UDC 81'34
УДК 81'34

Katarzyna Knoll
Institute of English, University of Silesia, Poland
Кнол Катаржина
Институт английского языка, Силезский университет
г. Силезия, Польша
e-mail: katarzyna.knoll@hotmail.com

THE PERCEPTION OF ENGLISH, MANDARIN AND POLISH WORD-INITIAL STOPS BY POLISH SCHOOLCHILDREN AND ADULTS*
ВОСПРИЯТИЕ АНГЛИЙСКИХ, КИТАЙСКИХ И ПОЛЬСКИХ СМЫЧНО-ВЗРЫВНЫХ СОГЛАСНЫХ В НАЧАЛЕ СЛОВ ПОЛЬСКИМИ ШКОЛЬНИКАМИ И ВЗРОСЛЫМИ*

Abstract
The present study concentrates on the perception of voicing contrasts in English, Polish and Mandarin Chinese stops by Poles. Three groups of students are compared: (1) advanced Polish learners of English and Mandarin: adults; (2) beginner Polish learners of English: children with no knowledge of Mandarin; (3) elementary-intermediate Polish learners of English: adults with no knowledge of Mandarin. The results show that the perception of voicing contrasts in English and Mandarin differs with language experience and age. Advanced users of English and Mandarin consistently divide stops into lenis (/b d g/) and fortis (/p t k/), whereas categorisation of voicing contrasts for short-lag stops (0 – 20 ms VOT) is more difficult for adults at low L2 and L3 proficiency levels. Children show the most chaotic pattern in reacting to English lenis stops, but as faced with Mandarin lenis stops, they tend to categorise them as fortis.

Аннотация
В настоящей статье рассматриваются особенности восприятия контраста по присутствию-отсутствию голоса у смычно-взрывных согласных в английском, польском и китайском (путунхуа) языках. Сравниваются перцептивные реакции трёх групп обучающихся: (1) взрослых польских студентов, имеющих продвинутый уровень владения английским и китайским языками; (2) польских детей, начинающих изучать английский без знания китайского языка; (3) студенты-поляки, владеющие английским языком на уровне между начальным и средним без знания китайского языка. Результаты исследования выявили отличия при восприятии указанного контраста на английском и китайском материале, обусловленные языковым опытом обучаемых и их возрастом. Обучаемые, находящиеся на продвинутом этапе владения английским и китайским языками, систематически отличали слабые (/b d g/) от сильных согласных (/p t k/), тогда как взрослые аудиторы с более низким уровнем владения иностранным языком испытывали трудности в различении стимулов с небольшой задержкой начала основного тона (от 0 до 20мс). Восприятие английских слабых согласных детьми было наиболее

хаотичным, однако при восприятии китайских слабых согласных они, в основном, интерпретировали их как сильные.

Keywords: VOT, speech perception, voicing contrasts, FL acquisition.

Ключевые слова: время начала основного тона, речевосприятие, контраст по наличию-отсутствию голоса, освоение иностранного языка.

1. Voice Onset Time in English, Mandarin and Polish

1.1. Voice Onset Time

Introduced by Lisker and Abramson in 1964, the Voice Onset Time (VOT) is 'the time interval between the burst that marks release and the onset of periodicity that reflects laryngeal vibration' [Lisker & Abramson, 1964, p. 422]. In other words, it can be defined as the time between the burst of air and the initiation of a vowel.

Graphically, 'VOT is usually represented as a continuum of time values' [Keating, 1984, p. 295]. Stop release serves as a reference point (0 ms) and Voice Onset is measured relative to it, thus negative values are assigned to measurements of VOT before the release of the stop, whereas positive values are assigned to Voice Onset occurring after stop release. Zero VOT means that Voice Onset (approximately) coincides with the release of the plosive [Lisker & Abramson, 1964; Keating, 1984].

1.1.1. VOT category

In most languages, stops can be characterized as produced with (1) voicing lead, (2) short voicing lag, or (3) long voicing lag.

(1) Voicing lead (negative VOT): voicing starts before the release of the stop, (approximately -30 ms or more VOT). It is present 'as low-frequency harmonics of a buzz source' [Keating et al., 1981, p. 1264].

(2) Short voicing lag (zero onset): voicing begins at or just after the release of the plosive (approximately 0 to +30 ms or up to +35 ms [Keating, 1984]).

