JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Comparing fundamental frequency of German vowels produced by German native speakers and Mandarin Chinese learners

Yingming Gao,[1,a] **Hongwei Ding,**[2,b] **Peter Birkholz,**[1] **and Yi Lin**[2]
[1]*Institute of Acoustics and Speech Communication, TU Dresden, Dresden 01069, Germany*
[2]*Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai 200240, China*
yingming.gao@mailbox.tu-dresden.de, hwding@sjtu.edu.cn, peter.birkholz@tu-dresden.de, carol.y.lin@sjtu.edu.cn

Abstract: This study compared the f_0 of 14 German vowels in monosyllabic words (/dVt/) embedded in carrier sentences produced by 30 native speakers and 30 Mandarin Chinese learners. Appropriate techniques were employed to robustly measure f_0 values and reliably analyze f_0 profiles. The results showed that Mandarin learners produced the vowels bearing sentence stress with significantly larger f_0 ranges and steeper f_0 slopes but comparable f_0 mean and maximum in comparison to German natives. Moreover, lax vowels produced by both groups demonstrated narrower ranges with faster f_0 changes than tense vowels, which was stronger for Mandarin learners.

[Editor: Charles C. Church] https://doi.org/10.1121/10.0005593

Received: 11 February 2021 **Accepted:** 22 June 2021 **Published Online:** 14 July 2021

1. Introduction

Compared with the extensive studies on Mandarin learners speaking English (Chen *et al.*, 2001a; Jin and Liu, 2013; Yuan and Liberman, 2014), acoustic analyses of their production of German as a second language (L2) are still limited. Studies on L2 speech learning usually start with the acquisition of vowel segments in the target language. In German, without consideration of schwa /ə/ or long tense /ɛː/, there are 14 monophthongs that can be grouped into seven pairs, the members of which differ exclusively with respect to tenseness. The phonetic-acoustic differences of the German tense-lax opposition in production may manifest through changes in vowel formants, duration, and f_0 (Schneeberg and Schlüßler, 2006). The former two aspects have been compared between German native speakers and Mandarin Chinese learners (Gao *et al.*, 2020), while the last factor (f_0 difference) remains to be investigated. Though previous studies have revealed that L2 German produced by Mandarin learners has higher f_0 mean and larger f_0 range on both sentence and phoneme levels compared with that produced by German natives (Ding *et al.*, 2006), they have rarely concerned the f_0 profiles associated with the tense-lax vowel contrast. Vowel intrinsic f_0 (IF0) was proved to be a language universal (Whalen and Levitt, 1995). It has been shown that high vowels have a higher intrinsic f_0 than low vowels and that intrinsic f_0 also plays an important role in distinguishing the vowel identity. Therefore, the tense-lax contrast of vowels should be evident not only in formants and duration, but also in f_0.

Unlike the role of f_0 in German vowels to signal stress and possibly tenseness, f_0 in Mandarin vowels is associated with a lexical tone and employed to distinguish lexical meanings. Moreover, Mandarin Chinese monophthongs are usually classified as tense vowels, that is, there are no tense-lax contrasts in Mandarin Chinese. Such f_0 contrasts are supposed to be employed by native German speakers to distinguish between tense and lax vowels (Schneeberg and Schlüßler, 2006), while L2 Mandarin Chinese learners may not use the same f_0-related strategy due to the different roles of f_0 in their native tone language. In addition, vowel intrinsic f_0 differences are also dependent on the prosodic context in running speech, and the intrinsic f_0 difference should be maintained when the vowels bear the main phrasal stress (Shadle, 1985). Regarding the interaction between the intrinsic f_0 and the prosodic environment, we predicted that Mandarin Chinese learners might have some difficulties in using f_0 properly when they speak the non-tone language German. To address this specific issue, the current study aims to compare the f_0 profiles of German vowels produced by German native speakers and Mandarin Chinese learners with a particular interest in the tense-lax contrast under sentence stress.

[a]ORCID: 0000-0001-5881-3723.
[b]Author to whom correspondence should be addressed.

