SPEECH RHYTHM CHANGE IN AN INTERNATIONAL SPOKEN ENGLISH ENVIRONMENT

Master thesis – 35 ECTS

Cognitive Artificial Intelligence

University Utrecht

Student:	Kimberley Giok Mie Cruden – 3704254
First supervisor:	Hugo Quené
Second reviewer:	Menno Lievers
Third reviewer:	Gerrit Bloothooft
Date:	February 20th, 2014

CONTENT

1. INTRODUCTION

The acquisition of a second language (SLA) can be influenced by the speech characteristics of a native language (Flege, Frieda, & Nozawa, 1997; Hansen, 2001). For non-native speakers this often results in having certain non-native segmental and prosodic speech characteristics, which can contribute to having a foreign accent (Mareüil & Vieru-Dimulescu, 2006). With the increase of linguistic experience and linguistic proficiency a speaker's speech characteristics might also change. In the present study the English speech of national and international students will be compared between two recording sessions, of which one is conducted at the beginning of the bachelor's study and the other is conducted at the end of the bachelor's study. In this way changes in speech rate and changes in speech rhythm will be analyzed.

Finding differences in speech rate and in speech rhythm would shed light on how these two speech characteristics can change overtime. Both speech rate and speech rhythm have shown to be important aspects of native (L1) and non-native (L2) speech. There is a relation of speech rate (and pauses) to linguistic proficiency (Giles, Coupland, & Coupland, 1991), fluency of L2 speech (Bosker, Pinget, Quené, Sanders, & De Jong, 2013) and to foreign accent (Trofimovich & Baker, 2006). Speech rhythm is important in speech perception (Quené & Port, 2005) and in native and non-native speech because the speech rhythm of the L1 can influence the speech rhythm of the L2, and so contribute to having non-native speech characteristics (Yuan, 2010; White & Mattys, 2007a, 2007b). The importance of speech rhythm to language was demonstrated for Dutch and English when it was found that words that contain stressed syllables were perceived better when they were aligned at regular time points (Quené & Port, 2005; Quené & Van den Berg, 2008). Therefore, regular speech timing improves spoken-word perception. Attempts to quantify speech rhythm have lead to several proposed quantification metrics (for a review see e.g. Loukina, Kochanski, Rosner, Shih, & Keane, 2011). These rhythm metrics were mostly and foremost developed to discriminate between different language rhythms, but they have also shown discriminative ability within the same language rhythms and within the same language (White & Mattys, 2007a, 2007b). In the present study, the rhythmic discrimination of several rhythm metrics will be compared within one language and between two recordings. The results of this study will show how well the rhythm metrics can discriminate within-language rhythm and will show which rhythm metric performance best. This study will specifically aim to find out whether international students will change their speech rate and their speech rhythm in an international English spoken environment during their 3-year bachelor's study. The participating students are from the same cohort and live on campus of University College Utrecht (UCU), in The Netherlands. The majority of participants are native Dutch and so it is likely Dutch speech characteristics will influence the English spoken at

UCU. This, variety of, English (or UCU English) and its speech rhythm (or UCU English speech rhythm) will be analyzed over time between new students (or non-native) and near-graduates (or native) speakers of UCU English. The speech rate will be analyzed from spontaneous informal speech and the speech rhythm will be analyzed from read-aloud sentences. This leads to the following research questions:

RQ1: Will there be change in speech rate in spontaneous informal L2 English speech over time?
RQ2: Will there be change in speech rhythm in read L2 English speech over time?

This thesis consists of a theoretical part followed by an experimental part. Chapter 2 and Chapter 3 will provide the theoretical background and the implications on the methodological options and choices for the present study will be explained. Firstly, in Chapter 2 two core concepts of speech rate and speech rhythm are discussed – as these form the basis of this study. Section 2.1 describes speech rate, its quantification methods and the factors by which speech rate is influenced most. Section 2.2 follows the same set-up as Section 2.1 for discussing speech rhythm – elaboration on speech rate, quantification methods, influential factors. Chapter 3 is focused on accommodation and acculturation, focussing on a sociolinguistic/sociological perspective. This is of high relevance for the understanding of the possible outcomes of long-term exposure to language in a new international environment. Accommodation strategies are described as part of the Speech Accommodation Theory (Beebe & Giles, 1984) in Section 3.1. Next, the Acculturation Model (Schumann, 1986) explicates the best cultural environmental conditions for optimal L2 learning. The accommodation strategy, which follows in case these conditions are being, is described in Section 3.3. Chapter 4 will serve as a bridge between the theoretical background and the experimental part of this thesis, and will summarize the methodological choices and state the hypotheses. The experimental part begins with the methodological section in Chapter 5. Chapter 6 is the result section, followed by the interpretation of the results and the discussion and conclusion in Chapter 7.

2. RATE AND RHYTHM IN L1 AND L2

2.1 SPEECH RATE

Speech tempo, or speech rate, is the core concept of the present study's first research question. Speech rate is one of the acoustic features important for speech intelligibility and for perceived foreign accent (Trofimovich & Baker, 2006; Bosker et al., 2013). It has also been indicated that faster speech tempo is likely to contribute to a decrease in comprehension especially the comprehension of heavily accented speech in non-native language (Anderson-Hsieh & Koehler, 2006). In other words, a "good" speech rate is highly important in order to have a comprehensible and understandable conversation, in both L1 and L2. The different ways of quantifying speech rate will be discussed in Section 2.1.1. Studies have applied these quantification measures and reported on several linguistic and paralinguistic factors as being influential to speech rate, which are discussed in Section 2.1.2. Lastly, the behavior of speech rate in an SLA environment and its relation to language proficiency as well as mechanisms by which it increases the perception of a stronger foreign accent is discussed in Section 2.1.3.

2.1.1 QUANTIFICATION

The two measures commonly used in studies that attempt to quantify speech rate are speaking rate and articulation rate (Laver, 1994). When speaking about speech tempo or speech rate without specification on whether speaking rate or articulation is meant, the default measure is speaking rate. Sometimes the inverse measure of syllables per second, called average syllable duration (Quené, 2008), is used but typically speaking rate and articulation rate are measures given in syllables per second or phonemes per second (Laver, 1994). The speaking rate denotes the number of speaking units during a certain time unit, the articulation rate is the speaking rate with the durations of pauses removed of the equation. Articulation rate and speaking rate capture different properties of speech, differing on their consideration of pauses.

Both the duration of speech units and the duration of pauses are important. Miller, Grosjean and Lomanto (1984) were the first to prove that variation in articulation rate was causing substantial variation in speaking rate. They found this was not just because of the use of pause durations as previous research had implied. This means both the durations of the speech units and the durations of the pauses are important. Some silent portions are intra-segmental, for example the Voice Onset Time or VOT (Lisker & Abrahamson, 1964), while others are inter-lexical and are intended as pauses (Zellner, 1994). The standard auditory threshold of pauses is usually 200ms-300ms as pauses are more easily perceived beyond this threshold and pauses of these lengths are rarely observed to occur intra-lexical. Different speaking styles have shown substantial differences in pause use. For example political speeches show longer and more frequent pauses when compared to casual interviews and spontaneous speech (Duez, 1982).

In the present study both speaking rate and articulation rate will be considered for the analyses of the spontaneous speech rate's changes. By calculating both measures, whilst keeping the speaking style consistent within this research, pauses and speech will be quantified separately which gives clarity on both speaking rate, pause use and pause durations in the particular speaking style. The standard auditory threshold of pauses of 300ms will be used to distinguish inter-lexical silent portions from pauses.

2.1.2 LINGUISTIC AND PARALINGUISTIC FACTORS

Extensive L1 research has been done with speech rate as the dependent variable which has lead to the detection of several linguistic and paralinguistic factors that can be of influence on the variable. The most prominent influencing factors of speech rate are mentioned in this section.

Physiological changes caused by emotional arousal can affect respiration, phonation and articulation (Scherer, 1986). Emotion is considered as a paralinguistic (intrapersonal or within-speaker) factor that can alter speech rate (e.g. Davitz, Beldoch, Blau, Dimitrovsky, Levitt, & Kempner, 1964; Scherer, 1986; Breitenstein, Van Lacker, & Daum, 2001). Davitz et al. studied this implication by letting participants rate speech which was spoken with 14 different kinds of emotions on an emotion adjective scale for prosodic characteristics. These prosodic characteristics included speech rate. For "affection" and "sadness" a slow speech rate was reported and for "boredom" a moderately slow speech rate was reported. Moderately fast speech rate was corresponding with "cheerfulness" and "impatience". Fast speech rate was corresponding with "anger" whereas normal speech rate corresponded with "satisfaction". Davitz et al. concluded from this that he found a relation between the activity of the emotion adjective scale and speech rate. Also Breitenstein et al. found support for this relation between emotion and speech rate. Their research, in accordance to the results from Davitz et al., showed that relatively slow speaking rate was associated with the speaker being "sad". The relatively fast rate was associated with an "angry", "frightened" or "neutral" state. Next to that they found that less pitch variation was associated with "sad" or "neutral", and more pitch variation was associated with "frightened", "angry", and "happy" states. However, speech rate manipulation had more influence on the labeling than pitch variation manipulation. These studies have shown the influence of speech rate on emotion perception and the influence of emotion production on speech rate.

Other paralinguistic (interpersonal or between-speaker) factors that have shown influence on speech rate were demographic background factors such as *age*, *gender, country* and *region*.

Verhoeven, De Pauw and Kloots (2004) studied speakers of one language with varying demographic backgrounds on their speech rate. By using a corpus containing 160 speech samples of 15 minute conversations from Dutch and Flanders speaking participants who varied in age, gender, country and region they found that these four predictors had significant effect on articulation rate. The average articulation rate was significantly faster in younger speakers (ages 21-40) than in older speakers (ages 45-59), higher in men then in women, and faster in The Netherlands than in Belgium. Between speakers in the Netherlands the articulation rate showed to be fastest in the western region of The Netherlands (called "Randstad") and slowest in the North and the South region. The middle region Verhoeven et al. call the transition region where the articulation rate was shown to be intermediate. Between the Flanders speakers the different regions showed no significant difference. Quené (2008) extended this research by including within-speaker effects and showed the importance of *phrase length* in articulation rate. This phenomenon is called "anticipatory shortening" which means that when speakers anticipate more syllables in a phrase they shorten their syllables. Taking this within-speaker effect into consideration, Quené concludes that the Flanders speakers show more variation within their phrase length than the Dutch speakers. Also, older speakers seem to produce relatively shorter phrases than younger speakers and with more variation. Conversely, younger speakers tend to speak faster with less variation. Anticipatory shortening causes the phrase lengths to be the highest influencing (within-speaker) factor in both the results of Verhoeven et al. and Quené.

The style of speech can also influence speech rate. Infant-directed speech or motherese has shown to be produced with exaggerated temporal characteristics like more and longer pauses, shorter sentences and slower speech (Fernald & Simon, 1984). Duez (1982) showed that in political speech significantly longer pause duration and significantly higher pause frequency are used. Kowal, Wiese and O'Connell (1983) showed similar results by comparing speech in storytelling to speech during an interview. The researchers investigated both speaking styles in English, Finnish, French, German and Spanish and found differences in articulation rate and speaking rate between the speaking styles within the same language. Kowal et al. conclude that the effect of language on speech rate was negligible compared to the effect of the speaking style of speech. Nevertheless, they did show that the speaking rate of storytelling was significantly lower than speech during an interview. Picheny, Durlach and Braida (1986) found a decreased speaking rate in clear speech compared to conversational speech for English. They also reported that in conversational speech the vowels were more modified and reduced than in clear speech and not all word-final consonants were released. This implies that conversational speech would be less rhythmical and faster than clear speech. This highlights the importance of speaking style for speech rate analyses even more.

As in any study the methodological choices should be chosen with aim of minimizing any undesired influences. Therefore, it is important to consider the prominent factors and possible influences on speech rate, which are emotion, demographic factors and speaking style, thoroughly. In the present study the participant group will consist of students from the same cohort. Presumably, the variability in age will not be of significant influence on the speech rate, yet this will be analyzed. To minimize any emotional influences on the results the experimenters will try to put the participants at ease and make them feel as comfortable as possible. In addition, gender will be taken into consideration, as it has shown that men and women can differ in articulation rate (Verhoeven et al., 2004; Quené, 2008).

2.1.3 SLA: LINGUISTIC PROFICIENCY

The factors that can be of influence on L1 speech rate were discussed in the previous section, Section 2.1.2. The current study involves an international L2 environment and speech rate analyses of two recording sessions during a 3-year bachelor. This section describes how L2 speech rate change can be connected to linguistic proficiency and perceived foreign accent.

