

EWA ZAJDLER

BETWEEN SOUND AND VOICE: TEACHING CHINESE TONES TO NON-TONAL LANGUAGE SPEAKERS

Abstract. The production of highly intelligible syllables in Mandarin Chinese entails a successful production of tones, which poses a challenge for learners of Chinese as a foreign language. The aim of the current paper is to address this issue by identifying the key tonal features contributing to tone intelligibility in the lexemes produced by Polish learners of Mandarin Chinese as a foreign language. Samples of Polish female students' tonal pronunciations at two stages of learning were selected and compared with productions made by a female native speaker of Mandarin Chinese from Taiwan. Four syllables produced by the students were selected from a corpus of read-out passages which had already been assessed for the intelligibility of mono-syllabic lexemes by native judges. The students' pronunciation samples (whose pronunciation improved from the A1 minus language level to A2) were analysed using pitch, fundamental frequency contour, and register span criteria, and then compared to the female native speaker's pronunciations of the same syllables. Importantly, before the results of this analysis are presented, the simplified model of tones widely used in language instruction is compared and contrasted with the acoustic analysis of tonal productions made by the native speaker. This is done to show to what extent the simplified, widely used model reflects real-life tonal productions.

Keywords: fundamental frequency; pronunciation; lexical tone; Mandarin Chinese; CFL.

INTRODUCTION

Chinese is a tonal language, which means that tonal phonemes make a distinction between morphemes or lexemes. “If we take a syllable *yi* and place the four tones on it, we obtain a paradigm of four different morphemes” (Li and Thompson 6). The syllable structure is constrained to an optional consonant onset followed by a rhyme containing a nucleus vowel (which can

Dr. habil. EWA ZAJDLER – Jagiellonian University, Faculty of Philology, Institute of Oriental Studies, Sinology Department; correspondence address – e-mail: ewa.zajdler@uj.edu.pl; ORCID: 0000-0002-2640-9029.

be realized in diphthongs or triphthongs) and an optional consonant (Li and Thompson 6). Tonal languages use sub-syllabic phonemic forms conveying the fundamental frequency (F0). Labelled as lexical tones, they incorporate the contrastive pitch and contour at the segmental level of a syllable, in fact – mostly in a vowel (Jeng 105-106). The present pilot study attempts to determine key tonal features leading to either minimal or maximal intelligibility of non-native tonal productions. To that end, the current work reports on an acoustic analysis of tones produced by Polish learners of Mandarin Chinese as a foreign language, which were recorded at two stages of the learning process. Within this analysis the learners' tonal productions, assessed earlier for intelligibility by Taiwanese judges, were juxtaposed with Taiwanese native speaker's pronunciations. However, before the results of this analysis are given, the simplified model of tones (Chao) typically used during Chinese language instruction is compared against acoustically analysed tonal pronunciations produced by a female native speaker of Mandarin Chinese from Taiwan. This is to demonstrate to what extent the simplified model so widely used during language instruction reflects the actual native production of nucleus vowel syllables.

Introduced by Chao Yuen Ren (25-26) the simplified model of tones on the five-point scale has been used in Chinese instruction for decades all over the world (Lu; NTNU MTC; Li) (see Fig. 1).

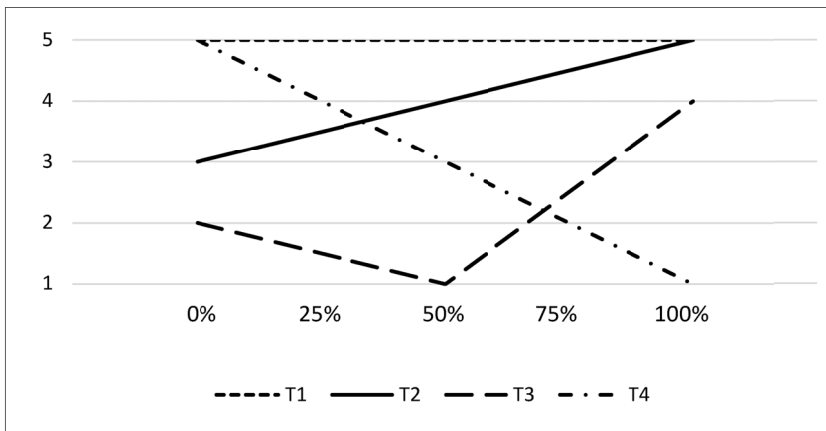


Figure 1. Chinese tones on a five-point pitch scale in time.

The digits 1-5 are used to refer to the pitch and contour of each tone: Tone 1 (T1) – 55¹, Tone 2 (T2) – 35, Tone 3 (T3) – 214 and Tone 4 (T4) – 51, with additional sandhi tones for T3 (the so-called “½ third”) – 21² and T4 – 53 (Zajdler 26-27, see Chao 25). Importantly, Chao’s model presents the characteristics and the relations between tones in a very simplified manner. A more accurate trajectory of tone contours has been obtained once the pitch of the fundamental frequency (F0)³ of the actual tonal pronunciation has been visualised (Xu 61-83). It is the vowel which is mostly associated with tonal pronunciation within a syllable.

Importantly, the F0 values are not fixed for a given language. The factors affecting F0 include e.g. the speaker’s sex, age, local pronunciation standard, contextual variation of the syllable or the intonation of the utterance. It also differs across languages. The individual intrinsic pitch affects the intonation and, in the tonal language, possible changes of the lexical tone pitch within a register span. F0 is lower for males and higher for females.

