



HAL
open science

Italian speakers learn lexical stress of German morphologically complex words

Maria Paola Bissiri, Hartmut R. Pfitzinger

► **To cite this version:**

Maria Paola Bissiri, Hartmut R. Pfitzinger. Italian speakers learn lexical stress of German morphologically complex words. *Speech Communication*, 2009, 51 (10), pp.933. 10.1016/j.specom.2009.03.001 . hal-00558517

HAL Id: hal-00558517

<https://hal.science/hal-00558517>

Submitted on 22 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

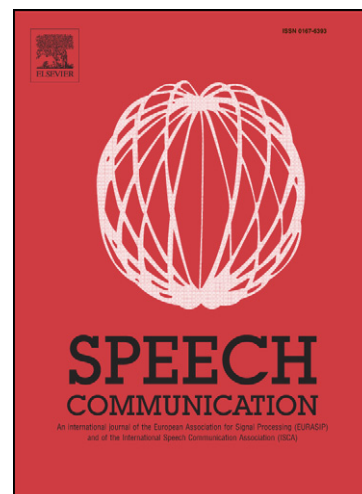
Italian speakers learn lexical stress of German morphologically complex words

Maria Paola Bissiri, Hartmut R. Pfitzinger

PII: S0167-6393(09)00036-3
DOI: [10.1016/j.specom.2009.03.001](https://doi.org/10.1016/j.specom.2009.03.001)
Reference: SPECOM 1790

To appear in: *Speech Communication*

Received Date: 2 July 2008
Revised Date: 6 January 2009
Accepted Date: 12 March 2009



Please cite this article as: Bissiri, M.P., Pfitzinger, H.R., Italian speakers learn lexical stress of German morphologically complex words, *Speech Communication* (2009), doi: [10.1016/j.specom.2009.03.001](https://doi.org/10.1016/j.specom.2009.03.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Italian speakers learn lexical stress of German morphologically complex words

Maria Paola Bissiri^{a,b,*} Hartmut R. Pfitzinger^c

^a*Dipartimento di Scienze dei Linguaggi,
Università degli Studi Sassari, Via Roma 151, 07100 Sassari, Italy*

^b*Institute of Phonetics and Speech Processing (IPS),
Ludwig-Maximilians-University Munich, Schellingstr. 3, 80799 Munich, Germany*

^c*Institute of Phonetics and Digital Speech Processing (IPDS),
Christian-Albrechts-University Kiel, Leibnizstr. 10, 24118 Kiel, Germany*

Abstract

Italian speakers tend to stress the second component of German morphologically complex words such as compounds and prefix verbs even if the first component is lexically stressed. To improve their prosodic phrasing an automatic pronunciation teaching method was developed based on auditory feedback of prosodically corrected utterances in the learners' own voices. Basically, the method copies contours of F0, local speech rate, and intensity from reference utterances of a German native speaker to the learners' speech signals. It also adds emphasis to the stress position in order to help the learners better recognise the correct pronunciation and identify their errors. A perception test with German native speakers revealed that manipulated utterances significantly better reflect lexical stress than the corresponding original utterances. Thus, two groups of Italian learners of German were provided with different feedback during a training session, one group with manipulated utterances in their individual voices and the other with correctly pronounced original utterances in the teacher's voice. Afterwards, both groups produced the same sentences again and German native speakers judged the resulting utterances. Resynthesised stimuli, especially with emphasised stress, were found to be a more effective feedback than natural stimuli to learn the correct stress position. Since resynthesis was obtained without previous segmentation of the learners' speech signals, this technology could be effectively included in Computer Assisted Language Learning software.

Key words: CALL, Computer Assisted Language Learning, prosody modification, intonation, local speech rate, intensity, performance assessment, lexical stress

* Corresponding author. Address: IPS, University of Munich, Schellingstr. 3, 80799 Munich, Germany. Tel: +49-89-2180-2758; Fax: +49-89-2180-5790

Email address: mariapa@phonetik.uni-muenchen.de; hpt@ipds.uni-kiel.de

1 Introduction

A Computer Assisted Language Learning (CALL) program should help second language learners to distinguish correct and incorrect pronunciations, as they often fail to do this without assistance. This applies to Italian learners of German who may not be able to recognise the correct lexical stress of German morphologically complex words without some form of feedback. The morphologically complex words we refer to in this work are affixed words stressed on the first syllable and compounds stressed on the first component. The challenge for CALL research consists in developing forms of feedback which raise awareness of the pronunciation characteristics, so that these can be noticed and learned by the users of a CALL software. This kind of feedback should be integrated in a comprehensive CALL program which also provides a description of the prosodic rules of the target language.

In both Italian and German, lexical stress is phonologically distinctive. In German, words are most frequently stressed on the first syllable (Mengel, 2000), while in Italian the most frequent case is stress on the penultimate syllable (Delmonte, 1981). In Italian, stress on the antepenultimate syllable is also possible, and there are several minimal pairs which contrast stress on the penultimate and on the antepenultimate syllable (e.g. 'compito "task" vs. com'pito "polite"). A transfer of their default stress pattern to German words could explain lexical stress mistakes of beginner learners who have not been exposed to a German speaking environment. However, if this kind of mistake persists in advanced learners of German who have been living in Germany for several years, then they must evidently still have difficulty in identifying the appropriate lexical stress pattern. In fact, misunderstandings for L2-learners can arise if in the two languages acoustic features are exploited by stress differently (Wang, 2008). Since learners create an interlanguage using elements of their native language (Gass, Selinker, 1994), they might also transfer the acoustic cues for stress identification from their native language to the second language. In Italian, the most prominent acoustic and perceptual correlate of stress is duration (Bertinetto, 1980, 1981), and duration seems very important also for the perception of sentence accent (D'Imperio, 2000). Besides, Italian has, in contrast to German, no distinctive quantity opposition between vowels. In German morphologically complex words with stress on the first component, the second component can keep long duration because quantity patterns of the components do not change in compounds (Kohler, 1995, p.188). For this reason Italians may perceive the second component as being stressed, especially if it contains a long vowel. In a study by Bissiri et al. (2008), Italians participated in a short-term memory sequence repetition task based on a method introduced by Dupoux et al. (2001). They performed significantly worse on a stress contrast than on a phoneme or a stress plus quantity contrast. It seems

(Hartmut R. Pfitzinger).

that Italians categorise German morphologically complex words as having the same stress pattern regardless of the position of stress.

The different realization of intonation patterns in Italian and in German may also contribute to confusions of stress position. Delmonte, Bortolato (2003) report differences in the intonation contours of German sentences produced by Italians compared to German natives. Italian learners did not produce upstep boundary tones ($H\text{-}\hat{H}\%$) in German compounds in yes/no-questions as German natives did, and made lexical stress errors. According to Ladd (1996), a different peak alignment in the two languages could cause misunderstandings. Since Italian is characterised by an early F0-peak alignment and German by a late peak alignment, L1-German speakers' productions of late peaks when they speak Italian, may be interpreted as an incorrect assignment of lexical stress to the following syllable by L1-Italian listeners (Ladd, 1996; Mennen, 2007). For the same reason, Italian learners of German might also interpret an antepenultimate stress as a penultimate stress in German.

However, lexical stress of German morphologically complex words is for Italians rather easy to pronounce, provided that they perceive it. The challenge consists in finding an appropriate form of feedback for CALL programs in order to help Italian learners to recognise the correct stress pattern. Simply highlighting the stressed syllable in the orthography would not be a suitable method since Italians would then probably shorten the full syllable in the second component even if it is phonologically long. For instance they might pronounce the German prefix verb 'abfahren' with a short instead of a long /a/ in the penultimate syllable, i.e. /'apfarən/ instead of /'apfa:rən/.

L2-prosody is often taught by means of a graphical display of intonation. First proposed by Vardanian (1964) and then by Lane, Buiten (1969), who also provided a graphical display of duration and intensity, this method was effectively implemented in the 1970's and 1980's (James, 1976; De Bot, 1983; Cranen et al., 1984). Since then, intonation displays are frequently implemented in CALL programs (Chun, 1989, 1998; Anderson-Hsieh, 1992, 1994; Delmonte et al., 1997; Delmonte, 2000; Kommissarchik, Kommissarchik, 2000; Hardison, 2004). An overview on feedback in Computer Assisted Pronunciation Training is given by Neri et al. (2002). Although they can be very useful for intonation, graphical displays might be less adequate for teaching lexical stress. Since the perception of stress usually depends on a combination of several acoustic properties, and since their weighting can differ between the native and the target language, the graphical representation of intonation, duration, and intensity might be confusing and not informative for learners. For lexical stress, auditory feedback would be more appropriate.

