The effect of correcting temporal patterns on the intelligibility of non-native Dutch speech

Lisette van Delft

(3224422)

Supervisor: dr. Hugo Quené

Article submitted as a final thesis for the degree of Master of Science

Clinical Language Speech and Hearing Sciences

Utrecht University

26 June 2009

Utrecht University

Abstract

This study examines the contribution of the temporal pattern of speech to the intelligibility of foreign accented utterances. Simple Dutch sentences, spoken by a Polish speaker of Dutch as a second language (D2), were artificially modified so as to match the temporal pattern of the same sentences of a native Dutch speaker (D1), and vice versa. Intelligibility was measured by a perceptual experiment according to the Speech Reception Threshold method (SRT, Plomp and Mimpen, 1979), in which native Dutch listeners repeated D1 and D2 sentences with an original and a modified temporal pattern. Results suggested that contrasts in intelligibility of native and non-native speech are partially explained by the temporal pattern. Additionally, the effect size of correcting temporal patterns of non-native speech presumably is related to other characteristics of speech, such as segmental errors.

Keywords

Intelligibility, non-native speech, temporal pattern, Speech Reception Threshold.

1. Introduction

The last few decades have brought worldwide increases in migration, telecommunication and global travel. As a consequence, conversations between interlocutors with different language backgrounds, in which at least one person uses a (foreign accented) second language (L2), become more and more common. Foreign accented speech could have consequences for communication, such as diminished speech intelligibility and misunderstandings (e.g. van Wijngaarden, 2001, 2003; Munro, 2008). It even can put the L2 speaker at social or professional disadvantage (e.g. Morley, 1991; Munro et al., 2006; Munro and Derwing, 1999;). The first concern, speech intelligibility, plays an important role in communicative effectiveness. Since communicating effectively is the main purpose of L2 teaching, rather than sounding accent free (Morely, 1991), it would be useful to assess the magnitude of certain acoustic-phonetic aspects of L2 speech that might affect speech intelligibility. Several domains could benefit from this knowledge, including L2 teaching, intelligibility improvement programs for individuals with affected speech (due to e.g. hearing impairment or dysarthria), development of synthetic speech devices, and theoretical reasoning.

In the present study it is attempted to investigate the effect of improving the temporal pattern of L2 speech on the intelligibility. The intelligibility of temporally (un)modified sentences spoken by a native (D1) and a non-native (D2) speaker of Dutch was examined by carrying out an intelligibility experiment with native Dutch listeners using the Speech Reception Threshold method (SRT, Plomp and Mimpen, 1979). The SRT method is assumed to be more appropriate than other methods used for similar purposes (see 1.5.).

1.1. Intelligibility of L2 speech

The perception of L2 speech has been investigated in various ways. Among previous investigations there is some inconsistency in the usage of terms for assessing accurate speech perception, such as acceptability, comprehensibility, intelligibility, fluency and degree of foreign accent. Although some of these terms are related to speech intelligibility, they are partially independent dimensions of L2 speech (Derwing and Munro, 1997; Munro and Derwing, 1999). In the current study the definition by Munro and Derwing (1999) for speech intelligibility is used, in which ‘intelligibility may be broadly defined as the extent to which a speaker’s message is actually understood by a listener’.

The degree to which L2 speech is intelligible depends partially on contextual factors, such as environmental noise or the pragmatic context of the message, as well as on listener characteristics, such as the familiarity with L2 speech or the listeners’ cognitive capacity to process a (complex) message (e.g. Bent and Bradlow, 2003; Munro, 2008; Munro et al., 2006; van Wijngaarden, 2003). Also various stimulus properties in L2 speech can affect speech intelligibility. Segmental errors, both at phonemic and allophonic/ sub-phonemic level, can disturb communication (e.g. Anderson-Hsieh et al., 1992; Derwing and Munro, 1997; Smith, 2004). At the same time, suprasegmental factors including lexical stress, intonation, and speech rhythm can help a listener to comprehend an utterance (e.g. Adams, 1979; Anderson-Hsieh et al., 1992; Chun, 2002; Holm, 2007). Furthermore, whether or not in combination with segmental and suprasegmental features, syntactic structure, gestures, visual cues, speaking rate, redundancy of the message, and voice quality could affect the intelligibility of L2 speech (e.g. Morley, 1991; Munro and Derwing, 2001; van Wijngaarden, 2003; Varonis and Gass, 1982).

Holm (2008, pp. 16-26), Neri et al. (2006) and Rajadurai (2007) discussed several investigations which have attempted to isolate and establish the (relative) contribution of certain phonetic-acoustic aspects on the intelligibility of L2 speech. In these reports it is clearly shown there is a lack of consistency in findings, which may be attributed to differences in definitions, methodologies, and samples used as well as divergent variables investigated. Consequently, results of studies assessing the (relative) effect of certain stimulus properties on the intelligibility of L2 speech are hard to compare, and are likely to be considered as explorative, rather than conclusive.

Despite the presence of conflicting results, it is now generally accepted for both teaching and research purposes that suprasegmental aspects (prosody), may have more impact on the speech intelligibility than segmental aspects (e.g. Anderson-Hsieh et al., 1992; Derwing and Munro, 1997; Derwing and Munro, 2005). Since prosody is partially defined by temporal features (Lehiste, 1977), it would be interesting to investigate how the improvement of temporal patterns enhance the intelligibly of L2 speech. This question suggests the idea that temporal patterns could carry a function which might be important to speech intelligibility. Additionally, it assumes the presence of temporal variation among languages, which consequently might affect the acquisition of temporal patterns in L2-speech. Evidence for these assumptions will be described in the following sections.