New second language learners often have shown slower speech tempo depending on their linguistic abilities and comprehension (Giles et al., 1991; Bosker et al., 2013). Nervousness and fear, are two emotions often conveyed with an increased speech rate (Breitenstein et al., 2001) and one could argue that these emotions could be present and the L2 speech production could be influenced as a consequence of for example insecurity in L2 linguistic ability. Nevertheless, a positive linear relation was found between speech rate and perceived competence (Giles et al., 1991). In general, Giles et al. define perceived competence as the judgment of an individual about his or her ability in any particular area. Giles et al. claim that with higher perceived competence comes higher speech rate. For non-native speakers this could mean that with increasing linguistic experience an increase in perceived competence could follow. Therefore, an increase in speech rate over a learning period would be likely.

Slow speech rate and non-native use of pauses are two speech characteristics proven to contribute to the perception of foreign accent, or the perceived foreign accent (Trofimovich & Baker, 2006). Trofimovich and Baker show the effect of the level of L2 experience on the production of the five suprasegmentals: stress-timing, peak alignment, speech rate, pause frequency and pause duration. The participants were adult Koreans with experience in L2 English for 3 months, 3 years or 10 years. Six English declarative sentences were spoken by 30 Korean learners of English and 10 native English speakers. The participants were judged through acoustic analyses and through listeners' judgment on the accurateness of the five suprasegmentals and how these contributed to foreign accent. The amount of experience was a big factor in the production of stress timing. The suprasegmentals, in particular pause duration and speech rate, added to foreign accent for all the learners regardless of the amount of L2 experience. It should be noted that L2 experience says little to nothing about L2 linguistic abilities and L2 comprehension, which were found to be related to speech rate according to Giles et al. (1991). Trofimovich and Baker results showed that the fluency-based characteristics (pause duration and speech rate) were therefore more associated with foreign accent then were the melody-based characteristics (stress timing, peak alignment and pause frequency). Bosker et al.(2013) found support for this claim by also finding speech rate and pauses to be more important than speech repairs in fluency judgments. Changes in L2 speech rate are connected to changes in linguistic ability. With a change towards a more native-like application of pause duration and speech rate the perceived foreign accent is likely to become less.

2.1.4 SUMMARY

The previous sections discussed speech rate. The two most common quantification measures are: articulation rate and speaking rate (Laver, 1984). Also, important linguistic and paralinguistic influences of speech rate have been discussed which are: age/gender, country/region (Verhoeven et al., 2004), emotions (Davitz et al., 1964; Breitenstein et al., 2001) and speaking style (Duez, 1982; Kowal et al., 1983; Picheny et al., 1986). This was followed by a section on speech rate in SLA highlighting the correlation between speech rate and linguistic proficiency: linguistic proficiency will probably increase accompanied by an increase in speech rate (Giles et al., 1991). For speech rate research on foreign accent it was shown that perceived foreign accent can decrease through change in speech rate and pause duration (Trofimovich & Baker, 2006). The next section concerns speech rhythm and will follow the same set-up of discussing the concept of speech rhythm, its quantification methods, and its influencing factors and biggest influence in SLA.

2.2 RHYTHM

Rhythm of speech can help the listener attend the speech signal and the linguistic content of the speech signal (Quené & Port, 2005; Quené & Van den Bergh, 2008). Quené and Port assessed the effect of listeners focusing on regular time points in speech events for English words with stressed syllables. These words were perceived better when they were aligned at regular time points in comparison with alignment at irregularly timed sequences. These findings show the listeners formed some rhythmical expectancy or relying on the timing of stressed syllables in spoken-word perception. This was also shown for Dutch (Quené & Van den Bergh, 2008).

On a perceptual level, different language rhythms can be distinguished. Attempts to classify these different rhythms started decades ago. The rhythmic class hypothesis (Pike, 1945;

Abercrombie, 1967) has received a lot of criticism as the rhythmic categories have not (yet) been consistently shown by any quantification method and thus the categorization appears to be only evident at the perceptual level (Lehiste, 1977). According to this hypothesis the stress-timed language category refers to languages that appear to have syllables of variable durations and a regularly recurring stress (e.g. English, Dutch). Syllable-timed languages refer to languages for which all syllables appear to have approximately the same duration, also often metaphorically referred to a machine-gun rhythm or staccato rhythm (e.g. French, Spanish). Mora-timed languages are languages that seem to have a regular recurrence of subsyllabic timing units (e.g. Japanese, Estonian). Support for this theory comes for example from a perception study which showed that 3-days-old infants were able to discriminate between two languages from different rhythmic classes, but not between two languages from the same rhythmic class (Ramus et al., 1999). However, the many rhythm metrics that have been proposed to quantify speech rhythm according to these classifications have not shown consistent success (Grabe & Low, 2002; Dellwo, 2006; Ramus, Nespor, & Mehler, 1999). Dauer (1983) claims the two most important aspects which distinguish between syllable-timed and stress-timed are syllable structure and vowel reduction. Stress-timed languages are claimed to have more variety in syllable types and syllable duration than syllable-timed languages do. With vowel reduction Dauer means that in stress-timed languages, unstressed vowels are consistently reduced in duration (e.g. schwa) or even absent. For the English language the most reliable of correlate of stress has indeed proven to be vowel duration (Sluijter & Van Heuven, 1996).

The next section will discuss several methods that have been proposed to quantify speech rhythm. The most prominent rhythm metrics (RMs) will be discussed which can be divided into durational measures and the pairwise variability indices. Next, the paralinguistic factors that influence speech rhythm will be discussed. Lastly, Section 2.2.3 discusses speech rhythm in second language acquisition and how L1 can influence the acquisition of L2 speech rhythm.

2.2.1 QUANTIFICATION: RHYTHM METRICS

Researchers have been attempting to quantify speech rhythm with the help of rhythm metrics (RMs). Comparisons between studies that apply the same RM have mostly shown inconsistent results (Gut, 2012). RMs are very sensitive to methodological choices e.g. read or spontaneous speech, inter-speaker variability and syllable complexity of the speech material (Arvaniti, 2012). Comparisons within studies, in which these methodological choices are kept consistent, have shown promising results in discriminating between rhythmically distinct and rhythmically similar languages as well as between native and non-native within one language (White & Mattys, 2007a, 2007b). The most broad categories of RMs are the durational measures (e.g. Ramus et al., 1999; Dellwo, 2006) and the pairwise variability indices (Low et al., 2000). The biggest difference between these RMs is the scope they operate on. The pairwise variability indices are local measures. Only local differences are considered, which are differences between adjacent rhythmical units. The durational measures are global measures and consider differences between rhythmical units taken over an entire utterance. Applied to different languages and different speaking styles each category shows its own advantages and disadvantages. Loukina et al. (2011) found that RMs differ from one another on three aspects. Some RMs can use durations of either vocalic and/or consonantal intervals while others build on syllables or pseudo-syllables. Another difference between RMs is normalization of the metrics or not. The third difference is the scope the RMs operate on, which can be local intervals or global intervals. The researchers also investigated how stable 15 well-known RMs performed by testing them on their discrimination between the three rhythms of five languages. There was not one RM that could significantly discriminate between all five languages nor between their rhythms and for discrimination between language pairs the RM performing best depended on the language pair. Loukina et al. pointed out that linguistic rhythm is a multidimensional system and that each RM captures different language properties. This research points out how hard it is to capture the essence of rhythm.

For the current study the RMs chosen are based on previous studies that reported promising results on the discrimination between native and non-native speech. RMs that have shown discriminative ability between rhythmically distinct languages and rhythmically similar languages could also possibly be promising in discriminating between native and non-native speech within one language. Discrimination between native and non-native UCU English is central to the research questions of the present study and probably Dutch speech rhythm will have a large influence on UCU English speech rhythm as most participants are Dutch natives. The next sections describe three promising RMs and studies that have shown discriminative ability between native and non-native English and Dutch.

2.2.1.1 DURATIONAL METRICS: %V AND VARCOV/VARCOC

Research done on comparing the rhythm of native speech with non-native English speech has shed light on two promising durational rhythm metrics: %V (Ramus et al., 1999) and VarcoV (Dellwo, 2006).

The %V (1) was proposed by Ramus et al. and denotes the percentage of total utterance duration comprised of vowel durations.

$$\%V = \frac{totalV}{totalduration} \times 100 \qquad (1)$$

Individual studies have shown promising results for this RM. However, results compared between studies show large inconsistencies. According to Gut (2012) this is largely due to inconsistent methodological choices between these studies. Gut sums up some of these inconsistencies: For British English Arvaniti (2012) found 45.7%, Grabe and Low (2002) found 41.1%, Ramus et al. (1999) found 40.1% and White and Mattys (2007a) found 38%. These results show the importance of methodological choices. For the present longitudinal research it is important to choose the RM best able to detect possible speech rhythm change between recordings (i.e. distinguishing between native and non-native English speech rhythm) while clarifying the methodological choices and keeping experimental factors as consistent as possible. This ensures that the main factor causing the discriminative ability of the RM will be the RM itself.

White and Mattys (2007a, 2007b) showed supporting results for the discriminative ability of V%, between rhythmic classes and within rhythmic classes, and for VarcoV (discussed below) between native and non-native speech. White and Mattys (2007a) investigated the local metrics nPVI-V and rPVI-C, the global rhythm metrics ΔV, ΔC and %V and the normalized global interval metrics VarcoV and VarcoC by analyzing prosodic sentences. The authors found that nPVI-V and VarcoV did fairly well in discriminating rhythms between rhythmic categories. Also, %V was able to discriminate between syllable-timed languages (Spanish and French) and, which could be relevant for this research, between stress-timed languages (English and Dutch). Another study from White and Mattys (2007b) again showed that %V showed a difference between English and Dutch. In addition VarcoV managed to discriminate between English learners of Spanish and Spanish natives and between Spanish learners of English and English natives. VarcoV showed to be the best predictor for the accent ratings of native and non-native English. From these results White and Mattys (2007a, 2007b) concluded that the combination of %V combined with VarcoV appears useful in discriminating between and within rhythmic classes for first en second languages. Another rhythmic study involving %V came from Sarmah, Gogoi and Wiltshire (2009) . The researchers compared the rhythm of Thai English speakers of 6 months and 18 months of living in the United States. Thai is considered as a syllable-timed language whereas American English is usually classified as stress-timed. They found differences in vowel durations. The learners with more exposure showed a higher %V which indicates a bigger percentage of vowel duration in the average utterance.

The second promising durational measure came from Dellwo (2006). Ramus et al. (1999) proposed the standard deviations of consonantal (or intervocalic) intervals and the standard deviations of vocalic intervals as RMs. Dellwo analyzed these measures and found that they were extremely sensitive to variables such as speech rate. Therefore, the normalized versions VarcoC

(2) and VarcoV (3) were introduced. These rhythm metrics compute the standard deviation of the consonantal/vocalic interval duration divided by the mean consonantal/vocalic duration and multiplied by 100:

$$VarcoC = 100 \times \frac{s(C)}{meanC} \tag{2}$$

$$VarcoV = 100 \times \frac{s(V)}{meanV} \tag{3}$$

Loukina et al. (2011) tested 15 RMs on speech samples of English, Greek, Mandarin, Russian, French. Their results show that to separate all 5 languages at once a combination of at most three rhythm measures are needed. Mandarin was identified most readily with vocalic measures. For English-Greek, English-Russian and English-French the best discrimination results were obtained with the global metric VarcoV (only for English-French the best results were obtained by combining VarcoV with another RM).

The two durational measures %V and VarcoV have shown promising results in finding significant differences between English and Dutch, which are both considered to be stress-timed, and between native and non-native English speech (Grabe & Low, 2002). In the present study the participants will be divided in four groups based on their native language to, in addition to the speech rhythm analyses between recordings, analyze if the RMs can discriminate between the UCU English of the different groups. The durational measures %V and VarcoV will be used in an attempt to (a) analyze the speech rhythm between the times of recording and to, in addition, (b) analyze the speech rhythm between groups based on the native language.

2.2.1.2 PAIRWISE VARIABILITY INDICES

The Pairwise Variability Indices (PVI) is proposed by Low, Grabe and Nolan (2000) and operates on a local scope. This RM has several variations depending on the speech unit chosen to be analyzed. The speech units are often vocalic intervals or intervocalic intervals. The researchers suggest to use the raw pairwise variability index (rPVI) on consonantal intervals. The rPVI (4) computes the difference in duration between each adjacent interval and then calculates the average of all these differences. For vocalic intervals it is suggested by Low et al. to use the normalized pairwise variability index nPVI (5) because vocalic intervals are supposedly highly influenced by speech rate because vowels are more elastic than consonants.

$$rPVI = \left[\sum_{k-1}^{m-1} |d_k - d_{k+1}| / (m-1)\right] \tag{4}$$

$$nPVI = 100 \times \left[\sum_{k-1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right] \tag{5}$$

One aspect of this RM, namely operating on a local scope, to be noted is that the adjacent units' sequence is not taken into consideration (Gibbon, 2003). A sequence of units with for example durational pattern 2s-6s-2s-6s will have the same nPVI value 100 as the pattern 2s-6s-18s-54s. Another possible problem that comes with analyzing on a local scope was reported by e.g. Arvaniti (2012) when she empirically applied the PVI to sentences in English, German, Greek, Italian, Korean and Spanish. The rhythm of all sentences was specifically designed as stress-timed by including much consonantal and vocalic variability, and syllable-timed sentences, which included little consonantal and vocalic variability. Arvaniti found significant differences using PVI between the sentences within the same language and concluded that the rhythmic variability within a language can be as high as rhythmic variability across languages. Other inconsistencies between studies have been found due to the use of varying speaking styles and the different ways of segmentation procedures (Gut, 2012).