An effective form of feedback for phonological mistakes are recasts (Lyster, 1998), which consist in "the teacher's correct restatement of a learner's incorrectly formed utterance" (Nicholas et al., 2001, p.720). Input — the language to which learners are exposed (Gass, 1997) —, whether produced by a CALL

program or by a human teacher, should be made comprehensible so that those pronunciation characteristics that are supposed to be learned are made salient (Gass, 1997; Chapelle, 1998). Recasts in a native speaker's voice might not be sufficient for the learners to notice the difference in word stress since the large number of differences between the native speaker's and the learners' speech — such as register, voice quality, and segmental differences — may be confounded with those cues for recognising the correct stress pattern. A method to make input comprehensible is to present the learners recasts in their own voice after correcting the prosodic parameters in their speech. Thus, by reducing the phonetic variability, it should be easier for them to compare the correct with the wrong pronunciation (Tillmann, Pfitzinger, 2004). Probst et al. (2002) showed that learners of English trained with the FLUENCY pronunciation training system (Eskenazi, Hansma, 1998; Eskenazi et al., 2000) by imitating the voice of a reference speaker similar to their own improved more than learners who imitated a dissimilar voice.

The effectiveness of resynthesised utterances in the learners' own voice was first shown by Nagano, Ozawa (1990). The system *WinPitch LTL* also provides resynthesised stimuli as feedback to the learners (Germain-Rutherford, Martin, 2000; Martin, 2004; WinPitch, 2005). WinPitch is an authoring tool for teachers which allows them to manipulate students' speech. Its advantages are that the teachers can manually insert commentaries and suggestions for the students and that they can check the resynthesised stimuli before they are given to the students as feedback. However the program requires phonetic expertise which many language teachers might not have. Furthermore, the program does not present the students with feedback immediately after their wrong pronunciation. However, immediate feedback is more effective since it allows the learners to compare directly their incorrect with the correct pronunciation (Nicholas et al., 2001).

Hirose et al. (2003) and Hirose (2004) developed a program to teach Japanese lexical accents to non-native speakers. The program automatically recognises whether an accent was pronounced correctly or incorrectly by the learner. In the latter case, the learner's speech was corrected by means of resynthesis with TD-PSOLA so that the program was able to offer as feedback the correct utterance in the learner's voice.

The aim of our study was to test the following hypotheses:

- (1) Concerning the perceptual evaluation of prosody correction by resynthesis
 - (a) Resynthesis of intonation and local speech rate of German morphologically complex words wrongly stressed by Italian speakers significantly improves the judgement of stress correctness by German native speakers.
 - (b) Resynthesis of intonation is more effective in correcting lexical stress than resynthesis of local speech rate.
 - (c) There is an interaction between resynthesis of the two prosodic pa-

rameters.

- (2) Concerning the effectiveness of different training methods
 - (a) Resynthesised stimuli in the learners' own voice — after correction of intonation, local speech rate, and intensity — are a more effective feedback than natural stimuli in a native speaker's voice for teaching Italian learners lexical stress of German morphologically complex words.
 - (b) Stimuli in which the stressed syllable received extra emphasis are a more effective feedback than normally stressed stimuli.
 - (c) Resynthesised stimuli with emphasis are more effective than without emphasis, and they are also more effective than natural stimuli with emphasis.

In the next section we examine lexical stress errors in a German speech corpus read by Italians. We then describe the design of our resynthesis method for prosodic manipulation. After that three experiments are presented. The first experiment investigates the influence of the modification and resynthesis quality of intonation and local speech rate contours on lexical stress correctness by means of a perception test. The second experiment is concerned with lexical stress training of German morphologically complex words for Italian speakers and the third experiment is a perceptual evaluation of learners' performances after feedback with resynthesis or emphasis. The final section summarises the results of this series of investigations.

2 Lexical stress errors in German utterances spoken by Italians

We recorded the German standard text “Die Buttergeschichte” (Appendix A). The speakers were six students of German linguistics at the University of Sassari in Italy and six Italian native speakers living in Munich. The students from Sassari were between 20 and 25 years old, and had learnt German for between one and seven years. The speakers living in Munich were between 24 and 42 years old, and had lived in Munich between two months and twelve years. The speaker who had lived in Munich for only two months had, however, studied German previously for two years. All were female native speakers of the variety of Italian spoken in Sardinia. Before the recording, participants received a printed version of the German text with translation footnotes. The students of Sassari performed a text analysis during a university course in order to guarantee that they understood the text while reading. The speakers living in Munich were recorded in an anechoic chamber at the Institute of Phonetics and Speech Processing of the University of Munich. The students from Sassari were recorded in a professional recording studio in Sassari. In both cases, we used a Neumann TLM 103 microphone with cardioid polar pattern at an approx. 50 cm mouth-microphone distance. Speech signals were digitised at a 96 kHz sampling rate and 24 Bit amplitude resolution, and

high-pass filtered at 79 Hz.

2.1 Lexical stress error analysis

All words pronounced by the Italian speakers were auditorily checked for lexical stress correctness by the first author and a German native speaker. It was found that all speakers, except one of the speakers living in Munich, very frequently stressed the second component of the morphologically complex words (or the third in words with three components) instead of the first component. Table 1 shows the number of speakers who incorrectly stressed morphologically complex words with initial stress.

Table 1

Morphologically complex words wrongly stressed on the second or third component instead of the first, ordered according to error frequency. Since each word appears only once in the text, maximally twelve errors — by all twelve speakers — are possible.

Word	Errors	Word	Errors
1. durchlassen	9	9. Lebensmittel	4
2. unruhiger	9	10. anzustellen	3
3. vorlassen	8	11. Buttergeschichte	3
4. aufschließen	7	12. meinetwegen	3
5. Menschenmenge	7	13. anschließen	2
6. ausverkauft	6	14. ebenfalls	2
7. eingetroffen	6	15. Ladentür	2
8. Schimpfwörter	6	16. rücksichtslos	1

Half of the words were wrongly stressed by at least half of the speakers. Even advanced Italian speakers of German, living in Munich for ten years or more, made this kind of error very frequently.

Table 2 shows that other kinds of words induced only a few lexical stress errors. Most of these errors were made by the students from Sassari probably because they have not had enough German language practice.

Words for which no stress errors were reported are not listed in Table 2.

The high frequency with which these kinds of stress errors occurred provides clear evidence that these words pose a particular problem for Italian speakers learning German lexical stress. For this reason, we focused on the resynthesis, production, and perception of the stress pattern in these kinds of morphologically complex words in the experiments reported below.

Table 2

Other wrongly stressed words: a) on the antepenultimate instead of the penultimate or final syllable, b) on the penultimate instead of the antepenultimate, and c) on the penultimate instead of the final syllable.

stressed antepenultimate instead of penult. or final		stressed penultimate instead of antepenult.		stressed penultimate instead of final	
Word	Errors	Word	Errors	Word	Errors
1. Drängelei	3	6. wartenden	3	9. Berlin	4
2. erhalten	2	7. beachtliche	2	10. bereits	1
3. genügend	1	8. forderten	1	11. empört	1
4. Polizei	1				
5. vorhanden	1				

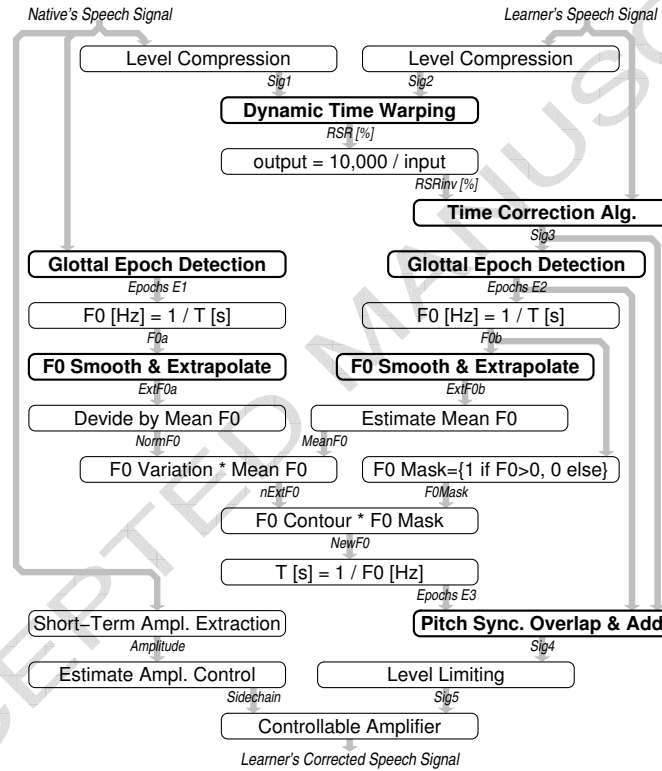


Fig. 1. Block diagram of the algorithm for copying the prosodic parameters local speech rate, intonation, and intensity of a native speaker to the learner's speech. Intermediate signals are shown in Fig. 2.

