German learners of English are not sufficiently salient for the native speakers of English because English and German are rhythmically similar languages, and possibly, the developmental changes in L2 English speech rhythm will have a larger impact on perception of accentedness if the native language of the learner is contrastively different in terms of speech rhythm from the target language. This could also be indirectly supported by the earlier results by Ramus and Mehler (1999) who showed that adults are sensitive to the rhythmic differences between rhythmically contrastive languages, i.e., between languages that are traditionally classified as stress-timed and as syllable-timed, or even better, between languages that are traditionally classified as stress-timed and as mora-timed. Therefore, the third perception experiment was set up. The stimuli were prepared with rhythmic patterns of French learners of English at different

proficiency levels. French and English are rhythmically contrastive (Ramus et al., 1999; White and Mattys, 2007).

Besides, it was found that rhythmic differences are perceived only in linguistic stimuli, but when listeners have to classify non-linguistic stimuli (SASASA stimuli, see experiment one) into perceptual categories, they do not use the differences in durational variability to perform classification task and base their judgments only on tempo-related features. If rhythmic differences are indeed perceived only in linguistic stimuli and ignored in non-linguistic stimuli, then making the stimuli more natural may supposedly enhance perception of durational variability. This possibility was tested by making two types of stimuli for the third experiment – flat stimuli with monotone F0 (similar to those used in experiment two), and intoned stimuli (with intonation contour that made the synthesized utterances sound more natural). If the assumption is correct, the differences in rhythmic patterns will be better perceived on intoned utterances than on flat utterances because intoned utterances are more speech-like than flat ones.

5.1. Speech material

For the perception experiments, the corpus of speech samples representing L2 English produced by monolingual French learners of English as a foreign language was used. The corpus was collected for a different project (see Ordin and Polyanskaya, 2015; Ordin, Polyanskaya and Wagner, 2015). For the convenience of the readers, I will provide the details of the corpus below.

The speech corpus contains the recordings of 48 monolingual French learners of English from Parisan metropolitan area. The recordings were done in the Laboratory of Phonetics and Phonology at Sorbonne Nouvelle Paris III University and were carried out using a sampling rate of 48kHz, a 32 bit quantization, mono. The recordings were done with each speaker individually, and each session was approximately 30 minutes in duration. Each session consisted of two parts. During the first part, the participants were asked ten questions – the same set of questions for every speaker – related to biography, music and reading preferences, educational choices, etc. This interview lasted 10-12 minutes for each learner. The interviews were anonymized and given to two teachers of English as a foreign language, native English speakers, certified TEFL specialists with at least four-year experience in language testing. The teachers were asked to

listen to the interviews 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. Cronbach alpha (< . 88 for each parameter) shows high agreement of assessments between teachers, i.e., if one teacher assessed fluency of one learner higher than fluency of another learner, the other two teachers also gave him or her higher ratings for fluency. Cronbach alpha (< .75 between parameter for each teacher) means that the assessments between parameters correlated, i.e., if the rater assessed fluency of one learner higher than fluency of another learner, he or she also gave higher rating to his or her grammatical accuracy and vocabulary.

The ratings for each learner were averaged across teachers and across parameters to obtain the overall assessment of learners' proficiency. These overall assessments were used to split the learners into three proficiency groups: beginners (with an overall rating below 5), intermediate learners (with the overall rating below 8 and above 5) and advanced learners (with an overall rating above 8).

The second part of the recording included a sentence elicitation task. For this purpose, picture prompts were used. We used the same elicitation procedure that is described by Bunta and Ingram (2007), and we used 26 picture prompts Bunta and Ingram used, and added 7 pictures drawn in the same style to get 33 picture prompts and 33 sentences in total. All picture prompts and the list of elicited sentences can be found in Appendix.

The pictures were presented to the participants, each picture on a separate slide with an accompanying descriptive sentence. The participants were instructed to look at the pictures and to remember the accompanying sentences. They participants could flap through the slides backwards and forwards at their own pace. Once the participants said they had memorized the sentences, 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. Using this methodology, 93% of sentences were produced correctly and without hesitations. In 7% of sentences there was a deviation between the expected and a 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. In all cases, one verbal prompt was sufficient to elicit the expected sentence. This procedure allowed 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 sentences was chosen as stimuli for the subsequent study. These were the same sentences that were used in experiment 2 because we wanted to have comparable results received with the stimuli based on L2 English utterances produced by German and French learners. For both experiment 2 and 3, the sentences were selected in a quasi-random fashion, but making sure that each of the 15 stimuli is produced once by a learner at advanced, intermediate and elementary proficiency level each. This resulted in 45 utterances in total which were used as speech material for the following perception experiments.

The selected utterances were annotated in Praat (Broersma, 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 by Stevens (2002) for C and V intervals. Syllabification was done following Wells (2006). Adjacent vowels and consonants were combined into the same V and C intervals, even when the consonantal cluster or a sequence of vocalic segments straddle the syllabic boundary. Pauses and hesitations within utterances were excluded, but vowels and consonants on either side of the pause were not combined into the same interval.

Traditional rhythm metrics were calculated on each utterance. The tempo-normalized rhythm metrics (nPVI, Varco) were used because they capture durational variability of speech intervals (i.e., rhythmic patterns) independent on mean duration of these intervals (i.e., independently of tempo). Thus, when speech rate is modified in the process of stimuli preparation, relative durations (i.e., rhythmic patterns) captured by these metrics will remain intact. The metrics were calculated on V and C intervals and on syllables. 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, which means that the

durational variability of speech intervals develops towards a higher degree of stress-timing as a function of proficiency growth in L2 (figure 5-1).

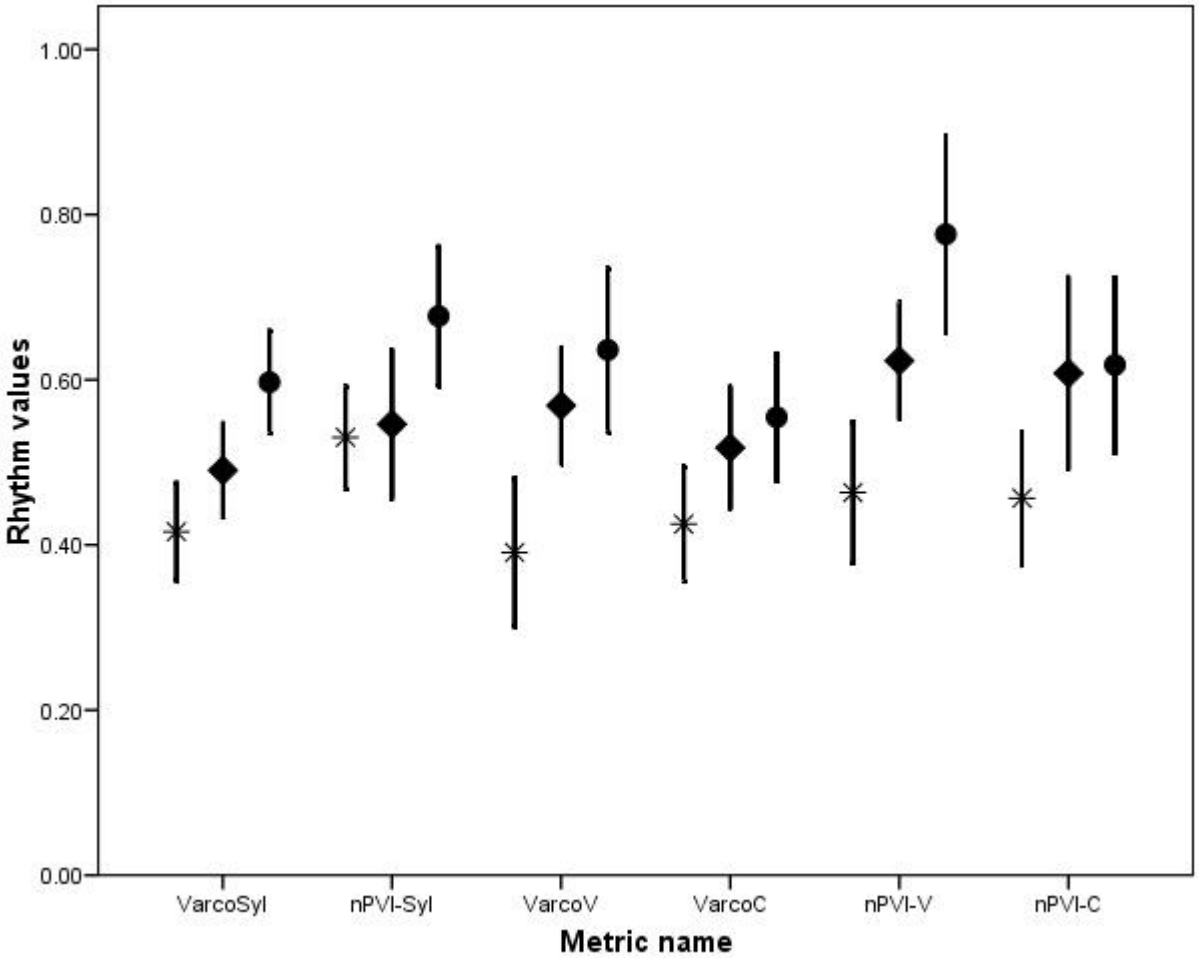


Figure 10**.** 5-1: Tempo-normalized metrics for utterances that differed only in speech rate (the same differences in the rhythm metrics are between the original utterances produced by learners at different proficiency levels). Stars identify the metrics calculated on utterances produced by beginners of English as an L2, squares identify intermediate learners, and circles – identify the metrics calculated on advanced learners' utterances. Error bars show ±2SE.

To check the statistical significance of this visual impression, a one-way MANOVA was performed using rhythm metrics as dependent variables and proficiency levels of the learners as a factor. The model turned out to be significant, $\Lambda = .398$, $F(12, 74) = 3.608$, $p < .0005$, $\eta^2 = .369$, which indicates that the proficiency level indeed has had a significant effect on the measured rhythm metrics. To test which metrics do differ between proficiency levels, the MANOVA was followed by a series of ANOVAs, one for each metric, to find out whether each metric differs between utterances produced by advanced and intermediate learners, and by intermediate learners and beginners (table 5-1).

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**

Table 5-1. ANOVAs testing the influence of speakers' proficiency levels on rhythm metrics.

The analysis shows that the values of the rhythm metrics indeed differ between utterances produced by English learners at different proficiency levels, with utterances from advanced learners exhibiting significantly higher durational variability of syllables, vocalic and consonantal intervals. This means that there are indeed durational cues present for the listeners to pick up and which may influence the perception of foreign accent independently of tempo. In a subsequent study, it was therefore analyzed whether and if yes, to what degree, this is the case.