(3) Long voicing lag (positive VOT): voicing begins well after the release of the stop (approximately +50 ms or more VOT). It is either accompanied by silence [Klatt, 1975] or aspiration. Aspiration is heard 'if the vocal tract resonates to turbulent air passing through the open glottis' [Lisker & Abramson, 1964, p. 416], and it is described as 'short glottal fricative sound which appears between the release of a syllable-initial voiceless plosive and the onset of vocal cord vibration for the following stressed vowel' [Sobkowiak, 2001, p. 100], 'friction noise generated at the still-open glottis by the flow of air through the vocal tract after the stop release' [Keating, 1984, p. 295], or 'noise (i.e. random stippling) mostly at frequencies of the second and third format' [Lisker & Abramson, 1964, p. 386].

In general, these three VOT categories correspond to the phonetic descriptions of plosives as (1) {voiced unaspirated}, (2) {voiceless unaspirated}, and (3) {voiceless aspirated}, and they are 'sufficient for descriptions of contrasts, and even for most cases of allophonic variation' [Keating, 1984, p. 296]. However, some other categories can also be distinguished, e.g. {voiced aspirated}, as well as 'tense', prenasalized, and implosive; 'but in each such case, the VOT value is the same as for one of the three basic categories' [Keating, 1984, p. 296].

The second category, short-lag stops, seems to be used almost universally across languages, and 51 languages surveyed by Keating et al. [1983] use at least some kind of short-lag stops in virtually every position. Across the surveyed languages, short-lag – long-lag contrasts and short-lag – voicing lead contrasts are equally common. While a number of languages contrast only two categories, i.e. short-lag – long-lag or short-lag – voicing lead, there are still languages which have three or more contrasts. Languages contrasting three categories include Korean, Thai and Eastern Armenian [Lisker & Abramson, 1964]. All three categories plus voiced aspirates, as already mentioned, are used in Hindi and other languages of India [Keating, 1984], and Owerri Igbo has six contrasting homorganic stops [Cho & Ladefoged, 1999]. However, such languages 'have only three VOT distinctions, and use some other action of the larynx, specified by one or more other features, to make these additional contrasts' [Cho & Ladefoged, 1999, p. 226].

Although no languages have more than three contrasts [Cho & Ladefoged, 1999], the categorisation suggested by Lisker and Abramson [1964] might be too general while comparing two or more languages. Thus, e.g. Cho and Ladefoged [1999] propose that four degrees for positive VOT should be distinguished: unaspirated, slightly aspirated, aspirated, and highly aspirated.

1.1.2. VOT determinants

A number of studies have been conducted to examine the effect of various factors on VOT, e.g. speakers' gender, speech rate, fundamental frequency (F0) and age. Most importantly, it has been found that the place of articulation and vocalic environment influence VOT values.

1.1.2.1. Place of articulation

Labial stops are consistently shorter than alveolar and velar stops. In the study of voicing in British English obstruents, Docherty [1992] finds a distinction between labials and non-labials, as well as a slight tendency for voiceless alveolars to have longer mean VOT than velar stops. Significant differences between velars and coronals were also reported in Cho and Ladefoged [1999]. The data presented by Lisker and Abramson [1964] suggest that velar stops always have a longer VOT, and in both aspirated and unaspirated stops, VOT is shortest after bilabial stops and intermediate after alveolar stops [Cho & Ladefoged, 1999]. More recently, Theodore et al. [2009] showed that VOT values for the English voiceless bilabial stop were shorter than those for the velar stop (which is in keeping with previous research), but they also indicate that contextual influence of place of articulation on VOT appears not to be talker-specific.

Some languages, however, do not conform to the abovementioned pattern, i.e. labial < alveolar < velar [Lisker & Abramson, 1964; Cho & Ladefoged, 1999]. Among them, Dahalo seems to be the most uncommon, as its alveolar stops come out with the longest VOT [Cho & Ladefoged, 1999]. Also, it can be observed that in the case of the unaspirated stops in Tamil and the aspirated stops in Cantonese and Eastern Armenian, the velar stop has the longest VOT, but the alveolar is shorter than the labial [Cho & Ladefoged, 1999]. The results of a number of studies on Mandarin Chinese VOT [e.g. Rochet & Fei, 1991; Ran, 2005; Chao et al., 2006; Chao & Chen, 2008] show the same trend, i.e. velar stops are consistently longer, but in voiceless aspirated stops the apical stop is correlated with slightly lower VOT values than the labial one. However, this is not the case in Shi and Liao [1986], where labial < alveolar < velar pattern can be found, and in Wu [1987], in which the apical /th/ is significantly longer than /ph/ and /kh/.