3 Resynthesis of local speech rate, intonation, and intensity

The basic justification for an automatic correction procedure via the three steps analysis, modification, and synthesis (subsumed under the term *resynthesis*) is that since the learner only speaks sentences which have already

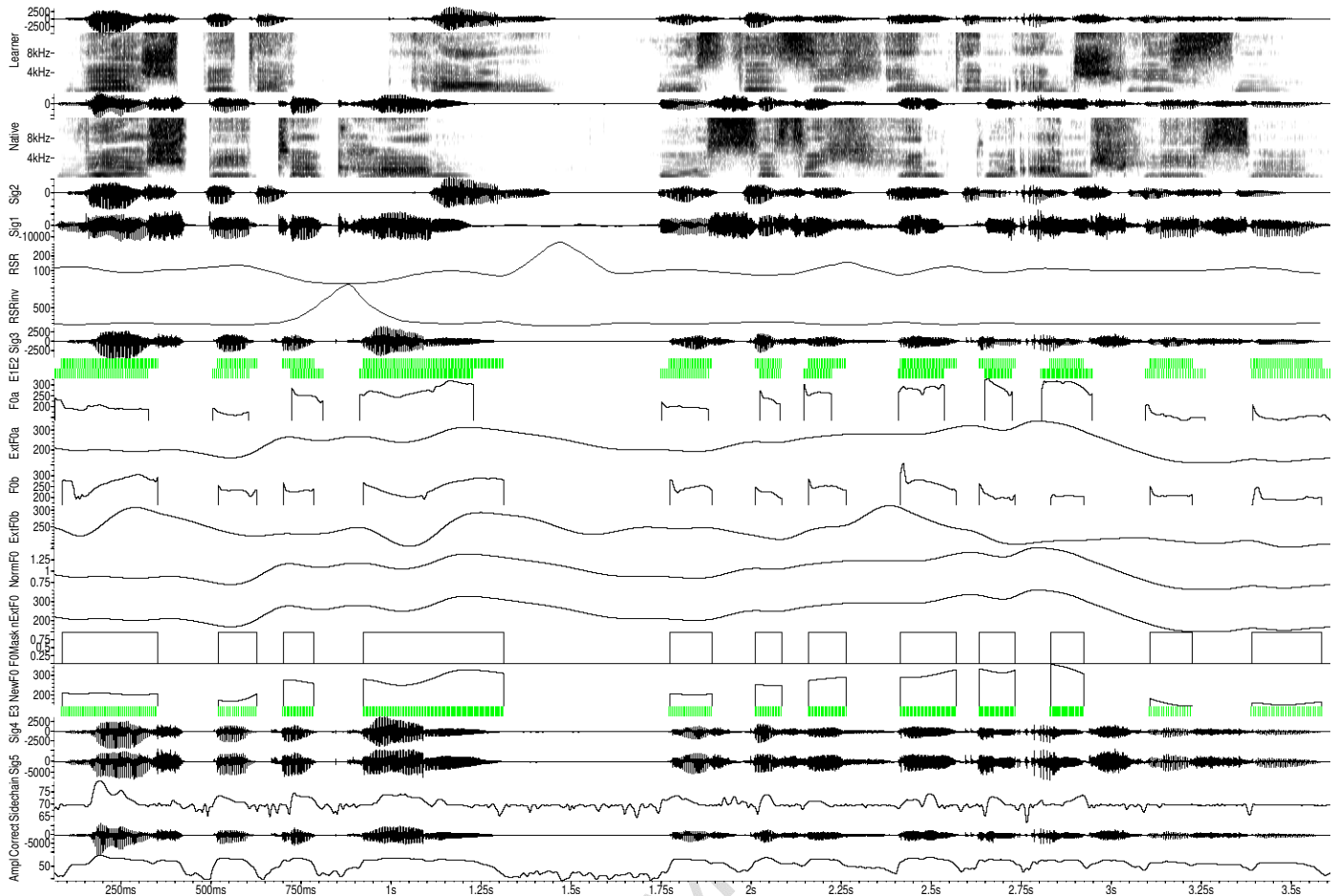


Fig. 2. Utterance „*Wer später kam, musste sich hinten anschließen.*“. *Top signal*: learner’s utterance. *2nd signal*: German native speaker’s utterance. *2nd last signal*: learner’s modified and resynthesised utterance. Fig. 1 shows the signal processing modules which produced and use all intermediate signals.

been spoken by the reference speaker, then the two speech signals are both sufficiently acoustically similar for automatic time alignment and sufficiently linguistically similar for prosodic congruency to be improved by increasing prosodic similarity.

However, in order to maintain speaker-individual properties, microprosodic and spectral features of the learner should not be changed. Consequently, resynthesis was carried out by copying the prosodic contours of 1) local speech rate, 2) intonation, and 3) intensity from a reference utterance of a German native speaker to a wrongly stressed utterance of an Italian speaker.

Fig. 1 shows the block diagram of the corresponding algorithm. Each module, represented as a small box and labelled with an arbitrary name, produces a signal as output, represented as an arrow. Each intermediate signal is an input for the next module. The signals produced by the modules are shown in the resynthesis example in Fig. 2. The main processing blocks are described in the following sections. Before carrying out the resynthesis, all speech signals were

downsampled to 32 kHz.

3.1 Local speech rate

The algorithm first normalises the amplitudes of the German speaker’s and Italian learner’s speech signals in order to facilitate the automatic alignment of the two speech signals. Dynamic Time Warping (DTW) using a Euclidean distance function was applied to temporally and spectrally smoothed feature vectors, i.e. they were extracted in 10 ms steps via 20 ms Kaiser windows ($\beta = 6$). Subsequently, autocorrelation-based LPC, FFT, and non-uniform spectral-domain windowing were applied resulting in 16 Bark-scaled spectral amplitudes between 200 Hz and 8 kHz for each frame. (Besides, this data reduction of $100\text{frames/s} \cdot 16\text{bytes} \cdot 8\text{bits} = 12.8\text{kBit/s}$ enables later CALL applications to store a huge number of various reference patterns.)

The resulting warping path was suitable for adjusting the segment durations of the learner’s signal to those of the German speaker’s signal. But since actual segment durations are to a certain degree speaker-specific, the change in their local relations should be kept to a minimum while the local speech rate that they reflect should be adjusted as accurately as possible. Therefore, the warping function was smoothed via a 300 ms triangular window which is a good compromise between reducing short-term peaks and dips in the warping function caused by microprosodic properties while maintaining variations on a syllable level caused by local speech rate differences between both speakers. Finally, the prosodic timing structure of the learner’s utterance was corrected according to the smoothed warping function by means of PSOLA.

In some cases, e.g. in the case of long sentences or if the learner spoke much more slowly than the German native speaker or produced additional long pauses, manual alignment of the pauses in the German’s and Italian’s signals was necessary. After DTW-alignment and PSOLA, the time-warped learner’s speech signal was synchronous with the speech signal of the German speaker.

3.2 Intonation

With the term *intonation* we refer to the smoothed and extrapolated fundamental frequency contour of the speech signal. F0 contours were extracted from the aligned signals by means of an island-driven AMDF-based (Average Magnitude Difference Function) pitch detection algorithm, smoothed to reduce microprosodic influences, and extrapolated to avoid disruptions to the mean F0 calculation caused by stretches of speech with undefined F0. In order to keep the voice register of the Italian speaker, the mean F0 of both signals were computed. To combine the learner’s mean F0 with the German speaker’s F0 modulation, we divided the F0 contour of the German native speaker by its mean value and multiplied it with the mean value of the learner’s con-

tour. At originally voiced stretches of speech, the positions of the new glottal excitation pulses were determined according to the new F0 contour. Finally, the resulting fundamental frequency contour was copied to the learner's time-warped speech signal by means of PSOLA to achieve the corrected F0 contour. To reduce F0 detection errors, the range of possible F0 values was reduced to 120–500 Hz for female voices and to 70–300 Hz for male voices. This way, male utterances could also be corrected without serious problems.

For Experiment 1 (Sect. 4), we created stimuli in which only intonation and not local speech rate was corrected. However, intonation can obviously be copied from a speech signal to another only if the two signals are synchronous. To solve this problem, we warped the intonation contour of the German speaker according to the previously described warping function so that it matched the timing structure of the Italian utterance and we then proceeded as described in the previous paragraph.