5.2. Stimulus preparation for testing the independent contribution of Durational Variation and Tempo on Foreign Accent perception

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

I) 45 original utterances, 15 sentences, each produced by a French learner of English at beginner, intermediate or advanced proficiency levels (Originals);

II) 45 intoned and 45 flat resynthesized utterances, 3 intoned and 3 flat versions of each of the 15 sentences that differ only in speech rhythm and speech rate (Rhythm and Tempo);

III) 45 intoned and 45 flat resynthesized utterances, 3 intoned and 3 flat versions of each of the 15 sentences that differ only in speech rhythm (Rhythm Only);

IV) 45 intoned and 45 flat resynthesized utterances, 3 intoned and 3 flat versions of each of the 15 sentences that differ only in speech rate (Tempo Only).

Creating the utterances of the same sentence that differed only in durational variability (when tempo is controlled for) or only in tempo (when durational ratios are controlled for) will allow us to disentangle rhythm and tempo and to estimate the unique contribution of each type of the timing pattern.

The first set of stimuli was comprised of 45 original utterances, 15 sentences, each produced by a beginner, an intermediate and advanced learner of English. The three versions of these sentences differed both at segmental as well as prosodic levels, i.e., in realization of vowel quality and quantity, consonantal features, intonation, lexical and sentence stress, speech rhythm, speech rate, voice quality etc.

The pre-selected 45 utterances were used to create further stimuli for the perception experiment. Praat was used to annotate and measure the durations of the original phone realizations (consonants and vowels). These durations were

used to resynthesize the original utterances using the MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) based on the native British English diphone database (en1). The utterances were resynthesized in two versions. First 45 utterances were resynthesized with a flat F0 set to 115 Hz throughout the utterance (average F0 for male European voices (Baken & Orlikoff, 2000). Besides, 45 intoned utterances were produced based on a native British English intonation contour. For this, a male native British English speaker was asled to pronounce the chosen 15 sentences. This “gold standard” F0 contour was resynthesized in the three versions of the same sentence based on the segmental durations of a beginner, intermediate and advanced learners.

Subsequently, the utterances resynthesized with the segmental durations of intermediate learners, and increased the overall duration of each utterance by 10% to create fast utterances, and decreased the overall utterance duration by 10% to create slow utterances (figure 5-2).

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 not different because tempo-normalized durational variability measures did not differ between the versions of the same sentence. The 10% manipulation was chosen because it slightly exceeds the 8% just noticeable difference (JND) in tempo for linguistic stimuli (Quene, 2007). Based on this JND it can be assumed that the 10% acceleration or deceleration will be sufficiently salient for the human auditory system. Therefore, if the differences in accent rating between faster and slow versions of the same sentence are indeed observed, these differences in the degree of the perceived foreign accent can be attributed to differences in speech rate.

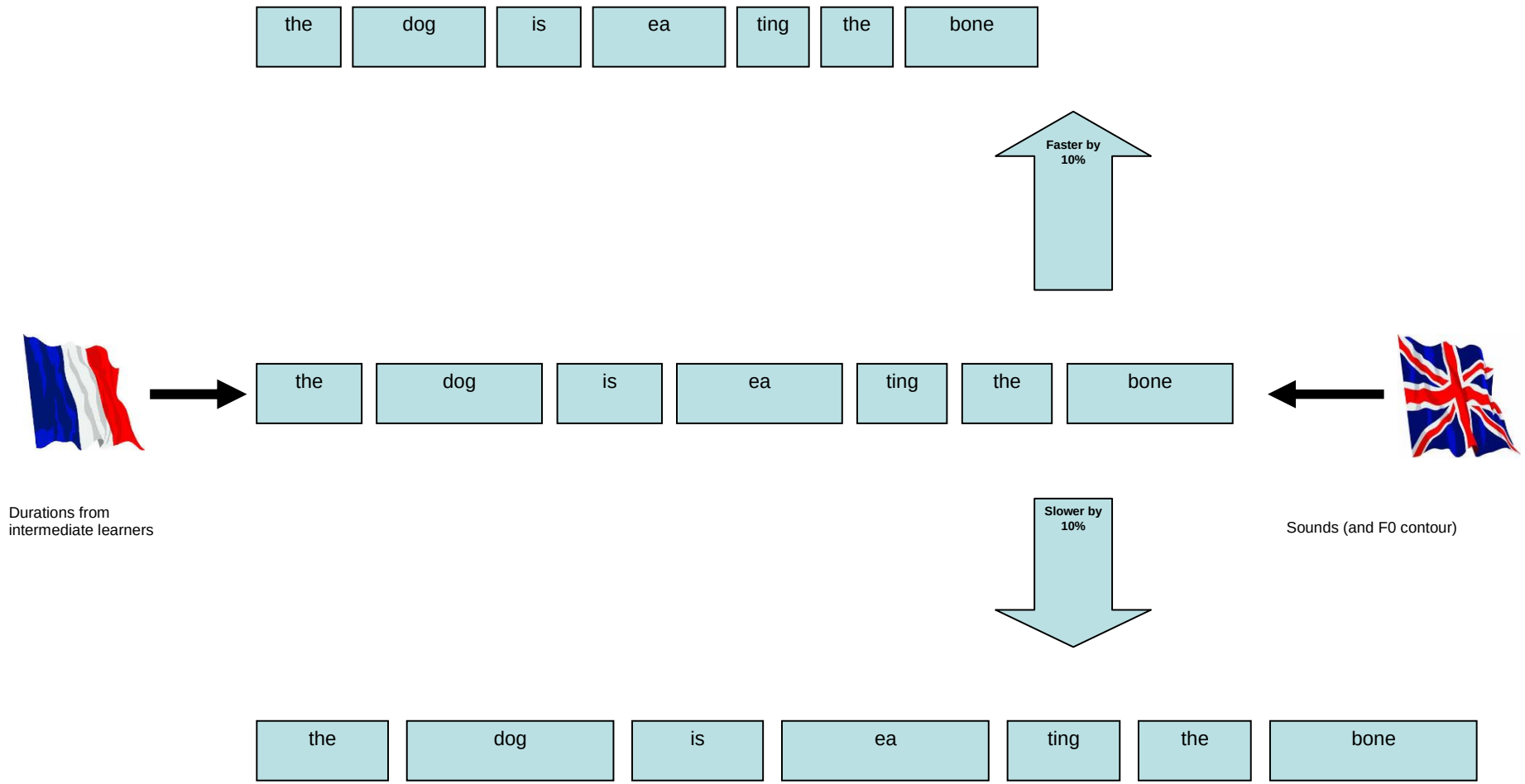


Figure 11**.** 5-2: Creating the utterances that differed only in speech rate.

To create another set of stimuli, I modified the overall duration of the resynthesized utterances with segmental durations from advanced and intermediate learners and from beginners was modified. I made the overall duration of the resynthesized utterances with the segmental durations from advanced learners equal to the overall duration of the corresponding fast utterances. The overall duration of the resynthesized utterances with segmental durations from beginners were made equal to the overall duration of decelerated utterances. The overall duration of the resynthesized utterances with segmental durations from intermediate learners was not modified and was left equal to the overall duration of the utterances at normal tempo (figure 5-3). 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 stretched or compressed proportionally, and the durational variability of speech intervals captured by the tempo-normalized rhythm metrics is not affected. As the result of such manipulations, the three versions of the same sentence were obtained. The versions were different in speech rate and in speech rhythm. Again, flat and intoned utterances were created.

To create the fourth set of stimuli (again, both intoned and flat versions), the overall utterance durations for three versions of the same sentence was equalized (figure 5-4), and the resulting utterances differ 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.

These stimuli were used in the subsequent perception experiments to verify the influence 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.

These stimuli were used in the subsequent perception experiments to verify the influence 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.

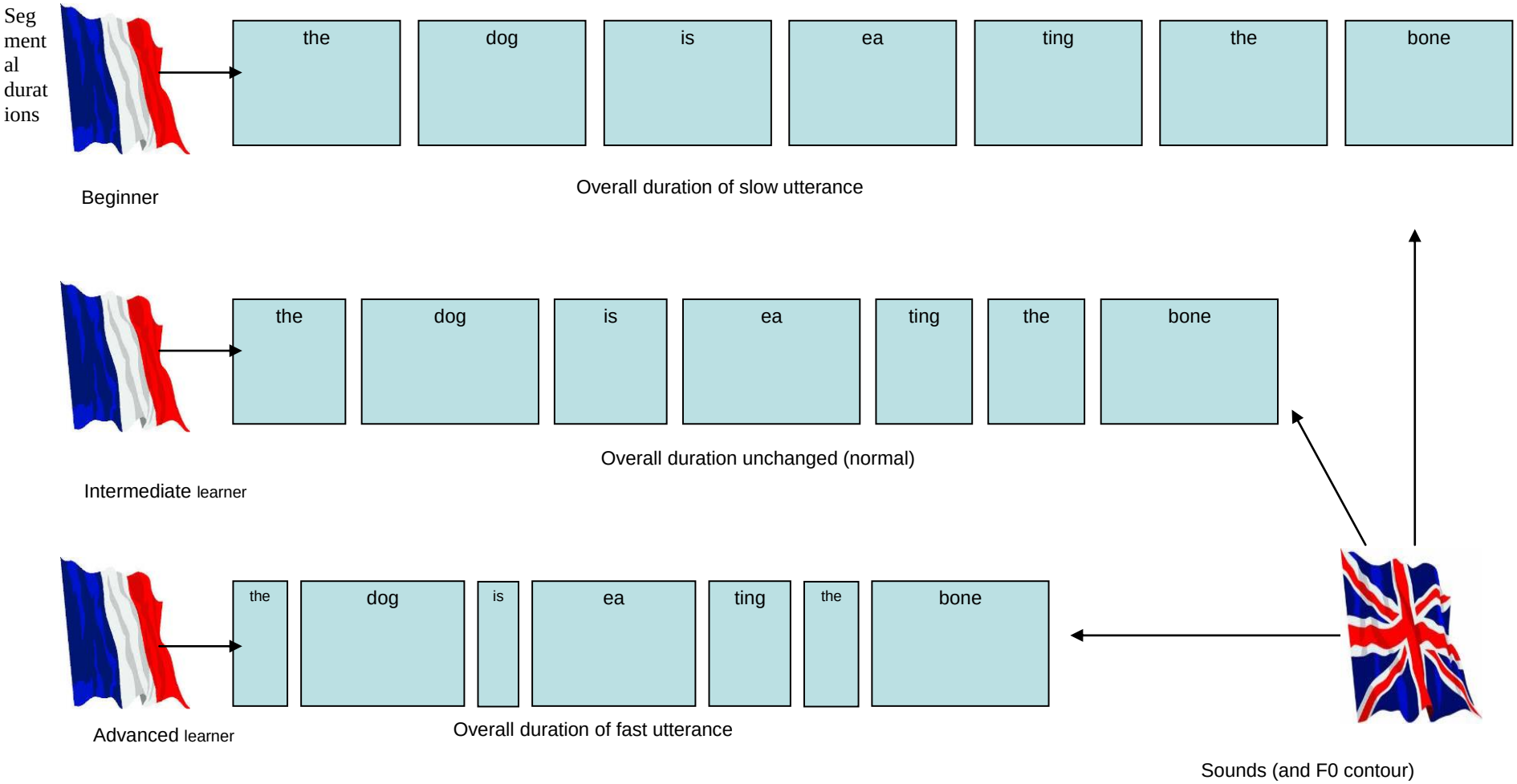


Figure 12**.** 5-3: Creating the utterances that differed only in speech rate and durational variability.