However, within studies the PVI has shown promising results in its discriminative ability (Low et al., 2000). The PVI was proposed and tested by Low et al. (2000) on the rhythm of Singapore English and British English. Singapore English is commonly described as more syllable-timed then stress-timed British English (Deterding, 1994). 10 British English speakers and 10 Singapore English speakers each read 10 sentences out loud. The PVI was tested on vocalic intervals and on intervocalic intervals. These results were compared to the results of the durational method VarcoV (for more information see Section 2.2.1.2) and the researchers concluded that the PVI is a better measure of rhythmicity than VarcoV. The PVI focused on the vocalic intervals performed best.

In the present study next to VarcoV and %V also the vocalic normalized pairwise variability index (nPVI-V) will be applied to the speech data. The discriminative ability of VarcoV and %V was shown by White and Mattys (2007a, 2007b) between English and Dutch and between native and non-native English. Low et al. (2000) showed the discriminative ability of the PVI between two varieties of English. Because vowel reduction could be an indication of a stress-timed language (Dauer, 1983) and the normalized PVI is not influenced by the rate of the speech, the nPVI-V seems to be favorable over the rPVI-C. To keep the speech style consistent and the pronunciation as clear as possible the speech data to be analyzed will consist of the read-aloud sentences (Nazzi, Bertoncini, & Mehler, 1998), which were also used in the studies of e.g. White and Mattys (2007a, 2007b).

2.2.2 PARALINGUISTIC FACTORS

It is important to know which factors can influence the outcomes of the speech rhythm analyses. According to Ladefoged (2004, p. 24), stress is mostly heard when an increase in pitch and an increase in length of a syllable are perceived. Ladefoged (2004, p. 72) describes how vowel length depends on different factors: the natural length of the vowel, whether the vowel is stressed or unstressed, the number of syllables in the word and the way the syllable ends. And in English the most reliable correlate of stress is vowel duration (Sluijter & Van Heuven, 1996). The rhythm of a sentence is formed by variations in vowel and consonant length. The stress in the speech of any specific language is described by its language-specific phonological rules. These factors cause the language rhythm, including the English language rhythm, to be restricted. The degree of freedom in speech rhythm is mostly caused by paralinguistic factors.

As previously mentioned before, Davitz et al. (1964) studied how *emotions* can influence the prosodic characteristics by analyzing ratings of emotions on emotion adjective scale. Amongst the outcomes of the study was that speech rhythm was consistently related to the emotion adjective scale (even though this relation was not as strong as with speech rate). He reported for "anger" that the speech rhythm was irregular. For "sadness" he found regular rhythm with irregular pauses. For "joy", "affection" and "tenderness" the speech rhythm stayed regular. Overall Davitz et al. concluded about rhythm that for "negative" feelings the speech rhythm was more irregular than for " positive" feelings.

The speaking style of the speaker can be an influencing factor on rhythm. One example is motherese, which has exaggerated temporal rhythmicity which sometimes occurs as metrical regularity of intervals between accented syllables over a span of several utterance (Fernald & Simon, 1984). Gut (2003) investigated the rhythm in read speech and storytelling speech by studying German speech rhythm from Polish, Italian and Chinese natives speakers who are learning German. These languages were chosen by Gut because of their typologically different language background in their use of vowels, accentedness and pitch. By analyzing vowel reduction and vowel deletion Gut found differences in speech rhythm compared to German native speech. These differences were more clearly pronounced in read speech than in storytelling, which was a re-telling of the read speech story. This indicates that speech rhythm is highly sensitive to the speaking style of the speech data. Studies that used a single speaking style for the rhythmic analyses came from e.g. White and Mattys (2007a, 2007b) and Grenon and White (2008) . The researchers used five prosodic sentences from a larger dataset created by

Nazzi et al. (1998). The English sentences were selected based on their exclusion of the approximants /l/, /w/, /j/ and /r/ as the boundary between these sounds and adjacent sounds are difficult to be established. White and Mattys investigated the L1-L2 interaction between rhythms of Castilian Spanish and British English, and rhythms of Dutch and British English. Grenon and White evaluated the L1-L2 interaction between rhythms of Canadian English and Japanese. Both studies showed significant discriminative performances between native and non-native speech by analyzing these five read aloud prosodic sentences.

Speaking style is an important factor to consider during speech rhythm analyses as it has shown to be an influencing factor (Gut, 2003). The choice of speaking style should be carefully chosen to minimize its influence and maximize its rhythmicity. As studies have shown, read speech seems a suitable speaking style for rhythmic analyses (White & Mattys, 2007a, 2007b; Grenon & White, 2008). In the present study the five prosodic sentences will be used (Nazzi et al., 1998). Next to the described paralinguistic factors in this section, also linguistic experience can have an influence on the production of stress-timing (Trofimovich & Baker, 2006), see Section 2.1.2 for more information. This leads to the next section that describes how transfer will occur when the SLA is facilitated.

2.2.3 RHYTHM AND SLA: TRANSFER

Studies done on L2 speech has shown that non-native speakers are likely to develop a rhythm that is intermediate between their first and second language, especially when the two languages are rhythmically distinct from each other (Benson, 2002; Yuan, 2010). The linguistic background can cause positive transfer or negative transfer (Benson, 2002). Positive transfer is defined as the use of implicit knowledge of the native language for the accurate performance in the second language and negative transfer has the opposite meaning and leads to inaccurate performance in the second language (Goldstein & Bunta, 2012). By comparing the first and second languages on duration and distribution of stressed syllables, vowel reduction and syllabic complexity, transfer showed to occur in multiple studies and for multiple languages. The rhythm of L1 Chinese has shown to some extend to transfer to L2 Singapore English (Low et al., 2000). Also the rhythm of L1 Chinese, English, French, Italian and Romanian can partly transfer to L2 German rhythm (Gut, 2003) as well as the rhythm of L1 Mexican Spanish to L2 American English rhythm (Carter, 2005). Next to that, the rhythm of L2 British English has shown transfer from L1 Castilian Spanish rhythm and L1 Dutch rhythm (White & Mattys, 2007a, 2007b). The L2 English rhythm has also shown influences of transfer from the rhythms of L1 German, Russian, French and Italian (Yuan, 2010). This highlights the importance of the native languages' rhythmical properties in second language learning. Whether the rhythmic class hypothesis is valid or not, when the L2 rhythmically differs from the L1 it is likely that there will be interaction between the L1 speech rhythm and the L2 speech rhythm and so transfer will occur. If the additional factor tested in the present study of grouping, based on the native language, turns out to be significant, transfer could also possibly be shown.

2.2.4 SUMMARY

This section, Section 2.2, has discussed speech rhythm. The most promising quantification measures of rhythm for the present study will be tested: %V (Ramus et al., 1999), VarcoV (Dellwo, 2006) and nPVI-V (Low et al., 2000). The most prominent possible influences on speech rhythm can come from emotion (Davitz et al., 1964) or speaking style (Gut, 2003). Also speech rhythm has been discussed in the context of SLA. Negative transfer is likely to occur when the L1 and L2 are rhythmically distinct, while positive transfer is likely to occur when the L1 and the L2 are rhythmically similar to each other (Benson, 2002; Yuan, 2010). To see if there will be change in speech rhythm over time the RMs will be applied to the speech data and the effect of recording session will be analyzed. In addition, the participants are grouped based on the rhythms of their native languages and the RMs will be applied to find if the non-native speech differs between these groups.

2.3 RATE-RHYTHM INTERACTION

In the previous section of this chapter it has been shown that speech rate and speech rhythm have separate methods of quantification, varying influencing factors and that they do not behave the same way in second language acquisition. In the present study the first research question is focused on speech rate while the second research question is focused on speech rhythm. However, a possible correlation may have been found between speech rate and rhythmicity (Pellegrino, Coupé, & Marsico, 2011). To be able to form the hypotheses to the two research questions of the current study it is important to elaborate on these findings to predict if the research questions and their results could be correlated.

Grabe and Low (2003) analyzed passages of "The North Wind and the Sun", a standard text from phonetic research, spoken in stress-timed languages English, Dutch, German, and syllable-timed French and Spanish. They used the rhythm measure of PVI. The researchers found that the durational variability in speech is larger in stress-timed languages. In other words, stress-timed languages have more variability between adjacent intervals. This strengthens the claim that syllable-timed languages have a more regular rhythm as opposed to stress-timed languages.

Another difference between syllable-timed languages and stress-timed languages would be between the number of phonemes per syllable and the number of syllables per clause (Fenk-Oczlon & Fenk, 2006). This relates to Dauer (1983), who described a more complex syllable structure in stress-timed languages. Fenk-Oczlon and Fenk compared speech rate and the rhythmic classes syllable-timed and stress-timed. They asked native speakers of 34 different languages to give written translations of 22 simple sentences from English to their native language. They also asked the participants to determine the number of phonological syllables for each sentence. Fenk-Oczlon and Fenk report to find for example that stress-timed Dutch has 5.05 syllables/clause and 2.97 phones/syllable and stress-timed English has 5.77 syllables/clause and 2.69 phones/syllables. Syllable-timed Spanish has 7.96 syllables/clause and 2.09 phones/syllable. Mora-timed Japanese has 10.23 syllables/clause and 1.88 phones/syllable. For the distinction between stress-timed and syllable-timed languages this could mean that the less phonemes per syllable and more syllables per clause would indicate a syllable-timed language. Conversely, more phonemes per syllable and less syllables per clause would indicate a stress-timed language.

Pellegrino et al. (2011) also did crosslinguistic research on speech rate. The authors hypothesized that the syllabic speech rate differs between languages depending on the rhythmic category of the language. They also attempted to point out the relation between speech rate and information density, which relates to the amount of information contained per speech unit. English read-aloud texts were translated to 6 languages: French, German, Italian, Japanese, Mandarin and Spanish. This way the text contained the same information and the speech rate was the dependent variable. 59 speakers read the texts aloud in their own native language. The rate of the information density, or the speed at which the information was conveyed, was assessed between languages. Pellegrino et al. found significant differences in syllabic speech rate between stress-timed English and syllable-timed Spanish, between English and syllable-timed Italian, and between syllable-timed French and syllable-timed Spanish. Although limited to the studied languages a strong negative correlation was found between information density and speech rate for the studied languages.

In conclusion, speech rate and rhythmicity, or the irregularity of speech rhythm show the tendency to be negatively correlated. Syllable-timed languages seem to have a more regular rhythm (Grabe & Low, 2003) and an overall higher number of phonemes per syllable (Fenk-Oczlon & Fenk, 2006). Syllable-timed languages also have a higher number of syllables per clause, whereas stress-timed languages seem to have a more irregular rhythm (Grabe & Low, 2003), more complex syllables and so an overall lower number of syllables per clause (Fenk-Oczlon & Fenk, 2006). Other research has shown significant differences in speech rate between stress-timed English and syllable-timed languages Spanish and Italian (Pellegrino, Coupé, & Marsico, 2011). These findings have been reported in studies that kept the speaking style consistent for both rate and rhythm analyses. This suggests there is a possible correlation between speech rate and the (ir)regularity of speech rhythm between languages. With the speaking style kept consistent, speech with a higher speech rate tends to be less rhythmical, and speech with a lower speech rate tends to be more rhythmical.

For the methodological choices concerning the speaking style of the speech data this possible relation needs to be considered. To avoid any possible correlative outcome in the current study, the speech rate and speech rhythm will be analyzed from two different speaking styles. For the speech rate analyses spontaneous informal speech will be analyzed. Speech rate has shown to be higher in an interview setting than in storytelling (Kowal et al., 1983) and speech was faster in conversational speech than in clear speech (Picheny et al., 1983). L2 experience will have more effect on spontaneous or conversational speech than on storytelling or clear speech. On the other hand, the speech rhythm in read speech was more clearly pronounced than the speech rhythm in storytelling (Gut, 2003) and vowels were less reduced and less modified (Picheny et al., 1983). Therefore, in the present study the speech rate will be analyzed from spontaneous informal speech and the speech rhythm will be analyzed from read speech. The read aloud text will consist of the five prosodic sentences (Nazzi et al., 1998), previously applied by e.g. White and Mattys (2007a, 2007b). Having treated the linguistic aspects of the research questions of the present study the next chapter will look at the research questions from a more sociolinguistic/sociological approach.