3.3 Intensity

We resynthesised intensity by copying the short-term amplitude envelope of the German speaker's signal to the amplitude-normalised signal of the Italian speaker via a short-term controllable amplifier. This operation resulted in a speech signal with the short-term amplitude envelope of the German speaker's utterance. All amplitude contours were the result of root-mean-square estimation during 15 ms Kaiser windows ($\beta = 5$).

The resynthesis of intensity was necessary for Experiment 2 (Sect. 5) since syllables produced with emphasis showed considerable intensity peaks in the reference utterances. However, in standard reference utterances without emphasis, F0 peaks at stressed syllables are also accompanied by above-average intensity.

4 Experiment 1: Influence of the resynthesis of intonation and local speech rate on lexical stress correctness

The aim of this experiment was to test i) if the resynthesis of intonation and local speech rate can significantly improve the judgements on lexical stress correctness by German native speakers, ii) which one of the two prosodic parameters is more relevant for the perception of correct stress, and iii) if there is any interaction between the correction of these two parameters.

The experiment consisted in a perception test with German native listeners, who evaluated lexical stress correctness in natural and resynthesised stimuli after manipulation of intonation, local speech rate or both parameters.

4.1 Method

4.1.1 Stimuli

14 sentences from “Die Buttergeschichte” (see Appendix A) containing 16 morphologically complex words stressed on the first component, and the same words read in isolation (words listed in Table 1) were selected as stimuli. In order to prevent Italian speakers from transferring the same stress pattern from one word to another, they read a list of 32 words consisting of the 16 complex words and 16 words from the text which were stressed on different syllables. The order of the words in the list was the same as in the text.

The material was read by one female German native speaker and eight female Italian native speakers. The German speaker was a 32-year-old trained phonetician from Munich who spoke standard German. The eight Italian speakers lived in Munich and three of them had already taken part in the previous recordings (Sect. 2), while the other five participants were from various Italian regions. The recordings took place in Munich in the same environment as described in Sect. 2, but with 48 kHz sampling rate and 16 Bit amplitude resolution.

The morphologically complex words spoken by the Italians were checked for stress correctness by the first author and by a German native speaker. The utterances containing words judged to be incorrectly stressed were submitted to resynthesis of local speech rate or intonation or both parameters by copying the prosodic parameters of the German native speaker to the Italians’ speech. This way, four different stimuli for each utterance spoken by the Italians were obtained: i) original, ii) with resynthesised local speech rate, iii) with resynthesised intonation, and iv) with both parameters resynthesised. We did not resynthesise intensity, since this is considered a less relevant parameter for lexical stress (Lehiste, 1970; Sluijter, van Heuven, 1996), and the additional intensity manipulation would have doubled the number of the stimuli for the perception test.

4.1.2 Participants

Lexical stress correctness of the morphologically complex word stimuli was evaluated by 31 German native speakers, between 20 and 49 years old. Most of them had no or scarce knowledge of Italian. They were all employees or students at the Institute of Phonetics and Speech Processing of the University of Munich.

4.1.3 Perception test

The number of stimuli was limited to 256 since the presentation of all stimuli would have lasted too long for the subjects to accomplish. We chose the most frequent wrongly stressed morphologically complex words, as the most

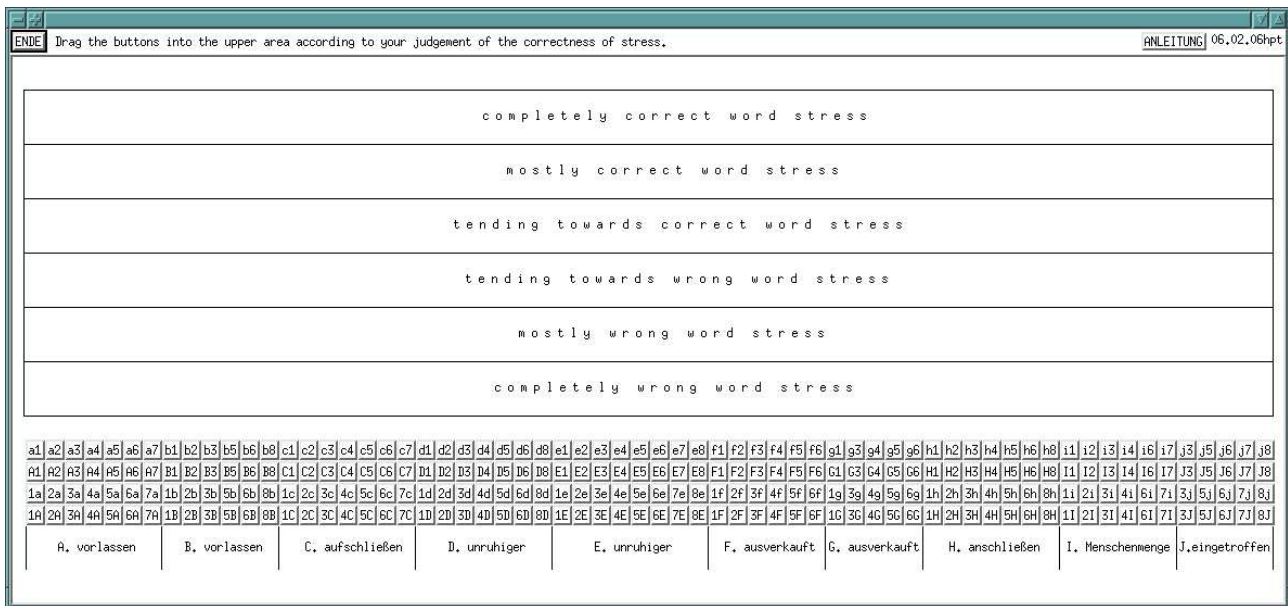


Fig. 3. Computer-aided graphical user interface for perception experiment 1.

frequent errors were probably the most relevant for this investigation. So, the stimuli consisted of six sentences and four isolated words.

On the computer-aided test interface, the stimuli were ordered in columns. Each column comprised four buttons coded with the same alphabet letter (see Fig. 3). By clicking on the buttons, the subjects could listen to the four versions of an utterance produced by the same speaker. They had to compare the four versions with each other and drag the buttons to the upper area according to their judgement of stress correctness. The upper area of the interface was divided into six small areas: 1. completely correct word stress, 2. mostly correct word stress, 3. tending towards correct word stress, 4. tending towards wrong word stress, 5. mostly wrong word stress, and 6. completely wrong word stress. The subjects could also place the buttons at different heights within each small area to express small differences in stress correctness.

4.2 Results

The 7936 judgements of the German natives were converted to a range between 0 and 100, 0 meaning the worst and 100 the best possible judgement.

Table 3 shows the mean values of the stress correctness judgements for the four versions of the stimuli. The highest values refer to the stimuli after resynthesis of intonation and after resynthesis of both intonation and local speech rate. Fig. 4 shows corresponding boxplots with medians and quartiles.

The Friedman and the Wilcoxon tests were employed to test the differences between the judgements of the four versions of the stimuli for significance. The Friedman test, a non-parametric test for the comparison of several dependent

Table 3

Mean judgements for the four versions of the stimuli.

	original	local sp. rate	intonation	local sp. rate + inton.
mean	40.08	44.24	58.26	62.32

samples, showed a highly significant difference between the judgements of the four versions of the stimuli ($p < 0.001$). The pairwise Wilcoxon signed rank yielded highly significant differences ($p < 0.001$) for each of the six comparisons. An ANOVA on the average judgements for each stimulus revealed that there was no significant interaction between local speech rate and intonation.

4.3 Discussion

The resynthesis of local speech rate and intonation significantly influenced the judgements of lexical stress correctness by the German native speakers. The resynthesis of both parameters turned out to have no interactive effects.

The resynthesis of intonation was more effective than the resynthesis of local speech rate in correcting lexical stress. The greater relevance of intonation can be explained by the fact that most judged words were in focus position. Several test participants reported that, instead of four different stimuli, they often heard two pairs of equal stimuli: one pair of wrongly stressed stimuli and the other pair of correctly stressed stimuli. In fact, while carrying out the resynthesis, we noticed that in most cases the manipulation of local speech rate alone did not change the stress position, while the manipulation of intonation was more effective. According to Jessen et al. (1995) and Dogil, Williams

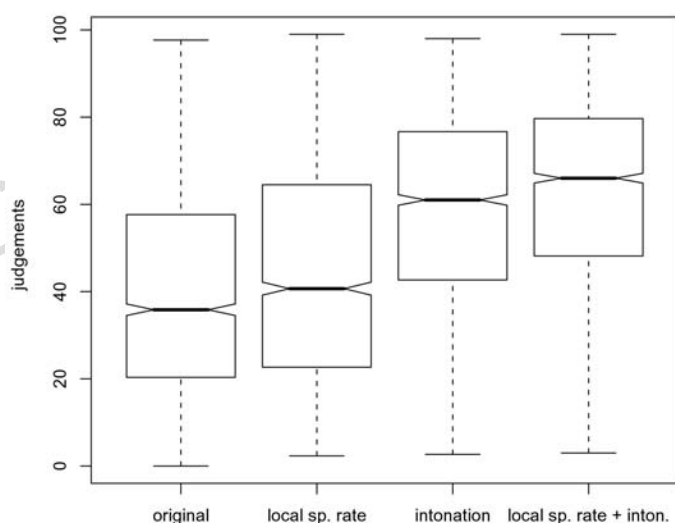


Fig. 4. Medians and quartiles of the 7936 judgements of the original and resynthesised stimuli.