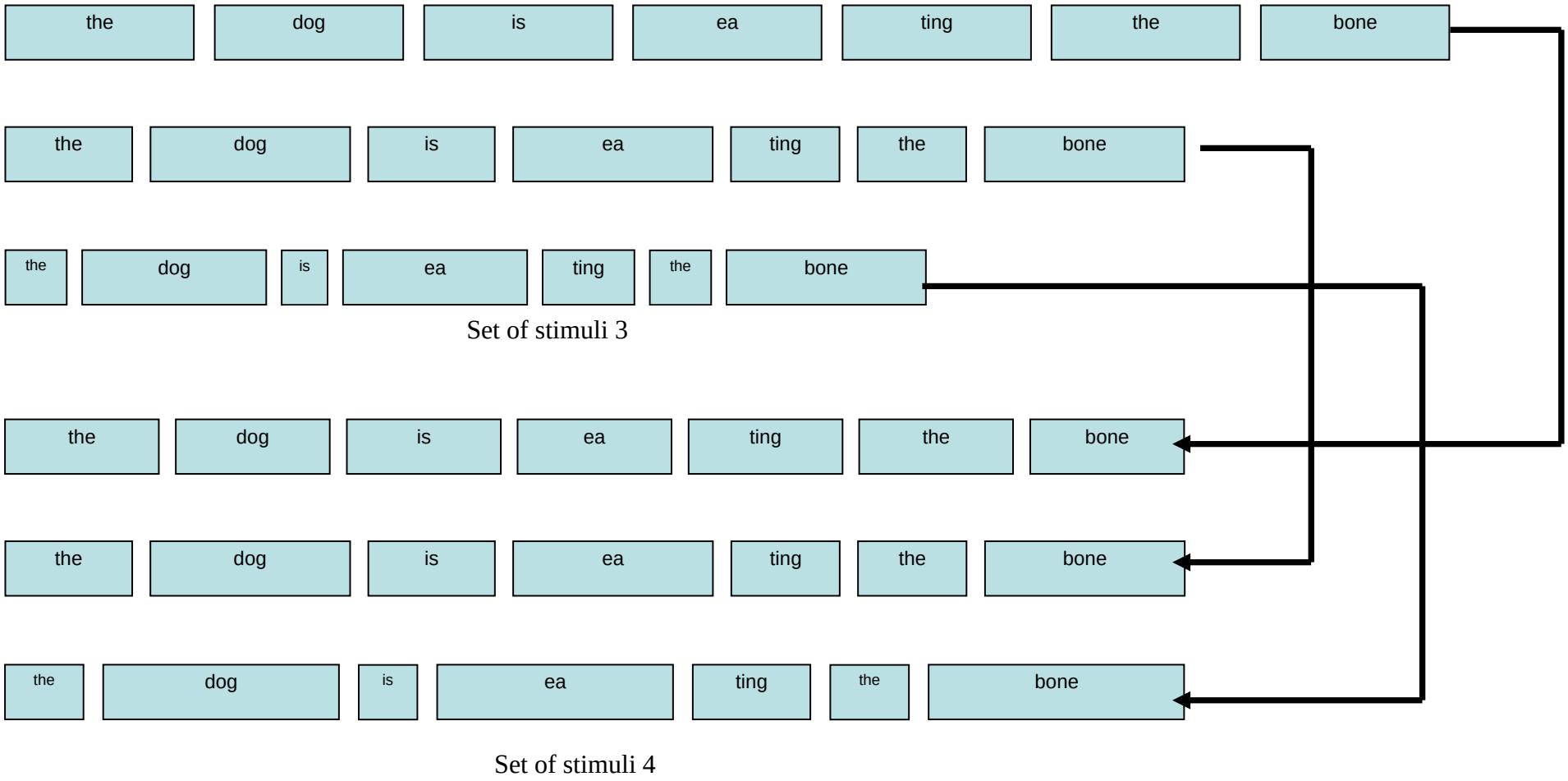


Figure 13. 5-4: Creating the utterances that differed only in speech rhythm.

5.3. Procedure

14 monolingual British English native speakers (age 25-45 years, 7 females) were recruited to participate in the experiment. Each participant had to come for the experiment four times, and the intervals between the listening sessions were 14 days. It was assumed that the effect of previous sessions will diminish over two weeks and the effect of habituation will not occur. Each time the participants had to listen to the stimuli of a certain type, first flat, then intoned (except for the original utterances, as there are no monotonized versions of the original utterances). The order in which the stimuli of different types were presented in different sessions was randomized.

During each session, 45 utterances of a certain 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 will be presented several times, and there were no pauses or experiment interruptions between the blocks. Thus, the total number of stimuli of each type the listeners had to evaluate was 135 (for the Originals) and 270 (for the resynthesized utterances, 135 flat stimuli followed by 135 intoned stimuli).

The listener had to listen to the stimuli, one by one, and evaluate the degree of the perceived foreign accent on a 6-point scale, from 1 (strongest accent) to 6 (native-like). Other degrees 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 listened to each stimulus and rated the degree of foreign accent by pressing a button on the computer screen using the mouse. The participants could replay the stimulus twice after the initial presentation by clicking the *replay* button. After the degree of the accentedness was evaluated, a new stimulus was played. On average, one session lasted 20 minutes for the original sentences and 40 minutes for the resynthesized stimuli.

5.4. Results

5.4.1. Assessment of Between-Rater and Within-Rater consistency

The first block in each session was considered as a familiarization phase and was not used in the subsequent analysis. It is assumed that during familiarization the

listeners made a reference line, the basis for comparison of the accentedness degree. The ratings given to the same stimulus when it was presented in the second and in the third blocks were used to evaluate inter-listeners consistency. Inter-rater consistency was calculated to make sure that the same stimulus receives the same or a comparable rating when it is evaluated several timed. We were also interested in agreement between listeners, i.e., whether an utterance that is rated as being more accented by one listener will also be rated as such by the others. Consistency of rating patterns between the listeners rather than the absolute agreement between listeners was in the focus. One of the listeners might be a stricter or a more tolerant rater than others. This should affect the mean ratings, but should not affect the consistency of the rating pattern. For example, if an utterance *A* is rated lower than an 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 the utterance *A* than a stricter listener.

To estimate the within- listeners and between-listeners reliability, a Cronbach alpha and a Guttman Split-Half test were used. The second and the third blocks (and the ratings given to the stimuli in the second and the third presentation) were considered as the same test administered to the same participants twice. Therefore, it is possible to estimate true test-retest reliability. The tests show that both inter-listeners and intra-listeners reliability was high (table 5-2) and justified using the obtained ratings for further analysis how speech rhythm and speech rate influence the degree of the perceived foreign accent.

Table 5-2. Assessment of reliability and consistency in foreign accent rating between and within listeners. The table presents Chronbach alpha for the ratings given to the stimuli presented to all 14 listeners in the second block (second table column) and in the third block (third table column). Split-Half coefficient shows averaged within-rater agreement for 14 raters, i.e., testing that the same stimulus receives the same rating when presented to the same person twice.

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

5.4.2. Differences in Foreign Accent rating between utterances produced by French learners of English at different proficiency levels

To verify the influence of presence of intonation, type of the stimuli and proficiency level of the speaker on the degree of the perceived foreign accent, a repeated-measures ANOVA was performed with *type* (intoned vs. flat), *stimuli* (Originals, Rhythm and Tempo, Rhythm Only and Tempo Only) and *level* (beginner, intermediate and advanced) as within-subject factors and with *accent rating* as the dependent variable. The differences in *accent rating* between *levels* (Λ = .204, $F(2, 208) = 406.99$, $p < .0005$, $\eta^2 = .796$) and between *stimuli* (Λ = .734, $F(3, 207) = 25.017$, $p < .0005$, $\eta^2 = .266$) were significant. A significant, though unsubstantial difference was also detected for *accent rating* depending on whether the stimuli were intoned or monotonized (Λ = .958, $F(1, 209) = 9.115$, $p = .003$, $\eta^2 = .042$). However, a significant and substantial interaction between *stimuli* and *level* (Λ = .245, $F(6, 204) = 105.041$, $p < .0005$, $\eta^2 = .755$) and unsubstantial but significant interaction between *type* and *level* (Λ = .947, $F(3, 207) = 406.99$, $p = .01$, $\eta^2 = .053$) were also revealed. This indicates that the *accent rating* is affected differently depending on how many and which cues are available to the listeners for making their judgement. The significant and substantial interaction between

stimuli and *level* might indicate that the perception of distinctions in rhythmic patterns between the utterances produced by learners at different proficiency levels is also affected by whether the differences in rhythmic patterns are accompanied by the differences in speech rate. The significant interaction of *type* and *level* indicates that the distinctions in rhythmic patterns between proficiency levels may be perceived differently in intoned and monotonized utterances. However, the presence of intonation may affect the perception of differences in speech rate as well. These interactions make interpretation of the main effects problematic.

Therefore, a repeated-measures ANOVA with *accent rating* as dependent variable and with *level* as factor was performed separately for each type of stimuli (Originals, Rhythm and Tempo, Rhythm Only and Tempo Only). The analysis was also performed separately on intoned and on flat stimuli. The ANOVAs were followed by controlled comparisons (Bonferroni correction). Ratings were compared within each category for the utterances from advanced learners of English with those from intermediate learners of English, and the ratings for the utterances from intermediate learners of English were compared with those from beginners. It was found that the accent ratings differ between proficiency levels for all types of stimuli, both for flat (table 5-3) and intoned (table 5-4) utterances.

When the listener rated the resynthesized utterances that differed only in speech rate, they perceived faster utterances as less accented and slower utterances as more accented compared to the utterances with the speech rate of an intermediate French learner of English. Interestingly, the difference in accent rating on Tempo Only utterances was smaller (figure 5-5) when the listeners had to rate intoned utterances, and the effect size on intoned Tempo Only utterances (η^2 = .193) was smaller compared to that on flat utterances (η^2 = .249).

When listeners evaluated the degree of foreign accent on the resynthesized utterances that differed both in rhythm and in tempo, with utterances preserving the rhythmic patterns of advanced learners being faster and the utterances preserving the rhythmic patterns of beginners being slower, we can see that the differences between the proficiency levels are significant, and the range of accent rating is higher than in the conditions when the participants evaluated Rhythm Only and Tempo Only utterances.