2. Methods

2.1 Participants

Two groups of speakers were recruited for the study, namely, a German native group (DEU) and a Chinese L2 learner group (CHN). The DEU group consisted of 30 German students studying at the TU Dresden with a mean age of 23.6 years (range: 18–38), while the CHN group included 30 Chinese L2 learners of German with an average age of 24.1 years (range: 18–31). Some CHN speakers were students majoring in German at Shanghai Jiao Tong University who had passed the nationwide unified examination for German students at the senior level of PGH (Prüfung für das Germanistik-Hauptstudium), and the others had passed the required German language examination (up to DSH-2 or DAF-16) before taking up their studies in German at TU Dresden. Speakers were exactly gender-balanced, i.e., 15 male and 15 female speakers in each group. Although they were born in different regions of their countries, the speakers in both groups had no strong regional accents. For example, all CHN speakers had achieved Grade Two Level B or above on the national standard Mandarin proficiency test (Putonghua Shuiping Ceshi), and most of them had less than one year of experience living in Germany after the age of 18 years. All participants, according to their self-reports, had normal speech and hearing functions with no history of any communication disorders.

2.2 Data collection

First, we embedded all 14 German vowels in monosyllabic words (/dVt/) to ensure that the speakers could produce the target vowels in a natural way. To create a systematic orthographic contrast, an "h" or an additional "t" was placed after the target vowel to indicate a tense or a lax vowel, respectively. Thus, we obtained 14 words, and most of them were nonsense but legal phoneme strings according to German phonotactic rules. These 14 words consisted of seven pairs with their International Phonetic Alphabet (IPA) transcriptions in parentheses as follows: *daht-datt (/daːt/-/dat/), deht-dett (/deːt/-/dɛt/), diht-ditt (/diːt/-/dɪt/), döht-dött (/døːt/-/dœt/), doht-dott (/doːt/-/dɔt/), düht-dütt (/dayyːt/-/dʏt/), duht-dutt (/duːt/-/dʊt/)*. The regular spelling patterns facilitated the correct grapheme-to-phoneme conversion for the speakers, so that the target vowels could be easily elicited. Moreover, we put each of the target words in a carrier sentence, "Ich habe /dVt/ gesagt (I have said /dVt/)," to ensure a stable prosodic context. By randomizing each set of the 14 sentences five times, we created a reading list of 70 sentences, and thus it was guaranteed that each vowel was produced five times with different intra-group orders by each speaker. The speakers were told that they should read all the sentences as naturally as possible with a short pause between them. After a period of familiarization and practice, all the speakers chose to place a pitch accent on the target word automatically. This way, f_0 values of target vowels under sentence stress could be elicited in an implicit way with well-controlled prosody. Though several CHN recordings were made in Shanghai and the others were in Germany, we ensured the same instructions and conditions. All recordings took place in a studio equipped with a recording console (Behringer Eurorack MX1602). The microphone (Microtech Gefell M930) was placed at a distance of approximately 20 cm from the speaker's mouth. All utterances were recorded with a sampling rate of 44.1 kHz and a quantization of 16 bits. The experiment lasted for about 5 min for each speaker, and they were financially compensated for their participation.

2.3 Acoustic measurements

In the first step, an automatic forced-alignment was carried out via the WebMAUS service (Kisler *et al.*, 2017) on both word and phoneme levels, outputting a TextGrid format annotation of Praat (Boersma and Weenink, 2019). Based on the derived word-level boundaries, we segmented all the recordings into 4200 individual sentences (14 vowels × 5 repetitions × 15 speakers × 2 genders × 2 language groups). Inaccurate alignments from the automatic phoneme annotation were manually adjusted by a phonetic expert by taking into account changes in both waveforms and spectrograms as well as perception cues if necessary.

A five-step procedure was applied to achieve a high accuracy of f_0 estimation. The first step was to extract the fundamental frequencies by the pitch-tracker developed by Shi *et al.* (2019), the analysis window of which was set at a length of 30 ms with 5 ms shift. A robust f_0 estimate and a voicing probability for each frame of speech were obtained. In the second step, we carried out the pitch tracking through a two-pass procedure following the strategy proposed by Hirst (2011) by calculating a more accurate f_0 range for f_0 estimation. In the first pass, we inspected our data and set a more accurate search range of 150–400 Hz and 75–300 Hz for female and male speakers, respectively, to cover all reasonable f_0 samples, and we extracted the f_0 with this range. Then we calculated the first and third quartiles (i.e., q_1 and q_3) across all f_0 samples for each speaker. In the second pass, the f_0 floor and ceiling for each speaker were set to $0.75q_1$ and $1.5q_3$, respectively. By using a personalized search range, we greatly reduced the estimation errors of f_0 extraction. This was confirmed by comparing speakers' f_0 histograms, in which long tails disappeared and samples were more centralized around the mean values. In the next step, a frame of speech with a voicing probability smaller than 0.5 was automatically removed from the data for f_0 extraction because these f_0 values were considered unreliable according to Shi *et al.* (2019). The fourth step was incorporated to deal with creaky voice that was frequently produced by several speakers. In glottalized periods, the corresponding f_0 was estimated by another pitch-tracker (Drugman and Alwan, 2011), which was more robust to