3. ACCOMMODATION AND ACCULTURATION

Accommodating speech is a naturally occurring process during communication. The Speech Accommodation Theory (Beebe & Giles, 1984) is a generally accepted theory and describes speech accommodation and the different speech accommodation strategies. In the present study the participants are subjected to a learning environment. The Acculturation Model describes the process of learning about and adapting to a new culture (Schumann, 1975, 1986). Acculturation refers to how successful second language acquisition relates to the blending in of the L2 learners into the target culture (Schumann, 1975, 1986). This chapter aims to clarify how speech can change as a consequence of intense social interactions over the research period. Describing different kinds of accommodation as well as variables for acculturation and second language acquisition will help in forming the hypotheses for the present study. The chapter starts off with a general explication of the Speech Accommodation Theory. The three different strategies of speech accommodation are speech accommodation. The next section explains the Acculturation Model (Schumann, 1986). Its social and psychological variables are discussed followed by a section on convergence.

3.1 SPEECH ACCOMMODATION THEORY

Extensive research has been done regarding methods by which people modify their speech style during conversations depending on the speaker they converse with, which has led to the Speech Accommodation Theory or SAT (Giles, 1973; Giles, Mulac, Howard, Bradac, & Johnson, 1987). SAT has been extended to the Communication Accommodation Theory (Giles, Bouhris, & Taylor, 1977) as it also turned to be applicable to communication in general. For the present study, accommodation of the speech characteristics speech rhythm and speech rate will be investigated. Therefore, this section only addresses the SAT.

Multiple strategies of speech modifications have been found which are influenced by a speaker's feelings, perceptions, values and attitudes during communication. Convergence, divergence and complementarity have been acknowledged as accommodation strategies (Giles, 1973; Giles et al., 1987). The three strategies, starting with the most common one, are:

Convergence (Giles, 1973) can be defined as the sociolinguistic strategy whereby a speech style changes towards the speech style of the other speaker by altering linguistic features including speech rates, pause lengths, utterance lengths, pronunciations, etc. Reasons for this strategy can be, for example, the desire for social acceptance, effective communication or integration. Three types of status-based convergence exist. Upward convergence occurs when the reference person, for example when a Scottish interviewee tones down his accent in an interview with an interviewer from London. If the reference person would be the interviewer downward convergence could appear when the interviewer would tone down his posh British accent. If both parties would adjust their speech “equally” to each other’s identity the accommodation is called mutual convergence. Section 3.3 elaborates more on convergence.

Divergence (Giles, 1973) is the process whereby a speech style changes away from the speech style of the other speaker. Reasons for this strategy can be maintaining positive social identities, attaining communicational efficiency between speakers, but also disapproval of the other party or as an attempt to make the other speaker change his or her style. Equivalent to the three types of convergence three types of divergence exist. If the reference group would be a subordinate group conversing with a native speaker group, the subordinate group could diverge from the speech of the native speaker group by applying e.g. more ethnic speech markers to obtain their social identity. This would be downward divergence. On the other hand, if the reference group consists of foreign learners conversing with a group of native speakers, the native speakers can wish to set themselves off from foreign learners by e.g. resort to a local accent. This is upward divergence. When both groups diverge in an “equal” manner mutual divergence occurs.

Complementarity (Giles et al., 1987) is a strategy applied during certain times when convergence and divergence are not suitable or desired. For instance in a situation when there is a social difference or power difference and these dissimilar speech patterns are expected and accepted. Over the years complementarity was added to the list of strategies**.**

According to Hamers and Blanc (2000) there are four social psychological processes on which the SAT has been developed and which cause either of the three SAT strategies to be applied: social exchange, causal attribution, intergroup distinctiveness and similarity-attraction. The social exchange process theory claims speakers try to assess the costs and rewards of possible actions and they often choose the course of action that will bring less costs and greater rewards. Causal attribution refers the way a listener interprets the speaker’s behavior according to its motives and intentions. The process of intergroup distinctiveness states that divergence is sometimes employed in an attempt to maintain intergroup distinctiveness and to be different from the other group. Similarity-attraction states that people like, and are attracted to, others who are similar to themselves. Convergence has shown to be the strategy that is positively evaluated by recipients and it will lead to high ratings for attractiveness, but also friendliness and solidarity (Feldstein, Dohm, & Crown, 2001). Therefore, when there is a need for social acceptance the tendency of convergence will be greater.

This section has discussed the SAT and different strategies that can be applied to different conversational situations. For the present study it needs to be related to an environment in which second language acquisition is taking place. Acculturation will be described in the next section in an attempt to relate speech accommodation to an SLA culture.

3.2 ACCULTURATION MODEL

Acculturation has been widely studied by multiple disciplines like sociology and psychology. The consequence of acculturation can be of a linguistic kind. Language greatly impacts acculturation, and acculturation can also have a big impact on language. The next section is derived from Schumann's work (1975, 1986) and explains the Acculturation Model.

The Acculturation Model describes how integrating into a new culture ensures successful second language learning, without the development of pidginization. Nowadays the model only serves as a rough outline of how social and psychological variables are related in second language acquisition. The eight social variables from Schumann's model for L2 learners' success are:

1) Social dominance: The learner can feel inferior/superior to the target group. The L2 learner is more likely to want to learn a target language from a culture that is not politically, culturally, technically, or economically superior to its own.
2) Assimilation, preservation and adaptation: The L2 group chooses (not) to adopt to the target language culture's lifestyle and/or to maintain its own lifestyle and values.
3) Enclosure: If the L2 group and the target group share the same social constructs (e.g. schools, churches, trades, recreational facilities) the second language learning is facilitated.
4) Cohesiveness: If the L2 group is not cohesive it will not tend to remain separate from the target language group. Also if the target language is not too culturally-bound the learning will be easier.
5) Size: If the L2 group is not so big there will be less contact within the L2 group and more contact with the target language group.
6) Congruence: If the L2 group culture is similar to the target language culture social contact is most likely to occur as well as L2 learning.
7) Attitude: If both groups have positive attitude to one another, L2 learning is more easily facilitated.
8) Intended length of residence: If the L2 learner is planning to stay longer they will be longer exposed to the target language and they will be more likely to feel the need to learn/master the target language.

Next to these social factors Schumann names four psychological factors are of influence during second language acquisition: Language shock, cultural shock, motivation and ego. Schumann argues the psychological factors are more important than the social factors. Language shock arises when the learner feels foolish or comical when speaking the new language. Culture shock relates to how comfortable the learner feels in the new culture. Language shock and cultural shock are two hurdles to overcome in order to be able to associate with the target language group. Two types of motivation can be distinguished in second language learning: integrative and instrumental. Integrative motivation comes from an interest in acquiring a second language because of practical reasons, for example communicate/meet valued members of target language group. Instrumental motivation comes from more self-oriented reasons, e.g. earn more money. Ego relates to the ability of a L2 learner to accept a new identity as a consequence of the new speech community.

According to the Acculturation Model, the degree in which a second language is acquired depends on social distance and psychological distance. The social variables lead to small or big social distance and the psychological variables lead to small or big psychological distance. When both the social distance and the psychological distance are small the best environmental circumstances are met for L2 learning. The result of this would be convergence between the L2 learner/group and the non-native speaker/target group.

By describing the Acculturation Model the best conditions have been clarified for optimal SLA in any new environment. When these conditions are met, there will be small psychological distance and small social distance and the result will be acculturation and convergence, in different communicative aspects including speech. That is the connection between speech accommodation and the optimal SLA environment described in this chapter. The next Section 3.3 will be on speech convergence as this will be one of the result of successful acculturation.

3.3 CONVERGENCE AND ATTRACTIVENESS

Features such as speech rate, pausing frequency and utterance length, but also head nodding, facial affect and posture converge towards each other during communication (Giles et al., 1991). Transfer is one type of convergence between L1 and L2 already mentioned. This next section discusses mutual convergence in different L1 speech properties. The studies mentioned related convergence to social acceptance and attractiveness judgments. This shows attractiveness is one of the most prominent motivations of convergence and support the similarity-attraction theory as one of the four social psychological processes that cause speech accommodation in L1 (Hamers & Blanc, 2000).

Studies on speech rate have shown how convergence relates to attractiveness. Speech rate can be altered during conversation depending on the speech rate of the other speaker. Buller et al. (1992) found social attractiveness is increased through the amount of speech rate similarity.

They also found increased speech rate gives the speakers an increased perceived dominance as well as increased perceived competence. This increases to 350 or 400 syllables per minute, after which the level of perceived dominance and perceived competence decrease. Street and Brady (1982) reported, however, that perceived competence does not depend on an absolute speech rate but is dependent on the difference relative to their own rate. Based on testing with a male speaker they found speakers were viewed by listeners as more competent when the speaker's rate was faster than the listener's rate. Conversely, speakers were viewed as less competent when their speech rate was lower than the speech rate of the listener. Street and Brady also found this for ratings on socially attractiveness. Speakers were considered to be more socially attractive when they had a similar or faster speech rate than the listeners' then speakers who had a lower speech rate than this listeners'. However, these speech rates – social attractiveness judgment are also influenced by the gender of the judges. Research on gender relating to judgments on social attractiveness as well as judgments on competence based on speech rate from Feldstein et al. (2001) pointed out gender can be an influencing factor on judgments. The speakers who had the speech rate most similar to the listeners' were perceived as most socially attractive as well as more competent. The ratings of competence were influenced by the listeners' gender, and the ratings of social attractiveness were influenced by both the listeners' gender and the speakers' gender. All female listeners rated all speakers as more competent listeners than the male listeners did. And all listeners judged male speakers to be more socially attractive than the female speakers.

Next to speech rate phonetic convergence or phonetic imitation has also been shown to be related to attractiveness judgments (Babel, 2012). Babel showed this in her research on spontaneous phonetic imitation . Her research found there is phonetic selectivity as well as social selectivity and the researcher related this to "liking". The phonetic selectivity relates to the phonetic freedom the phone has in its formants, which are the characteristic frequencies by which a particular vowel is recognized by. The social selectivity relates to "liking" judgment (of the other speaker/voice). Babel found these results by analyzing the participants' repeated word after hearing a voice saying the word or hearing a voice accompanied with a visual image of the speaker. The speaker was either a black or a white male. Phonetic imitation occurred especially in women. For women a high attractiveness rating of the voice corresponded with a high degree of phonetic imitation. For men this wasn't the case, possibly because an attractive male (voice) can be seen as a threat.

Multiple longitudinal studies have also showed phonetic convergence which can be related to attractiveness judgments. One example is from Pardo, Gibbons, Suppes and Krauss (2012) which involved 3 recordings during an academic year across native American English roommate pairs.

The phonetic convergence shown correlated with the roommates' self-report on how close they were. The students also showed longer-term adjustments in their phonetic repertoire as a consequence of continued contact with the same talker. Pardo et al. do mention that additional research is needed to see if these longer-term adjustments would apply outside of the individual pairs.

Evans and Iverson (2007) conducted longitudinal study, similar to the present study, on convergence in British English. However, this study does not look at the motivation behind convergence by using attractiveness judgments, but analyzes vowel convergence in British English and how this changes overtime. The researchers conducted a 2 year study on dialects of British English during which the participants read passages and target words in 4 different recording sessions. The recording sessions were before attending the university, 3 months after attending, 1 year into university and 2 years into university. This study shows that the Northern British English subjects' accent changed as a result of attending a university in Southern England. Some pronunciation of vowels shifted after only three months. After one and two years the rated degree of Southern British English accentedness increased. For most of the investigated words the vowel formants were more similar to the pronunciation of the Southerners, while few other vowels retained their Northern formant characteristics probably due to sociolinguistic influences.

The mentioned studies show the relation between attractiveness judgments and convergence in L1. In the present study, it will be likely that convergence will be shown, considering there will be accommodation and acculturation in the L2 environment.