1.2. Linguistic function of temporal patterns

The temporal pattern of speech is here considered as the relative duration of speech segments (or phonemes) in an utterance. Segmental durations are affected by a variety of factors in speech production, and affect in divergent ways the perception of speech (see Klatt, 1976; Nooteboom, 1997; van Santen, 1992 for an overview). Although speakers differ with regard to the production of relative segmental durations, several perceptual studies have demonstrated that durational cues, whether or not interacting with other acoustic cues such as pitch, spectral envelope and/or amplitude, might carry linguistic information. For example, segmental timing may serve as a cue to the phonetic identity of many segment types (e.g. Nooteboom, 1973; Cole and Cooper, 1975), it might reflect the position of the phoneme in a word or syllable (e.g., Klatt, 1974; Oller, 1973; Umeda, 1977) and it could play a role in the detection of word boundaries (e.g. Shatzman and McQueen, 2006; Quené, 1992). Those three factors (phoneme identity, phoneme position and word boundary) may contribute to word recognition in spoken language. The durational increase of phrase-final segments (e.g. Klatt, 1975; van Santen, 1992) and the occurrence of pauses (e.g. Zellner, 1994) play a role in the perception of phrasal structure, and may facilitate the listener's recovery of the meaning of an utterance. Also, lengthening of segments is one of the primary cues to the representation of lexical and phrasal stress and/or emphasis (e.g. Chun, 2002; Eefting, 1991; Sluijter, 1995). The alternation of stressed (prolonged) and unstressed (unaltered or shortened) syllables contribute to the rhythmic organization of speech, which plays an important role in the predictability and the segmentation of words in connected speech (Patel, 2008).

This brief overview of factors which may be responsible for the variation in the relative duration of speech segments indicates that the temporal pattern in speech has potential for carrying considerable linguistic information at the phonetic and prosodic level. Presumably this facilitates the lexical-semantic and syntactic interpretation of utterances and might contribute to the intelligibility of speech.

1.3. Cross-linguistic temporal variation in L1 and L2 speech

Cross-linguistic temporal variation is a consequence of language-specific phonotactic constraints, and has been widely described at different levels of analysis (segmental-, syllable- or supra-segmental level). For instance, at the segmental level some languages (e.g. Finnish), have a phonological contrast between long and short vowels with similar spectral properties, while others (e.g. Spanish) have not (Zampini, 2008). At the syllable level, increased vowel duration before voiced versus voiceless word-final consonants is seen in a number of languages (e.g. English), but the magnitude of this contrast differs across languages, and even some languages (e.g. Arabic) do not appear to exhibit it at all (e.g. Flege and Port, 1981; Laeufer, 1992).

Since L2 learners tend to apply phonological knowledge of their L1 when speaking a L2, it is assumed that cross-linguistic temporal variation causes potential difficulty in the acquisition of temporal features in L2. Results from several cross-linguistic investigations confirm this by reporting that temporal features in L2 speech, such as vowel duration, stress distribution and durational variation in vocalic and consonantal intervals, are different from native norms (e.g. Adams, 1979; Archibald, 1997; Bent et al., 2008; Flege, 1993; Flege et al., 1992; Gut, 2003; Rasier, 2006; White and Mattys, 2007). For example, at the segmental level it was found that vowels preceding voiced consonants were significantly shorter when produced by Chinese-Mandarin speakers of English (Bent et al., 2008). At the syllabic level is was argued that the duration of syllables in a spoken L2 can be affected by simplification (i.e.

deletion of phonemes) and/or epenthesis (i.e. inserting phonemes) in syllables, which is partly explained by constraints for syllable structure in the speaker's L1 (Young-Scholten and Archibald, 2000). At the supra-segmental level Adams (1979) asserted that the non-native speech rhythm in L2 learners of English was due to a variety of factors, such as insufficient durational difference between unstressed and stressed syllables, inappropriate pauses, and misplaced stress.

1.4. Perception of non-native temporal patterns in L2 speech

Several perceptual investigations suggest that both native and non-native listeners are sensitive to non-native temporal properties in speech, regardless of the more salient temporal influences of non-nativeness, such as hesitations, pauses and the amount of time spent in formulating sentences (e.g. Flege, 1993; Smith et al., 2003; White and Mattys, 2007). However, these investigations do not provide an answer to the question whether a non-native temporal pattern in L2 speech may cause diminished speech intelligibility, and hence, hamper communicative effectiveness. Few intelligibility-studies attempted to solve this issue.

Tajima et al. (1997) tried to quantify the effect of correcting the temporal pattern on the intelligibility of English sentences spoken by a L2 speaker with a Chinese-Mandarin L1. While retaining the spectral and source characteristics, sentences of a native and a non-native speaker of English were modified by aligning the duration of acoustic segments with tokens of the same phrase spoken by the other speaker. Intelligibility was measured by the percentage of correct responses of native English listeners' performance on a forced-choice identification test with four alternatives (one correct, three phonetically similar). Results showed that intelligibility of the native productions was high (94%), but declined significantly (to 83%) after temporal distortion. Performance of the unmodified foreign accented phrases was poor (39% correct), but improved significantly (to 58%) after temporal correction. Unfortunately, these results have to be interpreted with caution because several factors could have caused bias within the data. For example: the quality of some test phrases was distorted after temporal manipulation, there was an uneven phonetic divergence of the three distracting phrases from the correct response, and the small number of target stimuli (only 11 phrases) was presented three times which could have caused learning effects in the listeners. Furthermore, the speech materials were poorly comparable since the speakers varied in speaking rate, intonation pattern and segmental quality. The authors concluded that this study had raised more questions than it had answered. They also suggested that the intelligibility of L2 speakers may be enhanced if explicit training is provided on temporal properties of L2 speech.

Holm (2007, 2008) investigated the relative contribution of intonation and duration to the intelligibility of Norwegian as a L2. Norwegian sentences of a L1 speaker and seven pairs of L2 speakers with different L1 backgrounds were manipulated. Durational manipulation of the L2 sentences was obtained by lengthening and shortening each phoneme duration to match the duration of the corresponding phoneme in the same L1 sentence. Speech intelligibility was measured by the percentage of correctly perceived words per sentence, written down by native listeners. Results indicated that durational manipulation enhanced intelligibility in L2 speakers with a French, Tamil and Persian L1. In the other L2 groups (L1 was Russian, English, Chinese, German) the intelligibility did not increase after temporal correction. Results suggested that the relative importance of durational (and intonational) factors in intelligibility might be dependent on the L1 of the non-native speaker. Also this study suffered from several limitations, which created difficulty in interpreting Holm's findings as firm conclusions. For instance, manipulations led to speech signal degradation, the level of Norwegian ability varied among the L2 speakers, only three sentences per L2 speaker were tested, and the speakers varied in both intonation pattern and segmental quality.

