Cross-language Perception of Word Stress

Hansjörg Mixdorff and Yu Hu***

*Faculty of Computer Science and Media, TFH Berlin University of Applied Sciences, Germany
mixdorff@tfh-berlin.de
**University of Science and Technology, Hefei, China
yuhu@iflytek.com

10.21437/Interspeech.2005-274

Abstract

This paper presents a study of the perception of Mandarin di-syllabic words by native speakers of German. It examines how speakers of an accent language perceive word stress in words from a tone language. A corpus of 15 sets of words with all possible combinations of the four tones of Mandarin was recorded by a professional speaker. In addition monotonized versions of the words were created. In a forced-choice listening experiment native speakers of German were asked to assess whether they perceived the word stress on the first or second syllable. Results include, inter alia, that words with two high tones, as well as the monotonized stimuli were predominantly perceived as carrying the word stress on the first syllable. Words with a falling tone on the second syllable were mostly classified as carrying stress on the second syllable, with the combination of low and falling tone yielding the highest score. Many combinations of tones, however, could not be identified as any of the two kinds. This suggests that though some tonal configurations in Mandarin are similar to German two-syllable word accent patterns and can be associated with the latter, others might be rather interpreted as pertaining to two mono-syllabic words, both of which are stressed.

1. Introduction

As is commonly known each lexical item of German has one syllable that carries the primary lexical stress. Lexical stress position in German depends on the origin of a word and is rather flexible. Therefore a closed formulation for its assignment does not exist. In simple native words the stress usually falls on the first syllable of the stem, generally the penultimate syllable of a word [1]. Hence there exist di-syllabic words with either stress on the first or on the second syllable. Although stress on the first syllable is generally more frequent (79.5%, count from German dictionary used in the TTS system at TFH Berlin, omitting inflectional forms), many loan words (Arˈmee - army, Biˈlanz - result, …) as well as pre-fixed verbs in the past tense (verˈstand - understood, beˈsaß - possessed, …), for instance, are stressed on the second syllable. Empirical studies (see, for instance, [2][3][4]) showed that stressed syllables generally feature a co-occurrence (in descending order of importance) of a distinct *F0* movement, a longer duration than unstressed syllables, as well as a peak of intensity. Figure 1 displays typical *F0* contours of the two German di-syllabic words "ˈdaher'-*'therefore'* and 'daˈher' - *'along'* uttered by a male speaker. It can be seen that syllables are considerably shorter when unstressed.

Syllabic tones in tone languages are connected with distinct *F0* patterns (rising, falling, etc.). Mandarin has four different lexical tones: high (1), rising (2), low (3), and falling (4) (commonly used tone indices are given in brackets). A so-called neutral tone (0) is assigned to syllables that do not exhibit a proper tonal target, mostly in certain mono-syllabic function words. Furthermore and independent of the tone they carry, syllables of Mandarin can be strong or weak, though the concept of stress is not well defined in Mandarin and subject to change when a word is embedded in an utterance. It mainly depends on the type of nuclear vowel of the syllables involved. The predominant structure of Mandarin words is di-syllabic.

In the study presented in this paper we examine how speakers of a stress-timed language, German, interpret di-syllabic words from a syllable-timed tone language, Mandarin. We investigate whether and how the relatively complex tonal contours of Mandarin can be mapped on similar contours of German. This question is of interest with respect to language acquisition, for instance, as well as foreign language education.

Furthermore we want to determine whether the judgments are entirely based on the shapes of the *F0* contours, or whether secondary factors such as the syllabic durations as well as syllabic structures come into play. Besides, in the context of the cooperation between TFH Berlin and USTC we pursue this work with regard to mixed-language speech corpus annotation, for developing a unified prosodic annotation scheme for corpora of Mandarin with embedded foreign words.

