

This is a repository copy of *Methods for investigation of L2 speech rhythm: Insights from the production of English speech rhythm by L2 Arabic learners*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/194935/>

Version: Published Version

Article:

Algethami, Ghazi and Hellmuth, Sam orcid.org/0000-0002-0062-904X (2023) Methods for investigation of L2 speech rhythm: Insights from the production of English speech rhythm by L2 Arabic learners. *Second Language Research*. pp. 431-456. ISSN 0267-6583

<https://doi.org/10.1177/02676583231152638>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Methods for investigation of L2 speech rhythm: Insights from the production of English speech rhythm by L2 Arabic learners

Second Language Research

2024, Vol. 40(2) 431–456

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02676583231152638

journals.sagepub.com/home/slr**Ghazi Algethami** 

Taif University, Saudi Arabia

Sam Hellmuth 

University of York, UK

Abstract

Rhythm metrics can detect second language development of target-like speech rhythm but interpretation of the results from metrics in learners' speech is problematic because the mapping of metrics to underpinning phonological features is indirect. We investigate speech rhythm in first language (L1) Arabic / second language (L2) English, which differ in key properties contributing to the percept of rhythm: unstressed vowel reduction and syllable structure. Our production data are interpreted using additional measures, of stressed and unstressed vowels and of consonant cluster realization, alongside standard rhythm metrics; this combination facilitates disambiguation of competing interpretations of the metric results. The findings confirm the importance of using multiple rhythm metrics to study L2 speech rhythm and demonstrate how simple additional measures can guide interpretation of their results. In this study the metrics results showed that the speech produced by the L2 speakers, regardless of their length of residence in the UK, exhibited lower vocalic durational variability than the speech produced by the native Arabic and English speakers. However, closer inspection of the degree of vowel reduction by the native and nonnative groups confirms that no single metric captures the complex nature of the observed L2 rhythm patterns. Future L2 studies are advised not to draw firm conclusions about the degree of vowel reduction and consonant cluster realization in L2 speech based solely on the results of the rhythm metrics.

Keywords

Arabic, English, rhythm, metrics, unstressed vowels, vowel reduction

Corresponding author:

Sam Hellmuth, Department of Language and Linguistic Science, University of York, York, YO10 5DD, UK

Email: sam.hellmuth@york.ac.uk

I Introduction

Rhythm remains one of the lesser studied areas in second language (L2) speech research, despite evidence that L2 speech rhythm is usually characterized by non-target-like rhythmic patterns that potentially render it unintelligible or difficult to understand (e.g. Adams, 1979; Jones, 1962; Taylor, 1981). This research gap is perhaps due to the inherent difficulty in, or lack of consensus on, defining and measuring speech rhythm. Early studies of L2 speech rhythm relied either on impressionistic description, or on study of acoustic cues to stress which is only one component of rhythm. More recently, with the inception of rhythm metrics, a few studies have used these metrics to examine the production of L2 speech rhythm by various learners from different language backgrounds (e.g. Li and Post, 2014; Ordin and Polyanskaya, 2015; Stockmal et al., 2005; White and Mattys, 2007a). Despite the relative success of the acoustic metrics in capturing L2 speech rhythm, their results are hard to interpret, which is probably due to the fact that the mapping of rhythm metrics to phonological features, such as vowel reduction and syllable structure, from which the overall percept of speech rhythm is hypothesized to arise, is indirect.

The current study examines the production of English speech rhythm by L2 Saudi learners, with comparison to L1 English and L1 Arabic, for the first time. Although Arabic and English are both traditionally classified as stress-timed languages (Abercrombie, 1976; Miller, 1984), Arabic differs from English along three phonological parameters relevant to variation in speech rhythm: degree of unstressed vowel reduction, ratio of long/short phonemic vowel duration, and syllable structure. It is expected that the difference between Arabic and English along multiple parameters will make the results of the metrics hard to interpret. A simple methodological innovation is proposed whereby use of straightforward additional measures of relevant vocalic and consonantal properties, alongside the standard rhythm metrics, aid interpretation of rhythm metrics results. In addition, the current study examines the potential effect of length of residence as an index of language experience on the development of English speech rhythm among L2 Saudi learners.

Speech rhythm is both conceptually and empirically complex, so, in the next section, we first introduce the notion of rhythm in speech and how it has been operationalized and measured, before reviewing prior studies that examined L2 speech rhythm using rhythm metrics.

II Background

I Rhythm in speech

Rhythm in speech has long been debated among linguists: whether languages vary in their rhythmic properties and, if so, how that variation can be captured and measured. Early attempts suggested a strong tendency in English for stressed syllables to occur at regular or equal intervals (i.e. isochrony) (Jones, 1962). Pike (1945) coined the terms 'stress-timed' and 'syllable-timed' to describe rhythm in languages. Notably, he mentioned that all languages appeared to display both kinds of rhythm but differ in that they

may favor one more than the other. Subsequent studies did not find concrete evidence of isochrony in the speech signal (e.g. Dauer, 1983; Roach, 1982; Wenk and Wioland, 1982). For this reason, it was suggested that rhythm is instead a perceptual phenomenon (e.g. Allen, 1975; Lehiste, 1977). However, the question of how rhythmic variation between languages could be measured was not answered.

Roach (1982) and Dasher and Bolinger (1982) suggested that auditory classification of languages into 'stress-timed' and 'syllable-timed' might be attributable to differences that those languages exhibit in degree of complexity of syllable structure, and in existence of vowel length distinctions and/or reduction of unstressed syllables. It was suggested that 'stress-timed' languages, such as English, German and Dutch, tend to have more complex syllable structure and are more likely to exhibit vowel reduction than 'syllable-timed' languages, such as French, Chinese and Italian. This suggestion was elaborated further by Dauer (1983), who proposed that speech rhythm is not a phonetic feature or a phonological primitive, but rather a manifestation of multiple phonological features, namely, stress, vowel reduction and syllable structure. Dauer maintained that all languages are more or less stress-based and cannot be divided into two dichotomous rhythmic types. In the current research, we take Dauer's (1983) position that classifying languages into distinct rhythmic classes (stress-timed and syllable-timed) is untenable, and that the percept of rhythm is a consequence of various phonological features, among which vowel reduction and syllable structure play a major role.

An acoustic implementation of this phonological stance on speech rhythm was first put forth by Ramus et al. (1999), who support the phonological basis for classifying languages rhythmically, but propose that the resulting phonetic timing differences are independently measurable. Previous experiments had shown that neonates can discriminate between two languages conventionally classified into two different rhythmic types relying merely on rhythmic cues (Nazzi et al., 1998). Ramus et al. (1999) argued that infants cannot be using complex language-specific phonological concepts to segment speech, but rather the succession of vowels of variable durations separated by unanalysed speech segments; a similar view was expressed in Mehler et al. (1996). Thus, Ramus et al. (1999) combined the phonological explanation of rhythmic typology (Dauer, 1983) with the simpler task of segmenting speech into vowels and consonants (Mehler et al., 1996; Nazzi et al., 1998) to propose a new acoustic quantification of rhythmic typology.

Ramus et al. (1999) proposed three acoustic metrics of rhythm: %V, percentage of the total duration of vocalic intervals; ΔV , standard deviation of the duration of vocalic intervals; and ΔC , standard deviation of the duration of consonantal intervals. The authors hypothesized that 'syllable-timed' languages would display lower ΔV and ΔC values than 'stress-timed' languages because 'stress-timed' languages tend to show more durational variation between consonantal intervals (due to complexity of consonantal clusters) and between stressed and unstressed vowels (due to shortening of unstressed vowels). %V was hypothesized to be higher in 'syllable-timed' languages than in 'stress-timed' languages for the same reasons as for ΔV . Later additions and modifications to the metrics included normalization for speech rate and localization of measurements (e.g. nPVI-V, rPVI-C, VarcoV & VarcoC). Table 1 provides a summary of the most widely used rhythm metrics.

Table 1. Summary of the acoustic rhythmic measures.

Metric	Measurement	Related work
%V	Percentage of the total duration of vocalic intervals	Ramus et al., 1999
ΔV	Standard deviation of the durations of vocalic intervals	Ramus et al., 1999
ΔC	Standard deviation of the durations of consonantal intervals	Ramus et al., 1999
nPVI-V	Mean of the durational differences between successive vocalic intervals divided by their sum, and multiplied by 100	Low et al., 2000
rPVI-C	Mean of the durational differences between successive consonantal intervals	Grabe and Low, 2002
VarcoV	Standard deviation of the durations of vocalic intervals divided by the mean duration of vocalic intervals, and multiplied by 100	White and Mattys, 2007a
VarcoC	Standard deviation of the durations of consonantal intervals divided by the mean duration of consonantal intervals, and multiplied by 100	Dellwo, 2006

Arabic and English are both traditionally described as ‘stress-timed’ languages (Abercrombie, 1976; Miller, 1984). English is widely considered an archetypical ‘stress-timed’ language and was shown to exhibit relatively higher durational variability between vocalic segments and between consonantal segments (e.g. Grabe and Low, 2002; Ramus et al., 1999; White and Mattys, 2007a). A few studies have examined rhythmic variation in Arabic using some of the widely used acoustic metrics (e.g. Hamdi et al., 2004; Ghazali et al., 2002). The results generally show that Western Arabic dialects (e.g. Moroccan) are more ‘stress-timed’ than Eastern dialects (e.g. Jordanian). We are not aware of any study that makes direct comparison between Saudi Arabic and English using the rhythm metrics. In line with Tajima et al. (1999), who used a phrase repetition method, we expect Saudi Arabic to manifest less stress-timing (that is, less durational variability) than English because Saudi Arabic exhibits less complex syllable structure (maximum two-consonant clusters in onset and coda positions). In addition, stress seems to exert a lengthening effect on vowels in Arabic (for Jordanian Arabic, see de Jong and Zawaydeh, 2002), though it is unknown how unstressed vowels manifest phonetically in Saudi Arabic. Another key difference between English and Arabic is related to phonemic vowel length contrast. In Arabic, short vowels are around half the duration of long vowels, and quantity plays a major role in their contrast, while in English, vowel quality plays the major role in the contrast between short and long vowels (Alghamdi, 1998; Roach, 2009).