4. RESEARCH HYPOTHESES

The previous chapters have provided the theoretical background for this thesis and have discussed the methodological choices and options. It is the aim of this study to investigate whether there will be change in speech rate in spontaneous L2 speech and whether there will be change in speech rhythm in read L2 speech, in an international spoken English environment. Chapter 2 discussed speech rate and speech rhythm. Research has shown a positive correlation between speech rate and fluency (Bosker et al., 2013), and a positive correlation between speech rhythm and linguistic proficiency (Giles et al., 1991). For the speech rate analyses spontaneous speech is chosen over read speech, as free speech is faster than storytelling (Kowal et al. 1983) and faster than clear speech (Picheny, 1986). Therefore, spontaneous speech will be more sensitive to L2 experience and more suitable for the speech rate analyses. For the speech rhythm analyses read speech is chosen over spontaneous speech because read speech is more clearly pronounced and contains less vowel reduction and less vowel deletion (Gut, 2003). Chapter 3 showed how speech accommodation is a naturally occurring phenomenon during communication (Giles, 1973) and that the Acculturation Model described the prominent factors for the facilitation of SLA (Schumann, 1975, 1986). In an international environment where these conditions are met this results in convergence. This claim is supported by studies on accent change that showed accent convergence for dialects of e.g. British English (Evans & Iverson, 2007) and American English (Pardo et al., 2012). Other previous longitudinal research on phonetic change during the academic year across native roommate pairs has shown phonetic convergence (Pardo et al., 2012). These findings lead to the following hypotheses:

H1: The L2 spontaneous speech rate will collectively become faster over time
H2: There will be rhythmic convergence in read L2 speech between speakers over time

To test these hypotheses speech data will be used from a longitudinal (and still running) study by Orr, Quené, Van Beek, Diefenbach, Van Leeuwen and Huijbregts (2011). In this study the researchers analyze accent change for both native and non-native speakers of English in five recordings. The participants are students at University College Utrecht, or UCU, in The Netherlands and use English intensively as their lingua franca on campus. Every year around 64% of the incoming students are female and 64% of the students are Dutch[1]. In each recording session multiple speaking styles are recorded. In the present study spontaneous informal speech in English and the read-aloud prosodic sentences will be analyzed and compared between the recording session conducted at the beginning of the bachelor's and the recording session

---

[1] http://www.uu.nl/university/college/EN/whyUCU/Pages/FactsFigures.aspx?refer=/EN/faculties/universitycollege/whyUCU/Pages/FactsFigures.aspx

conducted at the end of the bachelor's. To analyze the speech rate, both articulation rate and speaking rate will be calculated and compared between recordings. As an additional factor the effect of gender on speech rate will also be analyzed. To analyze the speech rhythm of English, a language in which stress causes fluctuation in vowel duration (Sluijter & Van Heuven, 1996), the RMs %V, VarcoV and nPVI-V are chosen. The results will be compared between recordings. In addition, the effect of gender and the effect of grouping of the participants (based on the rhythmic category of their L1) will be measured in these analyses. A significant difference between the recording sessions and a decrease in standard deviation between the sessions would imply convergence.

Over the research period language experience will be gained and most likely linguistic proficiency will increase. Presumably there will be changes in the temporal pattern of the syllable durations as well as changes in the temporal pattern of vowel durations. The increasing linguistic proficiency over time will make it likely that fluency, and therefore speaking rate, will increase. The intensive exposure to the international environment and other students make it likely the speech rhythm will become more homogenous. Because the majority of the incoming students are Dutch natives, the characteristics of the Dutch language (e.g. phonology and stress) are likely to have a relatively large influence on UCU English. The change of rhythm will probably be directed towards an international variety of English or UCU English speech rhythm. As the accent will presumably converge towards one common UCU English accent, presumably also the speech rhythm of the participants will decrease in individual variability and converge towards an UCU English rhythm over the research period.

5. METHODS

5.1 PARTICIPANTS

The participants are students from the University College Utrecht. The data used will be from one cohort of which 59 students completed the two recordings, one conducted at the beginning and the other conducted at the end of their bachelor's study. The average age at the time of the last recording is 21.6 years old (*SD*=1.131). The <u>spontaneous informal speech</u> recordings were successfully completed by 55 participants of which 38 females (69.1%) and 17 males. The recordings of the <u>prosodic read-aloud sentences</u> were successfully recorded by 53 participants. Four groups were discerned, based on the participants' native language:

- Group 1: L1 Dutch (37 participants)
- Group 2: L1 English (5 participants)
- Group 3: L1 syllable-timed languages (6 participants)
- Group 4: L1 other languages (5 participants)

APPENDIX I describes the specific languages and number of females and males of each group. Almost 70% (69.8%) of the students were Dutch. None of the participants had a mora-timed language as native language so Group 4 consists of participants with a stress-timed native languages other than Dutch or English. All speakers had sufficient language proficiency of English before starting their bachelor's study as it was one of the admission requirements.

5.2 DATA COLLECTION

All participants were incoming students in September 2010 and participated in recordings during their bachelor's of 3 years. In the main project from Orr and Quené (2011) five recording sessions in 3 years are done with 3 different cohorts. The first and the last session will be compared in the present study. The first session is within six weeks of the students' arrival from the first cohort at University College Utrecht, and the final session was in their third year at the end of semester 6, just before graduation. Each original recording consists of 7 parts and for the present study only the five prosodic sentences (Nazzi et al., 1998) and the informal spontaneous speech will be relevant. The prosodic sentences are derived from a larger dataset from Nazzi et al. and previously used by multiple researchers (e.g. White & Mattys, 2007a, 2007b; Grenon & White, 2008). Depending on the participants' native language the spontaneous speech was in the native language and if the native language was not English also in the non-native language. Each spontaneous speech recording lasted around 2 minutes of which 20 seconds without any "abnormalities" will be analyzed in this experiment.

5.3 RECORDINGS

The speech recordings are made using a Saffire Pro 40 multichannel AD converter and preamplifier. Open source software Audacity[2] is used for recording and editing. Speech is recorded through a Sennheiser ME64/K6p microphone which was placed in front of the participant.

5.4 PROCEDURE

The interviews took place at a set-up table with the participant on one side and the interviewer diagonally across the table. Before the spontaneous speech takes place the participant is instructed by the interviewer to talk about one or multiple informal topics for 2 minutes. In order to keep the talk going, the participant is also suggested to write down several informal topics beforehand. Before starting the interviewer announces the task on recording.

The instructions for the prosodic sentences are to read the sentences aloud without naming the numbers and without making any mistakes. In case of any mistake the participant should repeat the whole sentence, either immediately after the mistake or at the end of five sentences. The five prosodic sentences to be read are:

1) The supermarket chain shut down because of poor management
2) Much more money must be donated to make this department succeed
3) In this famous coffee shop they serve the best doughnuts in town
4) The chairman decided to pave over the shopping center garden
5) The standards committee met this afternoon in an open meeting

5.5.1 SPEECH RATE ANALYSES

For the analysis of the speech rate 30 seconds of the spontaneous speech will be analyzed. The 30 seconds are selected after the first 20 seconds of speech. The selected 30 seconds were checked on abnormalities like background noises, the interviewer speaking and pauses due to topic changes. In case one of these occurred a different 30 seconds were selected.

To detect the speech rate (both the speaking rate and the articulation rate) the Praat script from De Jong and Wempe (2009) is used. This script takes as input a sound file and considers certain peaks as the possible nuclei of syllables. The script considers all intensity peaks above a certain threshold which depends on the median intensity of the speech file. Peaks that are not preceded and followed by dips of at least 2dB in intensity are discarded as well as unvoiced peaks, which could occur in case of loud fricatives.

[2] http://audacity.sourceforge.net/

5.5.2 SPEECH RHYTHM ANALYSES

Before applying the rhythm metrics the phonetic annotation needs to be obtained. To conduct the phonetic annotation automatically the Penn Phonetics Lab Forced Aligner Toolkit (P2FA) is used (Yuan & Liberman, 2008) and the software Praat (Boersma & Weenink, 2013). The toolkit takes a speech signal and a text file with orthographic transcription as input. Speech naturally contains pauses and errors, therefore any speech data can contain these events. The P2FA uses the CMU American English Pronouncing Dictionary[3] for the alignment and is trained on the SCOTUS corpus that containing American English speech. Therefore the best performance of the alignment will be on American English speech data. All mispronunciations, repetitions, unplanned pauses, false starts and other errors need to be denoted in the orthographic transcription or removed from the speech file as P2FA is a forced aligner and it will align every sound of the speech file with the transcript. Before feeding the audio into the P2FA all speech errors were removed from the sound file. The P2FA toolkit uses a pronunciation dictionary to translate the orthographic transcription (and align the sound file) to a sequence of phonetic symbols. The output is a TextGrid file containing a string of phonetic symbols. The TextGrid is compatible with Praat (Boersma & Weenink, 2013) and contains two tiers: a word tier and a phone tier. The phone tier contains the speech segments (phones) aligned according to the sound. The word tier contains the inserted words also aligned according to the sound. After retrieving the annotated speech files the alignments will be manually checked on errors and mistakes.

Using the Praat script adopted from Quené (2011)[4] the TextGrid is converted to a data table. An R script is created to calculate the values of %V, VarcoV and nPVI-V for each participant and for each group. The R script can be found in APPENDIX II.

---

[3] For more information see http://www.speech.cs.cmu.edu/cgi-bin/cmudict/
[4] http://www.let.uu.nl/~Hugo.Quene/personal/tools/tg2df.praat

6. RESULTS

6.1 RESULTS SPEECH RATE STUDY

The results of the speech rate analyses of 30 seconds of the informal spontaneous speech analyzed with a between-within subject ANOVA are shown in Table 1. No interactions were found between gender and recording session and they were not included in the analyses in this section. For the complete overview of *p*-values for number of pauses, articulation rate and speaking rate see APPENDIX III.

Table 1: Means spontaneous speech rate analyses

<table>
<tr><td>♂ (N=17)
♀ (N=38)</td><td colspan="2">Recording 1
Mean (SD)</td><td colspan="2">Recording 2
Mean (SD)</td></tr>
<tr><td rowspan="2">Speaking rate</td><td rowspan="2">3.525 (0.427)</td><td>♂: 3.419 (0.393)</td><td rowspan="2">3.896 (0.424)</td><td>♂: 3.703 (0.424)</td></tr>
<tr><td>♀: 3.573 (0.438)</td><td>♀: 3.982 (0.400)</td></tr>
<tr><td rowspan="2">Articulation rate</td><td rowspan="2">4.238 (0.408)</td><td>♂: 4.156 (0.354)</td><td rowspan="2">4.472 (0.474)</td><td>♂: 4.327 (0.433)</td></tr>
<tr><td>♀: 4.275 (0.429)</td><td>♀: 4.537 (0.482)</td></tr>
<tr><td rowspan="2">Number of pauses</td><td rowspan="2">8.018 (3.070)</td><td>♂: 8.235 (3.327)</td><td rowspan="2">6.655 (3.384)</td><td>♂: 7.647 (2.448)</td></tr>
<tr><td>♀: 7.921 (2.990)</td><td>♀: 6.211 (3.670)</td></tr>
</table>

Table 2 shows the correlations between speech rate (SR), articulation rate (AR) and the number of pauses (NP).

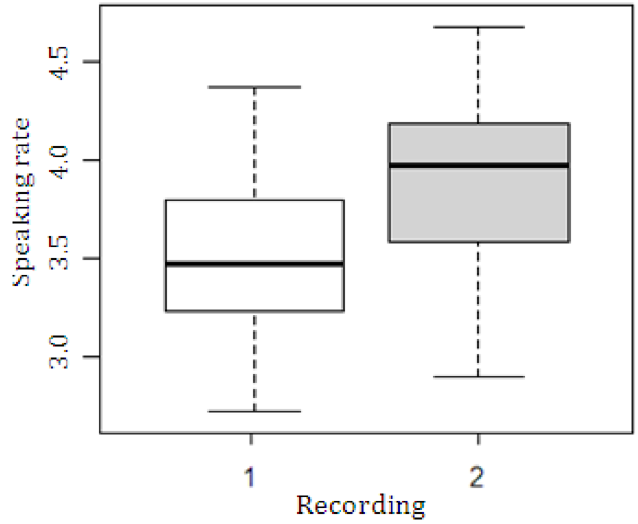
Table 2: Correlation matrix

	SR	**AR**	**NP**
SR	1.000		
AR	0.760	1.000	
NP	-0.404	0.226	1.000

Speaking rate.

The difference in speaking rate between the two spontaneous speech recordings was significant ($F(1, 53)=29.877$, $p<0.001$). Table 1 shows that the speaking rate was higher in Recording 2 and Figure 1 displays this difference in a box plot. Gender showed to have a significant effect on speaking rate ($F(1, 53)=4.373$, $p=0.041$). Men had an overall lower speaking rate than women.

Figure 1: Box plot of speaking rate of spontaneous speech per recording session



Articulation rate.

The difference in articulation rate between the two spontaneous speech recording was significant ($F(1, 53)=7.548$, $p=0.008$). Table 1 shows the articulation rate increased between the recording sessions. Gender was non-significant ($F(1, 53)=2.644$, $p=0.110$).

Number of pauses.

The difference in the number of pauses between the two spontaneous speech recordings was significant ($F(1, 53)=4.329$, $p=0.042$). Table 1 shows that the number of pauses was lowered in Recording 2. Gender was a non-significant factor ($F(1, 53)=1.321$, $p=0.256$).