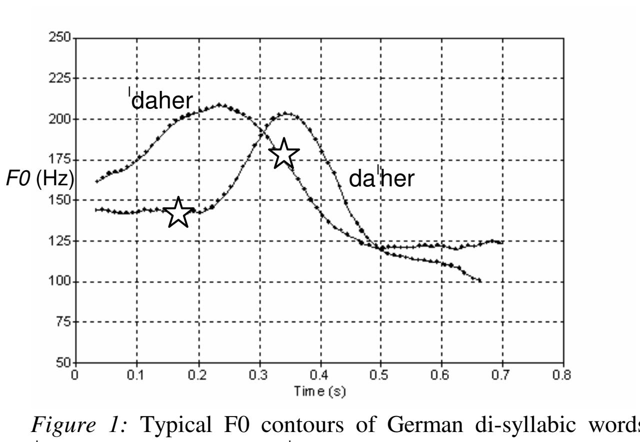

Figure 1: Typical F0 contours of German di-syllabic words "ˈdaher'-*'therefore'* and 'daˈher' - *'along'*. The stars denote the approximate locations of the inter-syllabic boundaries.

2. Speech Material and Method of Analysis

The speech material used in this study consists of 15 sets of words uttered in all possible tonal configurations yielding 4 x 4 = 16 tokens per set. We henceforth denote the tonal configurations with the numbers of tones involved. **32**, for instance, stands for a low tone 3 followed by a rising tone 2.

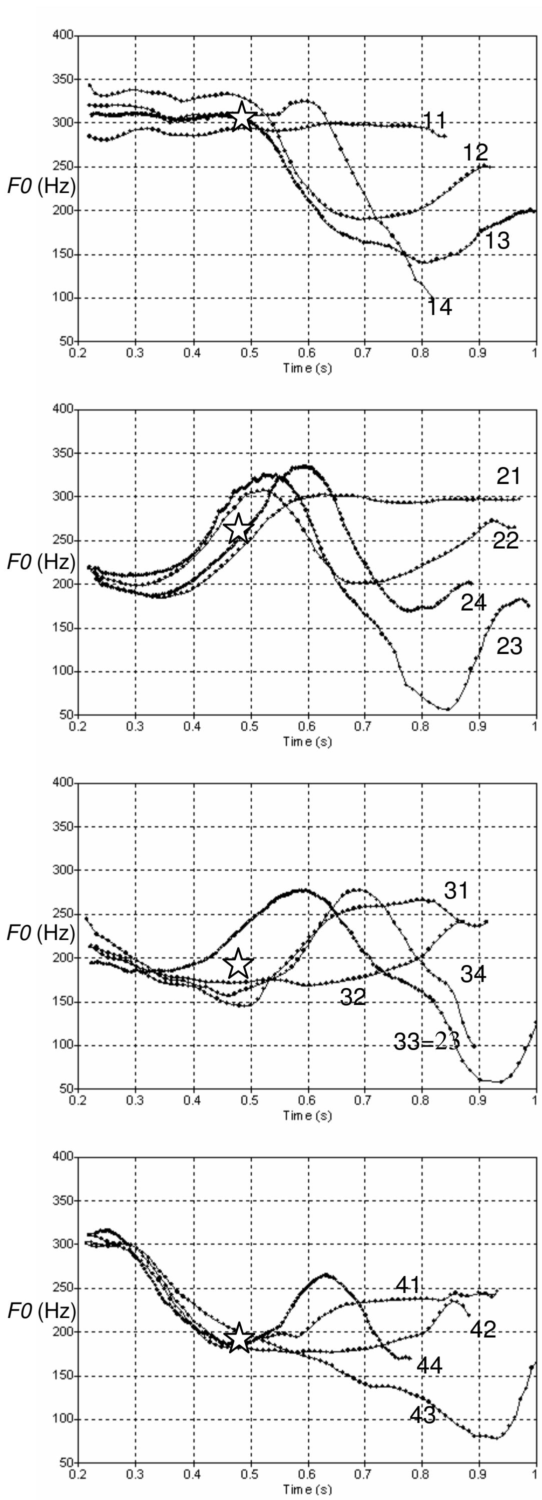

Figure 2: Example *F0* contours for all 16 tone combinations, word 'bengni'. The stars denote the approximate location of the inter-syllabic boundary.

It should be noted, that the combination **33** is replaced by **23** due to a tone sandhi rule.