Several studies have examined the success, stability and reliability of rhythm metrics (e.g. Arvaniti, 2012; Knight, 2011; White and Mattys, 2007a; Wiget et al., 2010). Due to the sensitivity of the metrics to variation in speech styles and samples, their potential to classify languages into traditional rhythm classes is generally agreed to be weak, but their capacity to distinguish languages and dialects is acknowledged. The rhythm metrics thus offer a potential solution to the original conundrum highlighted by Ramus et al.

(1999), whereby rhythmic differences can be perceived – even by newborns – but resist dichotomous categorization according to any single measurable acoustic parameter. Our stance follows Dauer (1983) in assuming a multi-source phonological basis to speech rhythm, which results in a continuum of surface variation. We hypothesize that the various phonological differences between Arabic and English will cause them to fall at different points along that surface rhythm continuum, and indeed at sufficient distance that the difference is detectable in perception and/or in rhythm metrics values. In this sense, the expectation that we will find differences in rhythm metric scores between languages, or between L1 versus L2 speech, remains consistent with rejection of a simple ‘stress-timed’ versus ‘syllabled-timed’ rhythm class dichotomy.

Overall then, the rhythm metrics provide a useful quantitative tool to study the acquisition and production of speech rhythm by L2 learners, especially in the absence of any other reliable rhythm measures, and given the importance of speech rhythm to L2 speech (e.g. Li and Post, 2014; Ordin and Polyanskaya, 2015).

2 L2 speech rhythm

A number of studies have used the acoustic rhythm metrics to examine the production and acquisition of L2 speech (Li and Post, 2014; Mok and Dellwo, 2008; Polyanskaya and Ordin, 2019; Stockmal et al., 2005; White and Mattys, 2007a, 2007b). The vocalic metrics (e.g. nPVI-V, VarcoV & %V) were generally found to be more successful than the consonantal ones (e.g. ΔC , VarcoC & rPVI-C) in capturing the rhythmic differences between native and nonnative speech (e.g. Li and Post, 2014; Stockmal et al., 2005; White and Mattys, 2007a). In particular, VarcoV was shown to be a robust measure for differentiating between native and nonnative speech (White and Mattys, 2007b). Later L2 research substantiated the power of VarcoV for examining L2 speech (Li and Post, 2014; Ordin and Polyanskaya, 2014).

One of the reasons for the relative weakness of the consonantal-based rhythm metrics in capturing the phonotactic nature of L2 speech is that they can be easily influenced by speaking rate (e.g. Grabe and Low, 2002; White and Mattys, 2007a). Grabe and Low (2002) tested a normalized Pairwise Variability Index (PVI) for measuring variability between consonantal intervals, but recommended using a non-normalized measure instead since consonantal intervals may comprise different segments which are affected differently by speaking rate. Normalizing the consonantal metrics thus potentially eliminates rhythmically important variation in the speech signal (White and Mattys, 2007a; Li and Post, 2014). Therefore, in the current study we only used the non-normalized consonantal metrics.

Previous research on L2 speech rhythm has widely focused on cross-linguistic transfer to explain non-native speech rhythm, based on the assumption that similarity between L1 and L2 rhythm would play a facilitative role in the acquisition of L2 rhythm. Results from some studies comparing the production of various L2 speaker groups belonging to rhythmically different L1 backgrounds support this L1 influence assumption. For example, White and Mattys (2007a) examined the production of English speech rhythm by native English speakers and nonnative Dutch and Spanish speakers. The speech produced by the English and L2 Dutch speakers exhibited similar VarcoV scores, which

were significantly higher than the scores obtained for the speech produced by the L2 Spanish speakers. One might attribute the relative success of the L2 Dutch speakers to the rhythmic similarity between Dutch and English, since both are traditionally classified as stress-timing languages, in contrast to Spanish, a syllable-timing language. In a similar and more recent study, Li and Post (2014) compared the rhythmic patterns of German and Chinese L2 learners of English, who were divided into lower intermediate and advanced level groups in terms of their English proficiency. Their results showed that L1 influence is not sufficient to account for L2 speech rhythm. Both the German and Chinese lower intermediate learners exhibited similar and significantly lower VarcoV scores than did the native English controls, despite their rhythmically different L1 backgrounds. In contrast, the VarcoV scores obtained for the advanced German and Chinese L2 learners approximated those of the native English speakers. The authors attributed the similar developmental trajectory followed by L2 learners from rhythmically different L1 backgrounds to a universal mechanism. A similar finding and conclusion were reached by Ordin and Polyanskaya (2015) with regard to L2 English learners who were native speakers of French and German, even though their advanced French learners did not approximate the rhythmic properties of native English speakers as closely as their advanced German learners did. Overall though, the observed developmental trend towards more stress-timing rhythm in the acquisition of English points to a universal developmental path (Li and Post, 2014; Ordin and Polyanskaya, 2015), which may also be mediated by a learner's L1 background (Ordin and Polyanskaya, 2015).

A few studies have examined the effect of language experience on the development of L2 speech rhythm (Lee and Song, 2019; Li and Post, 2014; Ordin and Polyanskaya, 2014, 2015). Language experience was operationalized either by measuring learners' length of residence in a target-language community or by dividing learners according to proficiency level. However, both length of residence and language proficiency level provide only approximate measures, and by no means capture the individual and complex nature of language experience. The results typically showed that L2 English rhythm progressed from syllable-timing towards stress-timing rhythm irrespective of learners' L1 backgrounds (Li and Post, 2014; Ordin and Polyanskaya, 2014, 2015), but the rhythm metrics used by Lee and Song (2019) did not reflect the different levels of proficiency of their L2 Korean learners of English. The effect of language experience, or indeed of other L2-acquisition related factors such as age of acquisition and mode of instruction, on the learning of L2 rhythm, remains largely uninvestigated.

III The current study

The current study examines the acquisition of L2 English speech rhythm by L2 Saudi learners. The study contributes to the growing literature on L2 speech rhythm in three ways. First, it examines English speech rhythm of an L2 population not examined before (Arabic learners of English), in languages which differ along key phonological parameters relevant to the global percept of speech rhythm regardless of the fact that they are both traditionally classified as stress-timed language. Second, it examines the effect of length of residence, as a rough index of L2 experience, on the production of English speech rhythm. Third, it goes beyond the use of rhythm metrics alone to also examine

vowel reduction and consonant cluster realization, as additional simple measures of the assumed phonological building blocks of speech rhythm.

Comparable speech samples were collected from native Saudi Arabic and native English speakers, as well as from L2 Saudi speakers of English divided into two groups based on their length of residence in the UK. Various vocalic and intervocalic rhythm metrics were calculated for all the collected samples, and post-hoc analyses of vowel reduction and syllable structure were also conducted to explain the metrics results.

Drawing on the results of previous research on L2 speech rhythm, we predict the Saudi learners with longer residence in the UK to show similar durational variability to that obtained for the native Arabic and English speaker groups, and those with shorter length of residence in the UK to show lower durational variability than that obtained for the native Arabic and English speaker groups.

IV Method

1 Speakers

The speaker participants in the current study were L1 and L2 English speakers. The L1 English group consisted of six native speakers of Standard Southern British English (SSBE) aged 20–40 years, drawn from students and staff at the University of York. The L2 English speakers were 12 native speakers of Najdi Saudi Arabic (NSA) who were of two groups, labelled ‘more experienced’ (ME) and ‘less experienced’ (LE), based on their length of residence in the UK. They were recruited from among international Saudi students in the UK, with length of residence in the UK ranging from one to five years. Six of the 12 L2 speakers also provided the native Najdi Saudi Arabic speech. There was no basis for their selection other than their availability at the time to provide the native Arabic speech. The participants were divided into four groups: SSBE, ME L2, LE L2 and NSA.

The ME L2 speakers were six university students in the UK, aged 27–32 years, who had spent from two and half to five years in the UK. The LE L2 speakers, aged 19–32 years, were six English language students, who had spent one year in English language schools in York, UK. Length of residence (LoR) was used in many previous studies as an index of L2 experience, even though it provides only a rough measurement of L2 experience, since longer residency does not always entail greater language experience (Piske et al., 2001). Nevertheless, LoR arguably provides a more objective measure of L2 experience than L2 speakers’ self-reported language use.