6.2 RESULTS SPEECH RHYTHM STUDY

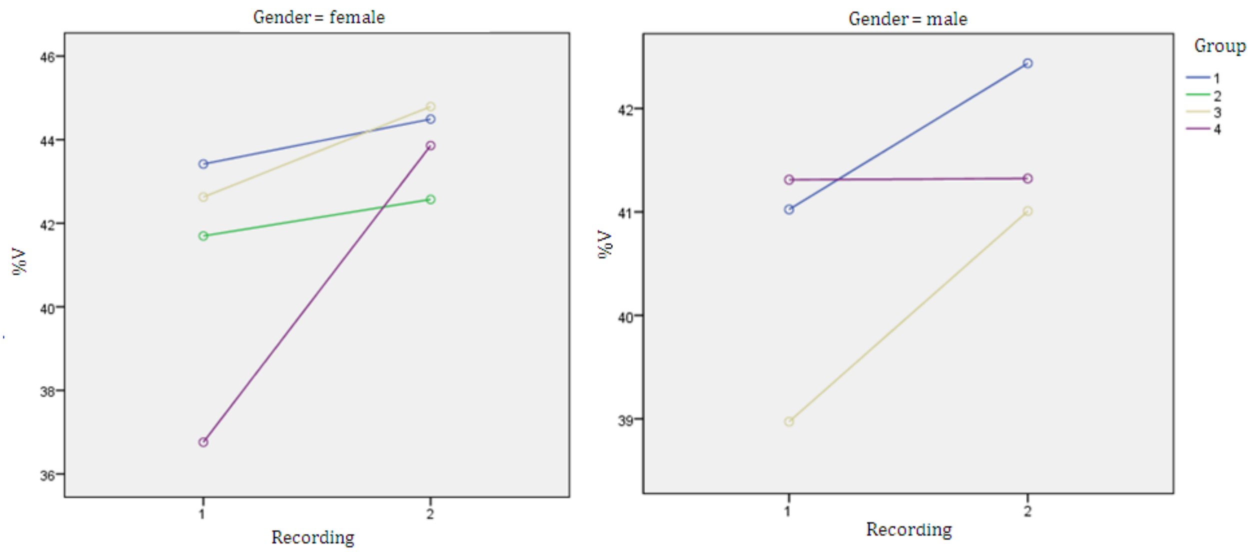
For each RM a mixed between-within subjects ANOVA was performed on the data. In APPENDIX IV all effects and interactions are described with their *p*-values. This section discusses the prominent effects and significant findings. Table 3 shows the means and the standard deviations of %V, VarcoV and nPVI-V by group.

Table 3: Means per RMs and standard deviations

	%V *Mean (SD)*	VarcoV *Mean (SD)*	nPVI-V *Mean (SD)*
Recording 1	**42.246 (3.023)**	**61.668 (5.398)**	**69.322 (5.916)**
1: L1 Dutch (*N=37*)	42.705 (2.762)	62.534 (5.564)	70.896 (5.538)
2: L1 English (*N=5*)	41.696 (4.706)	60.187 (2.330)	65.439 (4.819)
3: L1 syllable-timed (*N=6*)	41.409 (3.393)	57.328 (6.122)	64.075 (5.787)
4: L1 other (*N=5*)	40.401 (2.440)	61.942 (3.139)	67.855 (5.296)
Recording 2	**43.524 (2.719)**	**60.049 (4.690)**	**67.617 (5.454)**
1: L1 Dutch (*N=37*)	43.881 (2.706)	60.554 (4.572)	68.023 (5.333)
2: L1 English (*N=5*)	42.569 (3.063)	58.519 (4.482)	66.041(8.156)
3: L1 syllable-timed (*N=6*)	43.531 (3.198)	56.431 (4.748)	65.690 (5.725)
4: L1 other (*N=5*)	41.830 (1.472)	62.181 (4.345)	68.507 (3.465)

At first sight, Table 3 suggests that the standard deviations for the measures in Recording 1 are higher than for the measures in Recording 2 for the values obtained by RMs %V, VarcoV and nPVI-V. Therefore, this also suggests these RMs indicate less variability between speakers and more coherence, or convergence, in Recording 2. A statistical analysis is applied to the data to test these differences on significance for which for VarcoV and nPVI-V insignificant effects were found of recording session, group and gender. Therefore, only the results from %V will be discussed next.

Figure 2: %V per recording per group per gender



Recording by gender.

For %V the interaction recording*gender was significant for %V ($F(1, 46)=4.390$, $p=0.042$). There was a significant difference between male and female within Recording 1 ($t(46.249) = -2.821$, $p=0.007$) as well as an significant difference between male and female in Recording 2 ($t(32.853)= -3.051$, $p=0.004$). The estimated marginal means and the standard deviations are shown in Table 4.Women have a higher %V than men in both recordings. The interaction recording*group was not significant for %V ($F(3, 46)=1.655$, $p=0.190$).

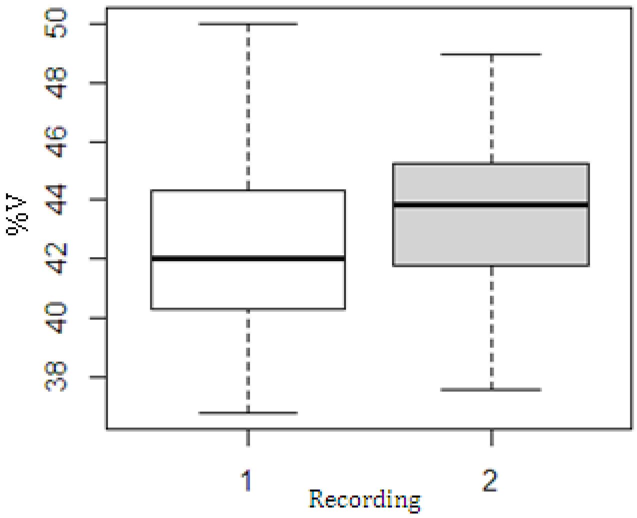
Table 4: Estimated marginal means for %V per recording per gender

	Gender	Mean	Std. Deviation	N
%V Recording 1	Female	42.905	3.206	36
	Male	40.850	2.043	17
	Total	42.246	3.023	53
%V Recording 2	Female	44.241	2.568	36
	Male	42.006	2.451	17
	Total	43.524	2.719	53

Recording.

There was a statistically significant main effect for recording for %V ($F(1, 46)=15.476$, $p<0.001$). Figure 3 is a box plot visualizing the results of %V showing the effect of the recording session and the standard deviations taken over all participants.

Figure 3: Box plot of %V of prosodic sentences per recording session



7. DISCUSSION AND CONCLUSION

In this study changes in speech rate and speech rhythm have been analyzed in an international spoken English environment. Data has been compared between two recording sessions. For the speech rate analyses the number of pauses, speaking rate and articulation rate were analyzed from 30 seconds of spontaneous speech. The effects of the recording session, participants' grouping (based on L1 rhythmic category) and gender were tested. The number of pauses showed a significant decrease between recordings and women had an overall higher speaking rate in their spontaneous speech than men. Most important to the research questions a significant increase was found in both speaking rate and articulation rate between recording sessions, as expected.

The speech rhythm analyses compared the discriminative ability of the three rhythm metrics: %V, VarcoV and nPVI-V. %V showed a significant effect of recording session on the speech rhythm while VarcoV and nPVI-V did not. No significant effect was found for group or gender on the speech rhythm measured by %V, nor by VarcoV or nPVI-V. Comparing results of %V between recording sessions the standard deviations were also shown to decrease. This implies that over the course of the bachelor's the participants showed less variability and thus showed convergence in their speech rhythm change.

The findings of this research support results found by previous studies on change in speech rate. The increase of speaking rate has in previous study been related to an increase in linguistic ability and comprehension (Giles et al., 1991). On the other hand, slow speaking rate, and the non-native use of pauses, have been indicated to contribute to perceived foreign accent (Trofimovich & Baker, 2006) and speaking rate and pauses have also shown to be important in fluency judgments (Bosker et al.,2013). Between the recordings, i.e. during the bachelor's, presumably the participants became more familiar and more fluent with UCU English which was shown trough their increased speaking rate.

The effect of gender on both speaking rate and articulation has also been studied before and the results of the speech rate analyses support these findings. Verhoeven et al. (2004) and Quené (2008) found men to have a higher articulation rate than women but in the present study articulation was insignificant and no effect was found of gender on articulation rate. In the spontaneous speech analyses men did have a significant lower speaking rate than women. Next to women having a higher speaking rate than men, the results also showed that women use more syllables. This supports the typical claim on the difference between men and women that women talk more (and faster).

The results of the speech rhythm analyses found support for the discriminative ability of %V. This was in accordance to the results from White and Mattys (2007a) who found a successful performance of %V in discriminating between stress-timed Dutch and stress-timed English. Discriminative ability was reported for VarcoV between rhythmically distinct languages (Loukina et al., 2011; White & Mattys, 2007a) and between (syllable-timed) Spanish speakers of English and native English speakers (White & Mattys , 2007b). The present study's results did not show any significant discrimination for VarcoV, neither between rhythmically distinct languages nor between native and non-native English. However, Post-hoc tests showed that VarcoV was the most successful of the three tested RMs in discriminating the syllable-timed native language group (Group 3) from the stress-timed Dutch native speakers (Group 1), p=0.051. This implies VarcoV could be a possible indicator for discrimination between rhythmic classes but more support is needed as no significant results were found. These non-significant results for Varco were in accordance with the results reported by Low et al. (2000). However, Low et al. found the PVI (both vocalic and intervocalic) to be a more successful RM than VarcoV. In the current study, for the pairwise variability index nPVI-V no discriminative ability was found within the English prosodic sentences. No significant effect between recording session, between groups nor between gender was shown. This could mean rhythm may be easier detectable on a global scope than on a local scope.

The shortcomings of the present study should be noted in order to take into consideration for future research and also to put results in broader perspective. The effect of the categorization into groups was not analyzed in the speech rate analyses. The relation between speaking rate and the (ir)regularity of speech rhythm had been previously analyzed by Pellegrino et al. (2011) and the researchers found a significant difference in speaking rate between stress-timed English and syllable-timed Spanish. Syllable-timed languages are often described as having a more regular rhythm (Grabe & Low, 2003), having an overall higher number of phonemes per syllable (Fenk-Oczlon & Fenk, 2006) and having a higher number of syllables per clause (Grabe & Low, 2003) compared to stress-timed languages. Considering the L1 rhythmic group in the speech rate analyses could have found results supporting or weakening the proposed correlation between speaking rate and rhythmic category of the native language.

The way the categorization was done into the four groups for the speech rhythm analyses may have attributed to the insignificance of the group effect. The different varieties of English were all analyzed in the same group while for example "English" can perceptually (Deterding, 1994) and rhythmically (Low et al., 2000) differ considerably in speech rhythm from Singapore English. This can be the reason for Group not having any significant effect on any measure. Also the number of participants and females and males differed greatly between groups. Group 2 consisted of five female participants, whereas Group 4 consisted of four males and one female. The number of Group 1 was much higher than the other groups. More homogenous groups could lead to more discriminative and reliable results between groups.

During the data preparation in the current study abnormalities of the speech in the speech data were excluded manually. For the speech rate analyses the speech was manually checked on abnormalities (e.g. background noises, the interviewer speaking, abnormal pauses). For the speech rhythm analyses the speech was checked to ensure the speech recording contained nothing but the five correctly read-aloud prosodic sentences. All values resulting from these analyses of the manually checked speech data were accepted as “normal” values. However, the speech could have still contained abnormalities, which could have resulted in outliers. By checking the data on outliers and/or checking the results on outliers could lead to overall more homogenous results. Which could, again, result in more discriminative and reliable results.

The speech data was not inspected on content. There was no transcription done on the spontaneous speech. The speech rate analyses was done through the detection of syllable nuclei by analyzing intensity and therefore the speaking rate was analyzed (De Jong & Wempe, 2009). However, the number of words were not analyzed. For the speech rhythm analyses the five recorded prosodic sentences were analyzed. In the speech rhythm analyses the sentences were not treated separately and the RMs were pooled over all five sentences. Therefore, all information on rhythmic differences between the sentences, or analyses on the consistency between the pronunciation of the five sentences could not be considered. Besides that, the forced aligner P2FA, used in the phonetic annotation of the prosodic sentences, was trained on a corpus that contains American English speech. In this study, no analyses has been done on UCU English. Analyzing the differences between UCU English and American English, but also British English, could highlight differences in phonetic annotation and in differences in results based on these annotations.

The speech data came from one cohort of UCU students and was fairly representative considering that on average 64% of the incoming UCU students are Dutch and 64% of incoming UCU students are female. The speech data used in the speech rate analyses came from 55 participants of which 38 (69.1%) were female. In the speech rhythm analyses data was analyzed from 53 participants of which 37 (69.8%) were Dutch. This proofs the analyzed data was quite representative and therefore the strongest influences in the overall means came from the female participants and the Dutch participants.

The first aim of the present study was to find how speech rate would change over time. By comparing the speaking rate and articulation rate from the spontaneous speech data at the beginning of the bachelor's to the speaking rate and articulation rate of the spontaneous speech data at the end of the bachelor's, a significant increase was found. In addition, the effect of gender was tested and it was significantly shown that women had a higher speaking rate. The second aim was to find if and how speech rhythm would change between recording sessions. Three rhythm metrics (%V, Varco and nPVI-V) were tested of which only %V showed a significant difference between the two recording sessions. VarcoV showed the best discrimination between typology different L1 rhythmic groups, however the discrimination was not significant. The means between recording sessions showed a significant increase in %V. In other words, the percentage of total utterance duration that was comprised of vowel durations increased over time. %V has shown to be able to discriminate between the recording sessions and thus its within-language discriminative ability. The lower standard deviation in the second recording session implies less variability in the speech rhythm of the participants over time. Therefore, we can conclude that convergence has been shown.