A professional female speaker at iFlyTek Corporation, Hefei, China recorded all 15 x 16=240 words twice at 16 kHz, 16 bit. The words were then checked for the correct tone combinations. It should be noted that not all of the words were meaningful, but since the German participants of the perception test envisaged were not proficient in Mandarin this was of minor importance.

Table 1 shows a list of the syllabic combinations used. The criteria for choosing the words were a good coverage of (1) Mandarin vowels, and (2) different syllable structures (monophthongs/diphthongs, open/closed syllables). Due to the large number of resulting tone combinations and the expected cognitive load on the participants the selection was limited to 15 sets. The right column lists the classification with regard to syllable weights. Essentially, syllables with the nuclear vowel [a] are regarded as heavy.

Table 1: List of 15 'words' used in the study and associated syllabic weights according to the Mandarin concept of syllabic weight.

bengni	light-light
dabeng	heavy-light
feijia	light-heavy
huanlei	heavy-light
hupiao	light-heavy
jiazhu	heavy-light
leifei	light-light
mengshi	light-light
nimeng	light-light
piaozha	heavy-heavy
qiaoda	heavy-heavy
shihuan	light-heavy
weihu	light-light
zhaqiao	heavy-heavy
zhuwei	light-light

The combinations with two high-tones (**11**) were employed for creating monotonous stimuli at *F0*=220 Hz using the PSOLA resynthesis functionality of the *Praat ManipulationEditor* [5]. Figure 2 displays *F0* contours calculated for the 16 tonal combinations of the word 'bengni' calculated at a step of 10 ms using the *Praat* pitch estimation. The contours are labeled with the corresponding tonal configuration. The approximate position of the inter-syllabic boundary is indicated by a star. As can be seen in all examples the second syllable is longer than the first one, especially when it carries a low tone 3. This tone is also often characterized by vocal fry and extremely low *F0* values. As mentioned before, the configuration **33** corresponds to **23**.

In order to quantify the effect of the syllable durations on the perceivers' judgment we manually labeled all tokens on the syllabic level.

3. Perceptual Experiment

The perception test was conducted during a lecture on Digital Audio Processing with 23 students (19 male, 4 female) in their second year of Multimedia Computer Science, aged 21-35. The 240 natural stimuli were augmented by two repetitions of the 15 monotonous stimuli yielding a total of 270 stimuli arranged in a randomized list, while ensuring that consecutive tokens pertained to different word sets. The experiment was preceded by a preparatory phase in which the participants where explained the task. To this effect, initially German examples were presented and participants asked to classify them with regard to their stress positions. Subsequently, the experimenter presented nonsense words and asked for the subjects' judgments. Then the Mandarin words were played back from a notebook over a high-quality stereo speaker system. Each stimulus was preceded by its number announced by the experimenter. In a forced decision participants were asked to note on a form which syllable they had perceived as stressed. After the first 135 tokens a 3 minute break was taken. The whole experiment required about 25 minutes to complete.

4. Results

An overview of the experiment results for the natural stimuli is listed in Table 2 which presents the participants' pooled judgments as percentages of the vote 'stress on second syllable' depending on the tonal configuration. The table also displays the number of judgments and the standard deviations. As can be seen from the total result there is a certain tendency towards the judgment 'second syllable'. All results are highly significant ($p < 0.01$). It comes a bit as a surprise that the configuration **11** is the one that yields the highest ratings for carrying stress on the first syllable. On the other end of the spectrum, the configuration **34** is predominantly classified as being stressed on the second syllable. If we compare the *F0* contour of **34** in Figure 2 with the contours of German words in Figure 1 we see that it is the best matching one.

Table 2: Result of perception test pooled from 23 subjects and 15 syllable sets. Percentage of judgment 'second syllable stressed' depending on the tone combination of the word presented (natural stimuli).