2 Materials and procedure

L2 speech elicited by direct pronunciation assessment tasks, such as sentence reading, may encourage L2 speakers to monitor their speech production. This could lead to underestimation of the extent of L1 transfer or phonetic variation observed in speech produced under more natural conditions. However, read speech tasks provide control over lexical content and phonetic/phonological features to be examined, and also maximize comparability of speech samples across speakers. One way to avoid self-monitored speech

while keeping the advantages of controlled speech is to place a moderate cognitive load on participants. In this way, they are more preoccupied with composing the message than with monitoring their pronunciation accuracy. The current study used an elicitation method (adopted from Algethami et al., 2011) which offers control over the content and lexical items in the utterances of the L2 speakers, but deflects them from monitoring their L2 speech production.

The L1 and L2 English speakers were asked to paraphrase 10 English sentences (for the full list of sentences, see Appendix 1). They were first asked to write a paraphrase in response to a written prompt word, then after writing each paraphrase they were asked to read it aloud twice into a microphone at natural speech pace. An example is given in (1).

- (1) Example Stimulus: One of the developed countries in the world is Japan.
 Prompt Word: Japan _____
 Paraphrase response: Japan is one of the developed countries in the world.

The second rendition was analysed only when hesitation or disfluency affected the first. Although the L2 speakers all had sufficient proficiency in English to engage in university studies, they were invited to stop the test at any time to ask what a certain word meant or how it should be pronounced. The paraphrase task was designed to be difficult enough to engage the L2 participants and deflect their attention from focusing on their pronunciation while reading. The test was also time-constrained, with 15 minutes per participant. Although the paraphrase task is not strictly needed for L1 English speakers, it was considered preferable to elicit all the English speech samples under the same conditions. Another advantage of the paraphrase task is that writing the paraphrase sentences out first familiarizes speakers with the sentence to be read, and should reduce the occurrence of pauses and hesitations that affect the rhythmic flow of utterances.

Elicitation of NSA speech samples was designed in the context of Arabic diglossia. Reading and writing in Arabic colloquial varieties such as NSA is unnatural to L1 Arabic speakers; reading/writing are associated with Standard Arabic, which is not used in daily conversation and is phonologically distinct from colloquial varieties. Therefore, when constructing Arabic sentences to be read by the L1 NSA speakers, one must consider the possibility that they may lean towards reading the sentences in Standard Arabic.

Ten NSA sentences were constructed by the first author who is an L1 speaker of Saudi Arabic (for the full list of the sentences along with their IPA transcription, see Appendix 2). Prior to recording, the sentences were checked verbally with three of the NSA speakers (from among the participants sample), who confirmed that the sentences sounded natural and acceptable in NSA. To avoid the sentences being read in Standard Arabic, the first author read the full set of sentences aloud in colloquial Saudi Arabic to each speaker at the start of each session, to provide an example of the speech register to be used. Reading all the sentences at once avoids biasing participants' responses towards imitation of the model rendition of the sentences; by the end of reading the last sentence they would have forgotten the acoustic detail of how the first one was produced. The NSA speakers then read each sentence twice at a normal speech rate in their own dialect.

Most of the speakers were recorded in a sound-treated phonetics laboratory at the University of York. Four NSA speakers were not resident in York, so were recorded in a

quiet furnished room in each participant's home, using a Marantz PMD660 digital recorder with Shure SM10A-CN headset condenser microphone. All recordings were digitized at 16 bit with 44.1 kHz sampling frequency, then transferred into computer memory for analysis.

The construction of the English and Arabic target sentences was not random. An attempt was made to make the sentences representative of the phonological and metrical features relevant to rhythm for both English and NSA. This was achieved by including all permissible syllable structures and all possible degrees of vowel reduction (secondary stress, function words, and schwas) within the full set of sentences for each language, and by avoiding consecutive syllables that carry primary stress. The latter step does not reflect natural speech, where two stressed syllables may follow each other, with the clash potentially resolved by assigning more prominence to one of them (Nespor and Vogel, 1989). However, because one of the objectives of the current research is to examine how L2 speakers temporally differentiate stressed and unstressed vowels, consecutive stressed syllables were avoided. In the English sentences, multisyllabic words expected to contain schwa were also included to examine how the L2 speakers produced target syllables containing schwa, and unstressed function words, in terms of degree of vowel reduction.

The total number of syllables in the English and the NSA target sentences was designed to be equal (155 syllables in each), following common practice when comparing across languages in studies of this kind (Ramus et al., 1999); although some metrics are normalized for speech rate they may not eliminate the effect of speech rate completely (White and Mattys, 2007a). Based on citation forms of the words in the sentences, the average number of syllables per sentence in each language was 15.5, ranging from 13 to 21 for NSA, and 9–17 for English. Mean sentence duration was 2.02 seconds for NSA and 2.5 for English. The difference in mean sentence duration between languages may be because NSA syllable structure is simpler than that of English (Ingham, 1994), with greater preponderance of CV syllables in NSA than in English.

3 Segmentation

Following generally accepted criteria (Peterson and Lehiste, 1960; Turk et al., 2006) (see Figure 1), all utterances were manually segmented and labelled by the first author into vocalic and intervocalic (i.e. consonantal) intervals, based on auditory impression and visual inspection of waveforms and spectrograms in Praat (Boersma and Weenink, 2010). Vocalic intervals are defined as the stretch of speech from the onset of a vowel to its offset, and consonantal intervals are defined as the stretch of speech from the offset of a vowel to the onset of the next vowel, regardless of the number of intervening consonants (Grabe and Low, 2002). The boundary between vocalic and intervocalic intervals was placed at the zero-crossing point on the waveform. Vowel–consonant boundaries were mainly delimited by the end of the pitch period preceding a break in the structure of the second vowel formant (F2) accompanied by a significant drop in the waveform amplitude; consonant–vowel boundaries were delimited by the start of a pitch period consistent with the beginning of the second vowel formant (White and Mattys, 2007a; Wiget et al., 2010).

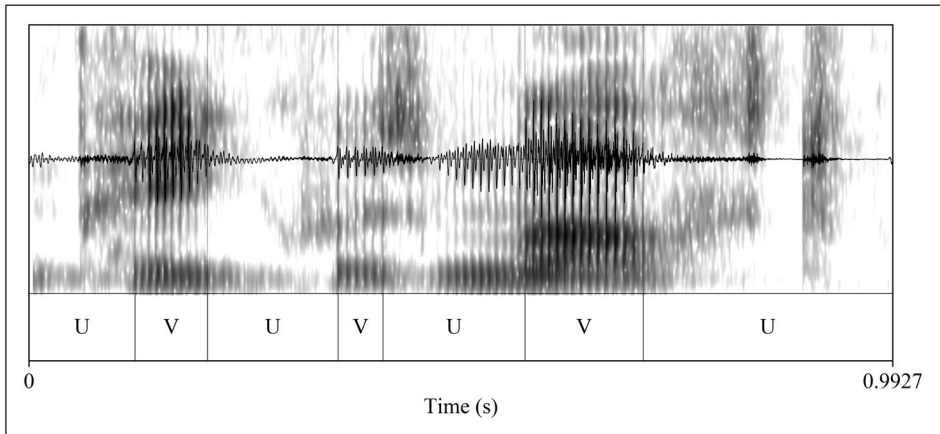


Figure 1. An illustration of the segmentation procedure, for the phrase 'peak was masked' produced by one of the Standard Southern British English (SSBE) speakers.

Additionally, stop consonants were identified by the end of a pitch period characterized by a significant drop in amplitude of the waveform and a break in the second formant. Fricative and nasal consonants were identified by the start of visible friction, and by abrupt spectral changes characterized by reduction of amplitude and spectrographic energy, respectively. Glottalized intervals, as often observed between two successive vowels (e.g. *be accepted*), were identified by changes in the pitch period such as reduction, doubling and lengthening (Dilley et al., 1996), and were labelled as consonantal intervals. Two successive vowels were labelled as one vocalic segment when there was no glottalization or pause separating them. The approach used for identifying glides and liquids followed Grabe and Low (2002), who based their judgements on acoustic, rather than phonological/phonemic, criteria. Where there were no clearly noticeable changes in the formant structure or amplitude of the signal, glides/liquids were treated as part of the vocalic interval. This strategy was also applied to segmentation of the Arabic pharyngeal /ʕ/, which has been shown to have vowel-like formant structure (Laufer, 1996). This was also deemed the best way of dealing with semivowels, since the rhythm metrics are fundamentally acoustic-based. For the same reason, any devoiced vowels or syllabic consonants were treated as (part of) intervocalic intervals.

The first consonant in all utterances was excluded from the measurements due to the sometimes extreme difficulty in demarcating its beginning. This holds particularly for stop consonants, but for consistency the exclusion was applied to all consonants. Due to possible final-syllable lengthening effects (Klatt, 1976; de Jong and Zawaydeh, 1999), final syllables were excluded from the measurements. Some previous studies (e.g. Grabe and Low, 2002; White and Mattys, 2007a) did include final syllables in their measurements. White and Mattys (2007a) argued that final-syllable lengthening may be language specific and might possibly contribute to the overall perception of rhythm. However, it was sometimes difficult to segment the final syllable, as in many cases the spectral energy decreases significantly, making it extremely hard to mark the boundaries of the

phonemes. It is also often difficult to delimit the end of utterance-final consonants (Deterding, 2001). Also, as the present research also examines durational differences between stressed and unstressed vowels, inclusion of utterance-final vowels might affect the results due to possible lengthening. Intervals of perceptible silent pauses within utterances were excluded from calculations. In the few cases where these silent pauses were preceded by a stop consonant, both the stop consonant closure and the pause silences were excluded, due to the difficulty of distinguishing the pause from the consonant closure (White and Mattys, 2007a).