Undergoing subglottal pressure during phonation, vocal folds produce vowels, i.e. voiced speech sounds. The vowels’ F0 provides variation of the pitch responsible for prosodic features, one of which is the lexical tone. Within a given syllable, F0 changes over time, in line with the expected tone contour. Thereby, the syllable gains distinctive acoustic cues and becomes a recognisable lexical unit. The basic pattern of the tone is maintained in the syllable nucleus (i.e. the vowel), unless the tone is affected by high-level prosodic factors (Sielska-Badurek and Domeracka-Kołodziej 110-111; Hirose et al. 43; Jeng et al. 67- 69). A vowel height (vertical position) contributes to the vowels’ tonal pronunciation (Jeng 102, 112-113).

A tonal language basically refers to the fundamental frequency change and its distinctive and contrastive function, thus calls for a wider register span than in a non-tonal language. Jeng (262) claims that the average F0 range of adult Mandarin Chinese speakers is 69 Hz to 185 Hz for males and 69 Hz to 227 Hz for females from Taiwan reading a passage (after Sheng Hwa). In turn, nucleus vowels with a zero onset and no coda show a span of 40 Hz to 157 Hz for a male (Howie) and 45 Hz to 250 Hz for a female (Tseng). Mentioned above Xu points to about 90 Hz to 140 Hz F0 of /ma/ for the native male speaker from Beijing (67).

¹ Here the first digit represents the beginning of the tone and the second one its end so that e.g. 55 indicates a level tone.

² T3, if not in an isolated or stressed lexeme, in Taiwan Standard Chinese is most often realized as a sandhi half- third (½ third) thus it is as such discussed in the paper.

³ “[T]he frequency at which vocal chords vibrate in voiced sounds” (Li and Jain).

Some studies discussing phonation pitch range of non-tonal L1 speakers learning Chinese point to a less extensive register span of F0, when compared to native speakers', in favour of the Chinese by 1.5 times (Jeng 261-262; Wang et al. 250; Chen; Sheng Hwa 79-86). The analysis of tonal errors suggests that learners of Mandarin as a foreign language tend to have a poor control of their voice ranges with respect to tones (Chang).

The morpho-syllabic structure of Chinese justifies using simple nucleus vowel syllables to map the tonal pronunciation patterns. The following section offers a mapping of F0 contour of four tones within simple nucleus vowel syllables recorded by a female native speaker of Mandarin Chinese from Taiwan. Of note, the Taiwanese female native speaker was chosen as a model for the purposes of the current analysis for the following two reasons. Firstly, the corpus of non-native pronunciations used for the analyses reported on in the second part of the paper contained productions obtained from female learners of Chinese as a foreign language who had been taught by Taiwanese teachers. Secondly, intelligibility assessments of the presently used corpus data had been performed by judges who were native speakers of Mandarin Chinese from Taiwan.

1. MAPPING A NATIVE SPEAKER'S F0 PATTERN

In order to obtain a more detailed model of the four tones than the one proposed by Chao, a female native speaker of Chinese from Taiwan was asked to record (at the sample rate 44100Hz) twenty syllables with [a], [ə], [i], [o] and [u] as nucleus vowels in four tones. Next, the Praat software (Boersma and Weenink) was used to perform a visual acoustic analysis of this material. The results of the analysis provided a visual representation of the F0, i.e. pitch track of the four tone contours for the nucleus vowel syllables, as well as min. and max. values of F0 for the female native speaker (see Table 1 and Figure 2-5).

The min. and max. F0 values extracted in the acoustic analysis of native speaker's productions allows for estimating her individual F0 register span. Table 1 presents this speaker's register span across vowels and tones (thereby designating the diapason of vowels for a given tone defined as the entire tone's range across all vowels and a given speaker's individual register span).

Table 1. A female native register span of vowels for the four tones.

	T1 F _{0min} – F _{0max}	T2 F _{0min} – F _{0max}	T3 F _{0min} – F _{0max}	T4 F _{0min} – F _{0max}
a	261.07 Hz – 275.26 Hz	183.81 Hz – 301.20 Hz	127.09 Hz – 219.83 Hz	140.67 Hz – 307.22 Hz
ə	256.62 Hz – 291.36 Hz	187.30 Hz – 295.63 Hz	138.70 Hz – 260.92 Hz	77.43 Hz – 309.04 Hz
i	241.09 Hz – 274.86 Hz	185.13 Hz – 302.45 Hz	145.42 Hz – 256.98 Hz	76.71 Hz – 326.86 Hz
o	254.38 Hz – 282.33 Hz	188.00 Hz – 286.06 Hz	75.41 Hz – 280.14 Hz	75.02 Hz – 324.56 Hz
u	247.18 Hz – 289.44 Hz	177.20 Hz – 282.67 Hz	90.21 Hz – 252.68 Hz	132.58 Hz – 367.69 Hz

For the female native speaker, F₀ oscillated between 75 Hz and 368 Hz, which shows that this speaker's F₀ register span is about 300 Hz. Importantly, each speaker has their own individual register span. In the presented study, the tone's range for T1 is about 50 Hz (from 241.09 Hz to 291.36 Hz), for T2 it is about 125 Hz (from 177.20 Hz to 302.45 Hz), for T3 it is even more than 200 Hz (from 75.41 Hz to 280.14 Hz) and for T4 it is about 290 Hz (from 75.02 Hz to 367.69 Hz). Since the pronunciation of each of the tones differs slightly across vowels, the diapason for T1 can fluctuate between 14 Hz and 42 Hz, for T2 between 98 Hz and 117 Hz, for T3 between 112 and 205 Hz, and for T4 between almost 170 Hz and 250 Hz.