When listeners evaluated the degree of foreign accent on the resynthesized utterances that differed both in rhythm and in tempo, with utterances preserving the rhythmic patterns of advanced learners being faster and the utterances preserving the rhythmic patterns of beginners being slower, it is evident that the differences between the proficiency levels are significant, and the range of accent rating is higher than in the conditions when the participants evaluated Rhythm Only and Tempo Only utterances.

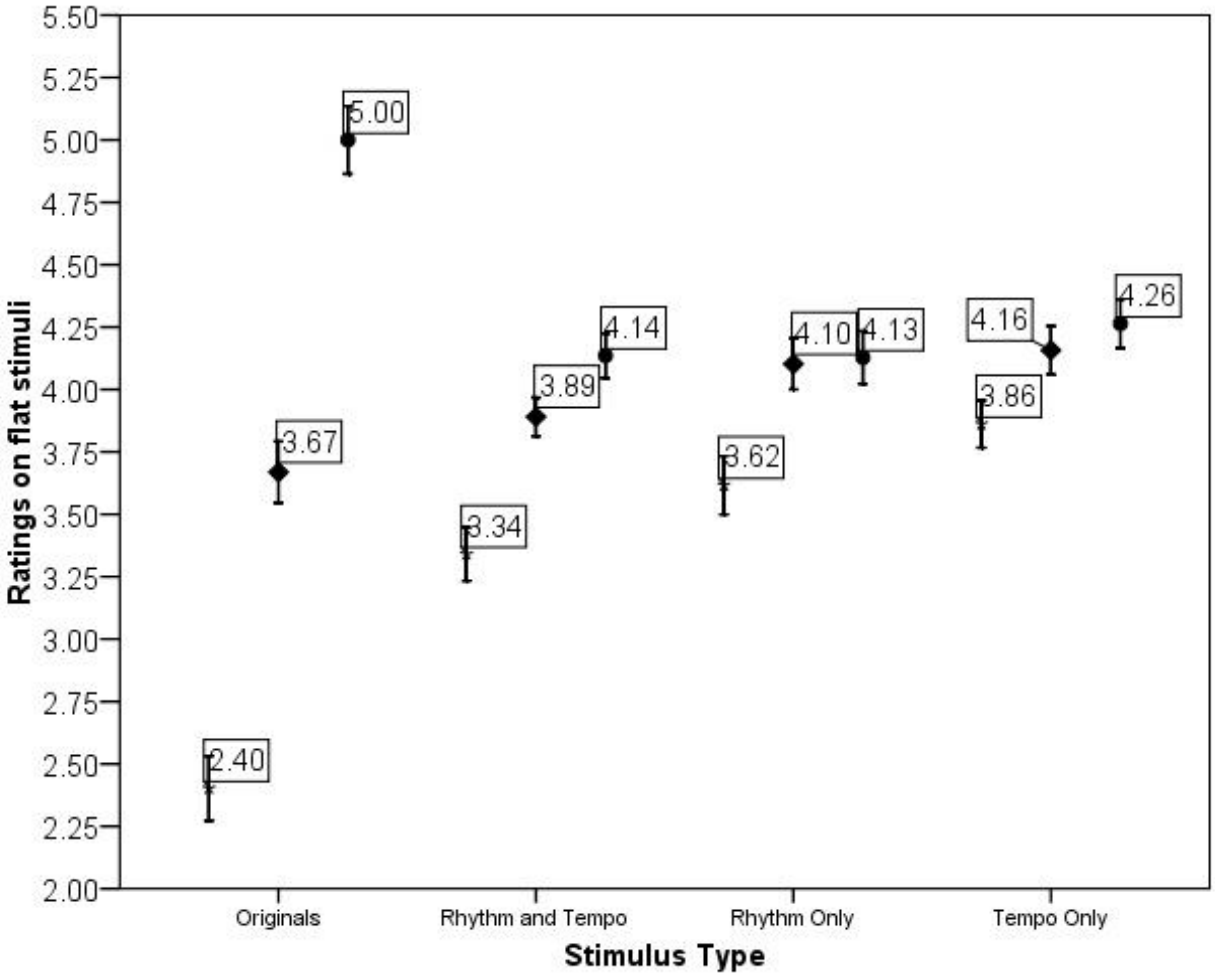


Figure 14. 5-5a: Accent rating on different types of flat (monotonized) stimuli. Error bars show ±2SE. Stars stand for the ratings received by beginners, squares intermediate and circles for advanced learners of English.

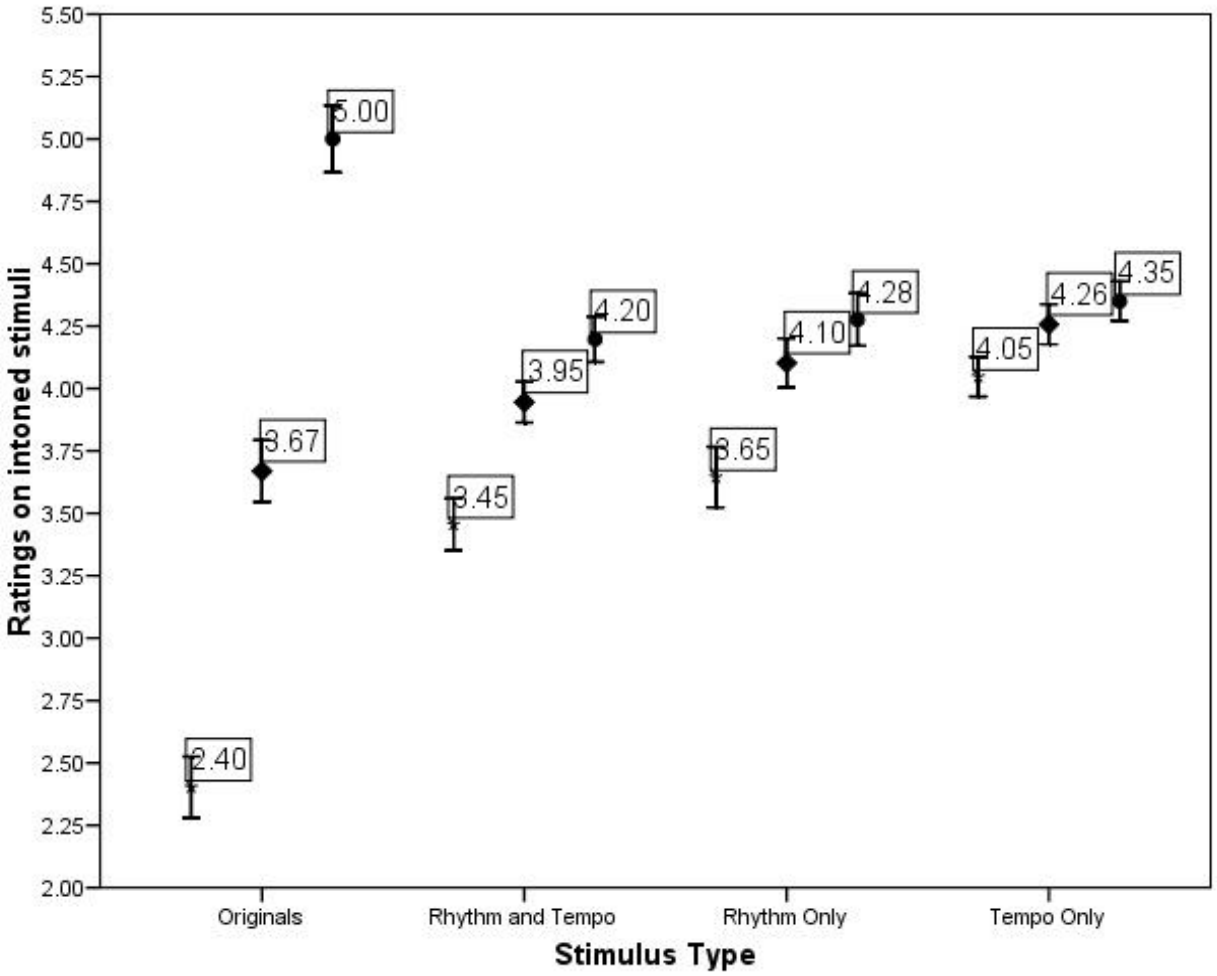


Figure 15**.** 5-5b: Accent rating on different types intoned stimuli. Error bars show ±2SE. Stars stand for the ratings received by beginners, squares intermediate and circles for advanced learners of English.

Table 5-3. Statistic data for the influence of *level* on the *accent rating* for different types of flat (monotonized) stimuli. The star in the significance column stands for $p < .0005$

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
Originals	.18	472.915	*	.82	325.649	*	.609	317.108	*	.603
Rhythm and Tempo	.595	70.781	*	.405	24.08	*	.103	73.872	*	.261
Rhythm Only	.68	49.049	*	.32	.352	.554	.002	84.212	*	.287
Tempo Only	.751	34.543	*	.249	5.64	.018	.026	42.891	*	.17

Table 5-4. Statistic data for the influence of *level* on the *accent rating* for different types of intoned stimuli. The star in the significance column stands for $p < .0005$

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
Originals	.18	472.915	*	.82	325.649	*	.609	317.108	*	.603
Rhythm and Tempo	.597	70.226	*	.403	27.783	*	.117	73.588	*	.26
Rhythm Only	.648	56.404	*	.352	12.885	*	.058	57.645	*	.216
Tempo Only	.807	24.847	*	.193	6.826	.01	.032	23.852	*	.102

5.5. Discussion

The third experiment dealt with individual and combined contribution of speech rate and speech rhythm into the perceived foreign accent, when the native language and the target language of the learner are rhythmically contrastive (French as the native and English as the target languages). The results confirm the outcome of the previous experiments that timing pattern pertaining to the durational variability of speech intervals and to speech tempo contain sufficient cues to reliably rate the degree of the foreign accent and to estimate the proficiency levels of the learner. Rhythmic patterns and speech rate together account for 43% (for intoned stimuli) and for 40% (for the flat stimuli) of variance in accent rating, when the segmental and intonational differences between utterances produced by learners at different proficiency levels are neutralized. It was also revealed that the differences in speech rhythm only, all other factors being equal, explains more variance (35% for intoned utterances and 32% for flat utterances) than the utterances that differ only in speech tempo (19% and 25% respectively). This means that the differences in durational variability of vocalic, consonantal and syllabic intervals make a bigger contribution on perceived foreign accent than the differences in the mean duration of the corresponding speech intervals (speech rhythm overweighs speech rate in perception of the degree of accentedness).

The analysis shows that accent ratings differ between proficiency levels for all types of stimuli, both for flat (table 5-3) and intoned (table 5-4) utterances. Original utterances and the resynthesized utterances produced by advanced

learners received better ratings than those produced by intermediate learners, and the utterances produced by beginners received the lowest scores. Thus, when the listeners had segmental, prosodic and indexical cues available, they assessed utterances from beginners as more accented and the utterances from advanced learners as less accented, and the differences between the pairs of proficiency levels were significant. When listeners had to evaluate the resynthesized utterances that kept only the rhythmic patterns of English learners at different proficiency levels in L2, a clear difference is evident in the result pattern 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, the listeners could hear the difference in rhythmic patterns between the utterances of advanced and intermediate learners, as well as between intermediate learners and beginners. Therefore, it is concluded that the presence of intonation enhances the perception of rhythm differences.