4 Analysis

After segmenting all the utterances, scores for %V, ΔC , rPVI-C, VarcoV, and nPVI-V (see Table 1 above) were calculated for each sentence produced by each speaker in the four groups. A measure of articulation rate (AR) is also included in the analysis because speech rhythm has been shown to be affected by speech rate (e.g. Dellwo, 2008; Meireles and Barbosa, 2008). Following some previous studies that have investigated speech rate in L2 speech (Munro, 1995; Towell et al., 1996; Trofimovich and Baker, 2006), AR was measured by dividing the number of syllables in an utterance by the total duration of that utterance. The number of syllables for each utterance was calculated based on the number of labelled vocalic intervals in the utterance (i.e. number of syllables in an utterance is equated to number of vowels produced).

The vocalic intervals segmented in each utterance from each speaker were labelled as either stressed or unstressed. Primary stressed vowels were labelled as stressed, and all other vowels were labelled as unstressed. Categorizing vowels only as stressed or unstressed is not the only way of dividing vowels in terms of the degree of stress they bear, since vowels, in English at least, can have more than two degrees of stress, e.g. primary, secondary, and weak (Fear et al., 1995; Roach, 2009). The current study, however, took a more general view of stress, dividing vowels into stressed and unstressed only, following Ladefoged (1975) who argues for two levels of phonetic stress in English. A few vocalic intervals contained two consecutive vowels (a stressed vowel preceded by an unstressed one, as in 'the outcome' [ði aʊt.kʌm]), and in these cases the interval was labelled as stressed.

For the identification of stressed and unstressed vowels, we first checked stress placement in English in dictionaries and reference books (Cambridge Dictionary Online, n.d.; Couper-Kuhlen, 1986; Pike, 1945). All function words were considered unstressed unless they were stressed by the speaker to express contrast (Couper-Kuhlen, 1986; Pike, 1945; Roach, 2009). All monosyllabic content words were labelled as stressed. Stress assignment in polysyllabic words was checked in the Cambridge Dictionary (n.d.). This was followed by auditory and visual inspection of the waveforms and spectrograms of all the vowels in Praat (Boersma and Weenink, 2010); the expectation was that stressed vowels would have longer duration, greater intensity and higher pitch than unstressed vowels (e.g. Fear et al., 1995; Fry, 1955; Roach, 2009). This procedure proved difficult to follow for the utterances of the L2 speakers of English. In many cases, they appeared to stress function words to the same degree as monosyllabic content words, and misplaced stress in polysyllabic words. For function words, the decision was made to

consider all function words as unstressed, since one of the main aims of the current study is to find out whether the L2 speakers make a durational difference between (what are expected to be) stressed and unstressed vowels. In the case of polysyllabic words, stress was assigned to vowels based on auditory judgement combined with visual inspection of the vowels' waveforms and spectrograms.

A parallel procedure was followed to segment and label the NSA vowels. Function words were labelled as unstressed, and monosyllabic words were labelled as stressed. Stress placement in polysyllabic words is fully predictable by phonological rules in NSA. Stress falls on the final syllable if the syllable has the shape CVCC or CV:C; stress falls on the penultimate syllable if it is CVC or CV: ; otherwise, stress falls on the antepenultimate (Ingham, 1994). Resyllabification sometimes occurs across word boundaries in Arabic (Kenstowicz, 1986) (e.g. *ʔal.mo:yah ʔa.li:* → *ʔal.mo:ya.li:*). As this might affect the weight of the syllable, and thus, possibly, the placement of stress, resyllabification was taken into consideration when identifying stress in polysyllabic words.

Having segmented and labelled all the vowels, we calculated the duration of all stressed and unstressed vowels, and measured their first and second formants (F1 and F2) at the midpoint. Because of the spectral transitions in diphthongs, their formant measurements were not included in the analysis. Formant tracking errors were checked and corrected manually. Formant values were LOBANOVA transformed prior to plotting (Adank et al., 2004). All duration measurements were then normalized for articulation rate by first dividing the duration of each sentence by its number of syllables to obtain an average syllable duration for each sentence, then dividing the duration of each vowel in each sentence by the obtained average syllable duration for that sentence (Kavanagh, 2012). The normalized vowel durations were then divided by 100 to give a more readable number than the large numbers obtained. Mean durations of normalized stressed and unstressed vowels were calculated for each sentence produced by each speaker. Finally, a ratio of the mean duration of stressed vowels to the mean duration of unstressed ones was calculated for each sentence.

A scatterplot of F1 and F2 for all stressed and unstressed vowels, corresponding to the acoustic vowel space, was drawn for each speaker group to visualize the extent to which each group centralizes unstressed vowels relative to the stressed ones. The current study focuses on temporal aspects of rhythm, so no attempt was made to further quantify vowel quality reduction or centralization of unstressed vowels.

All consonant clusters (i.e. CC, CCC, CCCC) in the speech of the L2 speakers were examined auditorily, supported by visual inspection of the waveform and spectrogram in Praat (Boersma and Weenink, 2010), to examine whether the speakers produced them in a target-like way. All consonant clusters were judged to be either correct or incorrect. No attempt was made to categorize alternate realizations of consonant clusters, but the production of a cluster was deemed incorrect if a vowel was inserted to break it up (e.g. /nst/ in 'against' realized as /nɪst/), or if one of the consonants was deleted (e.g. /ksts/ in 'texts' realized as /kst/ or /kɪst/). The L2 speakers' productions of consonant clusters were compared to canonical citation forms. In other words, their production was not compared to the SSBE speakers' production in the current study, but rather to the citation or dictionary forms of how the clusters are canonically produced by L1 English speakers. Although the position of a consonant cluster might have an effect on how the L2 speakers produced

Table 2. Mean scores and standard errors (in parentheses) for rhythm metrics for each speaker group (definitions of the metrics are in Table 1).

Metric	NSA	SSBE	ME L2	LE L2
%V	39 (0.6)	32 (0.6)	40 (0.6)	38 (0.5)
VarcoV	56 (1.2)	63 (1.7)	46 (1.4)	45 (1.2)
nPVI-V	60 (1.4)	77 (2.3)	48 (1.7)	49 (1.5)
ΔC	45 (1.6)	61 (1.8)	57 (1.7)	63 (2.3)
rPVI-C	54 (2.0)	67 (2.2)	65 (1.9)	68 (2.4)
Articulation rate	6.5 (0.1)	5.6 (0.1)	5.1 (0.1)	4.9 (0.1)

Notes. LE = less experienced. ME = more experienced. NSA = Najdi Saudi Arabic. SSBE = Standard Southern British English.

them, context was not considered in the analysis. An overall percentage of target-like production of consonant clusters was calculated for each speaker.

V Results

I Rhythm metrics

Table 2 provides the mean scores and standard errors (between parentheses) for all the rhythm metrics for each speaker group. For each rhythm metric, a mixed-effects model – with the rhythm metric as a dependent variable, Speaker Group as a fixed factor, and random intercepts for speakers and utterances – was run to examine whether the speaker groups differed significantly from each other in terms of the metric scores.

For %V, the results showed a significant main effect of Speaker Group, $F(3,236)=15.07$, $p < .01$. A post-hoc test showed that only SSBE was significantly different from all other speaker groups ($p < .01$). This means that the utterances produced by the SSBE speakers had a significantly lower percentage of total vowel duration, relative to overall consonant duration, than did the utterances produced by the NSA and the L2 speaker groups. It is not clear from the metric score alone whether the lower scores for %V on the part of L1 English speakers was because they shortened unstressed vowels to a greater degree than did the other groups, or because their utterances had longer consonantal intervals than the utterances produced by the other groups. LoR did not affect the L2 speakers' scores for %V, as the two L2 speaker groups were found to have similar scores.

The results for VarcoV showed a main effect of Speaker Group, $F(3,236)=25.66$, $p < .01$. Unlike %V, post-hoc tests showed no significant difference between NSA and SSBE ($p < .7$). The L2 speaker groups differed significantly from both NSA ($p < .02$ for the difference between ME L2 and NSA, and $p < .01$ for the difference between LE L2 and NSA) and SSBE group ($p < .01$). This means that the utterances produced by the NSA and the SSBE speaker groups exhibited significantly higher durational variability between vocalic intervals than did the utterances produced by the L2 speakers. The two L2 groups had similar scores for VarcoV, which indicates no significant effect of LoR on their VarcoV scores.