The register span exploited by the female native speaker from Taiwan was determined by the min. and max. values of T4's F₀, except for incidental short-time T3 phonation. The span is essentially in accordance with the Xu's description of tone contours (1997: 61-83),⁴ although it differs in height due to some factors related to the speaker. It confirms that the startpoint of the T4 for all five vowels (still elevated in the initial part of the T4 trajectory) surpasses the frequency of the T1 starting point, and thus extends the upper

⁴ Xu obtained mean F₀ contours of four Mandarin tones in the monosyllable /ma / produced with six repetitions by eight male native speakers from Beijing. The average span falls between about 90 Hz and 140 Hz. Tone production analysis showed that "Tone 1 starts with a high F₀ value (near 130 Hz) and stays around that level throughout the syllable. Tone 2 starts with a low F₀ (near 110 Hz), then falls slightly before rising (starting at 20% into the vowel) throughout the remainder of the syllable. Tone 3 starts with an F₀ value slightly lower than the onset of Tone 2, falls to the lowest f₀ of all the four tones (about 90 Hz) right at the vowel midpoint, then rises sharply to the end of the syllable. Tone 4 starts with the highest F₀ value of the four tones (140 Hz), continues to rise before reaching the maximum about one fifth of the way into the vowel,

threshold of the speaker's register span. It happens to the T2 ending for all simple syllables, too. The T3 tone realized both as the dipping-rising tone as well as the sandhi half-third ($\frac{1}{2}$ third) tone is obviously a low-register (dipping/low level) tone in its initial half.

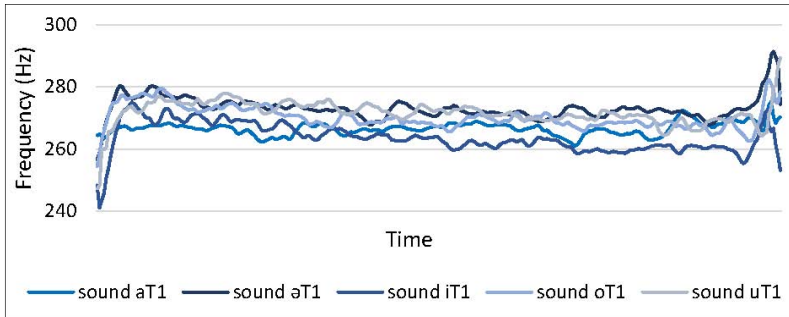


Figure 2. Span of F0 in T1.

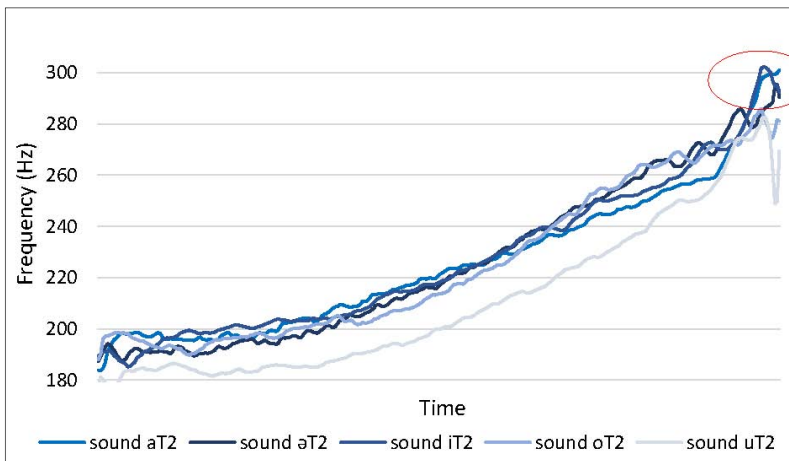


Figure 3. Span of F0 in T2.

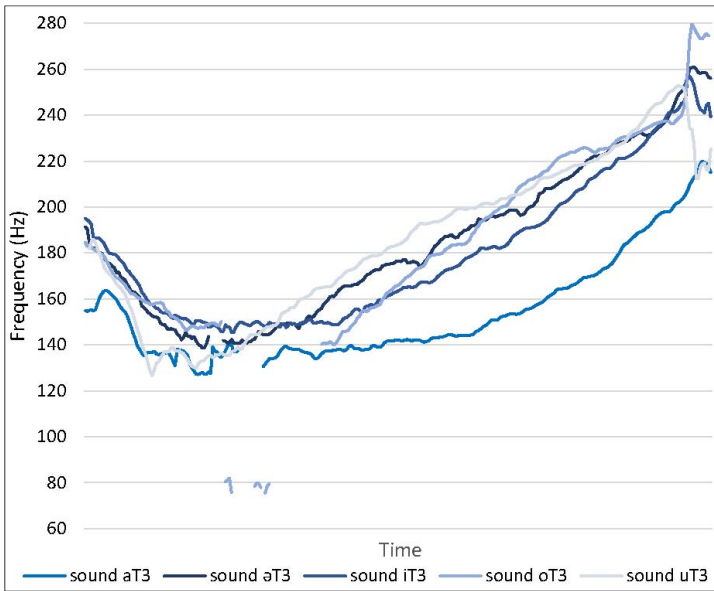


Figure 4. Span of F0 in T3.