When the listener rated the resynthesized utterances that differed only in speech rate, they perceived faster utterances as less accented and slower utterances as more accented compared to the utterances with the speech rate of an intermediate French learner of English. Interestingly, the difference in accent rating on Tempo Only utterances was smaller (figure 5-5) when the listeners had to rate intoned utterances, and the effect size on intoned Tempo Only utterances (η^2 = .193) was smaller compared to that on flat utterances (η^2 **=** .249). This might indicate that intonation impedes perception of small differences in speech rate. When intonation lacks, listeners have more resources to process the only parameter that differs between the utterances – speech rate – and these distinctions in speech rate become more salient in flat stimuli. However, this inference is inconclusive and requires further investigation.

The results show that faster sentences are perceived as less accented and slower sentences are perceived as more accented, all other factors being equal. Besides, significant differences were found in accent ratings given to the utterances that preserve the patterns of durational variability of advanced, intermediate and beginning learners of English. At the same time, the differences

in accent rating are bigger, if we combine faster tempo and patterns of durational variability of advanced learners, and slower tempo and durational ratios of beginning learners. Thus, we conclude that faster speech rate on the utterances with rhythmic patterns of advanced learners of English reduces the degree of the perceived foreign accent, while slower speech rate on the utterances with the rhythmic patterns of beginners increases accentedness. Consequently, speech rate and speech rhythm have a cumulative as well as a unique, independent effect on the perceived foreign accent. The unique contribution of speech rhythm into foreign accent, especially on intoned sentences, is bigger 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 Tempo Only utterances in intoned condition.

The major difference in result pattern between the second and the third experiments is the relative contribution of speech rate and speech rhythm into foreign accent. The results clearly show that if the native and the target language of the learner are rhythmically contrastive, the differences in rhythmic patterns between utterances produced by L2 learners at different proficiency levels become more salient than when the target and the native language of the learner are rhythmically similar. This clearly answers our fourth research question: the native language of the learner indeed influences the relative contributions of speech rate and speech rhythm into the degree of perceived foreign accent.

The result pattern of this experiment agrees with the result pattern of the second experiment in that the sentences produced by lower-level L2 learners received significantly lower scores in the original version compared to the resynthesized sentences.

The opposite was found in the comparison of original and resynthesized sentences produced by advanced L2 learners, i.e., the original sentences received higher scores that the manipulated sentences. Sentences produced by intermediate L2 learners received comparable scores in both original and modified conditions. This might indicate that the influence of segmental characteristics differs depending on the proficiency level. In other words, tempo and rhythmic patterns are overridden by foreign accented segmental characteristics in speech produced by L2 learners with a lower proficiency level. On the other hand, timing patterns of the target language become perceptually more salient for native

speakers with increased mastery of the segmental characteristics. This confirms our tentative pedagogical suggestion that maintaining timing patterns of the target language at the syllabic, consonantal and vocalic levels (durational ratios, i.e., timing patterns not related to phonemic differences between long and short vowels, for example) becomes more important at higher proficiency levels, while at lower proficiency levels it is more beneficial to the students to concentrate on maintaining phonemic realizations and stress patterns (unmarked stress location, phonetic correlates of stress their weighting, distribution of stress in the speech flow). Comparison of the accent ratings given to the flat and resynthesized utterances, however, casts some doubt that intonation will seriously reduce the foreign accent in non-emphatic declarative sentences with broad focus. The ratings given to the flat and intoned utterances do not differ. Intonation, on the other hand, interacts with perception of other prosodic systems, e.g., with rhythm, and enhances perception of rhythmic differences between utterances produced by learners of different proficiency levels. An interesting question for further investigation is whether the presence of intonation per se (i.e., the presence of F0 contour) enhances perception of rhythmic differences, or whether only an English-specific intonation contour will enhance the perception of rhythmic patterns by native English listeners. How native and non-native intonation contours will interact with other prosodic cues, and how a non-native intonation contour will affect the perception of speech rhythm is still to be discovered.

6. General Discussion

The results of the previous and the presented experiments clearly show that the average speech tempo and variation in duration of vocalic, consonantal and syllabic intervals change as L2 acquisition progresses. The direction of change is similar regardless of whether the native language of the learner is similar to or different from the target language in terms of speech rhythm and rate, however, the magnitude and the rate of change seem to be dependent on the L1 (Ordin and Polyanskaya, accepted; Ordin, Polyanskaya, & Wagner, 2015). For example, as acquisition of English as L2 progresses, both French and German learners speak faster and with higher degree of durational variability (Ordin & Polyanskaya, 2015; Ordin et al., 2015), and this can be schematically illustrated on figure 6-1.

If native and target language of the learner are rhythmically similar, native speakers of English classify timing patterns in L2 English based on tempo and ignore differences in rhythm. The second experiment shows that non-linguistic stimuli that preserve the tempo and rhythmic characteristics of advanced, intermediate and beginning German learners of English (three distinct groups) are classified into two categories based on the rate of consonant-vowel transitions, as shown in figure 6-2. Both categories include the stimuli with high and with low variation in duration of vocalic (“A”) and consonantal (“S”) intervals, but the first group features faster stimuli (i.e., stimuli with shorter vocalic and consonantal intervals), while the second group includes slower stimuli (i.e., stimuli with longer vocalic and consonantal intervals). This shows that differences in variation in duration are completely ignored in non-linguistic stimuli in a classification task.

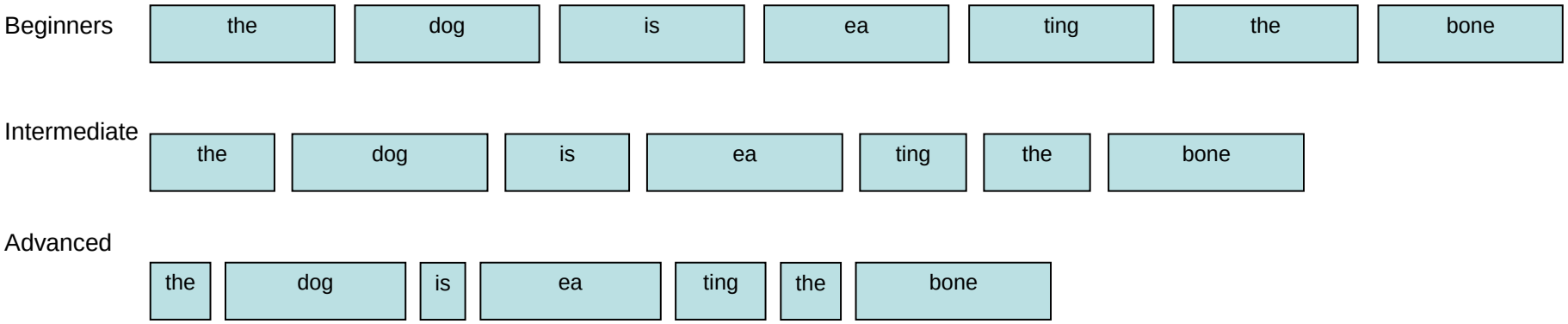


Figure 16. 6-1: Development of speech rate and speech rhythm in L2 acquisition

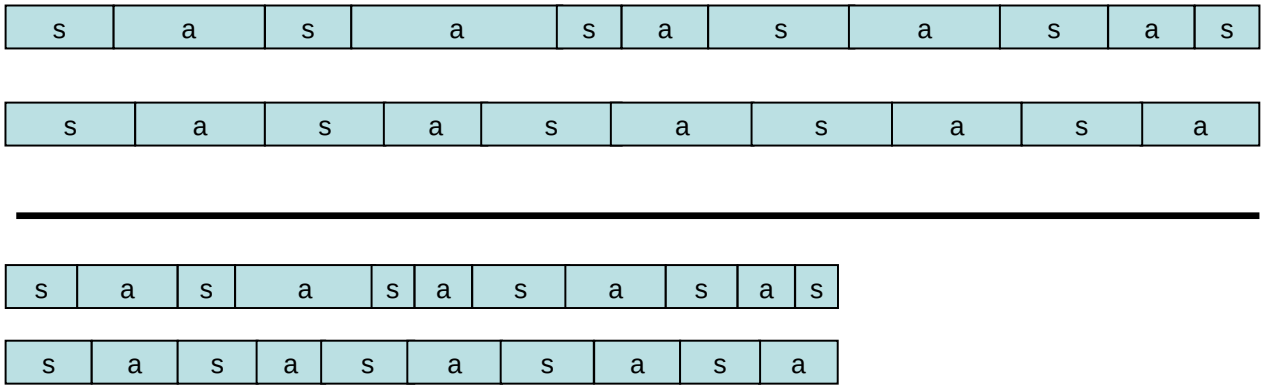


Figure 17**.** 6-2: Splitting the 'sasasa' stimuli into two categories based on durational variability of vocalic and consonantal intervals and speech rate.

The results of the second and the third experiments reveal that on linguistic stimuli both developmental differences in durational variability and in speech rate are perceivable and influence the degree of the foreign accent. It was found that both tempo and durational variability contribute to the perceived foreign accent, but in different degrees. The degree to which native speakers of English hear the differences in rhythm and tempo between proficiency levels of L2 learners depends on the native language of the learner. If native and target languages of the learner are rhythmically similar, tempo makes a bigger contribution into the degree of perceived foreign accent than rhythm (figures 4-2 and 4-3). If native and target language of the learner are rhythmically different, then rhythm makes a bigger contribution into the degree of perceived foreign accent (figure 5-5a and 5-5b). If the target and native language of the learner are rhythmically similar, then the effect of rhythm on accent rating will be smaller compared to the effect of rhythm if the native and target languages are rhythmically contrastive (figures 4-2, 4-3, 5-5a and 5-5b).

It was also found that intonation enhances perception of rhythm differences (figures 5-5a and 5-5b). This can be explained by the fact that the rhythmic differences become perceptually more salient in linguistic, speech-like stimuli. One can expect that intonation makes the resynthesized stimuli more natural than flat stimuli, and it is possible to hypothesize that this increased naturalness leads to a more prominent perception of variability. As rhythmic patterns in more natural stimuli are better perceived, the differences in durational variability of vocalic, consonantal and syllabic intervals are more prominent in intoned stimuli.

The presence of intonation allowed listeners to hear the differences in rhythmic patterns in L2 English produced by intermediate and advanced French learners. This difference, although acoustically present in speech, was not perceived when the listeners heard monotonized sentences. However, when I imposed **the same** intonational contours on utterances with rhythmic patterns of intermediate and advanced learners, the difference is speech rhythm becomes perceptually relevant for assessing the degree of foreign accent. This very important conclusion has been merely stated in the dissertation, but it definitely requires further interpretation.