glottalization. Finally, we applied a median filter with a window of seven f_0 samples to smooth the f_0 contour. Manual corrections were only applied when the values were still wrong during the final check. For example, if the f_0 samples with voicing probabilities slightly larger than 0.5 were actually voiceless, we had to make necessary corrections manually.

2.4 Acoustic analysis

Based on the optimized f_0 samples, we calculated the f_0 mean, f_0 range (maximal f_0 minus minimal f_0), and f_0 slope of each target vowel. Following Lehiste and Peterson (1961), we also measured the f_0 maximum as a complement of f_0 mean. We further measured the positions of the f_0 maximum and minimum of each target vowel. Each vowel variable for a specific speaker was the average of his/her five repetitions of this vowel. We adopted two approaches to make the acoustic analysis more precise and robust.

One statistic approach to alleviate the influence of physiological differences efficiently was to convert the physical measurement of f_0 to the perceptual variable of f_0 using speaker-specific bases, which made the f_0 produced by all speakers comparable across gender. In previous studies, f_0 was usually converted from the Hz scale to the semitone (St) scale with a fixed value as a reference, which did not change the relative relationship between them due to the monotonic property of the logarithmic function (Chen *et al.*, 2001b; Ding *et al.*, 2006; Zhang *et al.*, 2008). In the current study, we adopted the reference proposed by Yuan and Liberman (2014), where each f_0 in Hz was transformed to a St value according to Eq. (1), in which the $f_{0,\text{base}}$ was the speaker-specific 5th percentile of all f_0 in Hz scale

$$f_0[\text{St}] = 12\log_2\left(\frac{f_0[\text{Hz}]}{f_{0,\text{base}}}\right). \tag{1}$$

Another statistical approach was to measure the f_0 slope by conducting a linear regression with time as an independent variable and f_0 as the dependent variable, which was more robust than the usual practice of dividing the absolute f_0 range (difference between the maximal and minimal f_0 values in St) by the duration of the vowel. The slope we obtained could thus characterize the dynamic movements of f_0 contours, where a positive slope represented an overall rising pattern, and a negative slope indicated an overall falling one. The absolute value of the slope reflected the steepness of the rising or falling. In the case of an f_0 contour containing two parts (LH plus HL or HL plus LH), the value of the slope was dominated by the longer part or the relatively steeper part, i.e., the dominant part contributed more to the direction of the estimated slopes.

3. Results

A series of linear mixed-effects (LME) models were run in MATLAB (MathWorks, 2019), where SpeakerGroup, Gender, VowelIdentity, or Tenseness and their interactions were treated as fixed effects and Subject as a random effect for intercept, while the acoustic parameters (f_0 mean, maximum, range, or slope) were the dependent variables. We first fitted linear regression models to the data using the "fitlm" function and then computed the analysis of variance (ANOVA) statistics for each variable using the "anova" function. The variables of the best models were selected through the backward selection procedure using the "compare" function.

3.1 f_0 *mean and maximum*

Average group values for f_0 mean (in Hz), maximum (in Hz), and f_0 mean (in St) of the *target vowels* are shown in the first, second, and third rows, respectively, in Table 1. It can be observed that the female speakers produced higher f_0 mean (245 Hz) and maximum (261 Hz) (measured in Hz) than the male speakers (137 Hz; 147 Hz); also, high vowels (199 Hz for /iː ɪ yː ʏ uː ʊ/) were associated with higher f_0 than low vowels (179 Hz for /aː a/).