Another investigation in the intelligibility of temporally modified sentences was performed by Maassen and Povel (1984). In their study deaf children's speech was temporally corrected in six different ways, including a manipulation in which each phoneme in the deaf child's utterance was given the same relative duration as the corresponding phoneme in the norm sentence. Results demonstrated that most sentences showed the largest gain in intelligibility following a phonemic relative correction. However, the different corrections did not affect the intelligibility of all sentences in the same way. Some sentences profited most from a syllabic relative or phonemic absolute correction.

The investigation of Bent et al. (2008) examined native and non-native listener identification of English words that substantially differed in temporal pattern (i.e. relative vowel length before voiced versus voiceless obstruents) across L1 and L2 speakers. Results showed that contrasts in native and non-native intelligibility might be partially explained by temporal pattern differences in vowel duration, although other cues such as presence of stop releases and burst duration may also contribute.

1.5. Assessment of intelligibility: Speech Reception Threshold method (SRT)

The different methods used in the studies discussed above already indicate that there is no universally accepted way of assessing the intelligibility of speech. Each method has its own advantages and limitations, which depend on the questions and objectives of the study (see Atechi, 2004; Munro, 2008; Rajadurai, 2007 for a discussion of several of methods).

The experimental procedure used in the present study was based on the Speech Reception Threshold method (SRT, Plomp and Mimpen, 1979). This method gives a measure (SRT in dB) for sentence intelligibility in noise, corresponding to the signal-to-noise ratio (SNR) that gives 50% correct responses of a list of short, redundant, everyday sentences. Each list of 13 sentences yields one SRT-score. A lower SRT score means a higher intelligibility, because more noise is allowed to reach 50% correct perception of the sentences.

The SRT-method has several benefits in comparison to other methods which are mentioned in section 1.3. First, the SRT-method measures the perception of speech in noise. As the effects of non-nativeness on intelligibility are expected to be especially apparent with degraded speech due to noisy circumstances, SNRs should be an appropriate estimation of assessing speech intelligibility (Munro, 1998; Plomp, 1986). Second, the SRT measures a threshold based on the joint intelligibility of 10 sentences. Many investigators confine themselves to measure intelligibility scores for a few utterances in only one or a few SNRs per condition, so that it is not possible to derive a SRT. Only by using the SRT results can be expressed in a way that can be reliably compared from one condition to another (Plomp, 1986). Third, the SRT-method measures intelligibility at sentence level, this makes generalizations to conversational speech easier than experiments in which the intelligibility of words or syllables is assessed. Moreover, manipulation of temporal pattern might be more salient at sentence level than at lower levels of analysis.

1.5. Present study

Previous findings suggest that the intelligibility of (L2) speech might be enhanced when the temporal pattern is improved at both sentence and word level. However, as discussed above, evidence is not strongly convincing because of several methodological limitations which could have biased the outcomes. In order to supplement previous findings, the present study is performed in which the SRT-method of Plomp and Mimpen (1979) is used to examine speech intelligibility.

The primary purpose of this study was to quantify the effect of improving the temporal pattern in L2 speech on intelligibility. Secondly, it aimed to gain insight into the intelligibility-increasing and -decreasing factors associated with the temporal pattern of speech. The primarily question was whether non-native speech shows improved intelligibility when it is modified to resemble the temporal pattern of native speech. This questing was attempted to answer by testing the three hypotheses. The first hypothesis can be regarded as the main hypothesis. The latter two are more specific hypotheses and are derived from the first one.

1) Speech with a native temporal pattern has a higher intelligibility than speech with a non-native temporal pattern.

2) The intelligibility of sentences spoken by a D2 speaker increases after temporal correction with the native temporal pattern of the corresponding sentences spoken by a D1 speaker.

3) The intelligibility of sentences spoken by a D1 speaker decreases after temporal distortion with the non-native temporal pattern of corresponding sentences spoken by a D2 speaker.

These hypotheses were tested by carrying out an intelligibility experiment in which temporally modified and temporally unmodified Dutch utterances spoken by a D1 and a D2 speaker were presented to native Dutch listeners. The temporal pattern manipulation consisted of exchanging the relative phoneme-duration of both speakers in order to resemble the temporal patterns of the other one. By doing this, it was assumed that temporal modification affects temporal features at the segmental level, as well as the temporal pattern at higher levels such as syllable duration and speech rhythm.

2. Methods

2.1. Materials: original sentences and noise

The original Dutch SRT-method (Plomp and Mimpen, 1979) consists of 10 carefully selected lists of 13 Dutch sentences each, which describe everyday common situations in simple wording (see appendix). The sentences comprise eight or nine syllables each. Only sentences with approximately equal chances of correct recognition in noise are added to the SRT-method and as far as possible equal numbers of the various phonemes were divided into the 10 lists (Plomp and Mimpen, 1979).

Of each sentence two original versions were used, which were artificially manipulated for the experiment (see 2.2). One original version was spoken by a D1 speaker and one original version was spoken by a D2 speaker. The D1 sentences were spoken in very clear style by a female native speaker of standard Dutch, recorded by Plomp and Mimpen (1979). The D2 sentences were spoken by a female non-native speaker of Dutch, with Polish as her native tongue. The D2 samples were recorded by van Wijngaarden (2001, 2003). The temporal and intonation pattern in the D2 speech differed from the D1 speech and marked a foreign accent. The D2 speech was fluent and phonemes were pronounced near-native. This was assessed by the author, a phonetician and an experienced speech-language pathologist.