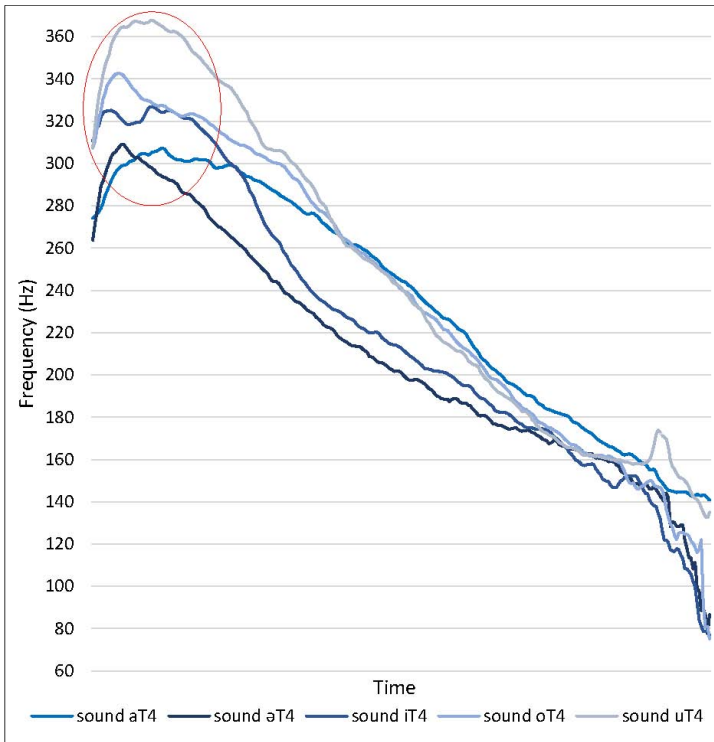


Figure 5. Span of F0 in T4.

As can be seen from Figures 2-5 and the acoustic analysis of the recorded tones presented in Table 1, the pattern of the actual tones produced by the native speaker is not as neat as could be expected on the basis of the simplified model presented in Figure 1. Unlike in the Chao's model, the native speaker's F0 patterns indicate that the starting point of T4 and the first quarter of T4's contour as well as the end point of T2 exceed the upper threshold determined by T1. This means that the digits traditionally used to indicate the tone contour (e.g. 55 indicating a level tone (T1) or 51 indicating a falling tone (T4)) require a more elaborate explanation during instruction of Chinese as a foreign language.

2. DETERMINING THE KEY FEATURES CONTRIBUTING TO TONE INTELLIGIBILITY IN NON-NATIVE SPEAKER'S SPEECH

This study investigates the production of Chinese monosyllabic words with respect to the four tone contours and pitch (with relevant sandhi tones) (see Jeng 259). The aim of this study was to identify those tonal acoustic features which most likely contribute to syllable intelligibility when Chinese syllables are being produced by Polish learners of Chinese as L2. To this end, the material taken from a corpus (Zajdler and Chu) was used. It consists of recordings of short texts produced by Polish learners of Chinese as L2 and subsequently assessed for intelligibility by Chinese native speakers from Taiwan as judges. More specifically, out of those texts, four monosyllabic lexemes were extracted (representing the four tones) which had been assessed by the native judges as maximally unintelligible at Stage 1 (when the students were at the A1 minus level of proficiency, according to the Common European Framework of Reference for Languages (CEFR)) and simultaneously as maximally intelligible at Stage 2 (once the students have achieved the A2 level of proficiency, CEFR). Next, the recordings were juxtaposed with the native speaker's F0 of the syllables in question. Due to the character of the reduced tone and the lack of a contour of the so-called neutral tone (T0) (Hirose et al. 43), T0 was not considered in the study. Also, since the timing was normalised for all the analysed samples, it is not discussed in the study either.

2.1 MATERIAL AND ANALYSIS

The materials used for the current analyses were extracted from a corpus of recorded texts read out by Polish learners of Chinese as L2 (Zajdler and Chu). All the speech material was recorded at the sampling rate 44100 Hz. The texts were recorded by 8 female and 3 male students in their 20s learning Chinese as their major. They recorded their production twice – at Stage 1 when the students were at the A1 minus level of Chinese proficiency (see CEFR) and at Stage 2, once they have achieved the A2 level of proficiency. These recordings were next segmented into single-syllable units whose intelligibility in terms of the tonal pronunciation was next assessed by 3 independent Taiwan native speaker judges on a 5-point scale (for details see Zajdler and Chu). In order to obtain the syllables for the purposes of the current analysis, the same text was next recorded by a female native speaker from Taiwan trained and experienced in teaching Chinese as a Foreign Language. Importantly, she had not been involved in language training of the Polish learners and she did not take part in the assessment of their productions.

The following criteria were used for the extraction of the analysed material. The syllables had to constitute monosyllabic lexemes in four different tones which received the lowest possible rating of intelligibility at Stage 1 and simultaneously the highest possible rating of intelligibility at Stage 2. These criteria had been established in order to maximise the chance of identifying those acoustic features which most likely contribute to the Chinese native speakers' perception of high intelligibility of the non-native pronunciation of Chinese. The adopted criteria resulted in selecting a set of four monosyllabic words spoken by female Polish learners of Mandarin Chinese. It was not possible to select a similar set for male learners as their pronunciations did not meet the required intelligibility assessment span at Stage 1 and 2.

Out of the collected speech material, the four monosyllable lexemes were selected, one in T1, T2, T3 (as a ½ third) and T4, all with the highest value of the spread between the first and the second scoring. The mean of the assessment score of the selected syllables' intelligibility in the first series (at A1 minus level) of assessment (Stage 1) was respectively:

都 <i>dōu</i>	1.00
難 <i>nán</i>	1.00
每 <i>měi</i>	1.33
就 <i>jiù</i>	1.00

The pronunciation of all four lexemes in the second series (Stage 2, at A2 level of proficiency) received score 5 on the five-point scale.