(1999), duration is the most prominent correlate of stress in German. However, durational differences may be less marked between the primary stressed syllables and the full syllables in other components of German morphologically complex words. Since Italians rely more on duration for identifying lexical stress (Bertinetto, 1980), this might explain why they have trouble learning the lexical stress of German morphologically complex words.

The resynthesis of the local speech rate did sometimes reduce the quality of the stimuli, especially when the Italian speaker spoke much more slowly than the German speaker. This might have influenced the judgement of lexical stress correctness of the stimuli in which local speech rate was manipulated. However, even for the words in isolation, in which the manipulation of local speech rate yielded high audio quality, the manipulation of intonation was more effective than local speech rate in correcting lexical stress.

5 Experiment 2: Pronunciation training for teaching Italian speakers lexical stress of German morphologically complex words

The aim of this experiment was to provide Italian speakers with pronunciation training so that they could detect the correct lexical stress pattern of German morphologically complex words. The feedback that formed part of this training was based on stimuli that had been resynthesised in their own voice with emphasised word stress after correcting for local speech rate, intonation, and intensity.

5.1 Method

5.1.1 Participants of the training

Twelve Italian native speakers participated in the training. Eight of them were the same speakers whose recordings were used for Experiment 1. The other four were two female and two male speakers coming from various regions of Italy. Thus, the participants were ten female and two male speakers, between 20 and 52 years old, and who had lived in Munich between two months and 23 years. Eight of them were advanced speakers of German, and four, including the two men, were beginners. The participants were randomly assigned to a test or to a control group. The participants of the test group were trained with resynthesised stimuli in their own voice, while the participants of the control group were trained with natural stimuli in the voice of the German native speaker. In order to avoid all beginner speakers being in the same group, one male and one female beginner speaker were randomly assigned to the test group and the other two to the control group.

5.1.2 *Stimuli*

The stimuli used in the training were the 14 sentences and 32 isolated words also recorded for Experiment 1. The same female German native speaker read the utterances again, this time putting emphasis on the stress position of the morphologically complex words. Her task was to imagine that a learner had stressed the wrong syllable, and that she should show him which syllable was lexically stressed. The recording conditions were the same as for Experiment 1.

The utterances by the test group containing wrongly stressed words were submitted to resynthesis of local speech rate, intonation, and intensity according to the method described in Sect. 3. Resynthesised stimuli with emphasised stress were also created by copying the prosodic parameters of the reference utterances with emphasised stress to the utterances of the Italian speakers.

The training material was different for each speaker, since it was based on the morphologically complex words each speaker had stressed wrongly. Other wrongly stressed words were also included in the training.

5.2 *Pronunciation training*

The pronunciation training was carried out about one week after the recording. It is most unlikely that the participants had learnt the correct pronunciation of lexical stress in the meantime. Firstly, they had no information about the aim and contents of the training. Secondly, even speakers who had lived in Germany for several years made numerous mistakes concerning lexical stress. They probably never noticed these mistakes in spite of their long stay in a full-immersion environment. The three Italian speakers who also took part in the first recording (Sect. 2) made exactly the same lexical stress mistakes in both recordings. This shows the errors were fossilised in their pronunciation.

The training consisted of two parts. Participants were first trained with normally stressed stimuli and then with stimuli with emphasised lexical stress. The training took place in the same environment as the recordings. A loudspeaker was placed in front of the participants for the feedback stimuli. Each participant received a list of the words and sentences as training material.

Before the training started, participants of the test group were told that they would hear their incorrect pronunciation and then a resynthesised utterance in their own voice after correcting for the intonation (melody), rate, and loudness of speech. They would hear each utterance twice. They were then instructed to pronounce the utterance twice again, adopting the correct pronunciation. At this point they would be recorded. This procedure would be carried out for all the utterances.

After the end of the first part of the training, the participants received the instructions for the second part. This time they were told that they would hear

for each utterance the same feedback as in the first part plus the resynthesised utterance in their own voice with emphasis placed on the stressed syllable, and that they would be recorded again. The training procedure for the control group was the same, except that the feedback consisted of the German native speaker's naturally produced utterances.

5.3 Results and Discussion

Feedback in the learners' own voice had a motivating effect. Several participants of the test group showed a keen interest in the training. They were pleased and surprised to hear themselves corrected, and some even asked if they could come back to do some more training of that kind. Obviously, the quality of the resynthesised stimuli was considered acceptable by the training participants. Self confidence and motivation are important for promoting learning (Eskenazi, 1999). The participants of the control group in contrast showed no particular interest in the pronunciation training. Some even declared that they were disappointed.

The resynthesised stimuli for the pronunciation training were created without previous segmentation of the speech data. This is important for using this technology in CALL-programs, since, as already mentioned in the introduction, users should receive their feedback immediately after pronouncing the incorrect utterance. Our study was carried out without immediate feedback for training participants for two reasons. Firstly, the material contained some long sentences that required manual alignment of speech pauses in the Italian speakers' signals with the signal of the German speaker. Secondly, we wanted to check the quality of all resynthesised utterances before offering them to the participants as feedback. Some resynthesised stimuli had to be rejected, because the presence of creaky voice or strong nasalization in the syllable to be stressed in the speech signals of the Italian speakers made stress correction more difficult. At this point we should mention that the Italian speakers had never heard a German native speaker reading the text before, and this is quite a different situation compared to the environment provided by a CALL-program. In a CALL-program, the users normally hear the stimulus pronounced by a reference native speaker of the target language before they have to pronounce it themselves. Thus, in an implementation of this technology in a CALL-program the resynthesis would be easier. Even though feedback was not offered to the learners during training immediately after their incorrect productions, it was because we played their incorrectly stressed utterances recorded in the previous session that the learners could directly compare the correct with their incorrect pronunciations just as if there had been immediate feedback.

completely correct word stress	
mostly correct word stress	
tending towards correct word stress	
tending towards wrong word stress	
mostly wrong word stress	
completely wrong word stress	
A. vorlassen	B. vorlassen
C. auf-schließen	D. auf-schließen
E. unruhiger	F. unruhiger
G. ausverkauft	H. ausverkauft
I. anschließen	J. eingetroffen
K. rücksichtslos	L. Menschenmenge
M. Schimpfwörter	N. durchlassen

Fig. 5. Interface of perception test for Experiment 3.

6 Experiment 3: Perceptual evaluation of learners' performance after feedback with resynthesis or emphasis

A perception test with German native speakers was carried out to assess the performance of the participants resulting from the training procedure presented in the previous section. The aim of the experiment was to find out: i) if feedback with resynthesised stimuli is more effective than feedback with natural stimuli, ii) if feedback with emphasised lexical stress is more effective than feedback with normally stressed stimuli, and iii) if feedback with resynthesised stimuli with emphasis is the most effective for learning lexical stress of German morphologically complex words.

6.1 Method

6.1.1 Stimuli

The stimuli presented in the perception test were the utterances produced by the Italian speakers recorded before and after the training in Experiment 2. There were therefore three versions of each utterance: i) pronounced before training, ii) pronounced after training with normally stressed stimuli, and iii) pronounced after training with stimuli including emphasised stressed syllables. The utterances were produced by the group trained with resynthesised stimuli in the learners' own voice and by the group trained with natural stimuli in the voice of the German native speaker.

Since the resulting number of stimuli was too high to be presented in a single perception test, we analysed only 255. As for Experiment 1, we chose the most frequent errors, leading to eight morphologically complex words in isolation

and four embedded in sentences.

6.1.2 Participants

The test participants were 37 German native speakers, between 21 and 53 years old. 25 of them were employees or students at the Institute of Phonetics and Speech Processing of the University of Munich. 19 had already participated in the perception test for Experiment 1. Approximately one month passed between both perception tests.