1.1.2.2. Contextual factors

Generally, it has been found that tense high vowels have longer VOTs than lax low vowels. Klatt [1975] reports that VOT values for voiceless stops are greater if followed by a high vowel, i.e. ‘VOT is 15% longer before the high vowels /i, u/ than before /aɪ, ɛ/’ [Klatt, 1975, p. 691]. Rochet and Fei [1991] in their study on the effect of consonant and vowel context on Mandarin Chinese VOT obtain similar results, reporting that ‘occlusives followed by the low vowel /a/ are always accompanied by a shorter VOT than those followed by the high vowels /u/ and /i/’ [Rochet & Fei, 1991, p. 105]. Summerfield [1975] finds that VOT values are longer in /ki/ than in /ka/, however, he does not report any significant differences between VOTs in /pi/ and /pa/, although there is a tendency for VOT values in /pi/ to be shorter than those in /pa/.

Moreover, some studies report vowel frontness to affect VOT as well. For example, Peng [2009] observes that /p/ in Taiwan Mandarin and /p, ph/ in Hakka have longer VOT before /u/ than before /i/, however, other studies on Mandarin [Chao et al., 2006] indicate that vowel frontness is insignificant for VOT variation.

Weismer [1979] finds that VOT is longer when the vowel is tense, as compared to lax, but he underlines that the difference, although significant, is quantitatively quite small. In the study analysing word-initial stops in English, Port and Rotunno [1979] reach similar results and they find that before lax vowels the mean VOT is 11% shorter (i.e. a difference of less than 10 ms, based on measurements accurate only to the nearest 10 ms) than before tense vowels. However, it should be mentioned that ‘there is some difficulty in evaluating these findings regarding the effect of tense/lax vowels on voice onset time. The chief problem is uncertainty about what the labels tense and lax signify in phonetic terms with respect to vowel production [and] neither Weismer nor Port address this matter in any detail’ [Docherty, 1992, p. 27].

Lisker & Abramson [1967] have also observed that stops which are followed by a stressed vowel have longer VOT than those followed by an unstressed vowel, equally in lists of words and words in senteces, however in the words in senteces the degree of unstressed shortening was smaller. Klatt [1975] investigates both a set of CVC words (with the first stop being investigated) and disyllabic test words read in a carrier sentence, and he finds that 'VOT for voiceless plosives was not greatly changed in two-syllable words. The VOT for /p, t, k/ in two-syllable words was on the average, 8% shorter than in the corresponding one-syllable word. The duration of the plosive closure interval was on the average, 45% shorter, and the duration of the syllable nucleus was on the average, 36% shorter in a two-syllable word' [Klatt, 1975, p. 691].

2. VOT values for English, Mandarin and Polish stops

English is known to divide up the VOT continuum with two categories: short lag vs. long lag for voiced and voiceless stops, however, prevoiced values for a voiced category and short-lag values for a voiceless category may also occur [Keating, 1984]. Table 1 shows measurements of English VOT means for word-initial stops as reported by Lisker and Abramson [1964], Klatt [1975], Kopczyński [1977] and Docherty [1992].

T a b l e 1. **Mean VOT values for English stops; all measurements in milliseconds (ms)**

	p^h	t^h	k^h	p	t	k	b	d	g
Lisker & Abramson	+58	+70	+80				+1/-101	+5/-102	+21/-88
Klatt	+47	+65	+70	+12	+23	+30	+11	+17	+27
Kopczyński	+82.5	+84	+71				+18	+14	+31
Docherty	+42	+64	+62				+15	+21	+27

Lisker and Abramson [1964] provide two sets of values for voiced stops /b, d, g/: with a short lag and a voicing lead, suggesting that only a single type is produced by each native speaker. Klatt [1975] presents VOT values not only for voiceless aspirated and voiced stops, but also for voiceless unaspirated /p, t, k/ occurring in word-initial /s/ clusters (/sp, st, sk/). In his study no clear distinction between voiced and voiceless unaspirated stops can be observed, i.e. having similar mean VOT values, English /p, t, k/ and /b, d, g/ belong to the same, short-lag, category. What is visible in all four studies, however, is the division of the VOT continuum into short-lag and long-lag categories. It can be clearly seen even in Klatt [1975] and Docherty [1992], who note higher values for voiced stops and lower values for voiceless aspirated stops, as compared to the original measurements obtained by Lisker and Abramson [1964], as well as in Kopczyński [1977], whose measurements of labial /ph/ and alveolar /th/ are the highest of all.