Then, the syllable lexical units (student's and native's production) were subjected to the F0 extraction. Praat software (Boersma and Weenink) for visual acoustic analysis was used to analyse all the studied material in order to provide information about the pitch in the requested detail (i.e. pitch listing for the 300 time-steps per view and minimum and maximum F0 values). It returned a visual representation of the fundamental frequency in the Praat spectrogram window (Boersma; Boersma and Weenink).

3.2 RESULTS

To investigate the differences in the F0 used by native speakers of Mandarin Chinese and Polish learners of Mandarin Chinese, the tonal pronunciation of the selected lexemes at the two stages of language learning was contrasted with that of the native speaker from Taiwan. Minimum and maximum F0 values were extracted for the tonal pronunciation of each syllable (T1, T2, T3 and T4) at two stages of learning (Stage 1 – A1 minus level of proficiency and Stage 2 – A2 level of proficiency) and then compared to the native speaker's production⁵ (see Table 2-5). While the extracted pitch values reflect the span of the particular tone, it needs to be remembered that the tonal production develops in time, hence time-steps should be considered. Even though the current acoustic analysis focuses on a limited number of lexical units (each representing one tone), the aim of the analysis is to showcase a tendency in the process of learning Chinese pronunciation. Figures 6-9 demonstrate the visual representation of the fundamental frequency of the compared monosyllabic lexemes.

Table 2. T1's min. and max. F0 values extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

	T1 <i>dōu</i> (Stage 1)	T1 <i>dōu</i> (Stage 2)	T1 <i>dōu</i> (NS)
F0 _{min}	194.78 Hz	355.64 Hz	314.35 Hz
F0 _{max}	202.61 Hz	483.30 Hz	335.00 Hz

⁵ It is referred to as (I), (II) and (Tw) in Figures.

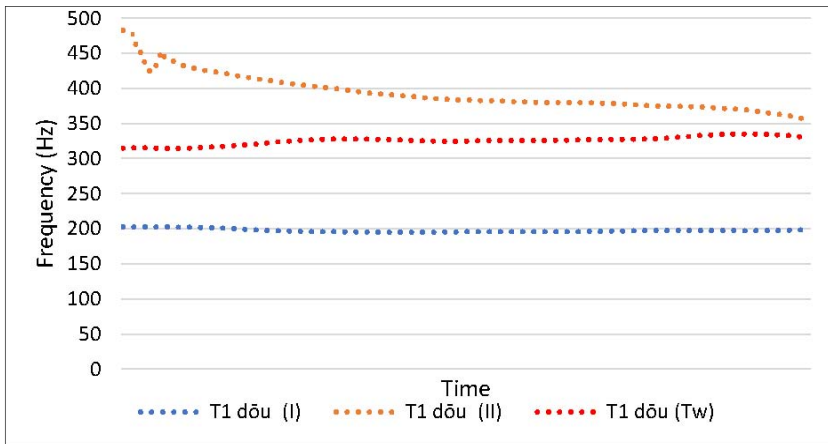


Figure 6. T1's F0 contour extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

Table 3. T2's min. and max. F0 values extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

	T2 <i>nán</i> (I)	T2 <i>nán</i> (II)	T2 <i>nán</i> (Tw)
F0 min.	179.23 Hz	199.35 Hz	131.97 Hz
F0 max.	350.00 Hz	259.83 Hz	267.40 Hz

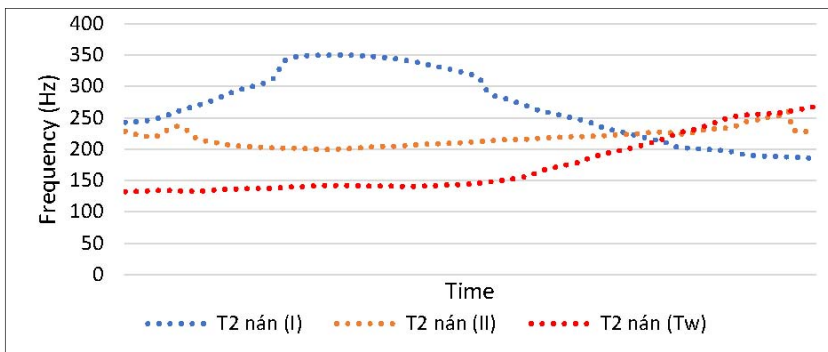


Figure 7. T2's F0 contour extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

Table 4. T3's min. and max. F0 values extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

	T3 <i>měi</i> (I)	T3 <i>měi</i> (II)	T3 <i>měi</i> (Tw)
F0 min.	93.40 Hz	109.23 Hz	131.14 Hz
F0 max.	293.13 Hz	252.47 Hz	212.49 Hz

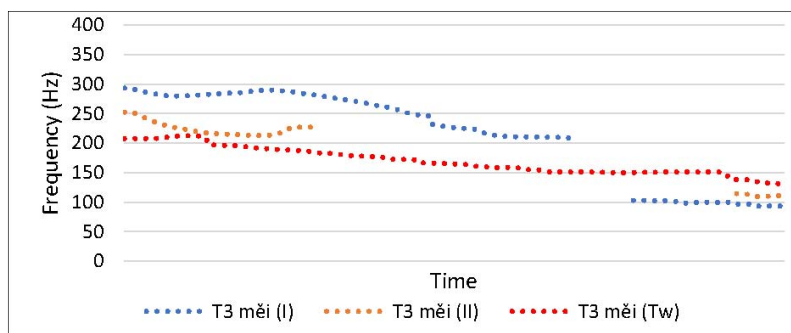


Figure 8. T3's F0 contour extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