6.1.3 Perception test

The test user interface was similar to the one used for Experiment 1. This time each column on the interface contained three buttons: one button for the version produced before the training and two buttons for the versions produced after feedback with and without emphasis on the stress position (see Fig. 5). As in the test interface for Experiment 1, the buttons in each column and the speakers for each utterance were randomly ordered.

The test procedure was the same as in Experiment 1 (Sect. 4.1.3). Fig. 6 shows the individual results of one of the 37 participants.

6.2 Results

The 9435 judgements were converted to a range between 0 (the worst possible judgement) and 100 (the best possible judgement).

Table 4 shows the mean results for the three versions of the utterances (before

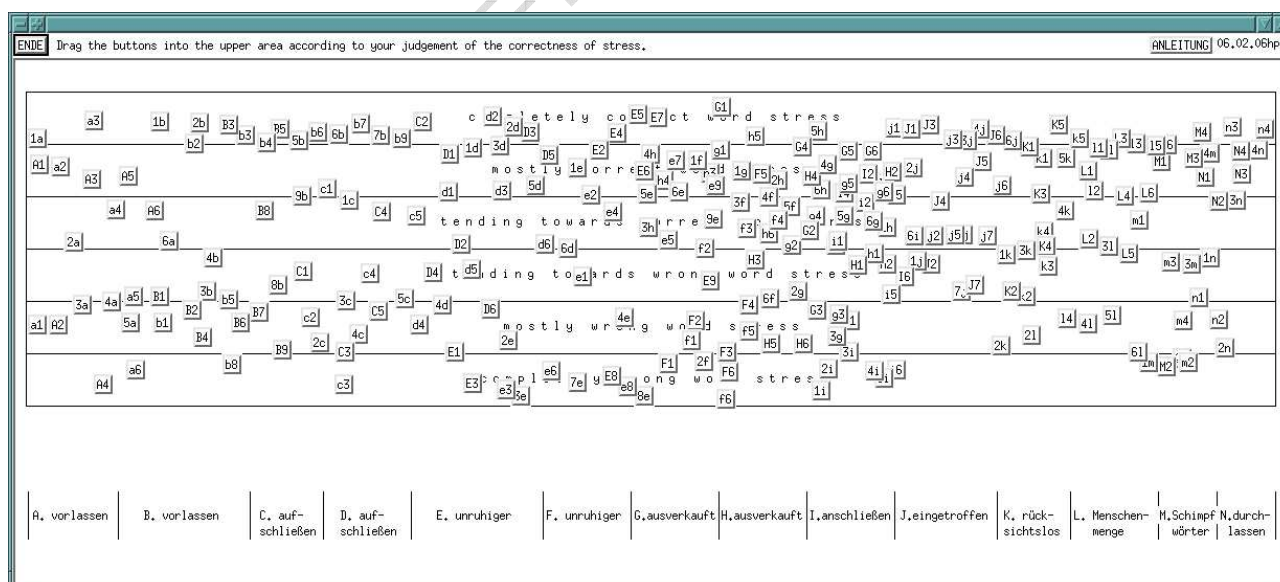


Fig. 6. Individual results of one of the 37 participants.

Table 4

Mean results for the three versions of the utterances by participants trained with natural or resynthesised stimuli.

	Participants trained with	
	natural stimuli	resynthesised stimuli
before training	37.35	38.13
after tr. with normally stressed stimuli	55.21	61.36
after tr. with stimuli with emphasis	62.70	63.80

training, after feedback with normally stressed stimuli, and after feedback with emphasis) for the participants trained with natural and with resynthesised stimuli. The utterances produced after feedback with resynthesised stimuli or after feedback with emphasis received the highest scores.

Fig. 7 shows the score difference before and after training. Again, utterances pronounced after feedback with resynthesised stimuli or after feedback with stimuli with emphasis on the stress position showed the greatest improvement.

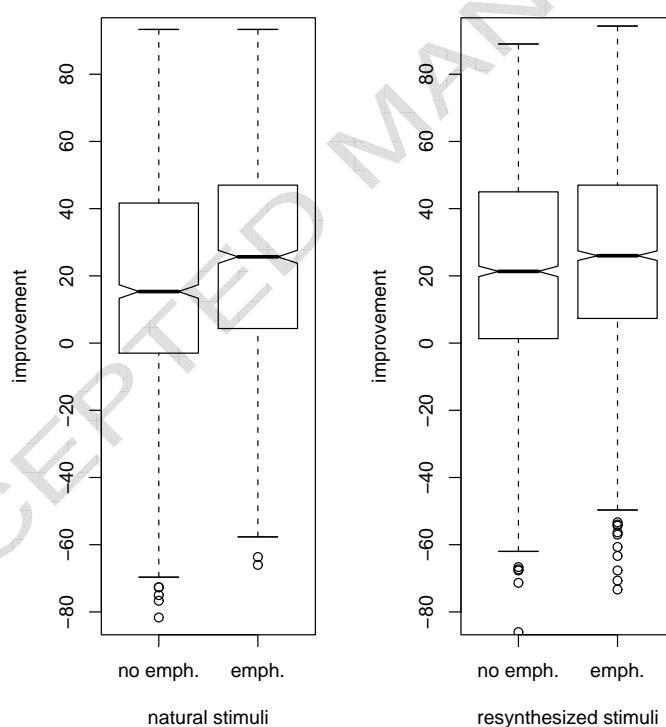


Fig. 7. Improvement (score difference before and after training) for the participants of the control group (trained with natural stimuli) and of the test group (trained with resynthesised stimuli) after feedback with normally stressed stimuli and after feedback with emphasis.

Since the data were not normally distributed, we used non-parametric tests for significance. We compared the utterances of the participants trained with resynthesised stimuli with the utterances of the participants trained with natural stimuli by means of the Wilcoxon rank sum test for unpaired samples. The difference between the pre-training utterances was not significant ($p = 0.1099$). The two groups can thus be considered to be of an equivalent level before training. We then compared the two groups trained with resynthesized and with natural stimuli on the extent of improvement in the post-training utterances after feedback without emphasis and found the group-difference to be highly significant ($p < 0.001$). The resynthesised stimuli without emphasis were thus more effective than the natural stimuli without emphasis. However, there was no significant difference between the groups on this measure for stimuli with emphasis.

To compare the pre- and the post-training utterances in each group, we used the pairwise Wilcoxon signed rank test. In the group trained with natural stimuli, we compared the pre- and post-training scores. Both post-training scores, after normally stressed feedback and after feedback with emphasised stress, were significantly higher than pre-training scores ($p < 0.001$ for both comparisons). The utterances after feedback with emphasis received significantly better judgements than those spoken after feedback without emphasis ($p < 0.001$). In the group trained with resynthesised stimuli, we compared pre- and post-training scores as well. As for the group trained with natural stimuli, both post-training utterances — after normally stressed feedback and after feedback with emphasis — received significantly better scores than the pre-training utterances ($p < 0.001$ for both comparisons). The difference between the post-training utterances (with and without emphasis) was almost significant ($p = 0.052$).

6.3 Discussion

As we predicted, feedback with natural stimuli was more effective if the stress position was emphasised, while for the normally stressed condition resynthesised were more effective than natural stimuli. We had not expected that resynthesised stimuli with emphasis would be as effective as natural stimuli with emphasis. It is possible that emphasis on the stress position was sufficient for the participants to recognise their error so that resynthesis in this case was not necessary, but there could be also another explanation. Several training participants imitated the stimuli with emphasis. We asked twelve of the German evaluators if they heard overstressed stimuli in the perception test. All of them reported that they did. Eight of them said although they judged them to be correctly stressed, since the Italian speakers stressed the right syllable, they gave these stimuli a lower score than if they had not been overstressed.

Since the Germans compared the two post-training versions of each utterance

— after feedback with and without emphasis — with each other (Sect. 6.1.3), their judgement that there was overstress in the version with emphasis could have been influenced by the correctness of the version without emphasis in the following two ways:

- (1) If the version after feedback without emphasis was stressed on the correct syllable, they might have judged a possible overstress in the version after feedback with emphasis more severely.
- (2) Alternatively, if the version after feedback without emphasis was stressed on the wrong syllable, it is possible that they considered overstress in the other version as a less serious mistake.

Since the group trained with resynthesised stimuli performed significantly better than the group trained with natural stimuli after feedback without emphasis, the explanation proposed in (1) might be why utterances after feedback with resynthesis and emphasis did not receive judgements as high as expected.

In a subsequent perception test 19 German natives (ages 22-48) were asked to assign the words produced by the Italians during the training to one of the following categories: i) stressed on the wrong syllable, ii) stressed on the correct syllable, and iii) stressed on the correct syllable but overstressed. Four of the 19 subjects had also participated in the previous perception test.