In Mandarin Chinese, there are no phonetically voiced stops, and aspiration is the only distinctive phonetic feature differentiating voiceless unaspirated and voiceless aspirated stops. According to the data provided by Rochet and Fei [1991], VOT duration for Mandarin voiceless stops ranges between +90 and +110 ms, while that of voiceless unaspirated stops ranges between +10 and +25 ms. Table 2 presents measurements of Chinese VOT means as reported by Shi and Liao [1986], Wu et al. [1987], Rochet and Fei [1991] and, more recently, by Ran [2005], Chao et al. [2006], and Chao and Chen [2008].

T a b l e 2. **Mean VOT values for Mandarin Chinese stops; all measurements in milliseconds (ms)**

	p^h	t^h	k^h	p	t	k	b	d	g
Shi & Liao	+94	+100	+103	+7	+7	+18			
Wu et al.	+72	+100	+85	+7	+9	+19			
Rochet & Fei	+99.6	+98.7	+110.3						
Ran	+106	+104	+112	+12	+13	+30			
Chao et al.	+82	+81	+92	+14	+16	+27			
Chao & Chen	+77.8	+75.5	+85.7	+13.9	+15.3	+27.4			

As can be seen in Table 2, mean VOT values for the voiceless unaspirated stop /k/ are consistently higher than for /p/ and /t/, and although mean VOTs for /p/ and /t/ are almost the same, a tendency for a labial < apical < velar pattern can be observed. In the case of voiceless aspirated stops, this pattern is visible only in the study of Shi and Liao [1986], however, it also occurred in one of two participants in Ran's [2005] study. In all but one of the studies (i.e. in Wu et al., 1987) a tendency for velar /kh/ being longer than /ph, th/ is visible, and also in Rochet and Fei [1991], Ran [2005], Chao et al. [2006], and Chao and Chen [2008] in voiceless aspirated stops the apical stop is correlated with slightly lower VOT values than the labial one. Regarding the set of voiceless aspirated stops, the results of the study conducted by Wu et al. [1986] differ much from the results obtained by other researchers. Apical /th/ is not the shortest but, on the contrary, the longest stop, and VOTs for /ph/ and /kh/ are relatively low. Measurements obtained by Chao et al. [2006] and Chao and Chen [2008] are also relatively low, however, this was probably due to using disyllables instead of monosyllables. Nevertheless, mean VOTs for Chinese voiceless aspirated stops in Table 2 are still significantly higher than VOTs for English voiceless aspirated stops shown in Table 1, except for the measurements of /ph/ and /th/ obtained by Kopczyński [1977]. Also, it can be seen that VOT values for Mandarin voiceless unaspirated stops correspond to VOTs of word-initial voiced stops in English.

According to the original division of the VOT continuum into voicing lead, short-lag and long-lag, English and Mandarin Chinese can be classified as phonetically the same, as they both exploit short-lag vs. long-lag patterns [Lisker & Abramson, 1964; Keating, 1983]. However, some researchers suggest that Cho and Ladefoged's [1999] classification is more suitable than three-way categorisation, especially for languages whose voicing contrast is the aspiration [Chao & Chen, 2008], and that for voiceless aspirated stops, Mandarin Chinese 'seems to fall into the highly aspirated region along the VOT continuum, while English falls into the aspirated region' [Chao & Chen, 2008, p. 228].

Polish, on the other hand, unlike Mandarin and English 'contrasts prevoiced stops with voiceless unaspirated or slightly aspirated stops, which corresponds to a contrast of voicing lead with short-lag VOT' [Keating, 1981: 1261]. In Table 3 measurements of Polish VOT means, as reported in Kopczyński [1977] and Keating et al. [1981], are presented.

T a b l e 3. **Mean VOT values for Polish stops; all measurements in milliseconds (ms)**

	p^h	t^h	k^h	p	t	k	b	d	g
Kopczyński				+37.5	+33	+49	-78	-72	-61
Keating et al.				+21.5	+27.9	+52.7	-88.2	-89.9	-66.1

As can be seen in Table 3, Polish voiceless stops are produced with moderate positive VOT values, slightly higher than mean VOTs for English voiced and Mandarin voiceless unaspirated stops, but also lower than VOT values for voiceless aspirated stops in both Mandarin and English.