A potential problem in the test would have been that the D2 speech was never be able to yield 100% intelligibility due to its non-native accent. To be sure this was not the case, the (near) 100% intelligibility of the D2 sentences was verified in native Dutch listeners at a very benign SNR of +15 dB (van Wijngaarden, 2001). In other words, the validity of the SRT results could not be degraded by the foreign accent, because the intelligibility of the 'clear' D2 speech was not approximating the 50% sentence intelligibility threshold.

The masking noise spectra were taken from the average spectra of the presented lists. By combining each specific list with its own matching noise spectrum, the lists were masked relatively equal.

2.2. Manipulation of the test sentences

In the experiment the lists were presented in a particular condition in which 'temporal pattern' (native= t1 or non-native= t2) and 'speaker' (native= D1 or non-native= D2) were considered as two variables. This resulted in four different test conditions (see table 1).

Table 1. *Ten lists consisting of 13 sentences each, were manipulated into four conditions. In condition D1t1 and D2t2 the temporal pattern of the sentences was unmodified; in condition D1t2 and D2t1 the temporal pattern of the sentences was modified.*

Temporal Pattern	**Speaker**	
	D1	**D2**
t1	**condition D1t1** Native speaker of Dutch with a native Dutch temporal pattern	**condition D2t1** Non-native speaker of Dutch with a native Dutch, modified temporal pattern
t2	**condition D1t2** Native speaker of Dutch with a non-native, modified temporal pattern	**condition D2t2** Non-native speaker of Dutch with a non-native temporal pattern

Artificial manipulation of the sentences was performed in Praat version 5.0.43 (Boersma and Weenink, 2008), and consisted of four steps: 1) segmenting the sentences into phonemes; 2) equalizing speaking rate; 3) manipulation of temporal pattern; 4) equalizing pitch contour. Step 1, 2 and 4 were accomplished for each condition. Step 3 was only required for the conditions with a temporal pattern manipulation, i.e. condition D2t1 and D1t2.

Step 1: The D1 and D2 version of each sentence were segmented into the same amount of intervals by visual analysis of the waveform and spectrogram, and by auditory perception of the signal. The interval borders were determined by the phoneme borders, for example '/s/t/r/e/p/' in the Dutch word 'streep' (/strep/, stripe). Also pauses between words were considered as an interval. Due to co-articulation the phonemes showed smooth transitions, in a few cases this impeded to place a border between two segments. For example, in the Dutch word 'nieuws' (/nius/, news) it was not possible segment the phonemes /i/ and /u/. In those cases two phonemes were taken together in one interval.

Step 2: The speaking rate was equalized among the speakers, in order to prevent this feature from influencing the outcomes. This was accomplished by multiplying each whole sentence, without initial and final pauses, by a factor which is the average duration of the D1 and D2 version divided by the duration of the sentence which was intended to modify. After doing this, the resulting duration of the modified sentence was intermediate between the D1 and the D2 version, while the temporal pattern of the sentence was preserved (i.e. the relative duration of the intervals was unaltered).

Step 3. For manipulation of the temporal pattern, each interval of a sentence was shortened or lengthened to match the duration of the corresponding interval in the other speaker-version of a sentence. In a few cases the difference in interval length between the D1 and D2 version exceeded a factor of two. This yielded a chirp in the signal which could degrade the sound quality and could bias the intelligibility scores. To counteract this, the durational manipulation of these intervals was adjusted manually into a near-natural sounding, temporally manipulated interval without undesirable acoustic distractions.

Step 4: The pitch contours (intonation patterns) of the D1 sentences were stylized and used to replace the pitch contours of the matching D2 sentences. This means, the intonation pattern was equalized among the speakers, and potential contrasts in intelligibility caused by various pitch contours among the speakers were eliminated. The automatic software occasionally made errors in finding the right pitch values (e.g. wrong octave, groundless pitch values), which could influence the intelligibility of a sentence. In these cases the stylized pitch contour was manually adjusted into a smooth native Dutch-sounding intonation pattern.

Finally, all test sentences were set at a mean intensity level of 70 dB and a sampling rate of 44.1 kHz.

2.3. Test procedure and data collection

During the experiment, the participants were seated in a quiet room wearing closed headphones in which the test sentences in noise were presented binaurally. The experimenter, who was seated next to the participant, watched a computer screen (which was shielded from the participant) displaying the target sentence and a correct/ incorrect button. Each participant was instructed to repeat sufficiently loud and well pronounced the Dutch sentences as (s)he perceived. The experimenter decided whether the response was correct or not by clicking on the correct/incorrect button. A correct reproduction of the entire sentence in the right word order was a condition for a correct score. When the experimenter doubted about the correctness of a reaction, for example when the subject pronounced unintelligible, the participant was asked to repeat the response.

Before the actual test session started, listeners accomplished a practice session with list I and II (see appendix) in condition D1t1 and D2t2 respectively. The other eight lists (III to X) were used for data collection, in which each test condition was represented in two lists. The combination of list and test condition was counterbalanced among the participants. This means, the lists were prevented from unequal distribution among the conditions in the experiment as a whole. The order of testing the lists was assigned at random, and each list was presented once to a participant.

The SNR in which the sentences were presented varied according to the simple up-and-down adaptive procedure developed by Plomp and Mimpen (1979). The first sentence of a list was repeated until the listener reproduced the entire sentence correctly, with an increasing SNR (starting at SNR -15 dB, step size +4 dB in the first and second presentation, subsequently step size +2 dB). This was done to quickly approximate the 50% intelligibility threshold. Then, the other 12 sentences of the list were presented successively only once. When a sentence was repeated correctly, the SNR of the next sentence was decreased by 2 dB; when the response was incorrect, the SNR of the next sentence was increased by 2 dB. Thus, the response to a sentence determined in which SNR the next sentence was presented. This procedure was repeated for each presented list. Data collection and adaptation of the SNR was performed automatically after a mouse click on the correct/incorrect button by the experimenter.