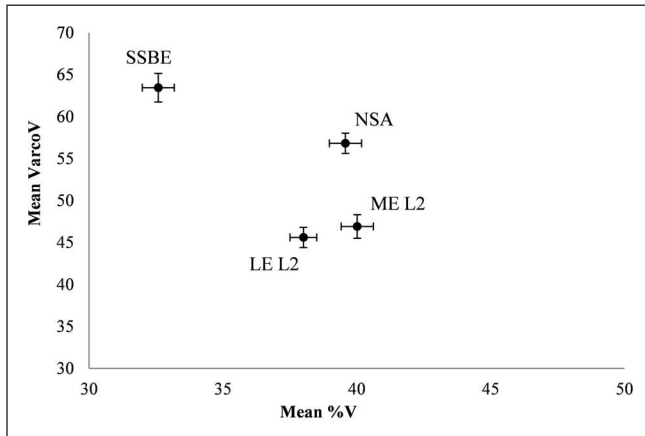


Figure 2. Scatterplot of mean scores for %V and VarcoV for all groups.

White and Mattys (2007a) suggested that %V and VarcoV are complementary and thus provide insights into the influence of L1 on L2. Figure 2 plots the average scores for %V and VarcoV for all speaker groups.

The graph shows the separation between the SSBE group, on the one hand, and all other speaker groups, on the other. For VarcoV, the NSA group appear to be intermediate between the SSBE and the L2 speaker groups. The graph also clearly illustrates the similarity of the results for the two L2 speaker groups.

For nPVI-V, the results showed a significant effect of Speaker Group, $F(3,236)=39.01$, $p < .01$. Post-hoc tests for pairwise comparisons revealed a significant difference only between the SSBE group, on the one hand, and all other speaker groups on the other ($p < .01$). The difference between the NSA and L2 speaker groups only approached significance ($p < .08$ for the difference between ME L2 and NSA, and $p < .07$ for the difference between LE L2 and NSA speaker groups). The utterances produced by the SSBE speakers displayed significantly greater durational variability between successive vowels than did the utterances produced by the NSA and the L2 speaker groups. A possible reason for this finding is that the SSBE speakers shortened unstressed vowels to a greater degree than the other groups. The results for nPVI-V showed a similar trend to those for %V, as only the SSBE group was found to differ significantly from the other groups. Unlike the scores for VarcoV, nPVI-V scores showed a significant difference between NSA and SSBE. LoR had no effect on the L2 speakers' scores for nPVI-V, as there was no significant difference between the two L2 groups.

The results of the consonantal rhythm metrics, ΔC and rPVI-C, showed significant differences only between the SSBE and L2 speaker groups, on the one hand, and the NSA group on the other (ΔC : $F(3,236)=5.98$, $p < .01$; rPVI-C: $F(3,236)=2.86$, $p = .05$). LoR did not affect the scores calculated for the L2 speakers, as there was no significant difference between the two L2 speaker groups in either measure. This suggests that the utterances produced by the L2 and SSBE speakers showed similar degrees of durational variability between consonantal intervals. However, previous studies have cast doubt on

Table 3. Means and standard deviations (in parentheses) of the durations of stressed and unstressed vowels and scores for SUR (durational ratio of stressed to unstressed vowel durations) for all speaker groups.

Speaker group	Stressed vowels	Unstressed vowels	SUR
NSA	5.64 (1.1)	2.67 (0.4)	2.14 (0.4)
SSBE	4.56 (0.8)	2.12 (0.4)	2.22 (0.5)
ME L2	4.84 (0.6)	3.19 (0.5)	1.56 (0.3)
LE L2	4.32 (0.6)	3.02 (0.3)	1.44 (0.2)

Notes. LE = less experienced. ME = more experienced. NSA = Najdi Saudi Arabic. SSBE = Standard Southern British English.

the reliability of the consonantal-based rhythm metrics, as their scores were shown to be easily affected by speech rate (e.g. Barry et al., 2003; Dellwo and Wagner, 2003; White and Mattys, 2007a).

The results for articulation rate showed a main effect of Speaker Group, $F(3,236)=16.46$, $p < .01$. Post-hoc tests showed that only NSA was significantly different from all other speaker groups ($p < .01$). The NSA speaker group spoke at a faster speaking rate than the SSBE and L2 speaker groups. This might be because NSA has simpler syllable structure than SSBE, as noted before. Previous studies have shown that languages with simple syllabic structures are usually spoken at a faster speaking rate (syllable/second) than languages with more complex syllabic structures (e.g. Dellwo, 2010). Although the L2 speakers spoke at a lower speaking rate than the SSBE speakers, the differences between the L2 and the SSBE speakers were not statistically significant ($p < .08$ for the difference between ME L2 and SSBE groups and $p < .4$ for the difference between LE L2 and SSBE groups). The difference between the L2 speaker groups was not significant, which indicates that LoR had no effect on the L2 speakers' production in terms of articulation rate.

2 Unstressed vowels

The mean durations of stressed and unstressed vowels, and the durational ratios of stressed to unstressed vowels (SUR), were calculated for all the utterances produced by all the speakers in the four speaker groups, and are reported in Table 3.

A mixed-effects model, with SUR as dependent variable, Speaker Group as a fixed factor, and random intercepts for utterances and speakers, was run to find out whether the four speaker groups differed significantly from each other in terms of SUR scores. The model showed a main effect of Speaker Group for SUR, $F(3,236)=32.41$, $p < .01$. Post-hoc tests for pair-wise comparisons showed significant differences only between the L1 speaker groups (NSA and SSBE), on the one hand, and the L2 groups, on the other (all differences were significant at $p < .01$). No significant difference was found between the NSA and the SSBE speaker groups, or between the L2 speaker groups.

The L2 speakers did not make as much durational difference between stressed and unstressed vowels as the NSA and SSBE speakers. As in the results for VarcoV, the NSA

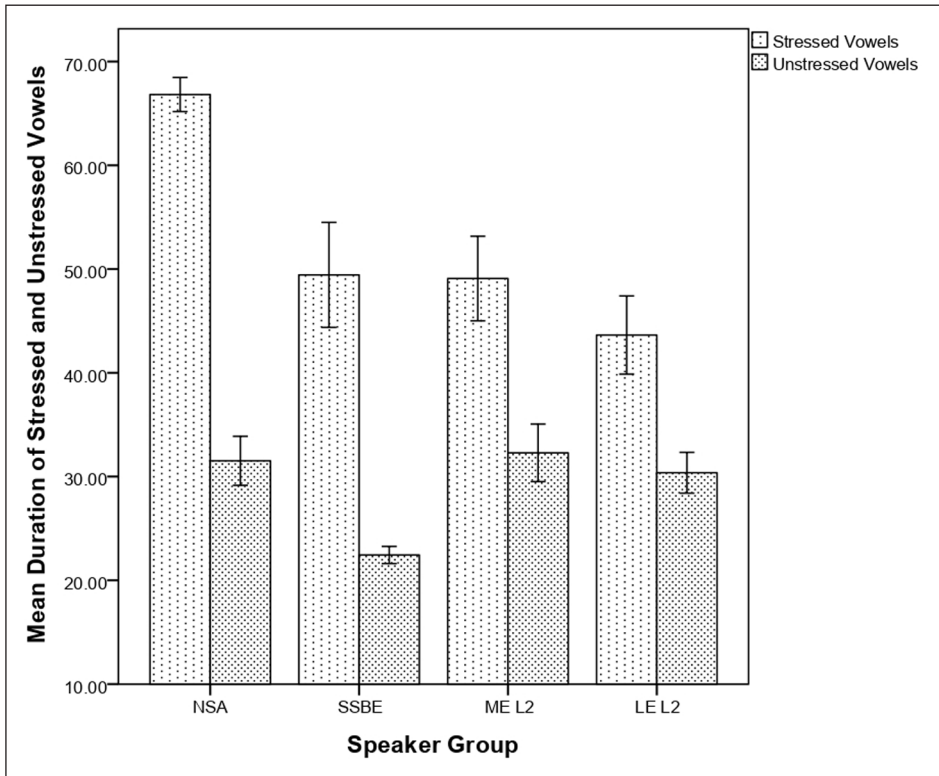


Figure 3. Mean durations of stressed and unstressed vowels for all speaker groups.

speakers had similar SUR scores to the SSBE speakers, which indicates that the speech of both groups had similar durational differences between stressed and unstressed vowels. LoR showed no effect on the results for the L2 speaker groups, as their SUR scores were not significantly different.

It is not clear from the SUR ratio measure whether the similar durational difference between stressed and unstressed vowels in the utterances produced NSA and SSBE are because both groups reduce unstressed vowels. Figure 3 illustrates durations of each type of vowel for all speaker groups.

A pair of mixed-effects models with durations of stressed vowels and unstressed vowels as separate dependent variables were run, with Speaker Group as a fixed factor and random intercepts for utterances and speakers, to find out whether the four groups differed in terms of the durations of stressed and unstressed vowels. Both models showed a significant main effect of Speaker Group, for mean durations of stressed vowels $F(3,236)=4.54, p < .01$ and for mean durations of unstressed vowels $F(3,236)=39.80, p < .01$. Post-hoc tests were run for pair-wise comparisons between the four groups. Although the NSA and the SSBE speaker groups show similar SUR scores, they differ in the mean durations of stressed and unstressed vowels independently, to a significant extent ($p = .05$ for stressed vowels and $p < .01$ for unstressed vowels), with NSA vowels

of both types longer than their SSBE counterparts. The L2 speaker groups differed significantly from the NSA and the SSBE speaker groups in terms of the mean durations of unstressed vowels ($p < .01$ for the differences between the L2 groups and SSBE, $p < .01$ for the differences between ME L2 and NSA, and $p < .05$ for the difference between LE L2 and NSA), with learners producing longer unstressed vowels than NSA and SSBE. There were no significant differences between the L2 speaker groups for mean durations of stressed and unstressed vowels.