Pearson's correlation coefficients were calculated between the stress correctness judgements, taken from the previous perception test and averaged for each stimulus, and the respective frequencies (between 0 and 19) of each of the three categories. As expected, we found a strong negative correlation between stress correctness judgements and the frequency of assignments of stress on the wrong syllable (-0.91 , $p < 0.001$). There was a positive correlation between stress correctness judgements and the frequency of assignments of stress on the correct syllable (0.78 , $p < 0.001$). We found a weak positive correlation between stress correctness judgements and the frequency of assignments of overstress (0.37 , $p < 0.001$). The Pearson's correlation coefficient was also calculated between the stress correctness judgements and the ratio of *overstressed* to *correct*, i.e. category iii) divided by the sum of categories ii) and iii). Each stress correctness judgement was assigned a weight equal to the sum of categories ii) and iii). A weak negative correlation was observed (-0.11 , $p < 0.001$) indicating that words with a high percentage of overstress assignments received lower stress correctness judgements in the previous perception test.

Fig. 8 shows the median judgements of the three versions of the 52 utterances produced by the group trained with resynthesised stimuli, ordered according to the post-training versions without emphasis. The vertical axis of the diagram corresponds to the test interface (see Fig. 5).

Most utterances produced after training with emphasis (crosses) have been placed above the 50 mark, which means that they were judged as correctly

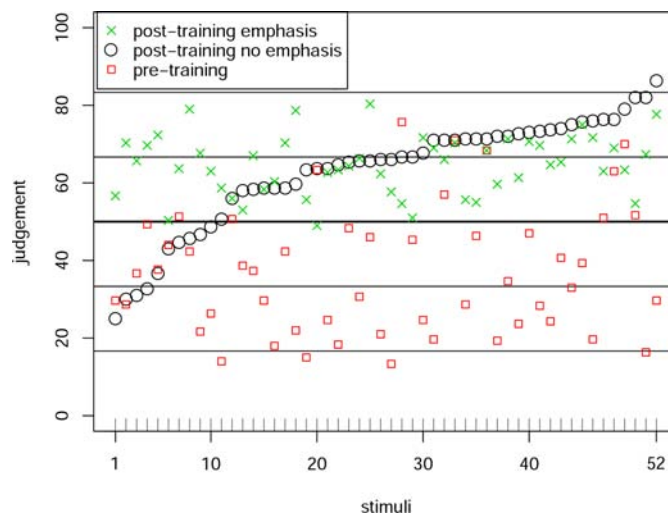


Fig. 8. Medians of the three utterance versions for the group trained with resynthesised stimuli. *Squares*: pre-training versions. *Circles*: versions after feedback with normally stressed stimuli. *Crosses*: versions after feedback with emphasis.

stressed. The first ten utterances on the left produced after feedback without emphasis (circles) are under the 50 mark, which means that in these cases feedback with resynthesised stimuli but without emphasis was not sufficient for the Italian participants to recognise the stress position. The corresponding utterances after feedback with emphasis (crosses) received better judgements, showing that emphasis helped the learners to recognise the lexically stressed syllable. However, about half of the utterances produced after feedback with emphasis received lower judgements than the corresponding utterances pronounced after feedback without emphasis, thus supporting the idea that the former were possibly “punished” by the evaluators because they were overstressed.

7 Conclusions

In order to learn a foreign language, learners should be made aware of its salient defining pronunciation characteristics. Even a full immersion in a foreign-language-speaking environment for several years is not sufficient if the learners do not receive indications about their own mistakes and suggestions on how to improve their pronunciation. This is also the case for Italian speakers of German. We found systematic lexical stress mistakes in German morphologically complex words even by Italian advanced learners of German living in Germany for ten years or more, who never noticed this mistake. Possibly, Italian speakers require a longer vowel segment duration in order to identify a stressed syllable, and misunderstand the second component of German morphologically complex words as being stressed, especially if it contains a long vowel. In this study, we tested two kinds of feedback to help Italian learners of

German notice the correct lexical stress position: resynthesised stimuli in the learner's own voice after manipulation of intonation, local speech rate, and intensity, and stimuli with emphasised stress position.

We developed a method for modifying and resynthesizing prosodic parameters that does not require manual segmentation of speech signals. Thus this method could be implemented in CALL programs since it allows the learner to receive immediate feedback and to compare directly the correct with the wrong pronunciation.

The effect of resynthesis on lexical stress perception was tested in a perception experiment with 31 German native speakers. The main outcome of this experiment is that the correction by means of resynthesis of intonation and local speech rate of wrongly stressed German morphologically complex words pronounced by Italian speakers significantly improves the judgement of stress correctness by German natives. Since lexical stress position can be corrected by means of resynthesis, we propose that resynthesised stimuli should be used as feedback in lexical stress pronunciation training. Another finding is that resynthesis of intonation is more effective in correcting lexical stress in the productions of Italian learners of German than resynthesis of local speech rate.

We tested the effectiveness of feedback consisting of resynthesis of the learners' own voices and of emphasized stress in training twelve Italian speakers to learn the lexical stress pattern of morphologically complex words in German. The learners' performance in the training was then assessed by means of a perception test with 37 German native speakers. The results of the test showed that feedback consisting of normally stressed resynthesised stimuli was more effective than normally stressed natural stimuli in the voice of the reference German speaker. In addition, resynthesised stimuli had a motivating effect on the learners. We also found that feedback consisting in natural stimuli was more effective with emphasised than with normal stress. Contrary to our expectations, there was no improvement in using synthesised compared with natural stimuli in the feedback condition with emphasis. There are two possible explanations for this result: i) emphasis was sufficient for the learners to notice the correct stress position so that resynthesis did not add new information, and ii) the overstressed utterances of the group trained with resynthesised stimuli received lower scores by the German evaluators than those of the group trained with natural stimuli (Sect. 6.3).

More investigations are necessary to test if the training with resynthesised stimuli or stimuli with emphasis also has long-term effects. This might be expected since the learners, once they become aware of a pronunciation rule such as lexical stress, should then be able to avoid systematic errors of this kind. It would be interesting to test the effects of these kinds of feedback on other prosodic factors, such as sentence accent or the reduction of function words compared to content words.

The results of our study show i) that resynthesised stimuli are a more effective form of feedback in teaching lexical stress than natural stimuli, and ii) that the resynthesis can be obtained without previous manual segmentation of the speech data. This technology can thus be recommended for implementation in CALL programs.

Acknowledgements

Portions of this work were presented at the 11th Australian International Conference on Speech Science & Technology 2006 in Auckland (Bissiri et al., 2006) and at the Italian Meeting of German Linguistics 2006 in Rome (Bissiri, 2007). Part of this work is also part of the PhD-thesis of the first author (Bissiri, 2008). The authors would like to thank J. Harrington, P. Hoole, E. O'Rourke, and three anonymous reviewers for ideas and comments on earlier versions of this article. The authors are also grateful to P.G. Bissiri for his valuable comments on the statistical analysis.

A Die Buttergeschichte

Es war in Berlin zu einer Zeit, als Lebensmittel nicht genügend vorhanden waren. Vor einem Laden stand bereits um sieben Uhr eine beachtliche Menschenmenge, denn man hatte dort am Abend vorher auf einem Schild schon lesen können, dass frische Butter eingetroffen sei. Jeder wusste, dass die Butter schnell ausverkauft sein würde, und dass man ganz früh kommen müsse, um noch etwas zu erhalten. Da das Geschäft erst um acht geöffnet wurde, stellten sich die Leute vor der Ladentür in einer Reihe an. Wer später kam, musste sich hinten anschließen.

Je näher der Zeiger auf acht kam, desto unruhiger wurden die Leute. Da kam endlich ein kleiner Mann mit grauem Haar und drängte sich ziemlich rücksichtslos nach vorn. Die wartenden Menschen waren empört über solches Verhalten und forderten ihn auf, sich ebenfalls hinten anzustellen. Aber auch als schon mit der Polizei gedroht wurde, ließ sich der Mann nicht beirren, sondern drängte sich weiter durch. Er bat, man solle ihn doch durchlassen, oder glaubte man, dass diese Drängelei für ihn vielleicht ein Vergnügen sei?

Das war für die Leute nun doch zu viel! Alle kochten bereits vor Wut, und der Mann konnte jetzt von allen Seiten Schimpfwörter hören. Er aber zuckte resigniert mit den Schultern und bemerkte: "Nun gut, wie Sie wollen. Wenn Sie mich nicht vorlassen, dann kann ich die Tür nicht aufschließen, und Sie können meinetwegen hier stehen bleiben, bis die Butter ranzig geworden ist."