The 50% intelligibility threshold (SRT) was calculated by taking the average over the SNRs which resulted from the responses to last 10 sentences of a list. The SNRs of the first three sentences of a list were not taken into the calculation. With these three sentences the listener could get used to the condition in which the stimuli were presented, which made the approximation of the 50% intelligibility threshold more reliable.

Before the experiment was carried out, a pilot session with two participants was accomplished. Results from this pilot session required no adaptations of the test procedure.

2.4. Participants

A total of 41 native Dutch listeners (3 male, 38 female) participated in the experiment, all of whom had self-reported normal hearing and no speech-language disorders (e.g. dyslexia or stuttering). The participants were aged between 18 and 28 years. They were recruited from a pool of volunteer participants of the Utrecht University and received a remuneration of €5.

3. Results

Data were collected from all the 41 participants who listened each to eight lists equally divided over four conditions. For each condition, 41*(8/4)= 82 SRTs were available. The SRTs were approximately normal distributed among the conditions. The observed mean SRT and SD of all participants in the four tested conditions are shown in the middle column of table 2.

The D1 speaker with a native temporal pattern (D1t1) was most intelligible, mean SRT (SD)= -3.122 (1.066) dB, the D2 speaker with a non-native temporal pattern (D2t2) was least intelligible, mean SRT (SD)= 0.717 (1.7416) dB. The mean SRTs of sentences in the temporally manipulated conditions (condition D2t1 and D1t2) were in between.

Table 2. *Middle column: observed mean SRT (SD) in dB for each condition. Right column: regression coefficients (β) in dB and standard errors (SE) for each condition, after both crossed-random factors 'list' and 'participant' were independently taken into account. Note that the lower the SRT, the higher is the intelligibility of the sentences in a certain condition.*

Condition	Mean SRT (SD) in dB	β in dB (SE)
D1t1	-3.122 (1.066)	-3.133 (0.190)
D2t1	0.337 (1.435)	0.348 (0.221)
D1t2	-2.063 (1.264)	-2.062 (0.201)
D2t2	0.717 (1.742)	0.716 (0.241)

The data were statistically analyzed by performing a mixed effects analysis in statistical package R, version 2.8.1(R Development Core team, 2008) and extension package 1me4 (Bates, 2005). The analysis was done within 'participants' and within 'lists', which were both considered as cross-random factors. The 'SRT' was the dependent variable and 'condition' was considered as a fixed factor. The mixed effects analysis with crossed-random effects was chosen for analyzing data because in this analysis the random effects of 'list' and 'participant' were not nested (as in a repeated measures ANOVA) but crossed. This means it is allowed to generalize effects across both crossed-random factors, which is an advantage over other within-subjects-analyses (Quené and van den Bergh, 2008). Backgrounds, computational models and other benefits of the mixed effects analysis are described more extensively in Quené and van den Bergh (2004, 2008).

To perform the mixed effects analysis the regression coefficient (β) and the standard error (SE) for each condition were computed, which are shown in the right column of table 2. The regression coefficient can be regarded here as the estimate of the mean SRT in a certain condition, in which the crossed-random factors 'list' and 'participant' were independently taken into account. The standard error (SE) is a measure for the variability of β. As well as the mean SRTs, the regression coefficients in table 2 indicate that sentences in condition D1t1 have highest intelligibility, and sentences in condition D2t2 have the lowest intelligibility.

The first hypothesis stated that speech with native temporal pattern (t1) has a higher intelligibility than speech with a non-native temporal pattern (t2). This hypothesis it was tested by examining whether a combination of condition D1t1 and D2t1 with observed mean SRT (SD)= -2.593 (1.281) dB significantly contrasted from a combination of condition D1t2 and D2t2 with observed mean (SD) SRT= 0.527 (1.602) dB. Mixed effects analysis using the χ^2 test for evaluating contrasts, showed a significantly lower SRT (higher intelligibility) for speech with t1, χ^2= 408.087, p<.001. Apparently, speech with a native temporal pattern was more intelligible than speech with a non-native temporal pattern when the effect of speaker was excluded from the analysis.

Additionally, the contrast in intelligibility between the D1 and D2 speaker was analyzed. Specifically, it was tested whether the SRT in the combination of condition D1t1 and D1t2 with observed mean SRT (SD)= −1.393 (4.597) dB significantly contrasted from the SRT in the combination of condition D2t1 and D2t2 with observed mean SRT (SD)= − 0.673 (2.061) dB. Mixed effects analysis using the χ^2 test for evaluating contrasts, showed a significantly lower SRT score (higher intelligibility) for D1 speech, χ^2= 22.142, p<.001. This indicates that the native speaker was more intelligible than the non-native speaker when the effect of temporal pattern was excluded from the analysis.

The second hypothesis stated that the intelligibility of the D2 sentences (condition D2t2), increases after temporal correction with the native temporal pattern of the corresponding D1 sentences (condition D2t1). Mixed effects analysis using the χ^2 test for evaluating contrasts did not show a significant contrast between the SRTs in condition D2t2 and D2t1, χ^2= 2.491, p= 0.1145 and d= 0.25. Although a small effect size (d= 0.25) is found, the intelligibility of D2 speech did not significantly increase after improvement of temporal pattern.

The third hypotheses stated that intelligibility of D1 sentences (condition D1t1) decreases after temporal pattern distortion with the non-native temporal pattern of the corresponding D2 sentences (condition D1t2). Mixed effects analysis using the χ^2 test for evaluating contrasts showed a significant contrast between the SRTs in condition D1t1 and D1t2, $\chi^2 = 48.124$, $p < .001$ and $d = -1.09$. This means, distortion of the temporal pattern in the D1 sentences leaded to a higher SRT, and hence to a decreased intelligibility.

5. Discussion and Conclusion

The results above show that the temporal pattern of spoken sentences contributes to the intelligibility of speech. Sentences of both the D1 and the D2 speaker with a native Dutch temporal pattern had a significantly higher intelligibility in comparison to sentences of the same speakers with a non-native temporal pattern. This confirms the general prediction for this study, namely that temporal patterns in utterances carry a function which may be important to speech intelligibility. The results also provide further support for earlier findings (Bent et al., 2008; Holm, 2007, 2008; Maassen and Povel, 1984; Tajima et al. 1997).