Although the NSA and the SSBE speakers had similar durational ratios of stressed to unstressed vowels (SUR), the NSA speakers did not shorten unstressed vowels to the same degree as the SSBE speakers. This suggests that the similarity between the SSBE and NSA scores for SUR is not because the NSA speakers shortened unstressed vowels to the same degree as the SSBE speakers, but instead most likely due to the fact that vowel length is phonemically contrastive in NSA, where all long vowels have short counterparts which are about half their lengths (Alghamdi, 1998). English also has long vowels but the durational difference between short and long vowels in Arabic is larger than in English (Mitleb, 1981). In addition, the Arabic quantity sensitive stress algorithm means that the overwhelming majority of long vowels will attract stress.

The higher score for SUR in NSA can thus be caused not only by shortening of unstressed vowels, but also by the phonemic length contrast between short and long vowels. Looking at the metric results above, it seems that %V and nPVI-V (which set SSBE apart from the NSA and L2 speaker groups) can account slightly better for unstressed vowel shortening; in contrast, VarcoV (which sets SSBE and NSA apart from the two L2 speaker groups) can account better for more general temporal differences between stressed and unstressed vowels, and is robust to the fact that not all languages shorten unstressed vowels.

Since unstressed vowel durational shortening is also associated with reduction in vowel quality (see Section I.1) (e.g. Flemming, 2004; Lindblom, 1963), plotting the formant values in stressed versus unstressed vowels for each speaker group should further illustrate the difference between the SSBE group and all other speaker groups in the degree of unstressed vowel shortening. Figure 4 shows scatterplots of LOBANOVA transformed F1 and F2 in all stressed and unstressed vowels, by speaker group.

Figure 4 shows that the SSBE speaker group made a clearer spectral distinction between stressed and unstressed vowels than all other speaker groups. The unstressed vowels, relative to the stressed ones, produced by the SSBE speakers are more clustered in the F1–F2 formant space than the unstressed vowels produced by the NSA and L2 speakers. In contrast, the NSA and the L2 speaker groups showed overlapping distributions of F1/F2 values for stressed and unstressed vowels. These patterns provide further support for interpretation of %V and nPVI-V scores as sensitive to degree of unstressed vowel reduction.

3 Consonant clusters

Both L2 speaker groups showed high percentages of target-like production of consonant clusters (ME L2: $M=87.61\%$, $SD=7.3$; LE L2: $M=80.47\%$, $SD=9.1$). The ME L2 group had a higher raw percentage of target-like productions than the LE L2 group; nonetheless, an independent samples *t*-test showed no significant difference between the

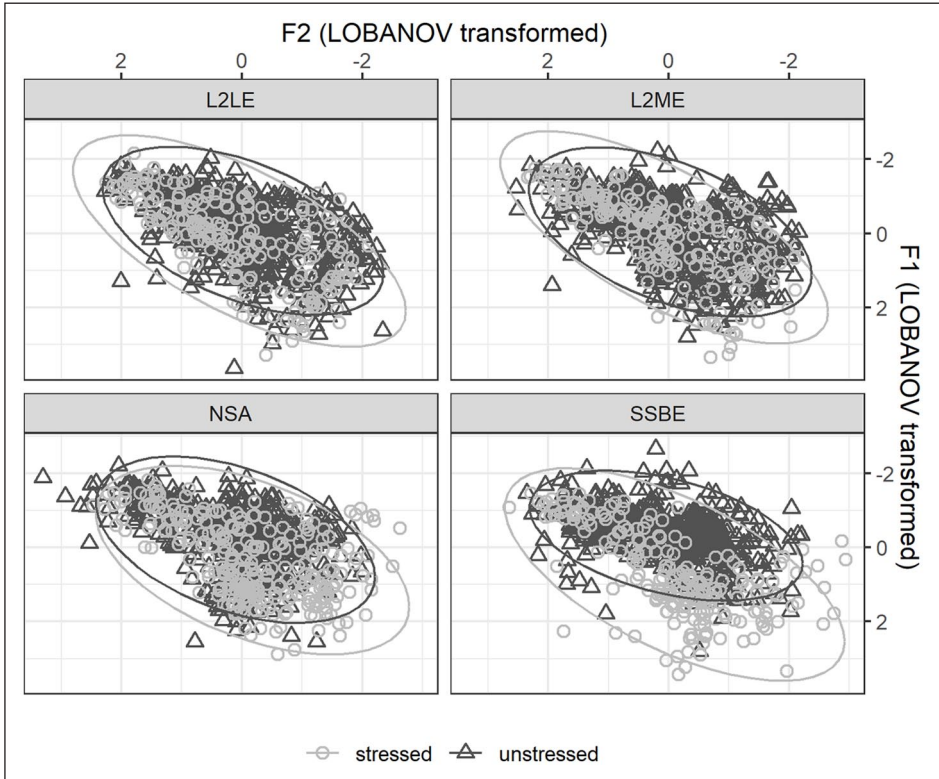


Figure 4. Scatterplot of F1 and F2 in all stressed and unstressed vowels, by speaker group.

two speaker groups, $t(10)=1.49$, $p>0.01$. To examine whether the results for the L2 groups differ significantly from a hypothesized L1 English group, with a mean of 100% correct production, a one-sample t -test was run for each L2 speaker group. Both ME L2 and LE L2 groups differed significantly from the hypothesized population, $t(5)=4.11$, $p<0.01$ and $t(5)=5.20$, $p<0.01$, respectively. The percentage of target-like production did vary considerably, however, according to the type of consonant cluster. As might be expected, the percentage decreased as complexity of consonant cluster increased.

The ME L2 speaker group had higher mean percentages of target-like productions of consonant clusters for all cluster types than the LE L2 speaker group (see Table 4). However, independent sample t -tests showed no significant difference between the two groups on any type of consonant cluster. The high percentage of target-like production consonant clusters by the L2 speakers may explain why the consonantal rhythm metrics did not show significant differences between the L1 and L2 English speakers.

VI Discussion

The current study used a range of rhythm metrics to examine the production of English speech rhythm by two groups of L2 Saudi learners: ‘more experienced’ (ME) and ‘less

Table 4. Mean and standard deviations (in parentheses) for percentage of target-like productions of English consonant cluster types by the second language (L2) speaker groups.

Speaker group	CC cluster	CCC cluster	CCCC cluster
ME L2	97.10 (2.2)	80 (20.9)	16.67 (25.8)
LE L2	92.75 (5.2)	66.67 (20.6)	8.33 (8.3)

Notes. LE = less experienced. ME = more experienced.

experienced' (LE), based on their length of residence in the UK. It also examined the speech rhythm of Najdi Saudi Arabic (NSA) and Southern Standard British English (SSBE) for comparison and to help in explaining the results. Similar to most previous L2 studies that have used the rhythm metrics (Ordin and Polyanskaya, 2014; White and Mattys, 2007b), all three vowel-based rhythm metrics used in the current study (%V, VarcoV and nPVI-V) showed significant differences between the native and non-native English speakers. Given that the vowel-based rhythm metrics were originally developed to capture the durational variability between vocalic segments arising from shortening of unstressed vowels, the initial conclusion would be that the L2 speakers did not shorten unstressed vowels to the same degree as the SSBE speakers. However, the rhythm metric results for NSA point to a more nuanced picture. While the NSA group showed significantly lower nPVI-V scores than the SSBE group, the two groups had similar VarcoV scores. This result gives further support to previous studies that have recommended use of more than one measure for studying rhythm in speech (e.g. Wiget et al., 2010).

To make more sense of the data, we analysed the duration of stressed and unstressed vowels for all the speaker groups. The durational ratio of stressed to unstressed vowels (SUR) showed similar results to VarcoV. However, a closer look at the durations of stressed and unstressed vowels independently showed that the durational variability between vocalic segments in the case of NSA must derive from another source of temporal variability, and not because of any shortening or reduction of unstressed vowels; we ascribe this result to the particular phonetic exponence in Arabic of the phonemic difference between short and long vowels. This finding was supported by the analysis of unstressed vowel quality reduction, as only the SSBE group was shown to clearly centralize unstressed vowels.

The results of the consonantal rhythm metrics showed similar results for both the L1 and L2 English speakers. This may either be due to the instability of the consonantal metrics, as shown in some previous studies (White and Mattys, 2007a; Wiget et al., 2010), or the success of the L2 learners in producing similar durational variability of consonant segments. The latter interpretation is supported by analysis of consonant cluster production by the L2 learners, the majority of which was target-like. The NSA speech exhibited less durational variability of consonantal intervals than the speech of the L1 and L2 English speakers. This result is consistent with the fact that although NSA permits CC clusters, overall it has a simpler consonantal structure than English (since NSA does not display CCC or CCCC).