3. Experiment

3.1. Subjects

3.1.1. Advanced Learners of English and Mandarin: Adults

A total of 20 Polish Advanced Learners of English and Mandarin (referred to here-after as AL) participated in the study: 4 males and 16 females. They were all fifth-year students of English Philology: Translation English/Chinese, University of Silesia. Their skills (both in English and Mandarin) had been repeatedly confirmed by annual practical examinations. They ranged in age from 22 to 25 years (Mean=22.6, Std. Dev.=0.75). They had had at least 9 years of experience with learning English and 4 years of experience with learning Mandarin Chinese. All subjects volunteered and were not paid for their participation. None of the participants reported any speech or hearing disorders. They were all naive to the object of the study.

3.1.2. Beginner Learners of English: Children

The group comprised 20 Polish beginner learners of English (referred to here-after as BL): 10 males and 10 females. They were recruited from the firstyear pupils at Stefan Żeromski Primary School no. 53 in Katowice. They had all had seven-month experience with learning English and no experience with learning Mandarin Chinese. All subjects were at the age of 8. None of the participants reported any speech or hearing disorders. They were all naive to the object of the study.

3.1.3. Elementary-Intermediate Learners of English: Adults

A total of 20 Polish Elementary-Intermediate Learners of English (referred to here-after as EL) participated in the study: 10 males and 10 females. They were all fourth- and fifth-year students of non-philological fields of study and had no sophisticated knowledge in linguistics. Two factors were taken into account while selecting subjects and assessing their level of proficiency in English: (1) level of the Matura exam in English; (2) self-assessment. All the subjects had taken the Matura exam in English at the basic level, which is compulsory in Poland, and which corresponds to CEFR level A2/B1. They claimed that they had not improved their English since then, they had not taken part in any English course, and had not studied or stayed abroad for more than a month. In the past, they had all been learning English for less than 9 years, and they had never had any experience with learning Mandarin Chinese. The participants ranged in age from 23 to 26 (Mean=23.4, Std. Dev.=0.94). None of them reported any speech or hearing disorders. They were all naive to the object of the study.

3.2. Stimuli

The stimuli used in the experiment were recordings of the selected lexical items. In the case of English and Mandarin words, recordings from course books and textbooks were used, so as to make sure that speakers' pronunciation could be considered to be standard British and Mandarin pronunciation. Polish items were recorded by an educated native Polish speaker who had not any reported history of a speech disorder or any detectable articulation impediment. English, Mandarin and Polish items were all produced by male speakers.

It should be mentioned that due to the fact that common words can be both produced [e.g. Dell, 1990] and recognised [e.g. Luce & Pisoni 1998] with greater facility than rare words, we selected such English and Mandarin items that were not similar to any Polish real-words. That is, the Polish items we used were nonsense words (with the exception of the outdated word tan /tan/), so as to avoid lexical influence, especially in the case of beginner learners and subjects with no knowledge of Mandarin. CV(V) and CVC patterns were used, i.e. each item had three sounds: (1) a stop consonant (2) a vowel (the low vowel /a/ or mid-low vowel /æ/); (3) a consonant or another vowel (i.e. second part in a diphthong).

Two Mandarin stimuli were produced in high-level (HL) tone, one in mid-rising (MR), two in falling-rising (FR), and one in high-falling (HF) tone. Then, VOT values of word-initial stops in each of 18 stimuli were measured using Praat 5.4.06 speech-analysis software package [Boersma & Weenink, 2015] by means of a spectrographic display and waveforms. VOT in each stimulus was measured between the first peak of the release burst to the onset of the second formant of the following vowel (e.g. Keating, 1980, 1981).