Although the (near) 100% intelligibility was verified in the D2 speech at SNR +15 dB, the D2 speech was significantly less intelligible than the D1 speech, even when the difference in temporal pattern was eliminated from the analysis. This contrast in intelligibility could not be due to the intonation pattern or speaking rate, since these features were equalized among the speakers. It is assumed that small pronunciation errors at the phonemic level caused the lower intelligibility in the D2 speech. Several studies support this assumption by concluding that segmental errors can hinder communication, for instance by slowing down word recognition speed (Anderson-Hsieh et al., 1992; Derwing and Munro, 1997; Smith 2004).

Remarkably, when the effect of temporal pattern on intelligibility was tested within the individual D1 or D2 speaker, the higher intelligibility of speech with a native (versus non-native) temporal pattern was only significant in the D1 speaker, and not in the D2 speaker. This suggests the possibility that it depends on certain stimuli characteristics to which degree temporal patterns affect speech intelligibility. In the present experiment a conceivable characteristic which probably reduced the effect of temporal pattern manipulation is the presence of segmental pronunciation errors in the D2 speech, as argued above.

Tajima et al. (1997) suggested that non-native speakers would benefit from training programs which focus on various temporal aspects of L2 speech. Although the effect of temporal correction the D2 speech was insignificant in the present experiment, a small effect size (d = 0.25) was measured. This supports to a lesser extent the recommendation for L2 education of Tajima et al. Moreover, from previous arguments it seems likely that improving temporal pattern, simultaneously with other speech degrading features (such as segmental errors), could enhance the intelligibility of L2 speech. Further work is needed to determine how to achieve this goal.

Several challenges in the manipulation of the stimuli have emerged in the course of this study. First, despite the segmenting process was done very carefully and as congruently as possible, in some sentences it was questionable where the segmental boundaries had to be placed. This might have unintentionally influenced the relative duration of each segment in a sentence. Second, in some cases durational and intonation manipulation interacted in various ways with other perceptually relevant acoustic cues, such as the spectral quality, which degraded the naturalness of the speech stimuli. This was partly counteracted by manual adjustment of some duration and pitch values in the manipulated sentences, but it is uncertain in which degree the acoustic interference influenced the intelligibility. Fortunately, it is expected that the SRT method compensates this small but potential bias, because the SRT is based on the average SNR of 10 sentences.

Skeptics might object to the main effect of temporal pattern manipulation reported here, claiming that the outcomes of this study may not generalize to other interlocutors in real conversations. Of course, having a foreign accent will not affect speech intelligibility equally for each L2 speaker. Factors such as age of L2 learning, length of residence in an L2 speaking environment, gender, formal instruction, motivation, language learning aptitude and amount of L1/L2 use, contribute to variation in the degree of foreign accent (see Piske et al., 2001 for a review). Also listener factors, environmental factors (see 1.1.) and the degree of dissimilarity between L1 and L2 influences the extent to which the L2 speech signal is affected (e.g. Major, 2008). Therefore, the extent to which similar results would be obtained in other speakers and listeners with different L1 and L2 backgrounds, has yet to be empirically determined in the future.

In conclusion, results from this study suggest that native listeners' perception of speech with a non-native temporal pattern in Dutch redundant sentences affects the intelligibility of speech. Correcting the temporal pattern in L2 speech does not necessarily improve intelligibility. An improved intelligibility of L2 speech is likely to occur with simultaneously enhancing the temporal pattern and other phonetic-acoustic features in speech, such as segmental quality. Future research which examines the intelligibility-effects of the interaction between temporal patterns and other phonetic-acoustic consequences of a foreign accent is necessary.

Acknowledgements

The study reported here was supported by the Utrecht Institute of Linguistics OTS. Thanks are due to dr. H. Quené for his enthusiastic and professional guidance during the course of this master thesis, to T.J.G. Veenker for his technical assistance in the experiment, to prof. dr. J.M. Festen and dr. S.J. van Wijngaarden for providing the D1 and D2 test materials respectively, to prof. dr. S.G. Nooteboom, L.J.E. Mus and D.L. van Dijk for their useful comments on previous versions of this paper, and to A. Lievaart for correcting my English. Finally, I thank Pieter, my family, friends, colleagues and fellow students who encouraged me in accomplishing this study.

References

Adams, C., 1979. English speech rhythm and the foreign learner. Mouton, Den Haag.

Anderson-Hsieh, J., Johnson, R., Koehler, K., 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. Language Learning 42 (4), 529-555.

Archibald, J., 1997. The acquisition of English stress by speakers of tone languages: Lexical storage versus computation of stress. Linguistics 35 (1), 167-181.

Atechi, S.N., 2004. The Intelligibility of Native and Non-Native English Speech: A Comparative Analysis of Cameroon English and American and British English. Doctoral dissertation Chemnitz University of Technology.

Bates, D., 2005. Fitting linear mixed models in R: Using the 1me4 package. R News 5 (1), 27-30.

Bent, T., Bradlow, A.R., 2003. The interlanguage speech intelligibility benefit. Journal of Acoustical Society of America, 114, 1600-1610.

Bent, T., Bradlow, A.R., Smith, B.L., 2008. Production and perception of temporal patterns in native and non-native speech. Phonetica 65 (3), 131-147.

Boersma, P., Weenink, D., 2008. Praat: Doing Phonetics by Computer, version 5.0.43, University of Amsterdam, Amsterdam. Software available at: http://www.praat.org/

Cole, R.A. and Cooper, W.E., 1975. The perception of voicing in English affricates and fricatives, Journal of Acoustical Society of America 58 (6), 1280-1287.

Chun, D.M., 2002. Discourse intonation in L2: From Theory and Research to Practice. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Eefting, W., 1991. Timing in Talking: Tempo Variation in Production and its Role in Perception. Doctoral dissertation Utrecht University.

Derwing, T.M., Munro, M.J., 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s. Studies in Second Language Acquisition 20, 1-16.