The utterances produced by the L2 speakers showed similar articulation rate to those produced by the SSBE speakers. This contrasts with the results of most previous studies,

where L2 English speakers have been found to speak at a lower articulation rate than L1 English speakers (e.g. Munro and Derwing, 2001). We ascribe this positive outcome to the fact that the speech elicited from the speakers was read, and the speakers were familiar with the utterances from the paraphrasing task before being asked to read them. In contrast, the NSA utterances were spoken at a faster rate than the L1 and L2 English utterances which we also ascribe to the simpler syllable structure of NSA.

Unlike Trofimovich and Baker (2006) and Ordin and Polyanskaya (2014), our results did not show an effect of length of residence on the results for the L2 speakers. This might be due to the relatively short difference in LoR between the two groups. Trofimovich and Baker (2006) used a larger time window (3 months to 3 years to 10 years) to examine the production of English stress timing by L2 Korean speakers. Only the L2 learners with 10-years stay in the USA achieved native-like stress-timing results. Ordin Polyanskaya (2014), however, used a shorter time window than the one used in the current study (6–30 months), and still showed a significant difference in progress towards stress-timing among four L2 English learners after spending 30 months in the UK. The acquisition of English rhythm by L2 learners seems to require an extensive language experience. Mennen and de Leeuw (2014) suggest that L2 prosody, of which rhythm is a component, is an extremely difficult aspect to learn, given that languages vary not only in what prosodic structures they exhibit but also in how these structures are implemented (Mennen and de Leeuw, 2014).

Apart from the reduction of unstressed vowel quality, none of the L2 results can be explained by L1 transfer. Language universal principles can provide a plausible explanation for the fact that the L2 learners showed lower durational vocalic variability than both the native Arabic and native English speakers. Previous studies have offered a similar explanation, suggesting that progressing from stress-timing to syllable-timing rhythm is a universal language developmental path (Li and Post, 2014; Ordin and Polyanskaya, 2014, 2015). L2 speech models which focus primarily on L2 segments, such as the Revised Speech Learning Model (Flege and Bohn, 2021), can in principle be extended to account for L2 prosody but there have been few attempts to do so without further adaptation (see Mennen, 2015). van Maastricht et al. (2019) and Ordin and Polyanskaya (2015) appealed to Eckman's (1977) Markedness Differential Hypothesis to explain the universal development path in the acquisition of speech rhythm, suggesting that stress-timing rhythm is more marked than syllable-timing rhythm, and hence more difficult to acquire. Our results support Li and Post's (2014) position that L2 rhythm is multisystemic; rhythm consists of several language-specific properties, such as vowel reduction and phonotactic rules governing consonant clusters, which can be subject to different acquisition paths and processes.

VII Conclusions

The current study showed that the L2 Saudi speakers, regardless of their length of residence in the UK, displayed lower durational variability of vocalic intervals than native English and native Arabic speakers. This might be the result of a universal constraint on the acquisition of stress-timing rhythm. As expected, and perhaps because of their susceptibility to speech rate, the results of the consonantal rhythm metrics were initially

difficult to explain. The L2 speakers were found to have similar durational variability of consonantal intervals to the SSBE speakers, but this is consistent with their largely target-like production of consonant clusters.

Overall, this study provides fresh support for the recommendation to use multiple rhythm metrics in investigation of L2 speech rhythm, since the different metrics are here shown to be sensitive to different phonological parameters that contribute independently to rhythmic variation. We demonstrated how the use of simple additional measures, such as durations of stressed and unstressed vowels and evaluation of consonant cluster realization, can aid in disambiguating between competing interpretations of rhythm metric scores.

Acknowledgements

Thanks to Dominic Watt and Laurence White for their valuable comments, and to Volker Dellwo for sharing the Praat script for calculation of the rhythm metrics.

Declaration of Conflicting Interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Taif University Researchers Supporting Project number (TURSP-2020/167), Taif University, Taif, Saudi Arabia.

ORCID iDs

Ghazi Algethami  <https://orcid.org/0000-0003-0575-9529>

Sam Hellmuth  <https://orcid.org/0000-0002-0062-904X>

References

- Abercrombie D (1976) *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Adams C (1979) *English speech rhythm and the foreign learner*. Berlin: De Gruyter Mouton.
- Adank P, Smits R, and Van Hout R (2004) A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116: 3099–107.
- Algethami G, Ingram J, and Nguyen T (2011) *The interlanguage speech intelligibility benefit: The case of Arabic-accented English*. In: Levis J and LeVelle K (eds) *Proceedings of the 2nd pronunciation in second language learning and teaching conference*. Ames, IA: Iowa State University, pp. 30–42.
- Alghamdi M (1998) A spectrographic analysis of Arabic vowels: A cross-dialectal study. *Journal of King Saud University* 10: 3–24.
- Allen G (1975) Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics* 3: 75–86.
- Arvaniti A (2012) The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40: 351–73.
- Barry W, Andreeva B, Russo M, Dimitrova S and Kostadinova T (2003) Do rhythm measures tell us anything about language type? In: *Proceedings of 15th ICPHS, Barcelona*, pp. 2693–96.

- Boersma P and Weenink D (2010) *Praat: Doing phonetics by computer: Version 5.1.30* [computer program]. Available at: <http://www.praat.org> (accessed March 2010).
- Cambridge Dictionary Online (n.d). Retrieved from <https://dictionary.cambridge.org/dictionary/english/> (accessed March 2010).
- Couper-Kuhlen E (1986) *An introduction to English prosody*. London: Hodder Arnold.
- Dasher R and Bolinger D (1982) On pre-accentual lengthening. *Journal of the International Phonetic Association* 12: 58–69.
- Dauer R (1983) Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–62.
- de Jong K and Zawaydeh B (1999) Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics* 27: 3–22.
- de Jong K and Zawaydeh B (2002) Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, 30: 53–75.
- Dellwo V (2006) Rhythm and speech rate: A variation coefficient for deltaC. In: Karnowski P and Szigeti I (eds) *Language and language processing: Proceedings of the 38th Linguistic Colloquium*. London: Peter Lang, pp. 231–24.
- Dellwo V (2008) The role of speech rate in perceiving speech rhythm. In: Barbosa P, Madureira S, and C Reis (eds) *Speech Prosody 2008*. Baixas: International Speech Communication Association (ISCA), pp. 375–78.
- Dellwo V (2010) Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence. Unpublished PhD thesis, University of Bonn, Bonn, Germany.
- Dellwo V and Wagner P (2003) Relations between language rhythm and speech rate. In: *Proceedings of 15th ICPHS, Barcelona*, pp. 471–74.
- Deterding D (2001) The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics* 29: 217–30.
- Dilley L, Shattuck-Hufnagel S, and Ostendorf M (1996) Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24: 423–44.
- Eckman F (1977) Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–30.
- Fear B, Cutler A, and Butterfield S (1995) The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America* 97: 1893–904.
- Flege J and Bohn O (2021) The revised speech learning model (SLM-r). In: Wayland R (ed.) *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press, pp. 3–83.
- Flemming E (2004) Contrast and perceptual distinctiveness. In: Hayes B, Kirchner R and Steriade D (eds) *Phonetically-Based Phonology*. Cambridge: Cambridge University Press, pp. 232–76.
- Fry D (1955) Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27: 765–68.
- Ghazali S, Hamdi R, and Melissa B (2002) Speech rhythm variation in Arabic dialects. In: *Proceedings of Speech Prosody 2002*. Baixas: International Speech Communication Association (ISCA), pp. 331–34.
- Grabe E and Low EL (2002) Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven C and Warner N (eds) *Papers in laboratory phonology 7*. Berlin: Mouton de Gruyter, pp. 515–46.
- Hamdi R, Barkat-Defradas M, Ferragne E, and Pellegrino F (2004) Speech timing and rhythmic structure in Arabic dialects: A comparison of two approaches. In: *Proceedings of Interspeech-2004*. Baixas: International Speech Communication Association (ISCA), pp. 1613–16.
- Ingham B (1994) *Najdi Arabic: Central Arabian*. Amsterdam: John Benjamins.