References

- Anderson-Hsieh, J., 1992. Using electronic visual feedback to teach suprasegmentals. *System* 20 (1), 51–62.
- Anderson-Hsieh, J., 1994. Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal* 11 (4), 5–22.
- Bertinetto, P.M., 1980. Perception of stress by Italian speakers. *Journal of Phonetics* 8, 385–395.
- Bertinetto, P.M., 1981. *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici.* Accademia della Crusca, Firenze.
- Bissiri, M.P., 2007. Einfluss von Intonation und lokaler Sprechgeschwindigkeit auf die Betonung von durch L2-Sprecher ausgesprochenen Komposita, in: Di Meola, C., Gaeta, L., Hornung, A., Rega, L. (Eds.), *Perspektiven Zwei. Akten der 2. Tagung Deutsche Sprachwissenschaft in Italien, Ist. Italiano di Studi Germanici, Roma*, pp. 353–364.
- Bissiri, M.P., 2008. Experimentelle Untersuchungen zum computergestützten Erlernen der Prosodie des Deutschen für italienischsprachige Lernende. PhD-thesis, Ludwig-Maximilians University, Munich.
- Bissiri, M.P., Pfitzinger, H.R., Tillmann, H.G., 2006. Lexical stress training of German compounds for Italian speakers by means of resynthesis and emphasis, in: *Proc. of the 11th Australasian Int. Conf. on Speech Science and Technology, Auckland*, pp. 24–29.
- Bissiri, M.P., Pfitzinger, H.R., Tillmann, H.G., 2008. Lexical stress perception of German morphologically complex words by Italian speakers, in: *Proc. of the 4th Conf. on Speech Prosody, Campinas*, pp. 639–642.
- Chapelle, C.A., 1998. Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology* 2 (1), 22–34.
- Chun, D.M., 1989. Teaching tone and intonation with microcomputers. *CALICO Journal* 7 (1), 21–46.
- Chun, D.M. 1998. Signal analysis software for teaching discourse intonation. *Language Learning & Technology* 2 (1), 61–77.
- Cranen, B., Weltens, B., De Bot, K., van Rosum, N., 1984. An aid in language teaching: The visualization of pitch. *System* 12 (1), 25–29.
- De Bot, K., 1983. Visual feedback of intonation I: Effectiveness and induced practice behaviour. *Language and Speech* 26 (4), 331–350.
- Delmonte, R., 1981. *L'accento di parola nella prosodia dell'enunciato dell'Italiano standard.* Studi di Grammatica Italiana, Accademia della Crusca, Firenze, 69–81.
- Delmonte, R., 2000. SLIM prosodic automatic tools for self-learning instruction. *Speech Comm.* 30, 145–166.
- Delmonte, R., Petrea, M., Bacalu, C., 1997. SLIM prosodic module for learning activities in a foreign language, in: *Proc. of EUROSPEECH, Vol. 2, Rhodes*, pp. 669–672.
- Delmonte, R., Bortolato, G., 2003. Apprendimento dell'Intonazione tedesca e annotazione GToBI, in: Marotta, G., Nocchi, N. (Eds.), *Atti delle XIII*

- Giornate di Studio del Gruppo di Fonetica Sperimentale (A.I.A.), Pisa, 28th–30th November 2002, pp. 169–176.
- D’Imperio, M., 2000. Acoustic-perceptual correlates of sentence prominence in Italian, in: *The Ohio State University Working Papers in Linguistics*, No. 54, The Ohio State University, Columbus, OH, pp. 59–78.
- Dogil, G.; Williams, B., 1999. The phonetic manifestation of word stress, in: Hulst, H. v. d. (Ed.), *Word prosodic systems in the languages of Europe*, Mouton de Gruyter, Berlin, pp. 273–334.
- Dupoux, E., Peperkamp, S., Sebastián-Gallés, N., 2001. A robust method to study stress ‘deafness’. *J. of the Acoustical Society of America* 110 (3), 1606–1618.
- Eskenazi, M., 1999. Using a Computer in Foreign Language Pronunciation Training: What Advantages? *CALICO Journal* 16 (3), 447–469.
- Eskenazi, M., Hansma, S., 1998. The Fluency Pronunciation Trainer, in: *Proc. of the STiLL Workshop on Speech Technology in Language Learning*, Marholmen, Sweden, pp. 77–80.
- Eskenazi, M., Ke, Y., Albornoz, J., Probst, K., 2000. The Fluency Pronunciation Trainer: Update and user issues, in: *Proc. of the InSTiLL Workshop on Speech Technology in Language Learning*, Dundee, Scotland, pp. 73–76.
- Gass, S., 1997. *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Gass, S., Selinker, L., 1994. *Second Language Acquisition: An Introductory Course*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Germain-Rutherford, A., Martin, P., 2000. Présentation d’un logiciel de visualisation pour l’apprentissage de l’oral en langue seconde. *Apprentissage des Langues et Systèmes d’Information et de Communication (ALSIC)* 3 (1), 61–76.
- Hardison, D.M., 2004. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8 (1), 34–52.
- Hirose, K., 2004. Accent type recognition of Japanese using perceived mora pitch values and its use for pronunciation training system, in: *Proc. of the Int. Symposium on Tonal Aspects of Languages (TAL)*, Beijing, pp. 77–80.
- Hirose, K., Gendrin, F., Minematsu, N., 2003. A pronunciation training system for Japanese lexical accents with corrective feedback in the learner’s own voice, in: *Proc. of EUROSPEECH*, Vol. 4, Geneva, pp. 3149–3152.
- James, E., 1976. The acquisition of prosodic features using a speech visualizer. *Int. Review of Applied Linguistics in Language Teaching* 14, 227–243.
- Jessen, M., Marasek, K., Schneider, K., Clahßen, K., 1995. Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German, in: *Proc. of the 13th Int. Congress of Phonetic Sciences*, Vol. 4, Stockholm, pp. 428–431.
- Kohler, K.J., 1995. *Einführung in die Phonetik des Deutschen*. Erich Schmidt Verlag, Berlin.
- Kommissarchik, J., Kommissarchik, E., 2000. *Better accent tutor — Analysis*

- and visualization of speech prosody, in: Proc. of the InSTiLL Workshop on Speech Technology in Language Learning, Dundee, Scotland, pp. 86–89.
- Ladd, D. R., 1996. *Intonational phonology*. Cambridge University Press, Cambridge.
- Lane, H., Buiten, R., 1969. A self-instructional device for conditioning accurate prosody: Trends, in: Valdman, A. (Ed.), *Language Teaching*, Academic Press, New York, pp. 159–174.
- Lehiste, I., 1970. *Suprasegmentals*, MIT Press, Cambridge, Massachusetts.
- Lyster, R., 1998. Negotiation of form, recasts and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning* 48 (2), 183–218.
- Martin, P., 2004. WinPitch LTL II, a multimodal pronunciation software, in: Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venezia, pp. 177–182.
- Malagoli, G., 1946. *Laccentazione italiana: Guida pratica*, Sansoni, Firenze.
- Mengel, A., 2000. *Deutscher Wortakzent. Symbole, Signale*, Books on Demand Gmbh.
- Mennen, I., 2007. Phonological and phonetic influences in non-native intonation, in: Trouvain, J., Gut, U. (Eds.), *Non-native prosody. Phonetic description and teaching practice*, Mouton de Gruyter, Berlin, New York, pp. 53–76.
- Nagano, K., Ozawa, K., 1990. English speech training using voice conversion, in: Proc. of ICSLP, Vol. 2, Kobe, pp. 1169–1172.
- Neri, A., Cucchiaroni, C., Strik, H., Boves, L., 2001. The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning* 15 (5), 441–467.
- Nicholas, H., Lightbown, P.M., Spada, N., 2001. Recasts as feedback to language learners. *Language Learning* 51 (4), 719–758.
- Probst, K., Ke, Y., Eskenazi, M., 2002. Enhancing foreign language tutors — In search of the golden speaker. *Speech Comm.* 37 (3–4), 161–173.
- Sluijter, A.M.C., van Heuven, V.J., 1996. Spectral balance as an acoustic correlate of linguistic stress. *J. of the Acoustical Society of America* 100 (4), 2471–2485.
- Tillmann, H.G., Pfitzinger, H.R., 2004. The development of an advanced SLP-based system for the individual learning and fast training of speaking skills in a new foreign language, in: Proc. of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning, Venezia, pp. 17–20.
- Vardanian, R. M., 1964. Teaching English through oscilloscope displays. *Language Learning* 3–4, 109–117.
- Wang, Q., 2008. L2 Stress Perception: The reliance on different acoustic cues, in: Proc. of the 4th Conf. on Speech Prosody, Campinas, Brazil, pp. 635–638.
- WinPitch, 2005. Pitch Instruments Inc, <http://www.winpitch.com>, 31.12.2005.