In total, 18 items in random order were given to the subjects:

English
(1) pie /paɪ/, /p/ +74 ms VOT
(2) bye /baɪ/, /b/ 0 ms VOT
(3) tan /tæn/, /t/ +60 ms VOT
(4) Dan /dæn/, /d/ +15 ms VOT
(5) cap /kæp/, /k/ +102 ms VOT
(6) gap /gæp/, /g/ +10 ms VOT

Mandarin
(7) 排 pái [a row] /phaɪ/, /ph/ +166 ms VOT
(8) 百 bǎi [a hundred] /paɪ/, /p/ +10 ms VOT
(9) 毯 tǎn [a blanket] /than/, /th/ +118 ms VOT
(10) 单 dān [a bill] /tan/, /t/ 0 ms VOT
(11) 看 kàn [to look] /khan/, /kh/ +155 ms VOT
(12) 干 gān [dry] /kan/, /k/ +20 ms VOT

Polish
(13) *paj /paj/, /p/ +20 ms VOT
(14) *baj /baj/, /b/ -148 ms VOT
(15) tan /tan/, /t/ +23 ms VOT
(16) *dan /dan/, /d/ -149 ms VOT
(17) *kan /kan/, /k/ +41 ms VOT
(18) *gan /gan/, /g/ -153 ms VOT

3.3. Procedures

The experiment took place in a quiet room. Prior to the experiment, each subject was instructed in Polish about the methodology of the study. They were told that they would hear 18 words read by native speakers of English, Mandarin and Polish. We asked beginner learners of English if they recognised the letters on the answer sheet and they were encouraged to read out loud each consonant, i.e. P, B, T, D, K, G, H, written in upper case since children are familiar with them earlier than with letters written in lower case.

Then, the subjects were asked to circle the word-initial sound they heard in each syllable. An example, i.e. a Polish word lala (/lala/, “a doll”), was provided for all subjects, so as to make sure that they understood the instruction. Originally, they were to choose among P, B, T, D, K, G. However, a preliminary study conducted on 8 subjects (five 7-year-old children and three adults) revealed that lack of ‘H’ on the answer sheet lead to such confusion among five subjects (four children and one adult) that the experiment could not be continued as we had expected.

The stimuli were presented via high-quality headphones built in the headset at a comfortable listening level. Special care was taken to provide the same acoustics for all subjects. Each stimulus was presented once and each presentation was followed by a three-second pause.

3.4. Measurements

VOT values of word-initial stops in each of the 18 stimuli were measured using Praat 5.4.06 speech-analysis software package [Boersma & Weenink, 2015] by means of a spectrographic display and waveforms. VOT was measured as a temporal span between the release burst and the beginning of regular vertical striations corresponding to the quasi-periodic voice pulses. The total number of measured target tokens was 1080 (3 groups x 20 talkers x 18 stops).

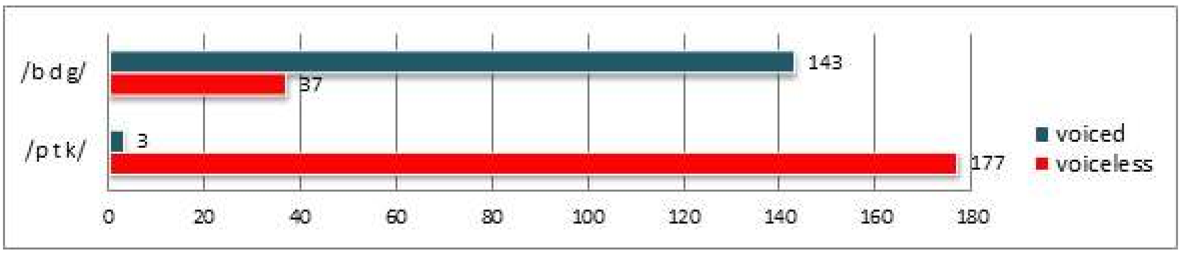
For testing the significance of the between-group effect, we used a Chisquare test, whereas Cochran Q test was used for testing significance of the within-group effect. All statistical analyses and graphical representations were made using MS EXCEL and STATISTICA v.10.

As already mentioned, in English and Polish, stops can be either phonetically voiced or voiceless, whereas in Mandarin there are only voiceless stops, which are differentiated by the presence or absence of aspiration. As it is the perception of voicing contrasts which is our primary concern, i.e. phonological rather than phonetic aspects, for the purposes of the present study, we will refer to Mandarin voiceless aspirated stops as /p t k/ and to their unaspirated counterparts as /b d g/, and abandon using 'voiced/voiceless' in favour of 'fortis/lenis'.