There are reasons for assuming that prominence may play a role in the interaction between perception of rhythmic differences and the presence of intonation. Prominence at different levels of prosodic hierarchy plays a role in segmentation (Ordin & Nespor, 2013), and listeners are sensitive to prominence fluctuations within utterances (Lee & Todd, 2004; Arnold, Wagner, & Möbius, 2011; Cole, Mo, & Hasegawa-Johnson, 2010). Some parts of the utterance (syllables in a word or words in a phrase) are made more prominent by syllable or vowel lengthening and by local fluctuations of pitch. Lengthening is used to mark lexical stress. Lexical stress is prominence at a word level. Stressed syllable and stressed vowels are generally longer (Dogil and Williams, 1999). Local fluctuations of pitch mark prominence at a higher level of prosodic hierarchy. In other words, local pitch modulations are correlates of phrasal prominence (Bolinger, 1958). When a syllable is made more prominent at a phrasal level, it often bears an additional degree of lengthening. As prominence is very important for speech segmentation, comprehension, lexical access, etc., people are sensitive to its acoustic correlates – pitch and duration. At the same time, fluctuations in duration (i.e., lengthening of syllables that bear lexical and phrasal stress) that are used to make one syllable more prominent than other also influence rhythmic patterns because they change the relative durational variability of vowels and syllables. This can explain why the presence of intonation (pitch contour over the utterance with local pitch peaks and valleys) can enhance perception of rhythm, i.e., perception of durational variability of speech units.

The integrative notion of prosodic prominence (Tamburini & Caini, 2005; Tamburini & Wagner, 2007) can be used to account for this interaction of F0 and durational variability in perception. This notion operationalizes prosodic

prominence as a weighted sum of F0 fluctuations and the overall fluctuation in syllabic energy related to pitch accents, and durational variation and the mid-to-high frequency emphasis related to lexical stress. These four prosodic parameters interact in prominence manifestation (Tamburini & Caini, 2005). However, weighting of these parameters is language-specific and differ between Italian, German and English (Tamburini & Caini, 2005; Tamburini & Wagner, 2007). Perception of the fluctuations in these acoustic parameters is also influenced by the linguistic expectations (Tamburini & Wagner, 2007). Listeners expect that F0 will interact with durational ratios, and the linguistic expectations adjust the "perceived prominence" even when acoustic manifestation of prominence is not exactly corresponding to the native norms. Thus adding F0 to the stimuli enables native English listeners in my experiments to perceive prominence fluctuations and thus also to perceive differences in durational ratios better, even though the acoustic measures in durational ratios are not always native-like.

These explanation for why the presence of F0 contour also enhances perception of rhythmic differences, and the idea of prosodic prominence as a weighted sum of acoustic parameters (Tamburini & Caini, 2005; Tamburini & Wagner, 2007) entails yet another question for further research, namely, what kind of F0 fluctuations enhance perception of rhythmic patterns: Local ones, i.e., those associated with pitch-accented syllables, or global ones, i.e., spanning over whole utterances, in other words, the presence of intonational contours per se. This line of research is to be pursued in the future.

It was shown that the original utterances produced by lower-level L2 learners received significantly lower scores compared to the resynthesized utterances that preserved timing patterns but included segmental characteristics and intonation of native English speakers. The opposite was found in the comparison of original and resynthesized utterances produced by advanced L2 learners, i.e., original utterances received higher scores that the manipulated utterances. Utterances produced by intermediate L2 learners and resynthesized utterances that preserved the timing patterns of intermediate learners but with the segmental and intonational contours of native English speakers received comparable scores. This might indicate that the influence of segmental characteristics on the perceived foreign accent differs depending on the proficiency level. In other words, perception of deviations in speech rate and

speech rhythm from the target patterns is overridden by accented segmental characteristics in speech of L2 learners at a lower proficiency level. As pronunciation of the segmental characteristics of the target language is improved, deviations in prosodic timing become perceptually more salient. Consequently, from the pedagogical perspective, maintaining timing patterns of the target language becomes more important on higher levels of L2 mastery, while on lower proficiency levels it is more worthwhile to concentrate on the phonemic realizations, and probably on such prosodic aspects as stress and intonation.

It was also established that the developmental changes in L2 speech rate and speech rhythm can be perceived – to a certain extent – separately, thus it makes sense to develop pronunciation training activities aimed at improving either speech fluency, or at speech rhythm. Practicing these two aspects of timing patterns separately may have a better effect on the overall pronunciation mastery of L2 learners.

It was also established that the developmental changes in speech rhythm (durational variability) are more salient for the native speakers of the stress-timed languages if the native language of the learners is rhythmically contrastive. On the other hand, if the rhythmic patterns in the native and the target languages of the learners are similar, then the developmental changes in durational variability contribute very little to the perception of accentedness. Consequently, it makes more sense to work on maintaining the durational ratios only if the learners are studying a prototypically stress-timed language and their native language is prototypically syllable-timed.

To put that all in a nutshell, at earlier stages of L2 learning it is more advisable to concentrate on acquisition of segmental characteristics of the target language and on stress patterns (location of stress, realization of stress, etc). At more advanced levels, it is advisable to practice the prosodic timing patterns, if the native and the target languages are rhythmically contrastive.

The obtained results might be used to expand the modern theories of second language learning and acquisition. Such theories, as it has been mentioned in the introductory part, have been almost exclusively developed and verified based on the data related to acquisition of segmental phenomena. The new data allows expanding the theories of L2 acquisition and learning from acquisition of segmentals onto the acquisition of L2 prosody. One of the proposals

of the Flege's (1995) speech learning model specified that L2 sounds that are different from a native-language (L1) sound will be easier to perceive than an L2 sound that is relatively similar to an L1 sound. We have shown that deviations in rhythmic patterns in L2 speech delivered by the French learners of English are easier to detect for the native English speakers than deviations in durational ratios in speech of German learners of English. This explains why the non-native rhythmic patterns of French learners have a greater influence on accentedness compared to the non-native rhythmic patterns of German leaners, whose native language rhythm is more similar to the target rhythm. Details of acquisitional dynamics of L2 rhythm in production do not fit so nicely into the whole picture because rhythmic patterns appear to be spread over the continuum rather than grouped into discrete categories. Moreover, rhythmic patterns in the languages that are traditionally referred to as stress-timed languages exhibit a large degree of within-language variability, with some utterances being more stressed-timed and the other utterances being more syllable-timed. So, this scattering of utterances in terms of durational ratios and the percentage of vocalic material is rather large, which makes it difficult for native speakers of English (and other stressed-timed languages) and for leaners of English to form the categories. And this makes it difficult to interpret the results in terms of the Speech learning model, which explains the dynamics of acquisition in terms of discrete categories for phonemes. Prosodic distinctions in general and rhythmic distinctions in particular are not discrete, but gradual (see, for example, Arvaniti, 2012 or Ordin and Polyanskaya, accepted for discussion of graduality in rhythmic distinctions within- and between languages). Therefore, rhythmic cross-linguistic and cross-interlanguage differences can hardly fit into the Speech Learning Model, the Native magnet language model or Assimilation model, which rely on the ideas of categorical perception. However, the details on ***perception***, and particular on the contribution of rhythmic patterns (and probably other prosodic patterns) into the perceived foreign accent are easily interpretable in terms of Flege's (1995) model that rests on the basic premises that it is possible to perceive phonetic (not categorical, i.e., phonological) differences between L2 speech sounds, and that the perception and production of sounds is guided by perceptual representations stored in long-term memory. It is possible to expand these premises to the phonetic, i.e., gradual distinctions in prosody and thus in gradual rhythmic distinctions, as well as the

distinctions between native and non-native patterns. The ability to perceive these gradual distinction is a pre-requisite for further successful formation of segmental categories, but for processing prosodic distinction between languages or between interlanguages (i.e., proficiency levels of L2 speakers) this further step is not necessary.

This dynamics of acquisition in production is more easily explained in terms of exemplars, which will probably bridge the gap between production and perception (Clopper and Pisoni, 2004). The major advantage of this theory is the premise that processing of linguistic structures is not relied on a single prototype (or several few prototypes), i.e., on a singe (or few) representative for a whole class. Rather than relying on a single representative, many samples of a pattern are stored in memory, and each new linguistic pattern that is perceived is compared against a whole cloud – with fuzzy boundaries – of clustered exemplars. This explains the graded membership of each separate pattern into a certain class much better. Considering that the idea of rhythmic classes is highly debatable, and that the rhythmic patterns within as well as between languages are spread in the continuum (see the introductory part of this thesis, or Ordin and Polyanskaya, accepted; Arvaniti, 2012 for discussion), it would be difficult to reconcile with the idea that there is a single prototype for a language-specific rhythmic pattern. Rather, native speakers of English have experience of processing both stress- and syllable-timed patterns, as the languages that are traditionally defined as stress-timed can exhibit a high degree of durational variability in some utterances and low degree of durational variability in other utterances produced by the same individual. Thus, there are no rhythmic prototypes, but a cloud of exemplars for rhythmic patterns in the long-term memory of native English speakers, and they need to compare each new pattern with the whole cloud when processing a new sentence in the course of speech perception. The exemplar theory does not pre-suppose the existence of formation of categories not only in perception, but also in production, it is even better than Flege's model to account for our results, because the Speech learning model does make the assumption that the learnt structures are categorized before they can be fully acquired by the learners for the purposes of production, although this assumption is not necessarily valid for perception. The idea of exemplars makes this assumption for categories neither in perception, not in production acquisition, thus it provides a better bridge between production and

perception. However, the detailed discussion of how the exemplar theory is applied to our results goes far beyond the scope of my work and needs a new research project that promises new insights into acquisition of non-categorical prosodic differences in production and perception, the line of research that is well worth pursuing in the future.

Major's (2001) Ontogeny-Phylogeny model of second language acquisition offers an adequate theoretical framework to account for the results (when combined with the expectations formulated based on Flege's model). When the assumptions of both the Speech Learning Model and the Ontogeny-Phylogeny models are combined, the resulting model might have a substantially increased explanatory power. The Ontogeny-Phylogeny model provides an account of language-specific and universal forces that drive and control the direction, magnitude and rate of changes of speech patterns in the course of L2 acquisition. The Speech learning model explains the variation in the extent to which individual learners acquire the patterns of the target language, and also in the extent to which individuals perceive the differences between native language norms and deviations from these norms. The Speech learning model, in other words, explains perception of phonetic details and mapping these details on phonological categories, while the Ontogeny-phylogeny model explain the driving forces that control the development of these abilities in the course of L2 acquisition. Therefore, when both models are put together, we have a very powerful theoretical framework to account for the development of prosodic patterns in L2 and, what is much more important for this thesis, to account for the *perception* of these patterns and for perception of deviation of L2 prosody from the native norms by the native speakers of the target language.