The results for f_0 mean or maximum (in Hz) showed similar patterns: significant effects were found for Gender and VowelIdentity but not for SpeakerGroup. The LME regression for f_0 in St revealed significant effects of VowelIdentity [$F(13,784) = 92.25$, $p < 0.001$] but non-significant effects of both SpeakerGroup [$F(1,784) = 0.02$, $p = 0.877$] and Gender [$F(1,784) = 0.28$, $p = 0.599$]. However, the effect of Gender was statistically significant for f_0 mean in Hz [$F(1,784) = 282.62$, $p < 0.001$]. In other words, by transforming f_0 from Hz to St with Eq. (1), we preserved the difference of f_0 mean due to vowel intrinsic f_0 effects but reduced the difference due to the gender effect. A significant interaction effect on f_0 mean in St was found between SpeakerGroup and VowelIdentity [$F(13,784) = 10.85$, $p < 0.001$]. For example, the f_0 mean (in St) of vowels /uː/ and /ʏ/ ranked as the first and seventh among 14 monophthongs for the DEU speakers, respectively, while they ranked as the sixth and fourth for the CHN speakers, respectively. Moreover, the effect of Tenseness was significant [$F(1,838) = 45.11$, $p < 0.001$], with lax vowels generally having a higher f_0 than their tense counterparts (6.06 St versus 5.6 St). This test was conducted for group mean differences between tense and lax vowels, and no *post hoc* test was applied to each individual pair.

The results further showed that, compared to the DEU speakers, the CHN speakers produced vowels with comparable f_0 mean (both in Hz and St) and maximum (in Hz), and they demonstrated similar intrinsic f_0 values among different vowels. As f_0 expressed in St could neutralize anatomy-based acoustic differences while retaining phonemic differences, we analyzed f_0-related parameters in St hereafter.

Table 1. f_0 mean in Hz, maximum in Hz (in parentheses), and mean in St (in brackets) of German vowels produced by DEU and CHN speakers, where F = female and M = male.

	iː	ɪ	yː	ʏ	uː	ʊ	eː	ɛ	øː	œ	oː	ɔ	aː	a
DEU-F	249	265	256	264	255	261	237	236	237	236	238	237	218	232
	(265)	(271)	(273)	(271)	(274)	(269)	(256)	(246)	(256)	(246)	(258)	(248)	(238)	(243)
	[5.91]	[6.96]	[6.37]	[6.87]	[6.29]	[6.67]	[5.02]	[4.93]	[5.02]	[4.88]	[5.03]	[4.97]	[3.51]	[4.62]
CHN-F	244	253	253	259	247	255	240	245	246	255	236	243	232	240
	(264)	(265)	(274)	(272)	(270)	(269)	(262)	(260)	(267)	(268)	(261)	(261)	(255)	(256)
	[5.74]	[6.42]	[6.35]	[6.8]	[5.91]	[6.57]	[5.42]	[5.88]	[5.84]	[6.53]	[5.15]	[5.73]	[4.8]	[5.46]
DEU-M	139	145	142	146	140	144	129	132	129	132	127	131	121	128
	(146)	(147)	(150)	(148)	(148)	(148)	(136)	(135)	(136)	(136)	(135)	(135)	(129)	(132)
	[6.84]	[7.51]	[7.21]	[7.59]	[6.89]	[7.35]	[5.53]	[5.89]	[5.5]	[5.98]	[5.3]	[5.79]	[4.49]	[5.41]
CHN-M	143	144	143	145	144	147	138	136	138	141	136	137	130	136
	(159)	(154)	(160)	(158)	(162)	(159)	(152)	(147)	(154)	(152)	(153)	(149)	(147)	(147)
	[6.03]	[6.3]	[6.02]	[6.31]	[6.15]	[6.63]	[5.41]	[5.15]	[5.54]	[5.86]	[5.15]	[5.42]	[4.48]	[5.17]