Table 5. T4's min. and max. F0 values extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

	T4 <i>jiù</i> (I)	T4 <i>jiù</i> (II)	T4 <i>jiù</i> (Tw)
F0 min.	196.15 Hz	204.86 Hz	187.09 Hz
F0 max.	282.36 Hz	338.89 Hz	358.54 Hz

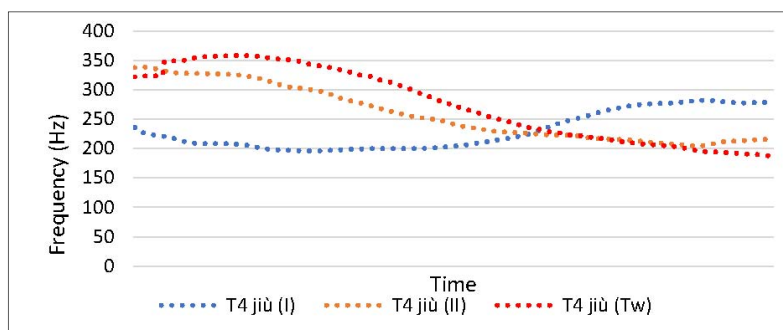


Figure 9. T4's F0 contour extracted from the learner's pronunciation at two stages of learning (Stage 1 and Stage 2) compared to the one produced by the native speaker (NS).

When the samples obtained from the learners at Stage 1 and 2 were subjected to visual acoustic analysis in the Praat software (Boersma and Weenink 2021) and compared to the native speaker's tonal pronunciation, it became evident for the tones in Stage 1 that:

- Learner's F0 contour and span of T1 were close to the native speaker's ones but pronounced over 100 Hz lower (Fig. 6.) than by the native speaker. The F0 span of T1 of the analysed lexeme was nearly 8 Hz for the learner, while it is about 20 Hz (Table 2.) for the native speaker. The learner's F0 minimum and maximum values ($F0_{\min}$, $F0_{\max}$, respectively), i.e. $F0_{\min}$ 194.78 Hz and $F0_{\max}$ 202.61 Hz did indicate a level tone, unfortunately they did not demonstrate the expected height of the tonal speech register, which most likely resulted in the low intelligibility assessment score for the learner.
- Linear trajectory of the fundamental frequency of T2 and T4 was not even close to the native speaker's tone contour of the syllables (Fig. 7., Fig. 9., Tab. 3., Tab. 4.). T2's $F0_{\min}$ was 239.48 Hz and $F0_{\max}$ was 350.00 Hz, and T4's $F0_{\min}$ was 196.15 Hz and $F0_{\max}$ was 282.36 Hz. The T2's $F0_{\max}$ and T4's $F0_{\min}$ appeared in the central part of the syllable course instead of at the end. As a consequence of these mispronunciations, T2 ran mostly as a falling tone (instead of a rising one), and T4 as rising one in the greater part of its contour (instead of a falling one along the contour). Upon a closer inspection, it could be noticed that the initial short period of the tones indicated a learner's attempt at a rising pronunciation of T2 and a falling one of T4, in about one-fourth of the contour of T2 (between 242.806 Hz and 350 Hz) and one-third of the contour of T4 (between 196.15 Hz and 228.97 Hz). Yet, these contour parts do not constitute syllable units and thus failed to contribute to the lexeme intelligibility in subsequent intelligibility assessments.
- The learner's T3's F0 contour trend was essentially correct, but the 200 Hz span (from $F0_{\min}$ 93.40 Hz to $F0_{\max}$ 293.13 Hz) applied to a the T3's ($\frac{1}{2}$ third) tonal pronunciation seems to have been excessively wide when compared to about 80 Hz span in the nativespeaker's pronunciation (Fig. 8., Tab. 4.). Considering the dipping sandhi tone, the learner's span must have seemed more like high-falling contour of T4, hence the subsequent intelligibility assessment score for this learner's pronunciation was low.

These results point to the few factors disqualifying the non-native tone pronunciation in the discussed samples, i.e. an inadequate span of a parti-

cular tone, an incorrect shape of the F0 contour (as in T2 in Figure 7. and T4 in Figure 9.), and, to some degree, a too low or too high starting point and/or ending point of a tone.

The same visual acoustic analysis of the syllables in Stage 2 revealed significant improvements of the learner's tone pronunciations at the A2 level.

- Learner's T1 was assessed as intelligible (score 5) in Stage 2, though only the 3rd and the 4th quarter of the F0 contour was close to the native speaker's pronunciation, falling to 355.64 Hz as – F0_{min}. The remarkably high F0 of the starting point made the initial one- third part of the learner's pronunciation excessively high. Consequently, the T1's span turned out to be about 100 Hz too wide in the 1st and 2nd quarter of the tone with F0_{max} 483.30 Hz (T1 in Figure 6, Table 2.). Nevertheless, the pronunciation was later assessed as highly intelligible (score 5) by native listeners. Therefore, it appears that the high frequency of T1 contributed to the meaningfulness and intelligibility of the tone and the lexeme.
- Learner's F0 contour of T2 was slightly rising within the span of 60 Hz (F0_{min} 199.35 Hz and F0_{max} 259.83), which might have been a cue for the native judges assessing the learner's pronunciation of the tone as highly intelligible (score 5). In fact, the initial part (the 1st quarter) of the T2 trajectory falls first by almost 30 Hz instead of being progressively rising as pronounced by the native speaker. Nevertheless, if compared to the mispronounced T2 at Stage 1 (A1 minus level), there was a significant improvement of T2 pronunciation observed within the sampled syllable (Figure 7, Table 3).
- The learner's F0 contour of T3 (½ third) and T4 were reasonably similar to the native speaker's pronunciation and were both assessed as highly intelligible (score 5) by the native judges. However, the following minor span differences between the learners' pronunciations and that of the native speaker need to be pointed out. The starting point of T3 was above 40 Hz higher and its ending point was about 20 Hz lower than the native speaker's F0. Thus, the learner's T3 (½ third) showed a relatively excessive span, i.e. about 144 Hz (F0_{min} 109.23 Hz and F0_{max} 252.47 Hz), compared to the native speaker's 81 Hz (Figure 8, Table 4). Still, the span of T3 (½ third) showed much more accurate tone contour than when recorded at Stage 1. In contrast, the learner's span of T4 pronunciation (between F0_{min} 204.86 Hz and F0_{max} 338.89 Hz) was somewhat narrow (134 Hz) when compared to about 170 Hz-wide span of the native speaker.