Tone Combination	Percentage "2nd syllable stressed"	Total N judgments	Standard deviation
11	18.4	368	14.6
12	41.4	368	13.9
13	50.7	368	18.8
14	57.4	368	14.6
21	58.3	368	15.2
22	54.9	368	15.0
23	51.2	368	13.9
24	73.3	368	14.0
31	68.7	368	11.2
32	69.8	368	17.8
33(=23)	52.8	368	15.7
34	81.7	368	7.5
41	55.4	368	14.6
42	44.9	368	12.9
43	47.2	368	15.8
44	58.9	368	11.5
Total	55.3	5888	19.9

Furthermore, in the groups with the same tone on the first syllable, the combinations with tone 4 on the second yield the highest scores of 'second syllable stressed'. This again can be explained by referring to the contours in Figure 2: Unless preceded by Tone 1, Tone 4 exhibits a characteristic rising and falling *F0* pattern similar to the one of a German word stressed on the second syllable.

31 and **32** which still yield almost 70% for carrying stress on the second syllable are also similar with respect to their *F0* contours. All remaining configurations lie in a range of +/- 10 percent around the decision boundary. In these cases participants were obviously unable to make a clear judgment, especially when the second syllable carries tone 3.

One could interpret this outcome in two ways: The listeners are unable to make a clear judgment because the tokens presented cannot be placed into any of their native language's categories. Secondly, the tokens might present conflicting patterns, that is, they cannot be interpreted as a di-syllabic word, but rather as a sequence of two mono-syllables, both of which are stressed. This kind of pattern, i.e. two consecutive stressed syllables, also called 'stress clash', may occur, but a strong tonal marking of both syllables is usually avoided, except in exaggerated speech, for instance. In this light, the question which of the syllables is stressed therefore would not make any sense.

Table 3 lists the pooled results for the natural stimuli depending on the word involved as percentages of the vote 'stress on second syllable'. These are as well highly significant ($p < 0.01$). As can be seen, judgments for most of the words lie in a range +/-10% around the decision boundary, and obviously depend on the particular combination of syllables in each word. Due to the small number of words generalizations are therefore problematic. The figures suggest that some words with open syllables in initial position were more often perceived as stressed on the first syllable (dabeng, nimeng, jiazhu) than others. When the same open syllables are in word-final position they also draw a high proportion of votes 'second syllable stressed'. This partly matches the Mandarin concept of heavy syllables, since many of these syllables contain the vowel [a]. The table also lists the ratios of duration between second and first syllables. Analysis shows that longer second syllables are not necessarily perceived as stressed.

Table 3: Result of perception test pooled from 23 subjects and 15 syllable sets. Percentage of judgment 'second syllable stressed' depending on the word presented (natural stimuli), and mean ratios between duration of second and first syllables.

Word	Percentage "2nd syllable stressed"	Total N judgements	Standard deviation	Mean Duration Ratio
bengni	57.6	368	16.2	1.30
dabeng	33.2	368	18.1	1.48
feijia	64.5	368	17.0	1.27
huanlei	57.0	368	17.7	1.03
hupiao	54.6	368	17.0	1.37
jiazhu	48.0	368	18.0	1.22
leifei	50.2	368	14.7	1.59
mengshi	63.2	368	16.5	1.78
nimeng	41.5	368	21.4	1.83
piaozha	60.5	368	17.2	1.10
qiaoda	66.4	368	13.6	0.98
shihuan	60.1	368	26.4	1.23
weihu	59.4	368	24.5	1.72
zhaqiao	53.5	368	15.5	1.75
zhuwei	60.2	368	18.4	1.08
Total	55.3	5888	19.9	1.38

It is unclear, however, why the syllable 'shi' in 'mengshi' which contains a schwa-like vowel is more often perceived as stressed than in the word 'shihuan', as schwas generally do not occur in lexical stress syllables of German. One could speculate that this is due to the fact that the vowel in 'meng' as well is schwa-like, but with a darker quality than the one in 'shi'.

Table 4 lists the pooled results for the monotonized stimuli depending on the word as percentages of the vote 'stress on second syllable'. As can be seen, all words are predominantly perceived as being stressed on the first syllable. This result matches that for tonal configuration **11** in the natural ones. Hence, when presented words with relatively flat *F0* contours, German subjects perceive the stress on the first syllable. This is somewhat surprising when considering that the durations of the second syllable in all words are longer than that of the first syllable, as can be seen in the right-most column of Table 4. One explanation for the German listeners' bias towards stress on the first syllable might be the fact, that stress on the second is generally connected with a shortened first syllable which is not the case in the Mandarin words. With respect to the influence of syllable structures and nuclear vowels the results are very different from those in Table 3, since all show a preference for the percept that the first syllable is stressed, though to varying degrees. The syllable 'ni' is still more often perceived as stressed, but the syllable 'da' is not. Once again, generalizations are difficult because every syllable combination is specific, and the participants' judgment is therefore probably influenced by several factors such as the duration ratio of the syllables and the vowels involved. Also in the monotonized cases the ratio of syllable durations is not correlated with the jugment of the listeners.