3.2 f_0 range

The f_0 ranges of each vowel produced by the DEU and CHN speakers are shown in Fig. 1. The effect of SpeakerGroup [$F(1,784) = 12.71$, $p < 0.001$] was significant, with the CHN speakers' target vowels having larger f_0 ranges than those of the DEU speakers (3 St versus 1.75 St), which reflected the same trend as with the Hz scale (32 Hz versus 20 Hz). Furthermore, the f_0 difference between maximum and mean in Hz for each vowel was consistently larger for the CHN speakers than that for the DEU speakers (Table 1). When f_0 added range was represented in St, there was a significant effect for VowelIdentity [$F(13,784) = 63.87$, $p < 0.001$] but no significant effect for Gender [$F(1,784) = 0.40$, $p = 0.526$]. Also, the f_0 ranges of the tense vowels were considerably larger than their lax counterparts (3.05 St versus 1.71 St), indicating the significant effect of Tenseness [$F(1,832) = 770.36$, $p < 0.001$]. In each German tense-lax vowel pair, the tense vowel has an inherently longer duration than its lax counterpart (215 ms versus 105 ms across pairs in this study). The smaller f_0 ranges of lax vowels were deemed related to their shorter duration. However, whether the tense and lax vowels had the same rate of f_0 change was still unclear. Therefore, we further compared the slope of f_0 contours between tense and lax vowels.

3.3 f_0 slope

The f_0 slope was used to represent the dynamic changes of the vowel f_0 contour, including the direction (rising or falling) and the steepness (i.e., f_0 change rate). To analyze the directions of f_0 contours, we measured the positions of the f_0 maximum and minimum for each vowel. Fig. 2 depicts the f_0 maximum/minimum position relative to vowel onset in percent,

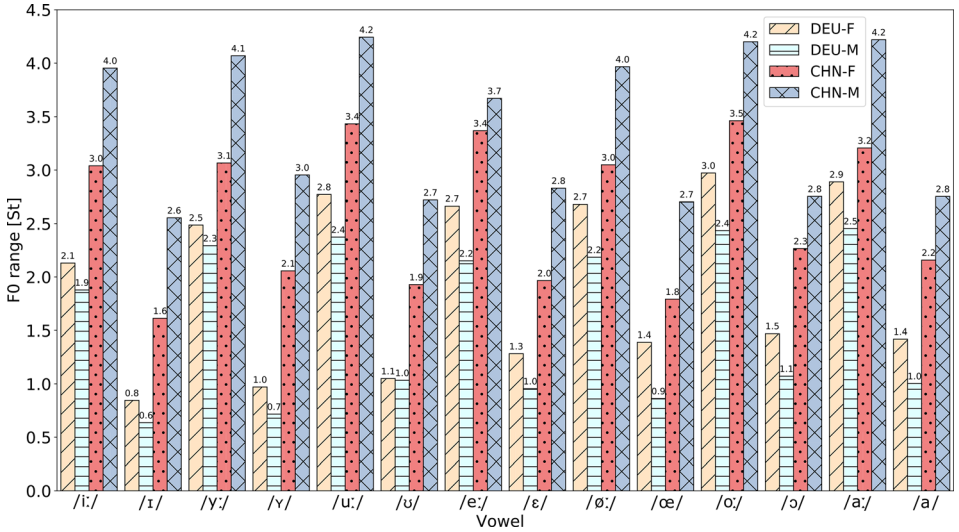

Fig. 1. Bar plots of average f_0 range (in St) of German vowels produced by DEU and CHN speakers. F, female; M, male.

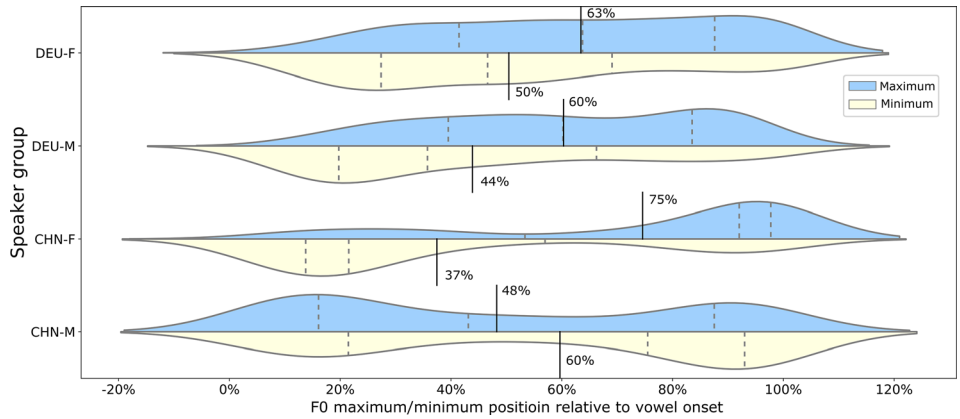

Fig. 2. Violin plots of f_0 maximum/minimum positions of German vowels produced by DEU and CHN speakers. F, female; M, male. The vertical solid and dashed lines show the mean and quartiles, respectively. Note that the plots used kernel density estimation to compute the distribution so that the range exceeds 0% (= onset) or 100% (= offset).