The progressively falling contour since the very beginning of T4 production and its slightly rising ending make the learner's pronunciation a bit flat but still assessed as highly intelligible (score 5) by the native judges (Figure 9, Table 5.).

3. DISCUSSION AND CONCLUSIONS

The present study aimed to identify tonal acoustic features which most likely contribute to syllable intelligibility when Chinese syllables are being produced by Polish learners of Chinese as L2. As presented in the frequency analysis and shown in the graphs and tables, it is a challenge for the learners of Chinese to control the tonal phonation. It was difficult for the learners to keep the voice at the required level, rise gradually or fall gently or even dip intensively when required in line with the Chinese native speaker's F0 contour. Having in mind the study research question, the following, more detailed observations on tonal pronunciation have been made.

It is the F0 contour and its realization within the register span that seem to contribute to a positive assessment of the intelligibility of the monosyllabic lexemes pronounced by the learners. Even though T1's F0 contour was pronounced as a level tone (similarly to the native speaker), it was pronounced too low in Stage 1. As a high level tone, T1's F0 contour has to maintain the required level throughout its pronunciation but it also needs to be pronounced at a certain height of the register span. The learner's pronunciation of T1 in Stage 2 was assessed as highly intelligible (score 5) when the tone was pronounced at sufficient height within the learner's register span, even though the level contour of T1 was initially not maintained. In turn, the F0 contour of T2 in Stage 1 was completely different from the contour produced by the native speaker which resulted in its low intelligibility score. In Stage 2, T2 had the required rising contour and it was assessed as highly intelligible despite its narrower diapazon than that of the native speaker. Learner's pronunciation of T3 (dipping/low-level $\frac{1}{2}$ third (T3)) in Stage 1 exhibited a too broad diapazon, which made the T3 actually resemble a falling tone – T4. This most likely induced its low intelligibility score. However, in Stage 2, $\frac{1}{2}$ third T3 had a narrower diapazon and was pronounced as a dipping tone, which resulted in its significantly better intelligibility. Finally, the F0 contour of T4 was completely mispronounced in Stage 1. Whereas, in Stage 2 the contour as well as diapazon of T4 became very much like the native speaker's thereby setting the speaker's register span.

In conclusion, the following two features of the learner's tonal pronunciation contribute to the complete unintelligibility of the produced lexemes. Firstly, the incorrect F0 contour of the tone makes the pronounced syllable close to incomprehensible. Secondly, the incorrect height of the initial or the final part of the tone contour affecting the diapazon of the particular tone leads to very low intelligibility of the produced lexeme. In contrast, high intelligibility of the pronounced lexemes is achieved by following the following three guidelines. Firstly, the F0 contour of the tone, so the change of its height in time, should be as close as possible to the native speaker's pronunciation. Secondly, the produced tone should keep its proper position within the speaker's register span. Finally, the diapazon of the particular tone should fall within the speaker's register span.

Therefore, the current analysis allows for formulating a few suggestions for teaching pronunciation during Mandarin Chinese instruction. Teaching T1 along with T3 as contrastive with respect to height might help the learners to identify the minimum and maximum of their individual register span. In turn, T2 should be taught with T4 since they contrast in F0 contour (rising vs. falling, respectively), which might be helpful in setting the diapazon correctly for each tone. During Mandarin Chinese pronunciation instruction it seems particularly important to teach the learners how to control their articulation of pitch changes and to assist the learners in the development of their own individual register span. Unfortunately, on the basis of the data of the current pilot study, it was not possible to determine whether or which features more heavily contributed to the intelligibility of the non-native tonal pronunciation of Chinese. Therefore, future research with the use of the corpus data will fill this gap.

BIBLIOGRAPHY

- Boersma, Paul, and David Weenink. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.40, retrieved 27 February 2021 from www.praat.org/. Accessed 14.11.2021.
- Boersma, Paul, and David Weenink. "Speak and unSpeak with Praat." *Glott International*, vol. 5 (9/10), 1996, pp. 341-347.
- Boersma, Paul. "Praat, a system for doing phonetics by computer." *Glott International*, vol. 5, 2001, pp. 341-345.
- Chang, Yung-hsiang Shawn. *Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners*. Jing-Schmidt, Zhuo (ed.). *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*. Vol. 1. University of Oregon, 2011, pp. 84-96.
- Chao, Yuen Ren. *A Grammar of Spoken Chinese*. University of California Press, 1968.
- Chen, G. T. "The Pitch Range of English and Chinese Speakers." *Journal of Chinese Linguistics*, vol. 2, no. 2, 1974, pp. 159-171.