Derwing, T.M., Munro, M. J., 2005. Second language accent and pronunciation teaching: A research based approach. TESOL Quarterly 39, 379-398.

Flege, J.E., 1993. Production and perception of a novel, second language phonetic contrast. Journal of the Acoustical Society of America 93 (3), 1589-1608.

Flege, J.E., Munro, M., Skelton, L., 1992. Production of the English word-final /t/-/d/ contrast by native speakers of Mandarin and Spanish. Journal of the Acoustical Society of America 91 (1), 128-143.

Flege, J.E., Port, R., 1981. Cross-language phonetic interference: Arabic to English. Language and Speech 24 (2), 125-146.

Gut, U., 2003. Non-native speech rhythm in German. In: Proceedings of the 15th International Congress of Phonetics, Barcelona, 2437-2440.

Holm, S., 2007. The relative contributions of intonation and duration to intelligibly in Norwegian as a second language. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, pp. 1653-1656.

Holm, S., 2008. Intonational and Durational Contributions to the Perception of foreign-accented Norwegian: An experimental phonetic investigation. Doctoral dissertation Norwegian University of Science and Technology.

Klatt, D. H., 1974. Duration of [s] in English words. Journal of Speech and Hearing Research 17, 51-63.

Klatt, D. H, 1975. Vowel lengthening is syntactically determined in a connected discourse, Journal of Phonetics 3 (3),129-140.

Klatt, D. H., 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America 59, 1208-1221.

Laeufer, C., 1992. Patterns of voicing-conditioned vowel duration in French and English. Journal of Phonetics 20 (4), 411-440.

Lehiste, I., 1977. Suprasegmentals. MIT Press, Cambridge, MA.

Maassen, B., Povel, D.J., 1984. The effect of correcting temporal structure on the intelligibility of deaf speech. Speech Communication 3 (2), 123-135.

Major, R.C., 2008. Transfer in second language phonology. In: Hansen Edwards, J.G., Zampini, M.L. (Eds.), Phonology and Second Language Acquisition. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp.63-94.

Morley, J., 1991. The pronunciation component in teaching English to speakers of other languages. TESOL Quarterly 25 (3), 481-519.

Munro, M.J., 1998. The effects of noise on the intelligibility of foreign-accented speech. Studies in Second Language Acquisition 20, 139–154.

Munro, M.J., 2008. Foreign accent and speech intelligibility. In: Hansen Edwards, J.G., Zampini, M.L. (Eds.), Phonology and Second Language Acquisition. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp.193-218.

Munro, M.J., Derwing, T.M., 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Language Learning 49 (s1), 285-310.

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

Munro, M.J., Derwing, T.M., Morton, S.L., 2006. The mutual intelligibility of L2 speech. Studies in Second Language Acquisition 28, 111–131.

Neri, A., Cucchiarini, C., Strik, H., 2006. Selecting segmental errors in non-native Dutch for optimal pronunciation training. International Review of Applied Linguistics in Language 44, 357-404.

Nooteboom, S.G., 1973. The perceptual reality of some prosodic durations. Journal of Phonetics 1 (1), 25-45.

Nooteboom, S.G., 1997. The prosody of speech: Melody and rhythm, in: Hardcastle, W.J., Laver J. (Eds.), The Handbook of the Phonetic Sciences, Blackwell Publishers, Oxford, pp. 640-673.

Oller, D.K., 1973. The effect of position in utterance on speech segment duration in English. Journal of the Acoustical Society of America 54 (5), 1235-1247.

Patel, A.D., 2008. Music, language, and the brain. Oxford University Press Inc., New York.

Piske, T., MacKay, I.R.A., Flege, J.E., 2001. Factors affecting degree of foreign accent in an L2: A review. Journal of Phonetics 29, 191-215.

Plomp, R., Mimpen, A.M., 1979. Improving the reliability of testing the speech reception threshold for sentences. Audiology 18, 43-52.

Plomp, R., 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. Journal of Speech and Hearing Research 29, 146-154.

Rajadurai, J., 2007. Intelligibility studies: A consideration of empirical and ideological issues. World Englishes 26 (1), 87-98.

Rasier, L., 2006. Prosodie en Vreemdetaalverwerving: Accentdistributie in het Frans en in het Nederlands als Vreemde Taal. Doctoral dissertation Université Catholique de Louvain.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing. Software available at: http://www.r-project.org

Shatzman, K.B., McQueen, J.M., 2006. Segment duration as a cue to word boundaries in spoken-word recognition. Perception and Psychophysics 68 (1), 1-16.

Sluijter, A.M.C., 1995. Phonetic Correlates of Stress and Accent. Doctoral Dissertation Leiden University.

Smith, R., 2004. The Role of Fine Phonetic Detail in Word Segmentation. Doctoral dissertation University of Cambridge.

Smith, B., Bradlow, A.R., Bent, T., 2003. Production and perception of temporal contrasts in foreign-accented English. In: Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, pp. 519–522.

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

Umeda, N., 1977. Consonant duration in American English. Journal of the Acoustical Society of America 61 (3), 846-858.

van Santen, J.P.H., 1992. Contextual effects on vowel duration. Speech Communication 11 (6), 513-546.

van Wijngaarden, S.J., 2001. Intelligibility of native and non-native Dutch speech. Speech Communication, 35, 103-113.

van Wijngaarden, S.J., 2003. The intelligibility of non-native speech. Doctoral dissertation VU Universitair Amsterdam.

Varonis, E.M., Gass, S., 1982. The comprehensibility of non-native speech. Studies in Second Language Acquisition 4 (2), 114-135.

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

Young-Scholten, M., Archibald, J., 2000. Second language syllable structure, in: Archibald, J. (Eds.), Second Language Acquisition and Linguistic Theory, Blackwell Publishers, Oxford, pp. 64-101.

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

Quené, H., van den Bergh, H., 2004. On multi-level modeling of data from repeated measures designs: A tutorial. Speech communication 43 (1-2), 103-121.

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, 413-425.