Putting these findings into a broader perspective, this study shows that the durational pattern and the temporal pattern can develop over time. Speech rate is an important aspect in non-native speech as it can add to foreign accent (Trofimovich & Baker, 2006). Speech rhythm is important because words that contain stressed syllables are perceived better when they are aligned at regular time points (Quené & Port, 2005; Quené & Van den Bergh, 2008). And as the rhythm of the native language can transfer to the L2 (Benson, 2002; Yuan, 2010) this could limit native-like L2 speech. This study showed that the speaking rate and the articulation rate increased over time, which is probably related to the increased linguistic experience (Trofimovich & Baker, 2006), increased fluency (Bosker et al., 2013) and perceived competence (Giles et al. 1991). The findings of this study on speech rhythm have shown that the native and non-native speakers of a lingua franca can show convergence over time.

8. REFERENCES

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

Anderson-Hsieh, J., & Koehler, K. (1988). The Effect of Foreign Accent and Speaking Rate on Native Speaker Comprehension. *Language learning*, *38*(4), 561-613.

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

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Beebe, L. M., & Giles, H. (1984). Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language*, *1984*(46), 5-32.

Benson, C. (2002). Transfer/Cross-linguistic influence. *ELT journal*, *56*(1), 68-70.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.3.57)(2013). *Computer program: http://www. praat. org*.

Bond, Z. S., & Stockmal, V. (2002). Distinguishing samples of spoken Korean from rhythmic and regional competitors. *Language Sciences*, *24*(2), 175-185.

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159-175.

Breitenstein, C., Lancker, D. V., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition & Emotion*, *15*(1), 57-79.

Carter, P. M. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. In *Theoretical and Experimental Approaches to Romance Linguistics: Selected Papers from the 34th Linguistic Symposium on Romance Languages (LSRL), Salt Lake City, March 2004* (Vol. 272, p. 63). John Benjamins Publishing.

Cesoniene, R. (2005). Phonetics Teaching and Learning Conference. In *English Phonetics and Phonology: Course for Future Interpreters* (pp. 1-4). Kaunas.

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

Davitz, J. R., Beldoch, M., Blau, S., Dimitrovsky, L., Levitt, E., & Kempner, P. (1964). Personality, perceptual, and cognitive correlates of emotional sensitivity. *The communication of emotional meaning*, 57-68.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, *41*(2), 385-390.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for∆ C. *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium* (pp. 231–241). Frankfurt: Peter Lang.

Deterding, D. (1994). The characteristics of Singapore English pronunciation. *Review of Educational Research and Advances for Classroom Teachers*, *1*.

Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, *25*(1), 11-28.

Elder, C., Golombek, P., Nguyen, T. T., & INGRAM, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, *39*(2), 309-319.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*(6), 3814-3826.

Feldstein, S., Dohm, F. A., & Crown, C. L. (2001). Gender and speech rate in the perception of competence and social attractiveness. *The Journal of social psychology*, *141*(6), 785-806.

Fenk-Oczlon, G., & Fenk, A. (2006). Speech rhythm and speech rate in crosslinguistic comparison. In *Proceedings of the 28 th Annual Conference of the Cognitive Science Society* (p. 2480).

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology*, *20*(1), 104-113.

Flege, J., Frieda, E., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169–186.

Gibbon, D. (2003). Computational modelling of rhythm as alternation, iteration and hierarchy. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2489–2492). Barcelona.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics , 15*, 87-105.

Giles, H., Bouhris, R., & Taylor, D. (1977). Towards a Theory of Language in Ethnic Group Relations. In H. Giles, *Language, Ethnicity and Intergroup Relations* (pp. 307-348). London: Academic Press.

Giles, H., Coupland, J., & Coupland, N. (1991). *Accommodation Theory: Communication, Context, and Consequence.* Cambridge: Cambridge University Press.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. *Communication yearbook*, *10*(1), 13-48.

Goldstein, B. A., & Bunta, F. (2012). Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism*, *16*(4), 388-401.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Papers in laboratory phonology 7(* 515–546). Cambridge University Press, Cambridge.

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* (pp. 155-166). Somerville, Massachusetts: Cascadilla Press.

Gut, U. (2003). Non-native speech rhythm in German. *Proceedings of the ICPhS conference* (pp. 2437-2440). Barcelona.

Gut, U. (2012). Rhythm in L2 speech. *Speech and Language Technology*, *14/15*, 83-94.

Hamers, J., & Blanc, M. (2000). *Bilinguality and bilingualism.* Cambridge: Cambridge University Press.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics, 22*(3), 338-365.

Kowal, S., Wiese, R., & O'Connell, D. C. (1983). The use of time in storytelling. *Language and Speech*, *26*(4), 377-392.

Ladefoged, P. (2004). *Vowels and consonants: An introduction to the sounds of languages* (2nd ed.). Oxford: Blackwells.

Laver, J. (1994). *Principles of phonetics.* Cambridge: University Press.

Lehiste. (1977). Isochrony reconsidered. *Journal of Phonetics , 5*(3), 253-263.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America*, *129*(5), 3258-3270.

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

Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, *63*(4), 247-267.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*(4), 215-225.

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

Orr, R., Quené, H., van Beek, R., Diefenbach, T., van Leeuwen, D. A., & Huijbregts, M. (2011). An International English Speech Corpus for Longitudinal Study of Accent Development. In *INTERSPEECH* (pp. 1889-1892).

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190-197.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, *87*(3), 539-558.

Peppé, S. (2012). Prosody in the world's languages. In S. McLeod, & B. Goldstein, *Multilingual Aspects of Speech Sound Disorders in Children* (pp. 42-52). Multilingual Matters.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434.

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

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America*, *123*(2), 1104-1113.

Quené, H., & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*(1), 10-13.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, *123*(2), 1104-1113.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413-425.

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

Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin , 99*(2), 143–165.

Schumann, J. H. (1975). Affective factors and the problem of age in second language acquisition. *Language learning*, *25*(2), 209-235.

Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual & Multicultural Development*, *7*(5), 379-392.

Siptár, P., & Törkenczy, M. (2007). *The phonology of Hungarian*. Oxford University Press.

Sluijter, A. M., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical society of America*, *100*(4), 2471-2485.

Street Jr, R. L., & Brady, R. M. (1982). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communications Monographs*, *49*(4), 290-308.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition*, *28*(01), 1-30.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, *47*(3), 297-308.

White, L. & Mattys, S. L. (2007a). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in romance phonology* (pp.237-257). Current issues in linguistic theory series. Amsterdam, Philadelphia: John Benjamins

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

Yuan, J. (2010). Linguistic rhythm in foreign accent. In *INTERSPEECH* (pp. 1848-1849).

Yuan, J. & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics* (p. 5687-5690).

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller, *Fundamentals of speech* synthesis and speech recognition (pp. 41-62). Chichester: John Wiley.

APPENDIX I: PARTICIPANT GROUPING

Participant group	Language	N		Classification (source)
1	Dutch	37	♀: 26 ♂: 11	Stress-timed (Grabe & Low, 2002)
2	English	5	♀: 5	Stress-timed (Dauer, 1983; Grabe & Low, 2002)
3	Indonesian	1	♀: 4 ♂: 2	Syllable-timed (Dauer, 1983)
	Vietnamese	1		Syllable-timed (Elder, Golombek, Nguyen, & Ingram, 2005)
	Latvian	1		Syllable-timed (Bond & Stockmal, 2002)
	Spanish	1		Syllable-timed (Dauer, 1983)
	Hungarian	1		Syllable-timed (Siptar & Torkenczy, 2000)
	Lithuanian	1		Syllable-timed (Cesoniene, 2005)
4	German	2	♀: 1 ♂: 4	Stress-timed (Grabe & Low, 2002)
	Russian	1		Stress-timed (Dauer, 1983)
	Hebrew	1		Stress (Peppé, 2012)
	Swedish	1		Stress (Pike, 1945)

APPENDIX II: R SCRIPT IMPLEMENTATION OF RHYTHM METRICS

```
###########################################
##      RMs: %V, VarcoV, nPVI_V     ##
##   input: dataframe (from Textgrid) ##
###########################################

folderpath = "PATH"

folderresults <- c(0,0,0,0)
results <- c(0,0,0,0)
sum_percV <- 0
sum_VarcoV <- 0
sum_nPVI_V <- 0

# for all files in directory
files <- list.files(path=folderpath, pattern=".df.txt", all.files=T, full.names=T)

for (file in files)
{
  tab <- read.table(file=file, header=T, sep=" ")

  # create table with only vowels
  vowels <-
tab[grep('AA1|AA2|AE1|AE2|AH0|AH1|AH2|AO1|AW1|AY1|EH1|ER0|ER1|EY1|EY2|IH0|IH1|IY0|IY1|OW1|UH1|UW1
', tab[,6]),]

  # sum and mean duration vowels
  sum_vowels <- colSums(vowels[5])
```

```
  mean_vowels <- colMeans(vowels[5])
  sd_vowels <- sapply(vowels[5], sd)

  # create table with total vowel/consonants
  totalphon <- tab[-grep('ns|sp|sil', tab[,6]),]
  consonants <- totalphon[-
grep('AA1|AA2|AE1|AE2|AH0|AH1|AH2|AO1|AW1|AY1|EH1|ER0|ER1|EY1|EY2|IH0|IH1|IY0|IY1|OW1|UH1|UW1',
totalphon[,6]),]

  # sum duration phon
  sum_totalphon <- colSums(totalphon[5])

  percV <- sum_vowels/sum_totalphon*100
  VarcoV <- 100*sd_vowels/mean_vowels

nPVI <- function(dur)
{
		x <- diff(dur)
		ub <- length(x) # upper bound of difference=x
		for (k in 1:ub)
				x[k] <- abs ( x[k]*2 / (dur[k]+dur[k+1]) )
		return( (100/ub)*sum(x) )
}#end nPVI

  nPVI_V <- nPVI(vowels[,5])

  # append result to table
  filename <- sub("\\-prosody.df.txt", "", basename(as.character(file)))
  new <- c(filename, percV, VarcoV, nPVI_V)
  results <- rbind(results, new)
  results.table <- as.table(results)

  # add mean_vowels tot total vowel means
  sum_percV <- sum_percV + percV
  sum_VarcoV <- sum_VarcoV + VarcoV
  sum_nPVI_V <- sum_nPVI_V + nPVI_V

} # end for


# mean of folder
mean_percV <- sum_percV/length(files)
mean_VarcoV <- sum_VarcoV/length(files)
mean_nPVI_V <- sum_nPVI_V/length(files)

newtot <- c(basename(as.character(folderpath)), mean_percV, mean_VarcoV, mean_nPVI_V)
folderresults <- rbind(folderresults, newtot)

colnames(results) <- c('participant', '%V', 'VarcoV', 'nPVI_V')
results <- results[-1,]

results.table <- as.table(results)
write.table(results, "PATH" , col.names=FALSE, row.names=FALSE)
```

APPENDIX III: ESTIMATED MARGINAL MEANS - SPEECH RATE

General Linear Model analysis.
Categorical between-subject variable: gender
Categorical within-subject variable: recording
Continuous dependent variable: number of pauses.

Number of pauses	Effect	df1	df2	F	*p*-value
Within-Subjects Effects	**recording**	1	53	4.329	**0.042**
	recording * gender	1	53	1.032	0.314
Between-Subjects Effects	gender	1	53	1.321	0.256

General Linear Model analysis.
Categorical between-subject variable: gender
Categorical within-subject variable: recording
Continuous dependent variable: speaking rate.

Speaking rate	Effect	df1	df2	F	*p*-value
Within-Subjects Effects	**recording**	1	53	29.877	**0.000**
	recording * gender	1	53	0.969	0.329
Between-Subjects Effects	**gender**	1	53	4.373	**0.041**

General Linear Model analysis.
Categorical between-subject variable: gender
Categorical within-subject variable: recording
Continuous dependent variable: articulation rate.