- “Fundamental Frequency, Pitch, F0.” Stan Z. Li, and Anil Jain, editors. *Encyclopedia of Biometrics*. Springer, 2009. Doi: 10.1007/978-0-387-73003-5_775; link.springer.com/referenceworkentry/10.1007%2F978-0-387-73003-5_775. Accessed 10.03.2021.
- Hirose, Keikichi, Qinghua Sun, and Nobuaki Minematsu. “Synthesis of fundamental frequency contours for standard Chinese based on superpositional and tone nucleus models.” *Archives of Acoustics*, vol. 32, no. 1, 2007, pp. 41-50.
- Howie, J.M. “On the Domain of Tone in Mandarin.” *Phonetica*, vol. 30, 1974, pp. 129-148.
- Jeng, Jing-Yi (鄭靜宜), Gary Weismer, and Raymond D. Kent. “Production and perception of mandarin tone in adults with cerebral palsy.” *Clinical Linguistics & Phonetics*, vol. 20, issue 1), pp. 67-87.
- Jeng, Jing-Yi (鄭靜宜). 語音聲學:說話聲音的科學 [Yuyin shengxue: Shuohua shengyin de kexue]. 心理出版社 [Xinli chubanshe]. 心理 [Xinli], 2011.
- Li, Charles N., & Sandra A. Thompson. *Mandarin Chinese. A Functional Reference Grammar*. University of California Press, 1989.
- Li, Xiao Qi (李晓琪), editor. *Boya Chinese: Elementary (1)* [博雅汉语 1] [Boya hanyu 1]. 北京大学出版社 [Beijing daxue chubanshe], 2004, 2013.
- Lu, Jianyi (魯健驥) et al., editors. *Modern Chinese Beginner's Course 1-2* [初级汉语课本] [Chuji hanyu keben]. 北京语言大学出版社 [Beijing yuyan daxue chubanshe], 1986, 1994.
- NTNU MTC. *Practical Audio-Visual Chinese (1)* [實用視聽華語 (1)] [Shiyong shiting huayu (1)]. 正中書局 [Zhengzhong shuju], 1999, 2008, 2017.
- Sheng Hwa (盛華 Sheng Hwa Chen). “Voice range profile of Taiwanese normal young adults: a preliminary study.” *聽語會刊* [Ting yu hui kan], 12, 1996, pp. 79-86.
- Sielska-Badurek, Ewelina, and Anna Domeracka-Kołodziej. “Rola podparcia oddechu w śpiewie operowym.” *Otolaryngologia*, vol. 8, no. 3, 2009, pp. 109-114.
- Tseng, Chiu-yu. *An acoustic phonetic study on tones in Mandarin Chinese*. Institute of History & Philology Academia Sinica. Special Publications No. 94, 1990.
- Wang, Yue, Joan A. Sereno, and Allard Jongman. “L2 Acquisition and Processing of Mandarin Tone.” Ping Li, Li Hai Tan, Elizabeth Bates, and Ovid J. L. Tzeng. *The Handbook of East Asian Psycholinguistics*. Vol. 1. Cambridge University Press, 2006, pp. 250-256.
- Xu, Yi. “Contextual tonal variations in Mandarin.” *Journal of Phonetics*, 25, 1997, pp. 61-83.
- Zajdler, Ewa, and Chu Man-ni (朱曼妮). “How Polish Students Develop Mandarin Pronunciation through Intensive Training.” *Applied Linguistics Papers*, vol. 26, no. 2, 2019, pp. 103- 115.
- Zajdler, Ewa. *Glottodydaktyka*. Wydawnictwo Akademickie Dialog, 2010.

MIĘDZY DŹWIĘKIEM A GŁOSEM.
NAUCZANIE CHIŃSKICH TONÓW
OSÓB NIEMÓWIĄCYCH JĘZYKIEM TONOWYM

Streszczenie

Poprawna realizacja tonów jest kluczowa dla czytelności sylaby w języku tonalnym, stanowi jednocześnie wyzwanie dla uczących się współczesnego języka chińskiego jako obcego. Celem niniejszego artykułu jest zidentyfikowanie dystynktywnych cech wymowy tonalnej wpływających na czytelność leksemów wymówionych w języku chińskim przez użytkowniczkę języka polskiego jako L1. Próbkę wymowy tonalnej realizowanej przez uczestniczki badania na dwóch etapach nauki języka (A1 minus i A2) porównano z materiałem językowym nagrany przez rodzimą

użytkowniczkę języka chińskiego z Tajwanu. Z korpusu wyrazów pochodzącego z czytanego fragmentu tekstu, który został wcześniej oceniony przez rodzimych sędziów pod kątem czytelności monosylabicznych leksemów, wyłoniono do analizy cztery sylaby. Uzyskały one wyraźną poprawę w tonalnej produkcji mowy na drugim etapie badania. Sylaby te zostały poddane analizie wysokości tonu podstawowego, przebiegu tonu i diapazonu i porównane z wymową tych samych sylab przez rodzimą użytkowniczkę. Co ważne, przed przedstawieniem wyników tej analizy szeroko stosowany w nauczaniu języka chińskiego uproszczony model tonów został skonfrontowany z analizą akustyczną produkcji tonów rodzimej użytkowniczki, aby pokazać, w jakim stopniu model uproszczony odzwierciedla rzeczywistą produkcję tonów.

Słowa kluczowe: podstawowa częstotliwość; wymowa; ton leksykalny; chiński mandaryński; CFL.