where 0% and 100% correspond to the onset and offset positions of vowels, respectively. As can be seen from the figure, the patterns of f_0 contours are similar between the DEU-F and DEU-M speakers, that is, minimums preceded maximums, suggesting a rising trend in general. The CHN-F speakers exaggerated this pattern, since their f_0 minimums and maximums were closer to the onset and offset of vowels, respectively, compared to the DEU speakers. The vowels produced by the CHN-M speakers showed a reverse pattern, in which f_0 maximums occurred generally earlier than f_0 minimums, resulting in a roughly overall falling direction.

We further examined the steepness of the f_0 slope. The proportions of negative slopes were 31.24% and 24.76% for tense and lax vowels produced by the CHN female speakers, respectively. The CHN male speakers produced even more negative slopes with 54.67% and 58.1% of tense and lax vowels, respectively. Averaging these slopes resulted in a "cancellation" effect in which the negative and positive slopes nullified each other. Therefore, we plotted the *absolute* values of slopes in Fig. 3(a), in which the average duration was also included to reflect the interactions between the f_0 slope and duration of the vowels. All tense or lax tokens were averaged over seven vowels of each group. All lines were normalized to the same starting point for convenient comparisons, and the end points represent the average duration and slope. The LME regression for the absolute f_0 slope revealed that there were significant effects of SpeakerGroup $[F(1, 832) = 8.64, p = 0.003]$ and Tenseness $[F(1, 832) = 51.63, p < 0.001]$. As can be seen from Fig. 3(a), the CHN speakers used a greater rate of f_0 change for the vowels bearing sentence stress than the DEU speakers (20.8×10^{-3} St/ms versus 11.8×10^{-3} St/ms). Also, the lax vowels generally had shorter duration but steeper slopes (18×10^{-3} St/ms versus 14.6×10^{-3} St/ms) than their tense counterparts. However, the effect of Gender $[F(1, 832) = 1.5, p = 0.222]$ was not significant, although the male speakers produced vowels with generally steeper f_0 (18.2×10^{-3} St/ms versus 14.4×10^{-3}

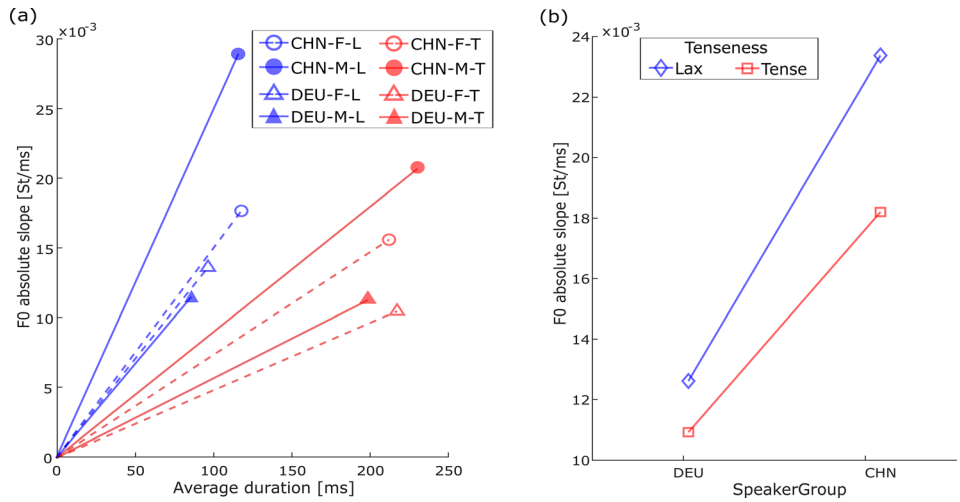

Fig. 3. Average f_0 slope (in St/ms) of German vowels produced by DEU and CHN speakers. F, female; M, male; T, tense; L, lax. (a) Absolute slope; (b) interaction plot of SpeakerGroup × Tenseness.