- Jones D (1962) *An outline of English phonetics*. Cambridge: Cambridge University Press.
- Kavanagh C (2012) New consonantal acoustic parameters for forensic speaker comparison. Unpublished PhD thesis, University of York, York, UK.
- Kenstowicz M (1986) Notes on syllable structure in three Arabic dialects. *Revue Québécoise de Linguistique* 16: 101–27.
- Klatt D (1976) Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208–21.
- Knight R (2011) Assessing the temporal reliability of rhythm metrics. *Journal of the International Phonetic Association* 41: 271–81.
- Ladefoged P (1975) *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Laufer A (1996) The common [ʃ] is an approximant and not a fricative. *Journal of the International Phonetic Association* 26: 113–18.
- Lee H and Song J (2019) Evaluating Korean learners' English rhythm proficiency with measures of sentence stress. *Applied Psycholinguistics* 40: 1363–76.
- Lehiste I (1977) Isochrony reconsidered. *Journal of Phonetics* 5: 253–63.
- Li A and Post B (2014) L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition* 36: 223–55.
- Lindblom B (1963) Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35: 1773–1781.
- Low E, Grabe E, and Nolan F (2000) Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43: 377–401.
- Mehler J, Dupoux E, Nazzi T, and Dehaene-Lambertz G (1996) Coping with linguistic diversity: The infant's viewpoint. In: Morgan J and Demuth K (eds) *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum, pp. 365–88.
- Meireles A and Barbosa B (2008) Speech rate effects on speech rhythm. In: Barbosa P, Madureira S, and C Reis (eds) *Speech Prosody 2008*. Baixas: International Speech Communication Association (ISCA), pp. 327–30.
- Mennen I (2015) Beyond segments: Towards a L2 intonation learning theory. In: Delais-Roussarie E, Avanzi M, and Herment S (eds) *Prosody and language in contact*. Berlin: Springer, pp. 171–88.
- Mennen I and de Leeuw E (2014) Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition* 36: 183–94.
- Miller M (1984) On the perception of rhythm. *Journal of Phonetics* 12: 75–83.
- Mitleb F (1981) Segmental and non-segmental structure in phonetics: Evidence from foreign accent. Unpublished PhD thesis, Indiana University, Bloomington, IN, USA.
- Mok P and Dellwo V (2008) Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In: Barbosa P, Madureira S, and C Reis (eds) *Speech Prosody 2008*. Baixas: International Speech Communication Association (ISCA), pp. 423–26.
- Munro M (1995) Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17: 17–34.
- Munro M and Derwing T (2001) Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23: 451–68.
- Nazzi T, Bertoni J, and Mehler J (1998) Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24: 756–66.
- Nespor M and Vogel I (1989) On clashes and lapses. *Phonology* 6: 69–116.

- Ordin M and Polyanskaya L (2014) Development of timing patterns in first and second languages. *System* 42: 244–57.
- Ordin M and Polyanskaya L (2015) Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology* 6: 1–15.
- Peterson G and Lehiste I (1960) Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 696–703.
- Pike K (1945) *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.
- Piske T, MacKay I, and Flege J (2001) Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.
- Polyanskaya L and Ordin M (2019) The effect of speech rhythm and speaking rate on assessment of pronunciation in a second language. *Applied Psycholinguistics* 40: 795–819.
- Ramus F, Nespore M, and Mehler J (1999) Correlates of linguistic rhythm in the speech signal. *Cognition* 37: 265–92.
- Roach P (1982) On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In: Crystal D (ed.) *Linguistic controversies*. London: Edward Arnold, pp. 73–79.
- Roach P (2009) *English phonetics and phonology: A practical course*. Cambridge: Cambridge University Press.
- Stockmal V, Markus D, and Bond D (2005) Measures of native and non-native rhythm in a quantity language. *Language and Speech* 48: 55–63.
- Tajima K, Zawaydeh B and Kitahara M (1999) A comparative study of speech rhythm in Arabic, English and Japanese. In *Proceedings of the XIV ICPHS*. San Francisco, pp. 285–88.
- Taylor D (1981) Non-native speakers and the rhythm of English. *International Review of Applied Linguistics in Language Teaching* 19: 219–26.
- Towell R, Hawkins R, and Bazergui N (1996) The development of fluency in advanced learners of French. *Applied Linguistics* 17: 84–119.
- Trofimovich P and Baker W (2006) Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30.
- Turk A, Nakai S, and Sugahara M (2006) Acoustic segment durations in prosodic research: A practical guide. In: Sudhoff S, Lenertová D, Meyer R, et al. (eds) *Methods in empirical prosody research (language, context and cognition)*. Berlin: De Gruyter, pp. 1–28.
- van Maastricht L, Krahmer E, Swerts M and Prieto P (2019) Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch. *Studies in Second Language Acquisition*, 41: 87–121.
- Wenk B and Wioland F (1982) Is French really syllable-timed? *Journal of Phonetics* 10: 193–216.
- White L and Mattys S (2007a) Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35: 501–22.
- White L and Mattys S (2007b) Rhythmic typology and variation in first and second languages. In: Prieto P, Mascaró J, and Solé M (eds) *Segmental and prosodic issues in Romance phonology*. Amsterdam: John Benjamins, pp. 237–57.
- Wiget L, White L, Schuppler B, et al. (2010) How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America* 127: 1559–69.

Appendix I. English target sentences (expected output of paraphrasing task).

-
1. The manager is the person in control of the city project.
ðə 'mæ.nɪ.dʒə ɪz ðə 'pɜ:.sən ɪn kən.'tɪəʊl ɒv ðə 'sɪ.ti 'pɹɒ.dʒekt
 2. It is against the law to bet on the outcome of the elections.
ɪt ɪz ə.'geɪnst ðə lɔ: tʊ bet ɒn ði 'aʊt.kʌm ɒv ði ɪ.'lek.ʃənz
 3. Japan is one of the developed countries in the world.
dʒə.'pæn ɪz wʌn ɒv ðə dɪ.'vɛ.ləpt 'kʌn.tɪz ɪn ðə wɜ:ld
 4. One of the government's commitments is to educate people.
wʌn ɒv ðə 'ɡʌv.ən.mənts kə.'mɪt.mənts ɪz tʊ 'ed.jʊ.keɪt 'pi:..pəl
 5. The mountain peak was masked by the clouds.
ðə 'maʊn.tɪn pi:k wɜz mɑ:skt baɪ ðə klaʊdz
 6. His parents gave him a present for solving the physics exercise.
hɪz 'peə.ɹənts geɪv hɪm a 'pɹɛ.zənt fɔ 'sɒvl.vɪŋ ðə 'fɪ.zɪks 'ɛk.sə.saɪz
 7. It was fun to read all the texts included in the reading pack.
ɪt wɜz fʌn tʊ ri:d ɔl ðə teksts ɪn.'klu:..dɪd ɪn ðə 'ri:..dɪŋ pæk
 8. It is not permitted to carry hairspray into the plane.
ɪt ɪz nɒt pə.'mɪ.tɪd tʊ 'kæ..ɹɪ 'heə.spɹeɪ 'ɪntʊ ðə pleɪn
 9. The policemen used electric sticks to break up the demonstrators.
ðə pə.'li:s.mən ju:zd ɪ.'lek.trɪk stɪks tʊ brɛk ʌp ðə 'dɛm.ən.stɹeɪ.təz
 10. You should have taken a diploma to be accepted for that job.
ju ʃʊd hæv 'teɪ.kən a dɪ.'pləʊ.mə tʊ bi ək.'sep.tɪd fɔ ðæt dʒɒb
-

Appendix 2. Arabic target sentences.

-
- | | |
|----|--|
| 1 | جرحت البنت اصبعها بالسكين وهي تطبخ
'ʒra.ħat al.biħnt isʕ. 'baʕ.ħa: bis.sa. 'ki:n whi: 'tatʕ.buħ
'The girl cut herself with a knife while she was cooking' |
| 2 | قبل امس ركبت الباص من الرياض إلى جدة
'ga.biħ ʔams rɪ. 'kɪbt al.ba:sʕ min ar.rɪ. 'ja:ðʕ ʔɪ.la 'dʒɪd.dah
'I took a taxi from Riyadh to Jeddah yesterday' |
| 3 | كتبت للمدير رسالة بشأن يعطيني اجازة
kɪ. 'tabt lɪl.mu. 'di:r rɪ. 'sa:.lah ʕa. 'ʃa:n jɪʕ. 'tʕi:.ni: ʔɪ. 'ʒa:.zah
'I wrote a letter to the manager to be given a vacation' |
| 4 | المدير هو اللي قال لا احد يطلع من الطلاب
ʔal.mu. 'di:r hu: 'a.li ga:l la: 'ʔa.ħad 'jatʕ.laʕ min atʕ.tʕu. 'la:b
'The school principal is the one who said no student is allowed to leave' |
| 5 | لقت المعلومات منشورة في كتاب قديم
lɪ. 'ge:t al.maʕ.lu:. 'ma:t man. 'ʃu:.rah fi: kɪ. 'ta:b gɪ. 'di:m
'I found the information published in an old book' |
| 6 | سكن أحمد في فندق قريب من الجامعة
'sɪ.kan 'ʔaħ.mad fi: 'fun.duħ gɪ. 'ri:b min al. 'dʒa:.mɪ.ʕah
'Ahmed has lived in a hotel near the university' |
| 7 | شربت الموية اللي كانت على الطاولة
ʃɪ. 'rɪbt al. 'mo:.jah 'a.li 'ka:.nat 'ʕa.la atʕ. 'tʕa:w.lah
'I drank the water which was on the table' |
| 8 | قابلت وليد بالصدفة في مطار الدمام
ga:. 'balt wa. 'li:d bisʕ. 'sʕud.fah fi: ma. 'tʕa:r ad.dam. 'ma:m
'I met Waleed by chance in Dammam Airport' |
| 9 | نسبة القبول في الجامعة كانت مرتفعة
'nis.bat al.ga. 'bu:l fi: al. 'dʒa:.mɪ.ʕah 'ka:.nat mɪr. 'taf.ʕah
'The percentage of admission to the university was high' |
| 10 | ظاهرة الكتابة على الجدران منتشرة في كل العالم
'ðʕa:.ħɪ.rat al.kɪ. 'ta:.bah 'ʕa.la al.ʒɪd. 'ra:n mun. 'ta.ʃɪ.rah fi: kul al. 'ʕa:.lam
'Writing on walls is a world-wide phenomenon' |
-