Articulation rate	Effect	df1	df2	F	*p*-value
Within-Subjects Effects	**recording**	1	53	7.548	**0.008**
	recording * gender	1	53	0.331	0.568
Between-Subjects Effects	gender	1	53	2.644	0.110

APPENDIX IV: ESTIMATED MARGINAL MEANS - RHYTHM METRICS

A. %V

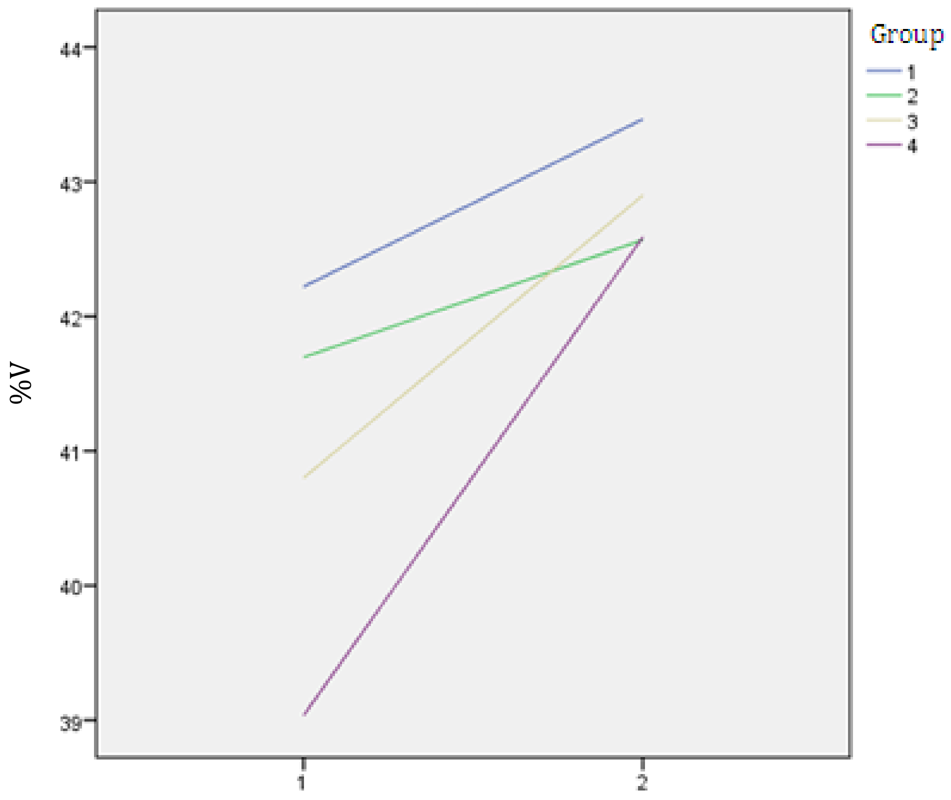
General Linear Model analysis.
Categorical between-subject variable: group, gender
Categorical within-subject variable: recording
Continuous dependent variable: %V.

	Group	Gender	Mean	Std. Deviation	N
%V Recording 1	1	Female	43.417	2.749	26
		Male	41.024	2.038	11

		Total	42.705	2.762	37
	2	Female	41.696	4.706	5
		Total	41.696	4.706	5
	3	Female	42.627	3.160	4
		Male	38.973	3.130	2
		Total	41.409	3.392	6
	4	Female	36.759	.	1
		Male	41.311	1.554	4
		Total	40.401	2.440	5
	Total	Female	42.905	3.206	36
		Male	40.850	2.043	17
		Total	42.246	3.023	53
%V Recording 2	1	Female	44.493	2.447	26
		Male	42.437	2.850	11
		Total	43.881	2.706	37
	2	Female	42.569	3.063	5
		Total	42.569	3.063	5
	3	Female	44.792	2.980	4
		Male	41.008	2.325	2
		Total	43.531	3.198	6
	4	Female	43.859	.	1
		Male	41.323	1.084	4
		Total	41.830	1.472	5
	Total	Female	44.241	2.568	36
		Male	42.006	2.451	17
		Total	43.524	2.719	53

%V	Effect	df1	df2	F	*p*-value
Tests of Within-Subjects Effects	**recording**	1	46	15.476	**0.000**
	recording * group	3	46	1.655	0.190
	recording * gender	1	46	4.390	**0.042**
	recording * group * gender	2	46	3.953	**0.026**
Tests of Between-Subjects Effects	**group**	3	46	1.321	**0.000**
	gender	1	46	1.902	0.279
	group * gender	2	46	0.939	0.175

Estimated Marginal Means of %V



Post-hoc test Tukey HSD

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	1.160838	1.1686397	0.754	-1.954166	4.275843
	3	0.823613	1.0794427	0.871	-2.053638	3.700863
	4	2.178182	1.1686397	0.258	-0.936823	5.293186
2	1	-1.160838	1.1686397	0.754	-4.275843	1.954166
	3	-0.337226	1.4851747	0.996	-4.295953	3.621502
	4	1.017343	1.5512144	0.913	-3.117413	5.152100
3	1	-0.823613	1.0794427	0.871	-3.700863	2.053638
	2	0.337226	1.4851747	0.996	-3.621502	4.295953
	4	1.354569	1.4851747	0.799	-2.604159	5.313297
4	1	-2.178182	1.1686397	0.258	-5.293186	0.936823
	2	-1.017343	1.5512144	0.913	-5.152100	3.117413
	3	-1.354569	1.4851747	0.799	-5.313297	2.604159

Based on observed means.

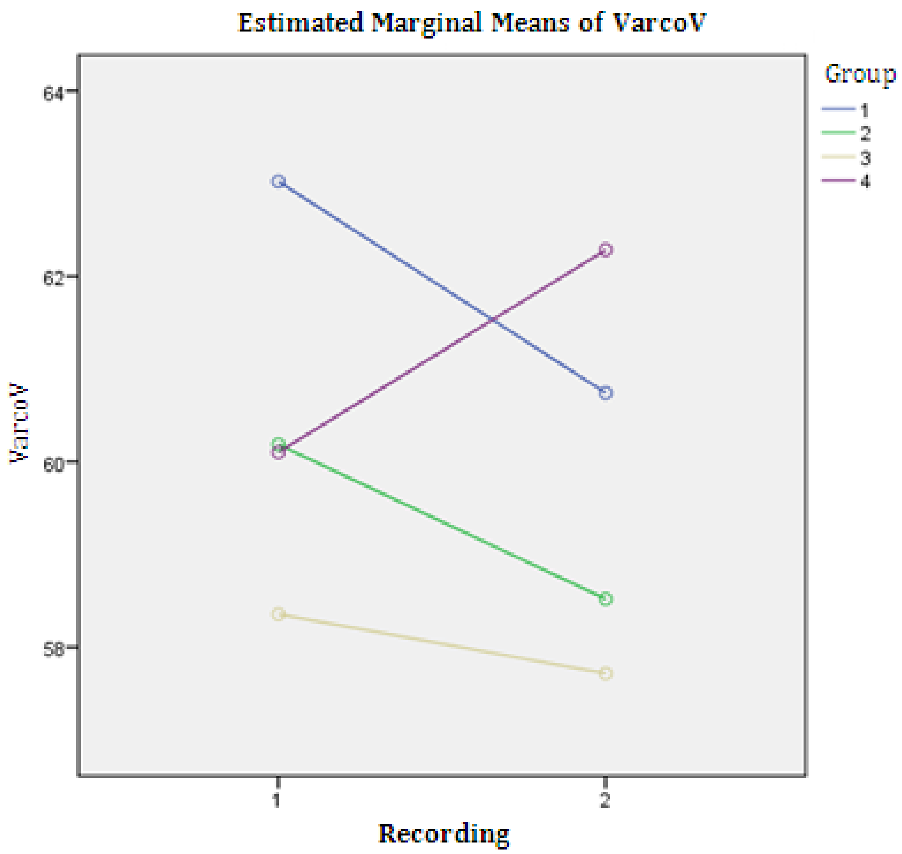
The error term is Mean Square(Error) = 6.016.

B. VARCOV

General Linear Model analysis.

Categorical between-subject variable: group, gender
Categorical within-subject variable: recording
Continuous dependent variable: VarcoV.

VarcoV	Effect	df1	df2	F	*p*-value
Tests of Within-Subjects Effects	recording	1	46	0.329	0.569
	recording * group	3	46	0.705	0.554
	recording * gender	1	46	0.629	0.432
	recording * group * gender	2	46	0.525	0.595
Tests of Between-Subjects Effects	group	3	46	1.413	0.251
	gender	1	46	3.900	0.054
	group * gender	2	46	0.996	0.377



Post-hoc test Tukey HSD

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
	2	2.191140	1.9017900	0.660	-2.878074	7.260354
1	**3**	4.664530	1.7566349	**0.051**	-0.017774	9.346834
	4	-0.516935	1.9017900	0.993	-5.586149	4.552279

	1	-2.191140	1.9017900	0.660	-7.260354	2.878074
2	3	2.473390	2.4169045	0.737	-3.968860	8.915640
	4	-2.708075	2.5243744	0.708	-9.436785	4.020636
	1	-4.664530	1.7566349	**0.051**	-9.346834	0.017774
3	2	-2.473390	2.4169045	0.737	-8.915640	3.968860
	4	-5.181465	2.4169045	0.155	-11.623715	1.260785
	1	0.516935	1.9017900	0.993	-4.552279	5.586149
4	2	2.708075	2.5243744	0.708	-4.020636	9.436785
	3	5.181465	2.4169045	0.155	-1.260785	11.623715

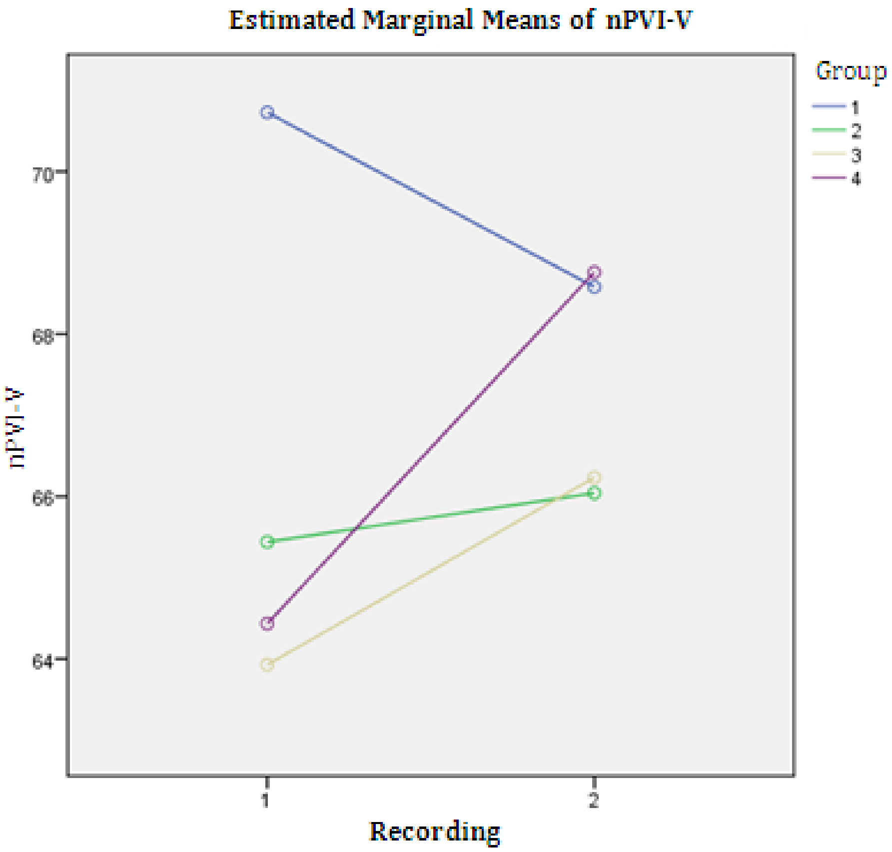
Based on observed means.

The error term is Mean Square(Error) = 15.931.

C. NPVI-V

General Linear Model analysis.
Categorical between-subject variable: group, gender
Categorical within-subject variable: recording
Continuous dependent variable: nPVI-V.

nPVI-V	Effect	df1	df2	F	*p*-value
Tests of Within-Subjects Effects	recording	1	46	0.760	0.388
	recording * group	3	46	1.525	0.221
	recording * gender	1	46	0.219	0.642
	recording * group * gender	2	46	2.121	0.131
Tests of Between-Subjects Effects	group	3	46	2.169	0.105
	gender	1	46	1.334	0.254
	group * gender	2	46	0.350	0.707



Post-hoc test Tukey HSD

(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
	2	3.719621	2.0985822	0.299	-1.874142	9.313384
1	3	4.576951	1.9384068	0.099	-.589865	9.743767
	4	1.278746	2.0985822	0.929	-4.315017	6.872509
	1	-3.719621	2.0985822	0.299	-9.313384	1.874142
2	3	0.857331	2.6669993	0.988	-6.251546	7.966207
	4	-2.440874	2.7855900	0.817	-9.865854	4.984105
	1	-4.576951	1.9384068	0.099	-9.743767	0.589865
3	2	-0.857331	2.6669993	0.988	-7.966207	6.251546
	4	-3.298205	2.6669993	0.607	-10.407082	3.810672
	1	-1.278746	2.0985822	0.929	-6.872509	4.315017
4	2	2.440874	2.7855900	0.817	-4.984105	9.865854
	3	3.298205	2.6669993	0.607	-3.810672	10.407082

Based on observed means.

The error term is Mean Square(Error) = 19.399.