St/ms) contours than the female speakers (with the exception of lax vowels for the DEU speakers). There was also a significant interaction effect of SpeakerGroup × Tenseness [$F(1, 832) = 13.21$, $p < 0.001$]. As shown in Fig. 3(b), both speaker groups had greater rates of f_0 change for lax vowels than for tense vowels, and the tendency was stronger for the CHN speakers than for the DEU speakers.

4. Discussion

In our current study of f_0 profiles of German vowels under sentence stress, apart from the general comparison of f_0 mean and maximum, range, and slope, a particular focus was the tense-lax contrast of German vowels produced by Mandarin learners of German compared to native German speakers. The following new findings emerged.

First, after transforming f_0 from Hz to St with the speaker-specific base, the effect of Gender on f_0 mean was no longer significant, whereas the effect of VowelIdentity remained significant. It is also clear that CHN learners demonstrate similar intrinsic f_0 patterns as German native speakers under sentence stress, namely, high vowels are produced with an intrinsically higher f_0 than low vowels. We have thus provided further evidence to support the universality of intrinsic f_0 pattern in L2 speech. Furthermore, we have found that lax vowels are also associated with a higher f_0 mean than their tense counterparts in the same stressed context, especially for peripheral vowels. However, Schneeberg and Schlüßler (2006) examined 14 German vowels produced by six speakers and found that only one tense-lax vowel pair showed a significant difference with the lax vowel having a higher f_0, while other pairs showed no significant differences or tense vowels had significantly higher f_0 than lax vowels. Other studies suggested that lax vowels had about the same/similar (i.e., statistically non-significant) f_0 as their tense counterparts, for example, in Fischer-Jørgensen (1990) examining six vowels (/iː ɪ eː ɛ aː a/) produced by six speakers and Pape and Mooshammer (2006) examining six vowels (/iː ɪ uː ʊ aː a/) produced by three speakers. Whether and how the mixed findings result from different reading materials, individual speakers, or measuring approaches of f_0 remains to be examined in the future.

Moreover, we have shown that both CHN learners and DEU speakers produce a larger f_0 range for tense vowels than for lax vowels, which probably results from the longer inherent duration of tense vowels. We have also found that CHN learners produce vowels with a larger f_0 range than DEU speakers, which echoes the findings in the previous studies for Chinese learners speaking German or English (Ding *et al.*, 2006; Zhang *et al.*, 2008). Due to negative first language (L1) transfer, many CHN learners may attach a lexical rising or falling tone to the vowel, which may enlarge the f_0 range at the syllable level. More specifically, male CHN learners tend to produce more vowels with negative f_0 slopes, while female CHN learners tend to produce more vowels with positive f_0 slopes, which are more likely realized as lexical falling and rising tones in their L1 language, respectively. This difference may also result in much larger f_0 ranges of vowels for male CHN learners than for female CHN learners (see the bar heights in Fig. 1), which could be explained by the fact that the Mandarin high-falling tone has the largest f_0 range among lexical tones in general. In addition, the CHN learners produce 14 vowels with longer duration than the DEU speakers, 10 of which are statistically significant (Gao *et al.*, 2020). The longer duration together with steeper slope may also contribute to the larger f_0 range of CHN learners' vowels.

Finally, we have found that both DEU and CHN speakers produce lax vowels with greater steepness than tense ones, and CHN speakers increase steepness more than DEU speakers when they produce lax vowels. Having inherently shorter duration than their tense counterparts, lax vowels may require a larger f_0 change rate to achieve sentence prominence. Besides, we have shown that CHN learners produce the target vowels bearing sentence stress with different directions of f_0 slope. Like DEU speakers, CHN female speakers produce target vowels with an overall rising f_0 contour, while CHN male speakers produce those with an overall falling f_0 contour, which could be attributed to the negative L1 transfer of Mandarin Chinese. Though all CHN learners recruited for the study had comparable L2 German proficiency, we observed that the L2 German speech produced by the males was more Chinese-accented than that produced by the females. Their Chinese accent may result from their frequent use of high-falling tones to achieve the prominence of the target vowel. This is in line with the previous finding that Chinese students tend to use a falling tone to signal an English stressed syllable (Juffs, 1990), which supports Ohala's argument that falling tones are more perceptually salient and can be accomplished quicker (Ohala, 1978). Similar explanations are also found in previous studies of L2 English, e.g., Mandarin learners use a sharply falling f_0 contour for strongly emphatic stress (Zhang *et al.*, 2008).