French learners of English need to acquire L2 rhythmic patterns that do not exist in their native language, and these target rhythmic patterns are marked. They are marked because they are less frequent cross-linguistically and the presence of stress-timing is not implied by the presence of syllable-timing, while the presence of syllable-timing is implied by the presence of stress-timing, and these two criteria are sufficient to claim the more marked status of a linguistic phenomenon (Eckman, 1977; 1991). Acquisition of marked phenomena is determined at earlier stages of acquisition by universal linguistic patterns. At earlier stages of acquisition the influence of native language on development of marked linguistic feature in

interlanguage decreases rapidly, but the influence of universals increases rapidly as well. Consequently, these two forces contradict each other. The influence of syllable-timing that is characteristic of French is decreasing rapidly, but the influence of universal pattern, which is also syllable-timing, is increasing rapidly. Thus, the speech produced by French learners is characterized by syllable-timing for quite a long time, before the influence of universals begins to decrease, and it only happens at later stages of acquisition, giving way to development of a marked speech feature – stress-timing – that is characteristic of the target language (English). Thus speech rhythm of French learners remains very different from the English native norms, and, according to the Flege's model, these differences are easily picked up and strongly influence the degree of perceived FA. Therefore, rhythmic deviations in L2 English produced by French learners make a substantial contribution into accentedness. When Germans acquire English rhythm, their acquisitional dynamics is determined by the universals at earlier stages of acquisition, which prevents them from merely transferring the rhythm, but when they achieve the intermediate proficiency level, the influence of universals and native language decreases rapidly, and rhythmic patterns are acquired quickly, and the native-like degree of stress-timing is achieved, diminishing the overall influence of durational ratios in L2 English spoken by German learners onto the foreign accent.

7. Conclusion

In response to the research questions that were set at the beginning of the quest for detecting and estimating the contribution of speech rate and rhythm into perceived foreign accent. It is possible to say that:

1) Native speakers of the target language hear the differences in the examined timing patterns between the sentences produced by L2 learners at different proficiency levels. However, they tend to ignore the differences in durational variability in classification task and in non-linguistic stimuli. The more natural and the more speech-like the stimuli are, the more perceptually prominent the differences in durational variability become.

2) The degree of the perceived foreign accent is indeed affected by the developmental changes in speech rate and speech rhythm. More advanced learners tend to speak with more rapidly and with higher degree of durational variability at the timescale of vowels and concomitants and at the timescale of syllables. Sentences with higher degree of durational variability and faster sentences are perceived as less accented by the native speakers of English, all other factors being equal.

3) Combined contribution of durational variability and speech rate into the perceived foreign accent is bigger than the unique contribution of either durational variability or speech rate. However, we have also found the contribution of durational variability when controlling for speech rate and the contribution of speech rate when controlling for speech rhythm. Thus, there is not only combined but also unique contribution of the interrelated characteristics of tempo and durational variability into the perceived foreign accent.

4) The relative contribution of durational variability and speech tempo into the perceived foreign accent depends on whether the native language of the learner is rhythmically similar to or different from the target language. The contribution of durational variability into foreign accent is bigger when the native language and the target language are rhythmically contrastive (e.g., English as the target language and French as the native language of the learner), than when the

languages are rhythmically similar (e.g., English and German). If the languages are rhythmically similar, then the native speakers pay much more attention to the speech rate (faster sentences are assessed as less accented) when evaluating the degree of accentedness.

8. References

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

2. 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

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

4. 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.

5. Arnold, D., Wagner, P., & Möbius, B. (2011). Evaluating different rating scales for obtaining judgments of syllable prominence from naive listeners. *Proceedings of the ICPhS 2011*, 252– 255.

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

7. Arvaniti, A., (2009). Rhythm, timing and the timing of rhythm. *Phonetica 66(1-2)*, 46-63.

8. Baken, R., & Orlikoff, R. (2000). *Clinical Measurement of Speech and Voice.* San Diego: Singular Publishing Group.

9. 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.

10. Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America 114*(3), 1600-1610.

11. Bertinetto, P. M., & Bertini, C. (2008) On modeling the rhythm of natural languages. *Proc. of the 4th International Conference on Speech Prosody*, Campinas 2008, 427-430.

12. Best, C. (1995). A direct realist viewo f cross-language speech perception research with adults. In W.Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 171–204). Timonium, MD: York Press.

13. Binghadeer, N. (2008). An acoustic analysis of pitch range in the production of native and nonnative speakers of English. *The Asian EFL Journal Quarterly 10*, 96-113.

14. Bion, R.A.H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech 54(1)*, 123-140.

15. Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

16. Boersma, P., Weenink, D., "Praat: doing phonetics by computer" (Version 5.1.22) Retrieved from http://www.praat.org/

17. Bolinger, D. (1958). A theory of pitch accent in English. *Word, 7*, 109-149.

18. Bond, Z. (1999). *Slip of the Ear: Errors in the Perception of Casual Conversation*. San Diesgo, CA: Academic Press.

19. Bond, Z., & Fokes, J. (1985). "The timing of English words by non-native speakers. *J. Acoust. Soc. Am. Suppl. I 77*, S53

20. Bosch, L., & Sebastian-Galles, N. (2001). Early language differentiation in bilingual infants. In: J. Cenoz, F. Genesee (Eds.), *Trends in Bilingual Acquisition* (pp.71–93). Amsterdam: Benjamins

21. Boula de Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica 63*, 247–267.

22. Brahimi, B., Boula de Mareuil, P., & Gendrot, C. (2004). Role of segmental and suprasegmental cues in the perception of maghrebian-acented French. In *Interspeech 2004, proceedings* 341-344.

23. 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.

24. Bybee, J., Chakraborti, P., Jung, D., & Scheibman, J. (1998). Prosody and segmental effect: Some paths of evolution for word stress. *Studies in Language, 22*, 267-314.

25. Cambier-Langeveld, T., Nespor, M., & van Heuven, V., J. (1997). The domain of final lengthening in production and perception in Dutch. *EUROSPEECH-1997*, 931-934.

26. 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.

27. Christophe, A. & Dupoux, E. (1996). Bootstrapping lexical acquisition: The role of prosodic structure. *Linguistic Review* 13(3-4), 383-412.

28. Christophe, A., Gout, A., Peperkamp, S., & Morgan, J.L. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics 31*, 585-598.

29. Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics 9*, 237-245.

30. Clopper, C., & Pisoni, D. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics 32*, 111–140

31. Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation based factors in the perception of prosodic prominence. *Journal of Laboratory Phonology 1*, 425–452.

32. Cook, V. J. (2009). Language user groups and language teaching. In V. J. Cook & Li Wei (Eds) *Contemporary applied linguistics: Volume 1 Language Teaching and Learning*.(pp.54-74) London: Continuum.;

33. Cook, V. J. (2010). The relationship between first and second language acquisition revisited. In E. Macaro (Ed.) *The Continuum Companion to Second Language Acquisition* (pp.137-157). London: Continuum

34. Crookes, G., Davis, K., 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.

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

36. D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2005). Intonational Phrasing in Romance: The role of prosodic and syntactic structure. In S. Frota, M. Vigário & M. J. Freitas (Eds) *Prosodies.* Phonetics & Phonology Series. Berlin: Mouton de Gruyter, pp.59-97.

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

38. Dauer, R. (1987). Phonetic and Phonological Components of Language Rhythm. *Proceedings of the11th International Congress of Phonetic Sciences, August 1-7, 1987, Tallinn, Estonia,* 447-450.

39. Dellwo, V., & Wagner, P. (2003). Relationships between Rhythm and Speech Rate. In *Proceedings of the 15th ICPhS,* 471-474.

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

41. Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition 19*, 1-16.

42. 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.

43. Diez, F., Dellwo, V., & Gavalda, N. (2008) The development of measurable speech rhythm during second language acquisition. *Journal of Acoustical Society of America 123(5)*, 3886-3886.

44. Dogil, G., & Williams, B. (1999). The phonetic manifestation of word stress in Lithuanian, Polish, German, and Spanish. In H. van der Hulst (Ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 273–311). Berlin: Mouton de Gruyter.

45. 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.

46. Eckert, H., & Laver, J. (1994). *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation*. Weinheim: Psychologie Verlags Union.

47. Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning 27(2)*, 315–330.

48. Eckman, F. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition 13(1)*, 23-41.

49. Elordieta, G., Frota, S., & Vigário, M. (2005). Subjects, objects and intonational phrasing in Spanish and Portuguese. *Studia Linguistica* 59 (2-3), 110-143.

50. Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly 39*, 399-423.

51. Flege, J. & Eefting, W. (1987B). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics 15*, 67-83.

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

53. 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.

54. Flege, J., & Eefting, W. (1987a). Cross-language switching in stop consonant production and perception by Dutch speakers of English. *Speech Communication 6*, 185-202.

55. 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.

56. Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication 40*, 467-491.

57. Flege, J., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech 38*, 25-55.

58. Frota, S. (2000). *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. ed. 1, 1 vol.. New York: Garland Publishing.

59. Fujisaki, H., Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan 5*, 233-242.

60. Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition 36(2)*, 257-281.

61. Gabriel, C., Stahnke, J., & Thulke, J. (2015). Assessing foreign language speech rhythm in multilingual learners: An interdisciplinary approach. In Peukert, H. (Ed). *Transfer Effects in Multilingual Language Development*. Amsterdam: Benjamins, 191-219.

62. Gabriel, C., Stahnke, J.; & Thulke, J. (2015). Acquiring English and French speech rhythm in a multilingual classroom: A comparison with Asian Englishes. In Gut, U., Fuchs, R., Wunder, E.-M. (Eds). *Universal or diverse paths to English phonology*? Berlin: De Gruyter, 135-163.

63. Gibbon, D. (1998). German intonation. In Daniel Hirst & Albert di Cristo (Eds.), Intonation Systems (pp.78-95). Cambridge University Press, 1998.

64. Gluszek, A., & Dovidio, J. F. (2010). The way they speak: Stigma of non-native accents in communication. *Personality and Social Psychology Review 14*, 214–237.

65. 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.

66. Greenberg, S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication, 29*(2-4), 159-176.

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

68. 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.

69. Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language 36(1)*, 3-15.

70. 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.

71. Gussenhoven, C. (2004). *Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

72. Gut, U. (2007). Foreign Accent. In C. Muller (Ed.) *Speaker Classification I: Fundamentals, Features, and Methods*. *Lecture Notes in Computer Science* 4343, 75-87.

73. Hahn, L.D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly 38*, 201–223.

74. Horne, M. (1989). The Clitic group as a prosodlc category in Old French. *Lund Working Papers in Linguistics*, *35*, 99-111.

75. Hughes, A. (2002). *Testing for Language Teachers*. Second Edition. Cambridge: CUP.

76. Hughes, A., Trudgill, P., & Watt, D. (2012). *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles*. London: Hodder Education.