Zampini, M.L., 2008. L2 speech production research, in: Hansen Edwards, J.G., Zampini, M.L. (Eds.), Phonology and Second Language Acquisition. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 219-252.

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

Appendix: Lists of Dutch SRT test-sentences (Plomp and Mimpen, 1979)

List I
1 De bal vloog over de schutting
2 Morgen wil ik maar een liter melk
3 Deze kerk moet gesloopt worden
4 De spoortrein was al gauw kapot
5 De nieuwe fiets is gestolen
6 Zijn manier van werken ligt mij niet
7 Het slot van de voordeur is kapot
8 Dat hotel heeft een slechte naam
9 De jongen werd stevig aangepakt
10 Het natte hout sist in het vuur
11 Zijn fantasie kent geen grenzen
12 De aardappels liggen in de schuur
13 Alle prijzen waren verhoogd

List II
1 Zijn leeftijd ligt boven de dertig
2 Het dak moet nodig hersteld worden
3 De kachel is nog steeds niet aan
4 Van de viool is een snaar kapot
5 De tuinman heeft het gras gemaaid
6 De appels aan de boom zijn rijp
7 Voor het eerst was er nieuwe haring
8 Het loket bleef lang gesloten
9 Er werd een diepe kuil gegraven
10 Zijn gezicht heeft een rode kleur
11 Het begon vroeg donker te worden
12 Het gras was helemaal verdroogd
13 Spoedig kwam er een einde aan

List III
1 Ieder half uur komt hier een bus langs
2 De bel van de voordeur is kapot
3 De wind waait vandaag uit het westen
4 De slang bewoog zich door het gras
5 De kamer rook naar sigaren
6 De appel had een zure smaak
7 De trein kwam met een schok tot stilstand
8 De koeien werden juist gemolken
9 Het duurt niet langer dan een minuut
10 De grijze lucht voorspelt regen
11 Hij kon de hamer nergens vinden
12 Deze berg is nog niet beklommen
13 De bel van mijn fiets is kapot

List IV
1 De auto heeft een lekke band
2 Het moeilijke werk bleef liggen
3 Het vliegtuig vertrekt over een uur
4 Die jongens vechten de hele dag
5 De schoenen moeten verzoold worden
6 In de krant staat vandaag niet veel nieuws
7 Door de neus ademen is beter
8 Het kind was niet in staat te spreken
9 De witte zwaan dook onder water
10 Hij nam het pak onder zijn arm
11 Gelukkig sloeg de motor niet af
12 De leraar gaf hem een laag cijfer
13 Het huis brandde tot de grond toe af

List V
1 De foto is mooi ingelijst
2 Mijn broer gaat elke dag fietsen
3 Een kopje koffie zal goed smaken
4 De schrijver van dit boek is dood
5 Zij heeft haar proefwerk slecht gemaakt
6 De sigaar ligt in de asbak
7 De appelboom stond in volle bloei
8 Er wordt in dit land geen rijst verbouwd
9 Hij kan er nu eenmaal niets aan doen
10 De kleren waren niet gewassen
11 Het gedicht werd voorgelezen
12 Haar gezicht was zwart van het vuil
13 De letters stonden op hun kop

List VI
1 De groene appels waren erg zuur
2 In het gebouw waren vier liften
3 Lopen is gezonder dan fietsen
4 Het lawaai maakte hem wakker
5 Mijn buurman heeft een auto gekocht
6 Als het flink vriest kunnen we schaatsen
7 De kast was een meter verschoven
8 Oude meubels zijn zeer in trek
9 De portier ging met vakantie
10 De lantaarn gaf niet veel licht meer
11 Door zijn snelheid vloog hij uit de bocht
12 Het is hier nog steeds veel te koud
13 De oude man was kaal geworden

List VII
1 De bomen waren helemaal kaal
2 Rijden onder invloed is strafbaar
3 Onze bank geeft vijf procent rente
4 Het verslag in de krant is kort
5 In de vijver zwemmen veel vissen
6 Honden mogen niet in het gebouw
7 Een flinke borrel zal mij goed doen
8 Gisteren waaide het nog harder
9 Het meisje stond lang te wachten
10 De volgende dag kwam hij ook niet
11 Het geschreeuw is duidelijk hoorbaar
12 Eindelijk kwam de trein op gang
13 De grote stad trok hem wel aan

List VIII
1 De bus is vandaag niet op tijd
2 Onze dochter speelt goed blokfluit
3 Ook in de zomer is het hier koel
4 Zij moesten vier uur hard werken
5 Niemand kan de fransman verstaan
6 Eiken balken zijn erg kostbaar
7 Het aantal was moeilijk te schatten
8 Er waaide een stevig briesje
9 De vis sprong een eind uit het water
10 Iedereen genoot van het uitzicht
11 Het regent al de hele dag
12 Het tempo was voor hem veel te hoog
13 In juni zijn de dagen het langst

List IX
1 De bakkers bezorgen vandaag niet
2 Het licht in de gang brandt nog steeds
3 De wagen reed snel de berg af
4 Lawaai maakt je op den duur doof
5 In de kerk wordt mooi orgel gespeeld
6 De schaatsen zijn in het vet gezet
7 Toch lijkt me dat een goed voorstel
8 Hij probeerde het nog een keer
9 De zak zat vol oude rommel
10 Zij werd misselijk van het rijden
11 Door zijn haast maakte hij veel fouten
12 De nieuwe zaak is pas geopend
13 Dat is voor hem een bittere pil

List X
1 Op het gras mag men niet lopen
2 Steile trappen zijn gevaarlijk
3 De zon gaat in het westen onder
4 De hond blafte de hele nacht
5 De kat van de buren is weg
6 De trein vertrekt over twee uur
7 Het was heel stil in de duinen
8 Hij rookte zijn sigaret op
9 De rivier trad buiten haar oevers
10 De jongens gingen er gauw vandoor
11 Moeizaam klom de man naar boven
12 De biefstuk is vandaag erg mals
13 De kat likt het schoteltje leeg