Acknowledgments

This research work is partially sponsored by the China Scholarship Council and Shanghai Social Science Project (Grant No. 2018BYY003). We would like to express our gratitude to the anonymous reviewers and the editor for their constructive comments and suggestions to improve this work.

References and links

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.1.16) [computer program]," http://www.praat.org/ (Last viewed 20 July 2020).

Chen, Y., Robb, M., Gilbert, H., and Lerman, J. (**2001a**). "Vowel production by Mandarin speakers of English," Clin. Linguist. Phon. **15**(6), 427–440.

Chen, Y., Robb, M. P., Gilbert, H. R., and Lerman, J. W. (**2001b**). "A study of sentence stress production in Mandarin speakers of American English," J. Acoust. Soc. Am. **109**(4), 1681–1690.
Ding, H., Jokisch, O., and Hoffmann, R. (**2006**). "F0 analysis of Chinese accented German speech," in *Proceedings of the 5th International Symposium on Chinese Spoken Language Processing (ISCSLP 2006)*, December 13–16, Singapore, pp. 49–56.
Drugman, T., and Alwan, A. (**2011**). "Joint robust voicing detection and pitch estimation based on residual harmonics," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)*, August 27–31, Florence, Italy, pp. 1973–1976.
Fischer-Jørgensen, E. (**1990**). "Intrinsic F_0 in tense and lax vowels with special reference to German," Phonetica **47**(3), 99–140.
Gao, Y., Ding, H., and Birkholz, P. (**2020**). "An acoustic comparison of German tense and lax vowels produced by German native speakers and Mandarin Chinese learners," J. Acoust. Soc. Am. **148**(1), EL112–EL118.
Hirst, D. (**2011**). "The analysis by synthesis of speech melody: From data to models," J. Speech Sci. **1**(1), 55–83.
Jin, S.-H., and Liu, C. (**2013**). "The vowel inherent spectral change of English vowels spoken by native and non-native speakers," J. Acoust. Soc. Am. **133**(5), EL363–EL369.
Juffs, A. (**1990**). "Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors," Int. Rev. Appl. Linguist. Lang. Teach. **28**(2), 99–118.
Kisler, T., Reichel, U., and Schiel, F. (**2017**). "Multilingual processing of speech via web services," Comput. Speech Lang. **45**, 326–347.
Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**(4), 419–425.
MathWorks (**2019**). "Statistics and Machine Learning Toolbox (version 11.6) [computer program]," https://www.mathworks.com/help/stats/index.html?s_cid=doc_ftr (Last viewed 05 February 2021).
Ohala, J. J. (**1978**). "Production of tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 5–39.
Pape, D., and Mooshammer, C. (**2006**). "Intrinsic F0 differences for German tense and lax vowels," in *Proceedings of the 7th International Seminar on Speech Production (ISSP 2006)*, December 13–15, Ubatuba, Brazil, pp. 271–278.
Schneeberg, J., and Schlüßler, B. (**2006**). "Relations between intrinsic f0, voice quality and the tenseness contrast for German vowels," Arbeitsberichte des Instituts Phonetik und digitale Sprachverarbeitung der Universität Kiel **37**, 27–37.
Shadle, C. H. (**1985**). "Intrinsic fundamental frequency of vowels in sentence context," J. Acoust. Soc. Am. **78**(5), 1562–1567.
Shi, L., Nielsen, J. K., Jensen, J. R., Little, M. A., and Christensen, M. G. (**2019**). "Robust Bayesian pitch tracking based on the harmonic model," IEEE/ACM Trans. Audio Speech Lang. Process. **27**(11), 1737–1751.
Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic f0 of vowels," J. Phon. **23**(3), 349–366.
Yuan, J., and Liberman, M. (**2014**). "F_0 declination in English and Mandarin broadcast news speech," Speech Commun. **65**, 67–74.
Zhang, Y., Nissen, S. L., and Francis, A. L. (**2008**). "Acoustic characteristics of English lexical stress produced by native Mandarin speakers," J. Acoust. Soc. Am. **123**(6), 4498–4513.