77. James, C. (1998). *Errors in Language Learning and Use*. London: Longman.

78. 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.

79. Kachru, B. (1992). World Englishes: approaches, issues and resources. *Language Teaching* 25: 1-14.

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

81. 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.

82. Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *International Journal of Bilingualism 6(3)*, 315-334.

83. Klatt, D. (1976). Linguistic uses of segmental durations in English: Acoustic and perceptual evidence. *Journal of Acoustical Society of America 59*, 1208-1221.

84. 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.

85. Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: U. of Pennsylvania Press.

86. Ladd, R. (2008). *Intonational Phonology.* Second Edition, Cambridge: CUP.

87. Lee C. S., & Todd, N. P. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory ‘primal sketch’ to two multi-language corpora. *Cognition 93*, 225–254.

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

89. Lepetit, D. (1989). Cross-Linguistic Influence in Intonation*:* French/Japanese and French/English. *Language Learning 39*, 397–413.

90. Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition 12*, 251–285.

91. Loukina, A., & Kochanski, G. (2010). Patterns of durational vvariation in British dialects. Presentation at PAC colloquium 2010 on 13 september 2010, Montpellier, France.

92. 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.

93. Magen, H. (1998). The perception of foreign-accented speech. *Journal of Phonetics 26*, 381-400.

94. Major, R. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology.* London: Lawrence Erlbaum Associates.

95. Major, R., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning 46*, 465-496.

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

97. Marcus, S. M. (1981). Acoustic determinants of perceptual centre (P-centre) location. *Perception and Psychophysics 30,* 247-256.

98. Mazuka, R. (1996). How can a grammatical parameter be set before the .rst word? In J.L. Morgan & K. Demuth (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, (pp.313–330). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

99. McNamara, T. (2000). *Language Testing*. Oxford: OUP.

100. Mehler, J., Dupoux, E., Nazzi, T. & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. Morgan & Demuth K.D. (Eds) *From Signal to Syntax: Bootstrapping from speech to grammar in early acquisition*, (pp 101-116) Hillsdale, NJ: Erlbaum.

101. Mehler, J., Jusczyk, P., Dehaene-Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition 29*(2), 143–178.

102. Mehler, J., Sebastian-Galles, N., & Nespor, M. (2004). Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. In: M. Gazzaniga (Ed.). *The New Cognitive Neuroscience* (825–36). Cambridge, MA: MIT Press.

103. Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics 32*, 543-563.

104. Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In Trouvain, J. & Gut, U. (Eds.) *Non-native Prosody: Phonetic Descriptions and Teaching Practice (Nicht-muttersprachliche Prosodie: phonetische Beschreibungen und didaktische Praxis)(53-76)*. Mouton De Gruyter.

105. Mennen, I. Speech production in simultaneous and sequential bilinguals (2011). In P. Howell and J. van Borsel (Eds.) *Multilingual aspects of fluency disorders.* (pp.24-42)*.* Bristol: Multilingual Matters.

106. Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language difference in f0 range: a comparative study of English and German. *Journal of the Acoustical Society of America 131(3)*, 2249-2260.

107. Mok, P. (2011). The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition 14*(4), 458-472.

108. Morgan, J.L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language 35*, 666–688.

109. Morton, J., Marcus, S. M., & Frankish, C. (1976). Perceptual centres (P-centres). *Psychological Review 83*, 405-408.

110. Munro, M. J. (1995). Nonsegmental Factors in Foreign Accent. Ratings of Filtered Speech. *Studies in Second Language Acquisition 17*, 17-34.

111. Munro, M. J. (1995). Non-segmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17, 17-34.

112. 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.

113. 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.

114. Nakai, S., Turk, A., Kari, S., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics 40(6), 796-807*.

115. 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

116. Nespor, M., Guasti, M.T., & Christophe, A. (1996). Selecting word order: The Rhythmic Activation Principle. In U. Kleinhenz (Ed.), *Interfaces in Phonology*, pp. 1–6. Berlin: Akademie Verlag.

117. Ordin, M. & M. Nespor (*2013*) Transition probabilities and different levels of prominence in segmentation. *Language Learning*, 63(4), 800-834.

118. Ordin, M., Polyanskaya, L, Wagner, P. (2015). Acquisition of speech rhythm in second language: A cross-linguistic study. In A. Leemann, Kolly, M.-J., Schmid, S., & Dellwo, V. (Eds.) Trends in Phonetics & Phonology in German-speaking Europe (pp. 331-347). Zurich: Peter Lang.

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

120. Ordin, M., Polyanskaya, L. (2015). Production and perception of speech rhythm in English as the second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology* 6, 1-15.

121. Ordin, M., Polyanskaya, L. (accepted). Acquisition of speech rhythm in second language by learners of English with rhythmically different native languages. *Journal of Acoustical Society of America*.

122. Ordin, M., Polyanskaya, L., Ulbrich, C. (2011). Acquisition of Timing Patterns in Second Language*. Interspeech 2011*, 1129-1132.

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

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

125. Pierrehumbert, J. & Steele, S. (1990) Categories of Tonal Alignment in English. *Phonetica 47*, 181-196.

126. Pierrehumbert, J. (1990) Phonological and Phonetic Representation. *Journal of Phonetics 18*, 375-394.

127. Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational contours in the Interpretation of Discourse. In P. Cohen, J. Morgan. & M. Pollack. (Eds*). Intentions in Communication* (271-311). Cambridge MA: MIT Press.

128. Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics 29*, 191-215.

129. Polyanskaya, L., Ordin, M., & Ulbrich, C. (2013). Contribution of timing patterns into perceived foreign accent. In P. Wagner (Ed.). *Elektronische Sprachsignalverarbeitung 2013, 71-79.* Dresden: TUDpress.

130. Port, R., & Mitleb, F. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics 11*, 219–229.

131. 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.

132. Quene, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics 20(3)*, 331-350.

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

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

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

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

137. Scharff –Rethfeldt, W., Miller, N., Mennen, I. (2008). Unterschiede in der mittlere Sprechtonhöhe bei Deutsch/Englisch bilingualen Sprechern (Speaking Fundamental Frequency Differences in Highly Proficient Bilinguals of German/English). *Sprache Stimme Gehör 32*, 123 – 128.

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

139. Scott, S. (1998). The point of p-centres. *Psychological Research 61*, 4-11.

140. Scovel, T (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning 20*, 245–253.

141. Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics 20*, 213–223.

142. Shport, I. (2008). Acquisition of rhythm: Evidence from spontaneous L2 speech. *Journal of Acoustical Society of America 123(5)*, 3328-3328.

143. t'Hart, J.H., & Cohen, A. (1973). Intonationa by rule: A perceptual quest. *Journal of Phonetics 1,* 309-327.

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

145. Tamburini, F., & Caini, C. (2005). An automatic system for detecting prosodic prominence in American English continuous speech, *International Journal of Speech Technology 8*, 33–44.

146. Tamburini, F., & Wagner, P. (2007). On Automatic Prominence Detection for German. *Proceedings of Interspeech 2007*, 1809–1812.

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

148. Taylor, L. (2011). *Examining Speaking: Research and Practice in Assessing Second Language Speaking*. Cambridge: CUP.

149. Thiessen, E., & Saffran J. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development 3(1)*, 73-100.

150. Thomas, K. (2007). Just noticeable differences and tempo change. *Journal of Scientific Psychology* 14-20.

151. Turk, A., & Shattuck-Hufnagel, S. (2007). Multiple targets for phrase-final lengthening in American English words. *Journal of Phonetics 35*, 445-472.

152. Ulbrich, C., & Mennen, I. (in prep). When prosody kicks in! The intricate interplay between segments and prosody in perceptions of foreign accent. *Second Language Research*.

153. Underhill, N. (1987). *Testing Spoken Language: A Handbook of Oral Testing Techniques*. Cambridge: CUP.

154. Venneman, T. (1988). Systems and changes in Early Germanic phonology: A search for hidden identities. In D. Calder, & T. Christy (Eds.), *Germania: Comparative studies in the Old Germanic languages and literatures*, (pp. 45-65), Woodbridge: Suffolk.

155. Wells, J. C. (1982). *Accents of English*. 3 volumes. Cambridge: Cambridge University Press.

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

157. White, L., & Turk, A. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics 38*, 459-471.

158. White, L., Mattys, S., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythmic class. *Journal of Memory and Language. (in press).*

159. White, L., Payne, E., & Mattys, S. (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigario, S. Frota & M.J. Freitas (Eds.). *Phonetics and Phonology Interactions and Interrelations* (137-155). Amsterdam: John Benjamins.

160. Whitworth, N. (2002). Speech rhythm production in three German-English bilingual families. *Leeds Working Papers in Linguistics & Phonetics 9*, 175-205.

161. 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.

162. Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of Acoustical Society of America 91*, 1707-1717.

9. Appendix

1. **The dog is eating a bone**

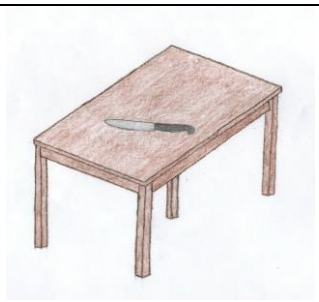
2. **The book is on the table**

3. **The girl is eating an apple**

4. **The ball is on the chair**

5. **The boy is kicking the ball**

6. **The man is catching a fish**

7. **The knife is on the table** 

8. **The children are watching the giraffe** 

9. **The boy is drinking juice** 

10. **The bread is on the table** 

11. **The boy is eating ice-cream** 

12. **The boy is under the tree** 

13. **The cat is drinking milk** 

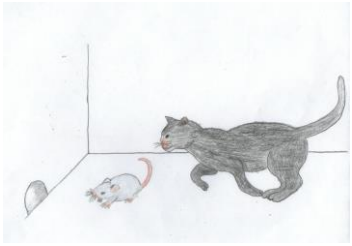
14. **The girl is talking on the phone** 

15. **The girl is buying a balloon** 

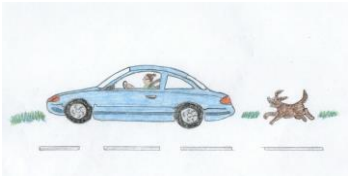
16. **The woman is eating an orange** 

17. **The baby is crying** 

18. **The baby is sleeping** 

19. **The cat is chasing the mouse** 

20. **The man is walking** 

21. **The dog is chasing the car** 

22. **The baby is taking a bath** 

23. **The boy is riding a bike** 

24. **It's raining outside** 

25. **The girl is eating candy** 

26. **The children are going to school** 

27. **It's snowing outside** 

28. **The dog is chasing the cat** 

29. **The spoon is on the table** 

30. **The boy is running to the car** 

31. **The boy is talking on the phone** 

32. **The boy is chasing the girl** 

33. **The computer is on the table**