Table 4: Result of perception test pooled from 23 subjects and two presentations. Percentage of judgment 'second syllable stressed' depending on the syllables of the word presented (monotonous stimuli), and mean ratios between duration of second and first syllables.

word	Percentage "2nd syllable stressed"	Total N judgements	Standard deviation	Mean duration ratio
bengni	20.3	46	16.4	1.60
dabeng	13.0	46	12.3	1.53
feijia	19.6	46	9.2	1.22
huanlei	21.7	46	6.1	1.11
hupiao	23.0	46	7.9	1.62
jiazhu	17.4	46	6.1	1.20
leifei	10.9	46	3.1	1.62
mengshi	21.7	46	6.1	1.83
nimeng	0.0	46	0.0	1.89
piaozha	10.9	46	9.2	1.23
qiaoda	8.7	46	2.9	1.04
shihuan	4.3	46	2.1	1.48
weihu	6.5	46	3.1	1.77
zhaqiao	19.6	46	3.1	1.71
zhuwei	19.6	46	3.1	1.00
Total	14.5	690	8.8	1.46

5. Discussion and Conclusions

This paper presented a study of stress perception in Mandarin words by German native listeners. We examined how subjects without prior knowledge of a tone language assess the tonal combinations on di-syllabic words in the light of their word stress system. To this effect natural recordings of Mandarin words, as well as monotonized versions of some of these were presented and classified with respect to the position of the stressed syllable.

Our results indicate that words with a falling tone on the second syllable are perceived as stressed on the second syllable. This result can be explained by the configurational similarities of *F0* patterns. Words with two high tones, as well as those with monotonous *F0* contours were most often classified as stressed on the first syllable. This suggests that in the absence of tonal cues the durational characteristics come into play, and although the Mandarin words generally have longer second syllables, stress on the second syllable in German implies considerable shortening of the first syllable. In all cases we did not find any influence of the ratio of syllabic durations on listeners' judgments.

Many tonal configurations, especially those with low tones on the second syllable, can not be interpreted by the German listeners, either because they do not match any of the available categories, or as we assume, because they actually present examples of stress clash, that is, must be interpreted as two mono-syllables, both of which stressed. Due to the small number of words employed only tentative judgments with respect to the influence of vowels and syllable structures were possible.

Therefore, we plan to investigate more in detail the effect of these factors, as well as the perception of monotonized versions other than configuration **11**. Furthermore, we will conduct an inverse experiment, in which native speakers of Mandarin will be presented German words and asked to classify them according to their tonal system.

6. Acknowledgements

This study was supported by a grant from the "Germany/China Joint Research Programme" of the German Academic Exchange Service and the China Scholarship Council. The assistance of Tasso Stumpe during data acquisition and the statistical lexicon survey is gratefully appreciated.

7. References

[1] Kohler, K.J., , "Einführung in die Phonetik des Deutschen", Erich Schmidt Verlag, Berlin, 1977.

[2] Isačenko, A.V., Schädlich, H.J., "*Untersuchungen über die deutsche Satzintonation*", Akademie-Verlag, Berlin, 1964.

[3] Altmann, H., Batliner, A., and Oppenrieder, W. "*Zur Intonation von Modus und Fokus im Deutschen*", Reihe Linguistische Arbeiten. Niemeyer-Verlag, Tübingen, 1989.

[4] Mixdorff, H., Fujisaki, H. "Word Accent in German and Japanese - A Comparative Study Using a Quantitative Model." In *Reports of the Spring Meeting of the Acoustical Society of Japan*, vol. 1, pages 313, 314, 1995.

[5] http://www.praat.org