4. Results

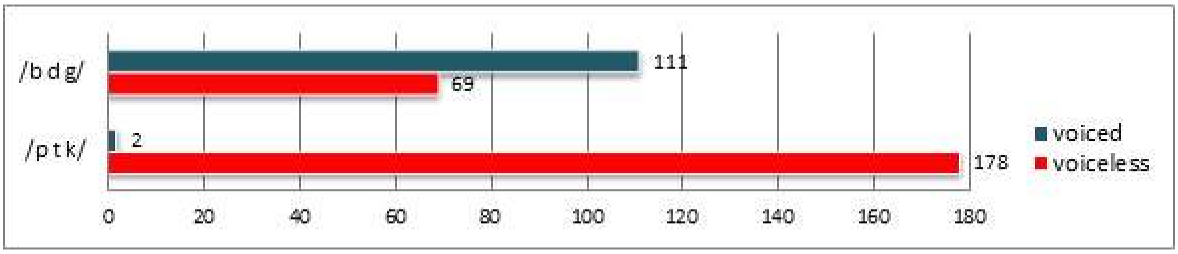
4.1. Perception of English, Mandarin and Polish stops

The results show that English word-initial /p t k/ were recognised as fortis in 177 out of 180 instances (3 stops x 20 subjects x 3 groups), i.e. in 98.3%, whereas their lenis counterparts were identified as lenis in 143 instances (79.4%) (Figure 1).



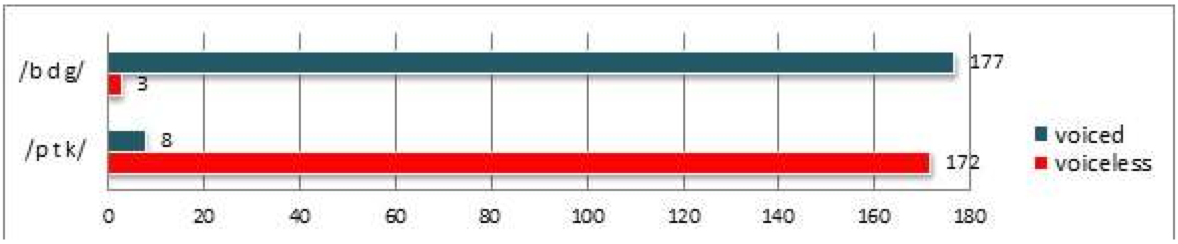
F i g u r e 1. **Perception of English stops**

Mandarin fortis stops were identified as fortis in 178 instances, i.e. 98.9%, and /b d g/ were recognised as lenis in 111 instances, i.e. 61.7% (Figure 2).



F i g u r e 2. **Perception of Mandarin stops**

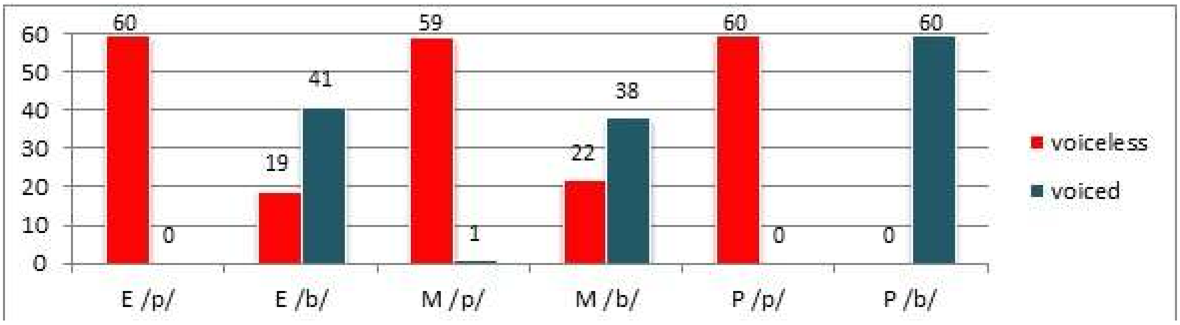
Polish /p t k/ were identified as fortis in 172 instances (95.6%), whereas /b d g/ were recognised as lenis in 177 instances, i.e. 98.3% (Figure 3).



F i g u r e 3. **Perception of Polish stops**

4.1.1. Bilabial stops

The results show that the English, Mandarin and Polish /p/ were perceived as fortis by 59–60 participants (98.3–100%), and all subjects identified the Polish /b/ as lenis. Views on whether the English and Mandarin /b/ were lenis or fortis, however, were divided: they were labelled as lenis sounds by 41 and 38 subjects respectively (68.3% and 63.3%) (Figure 4).



F i g u r e 4. **Perception of bilabial stops (overall results)**

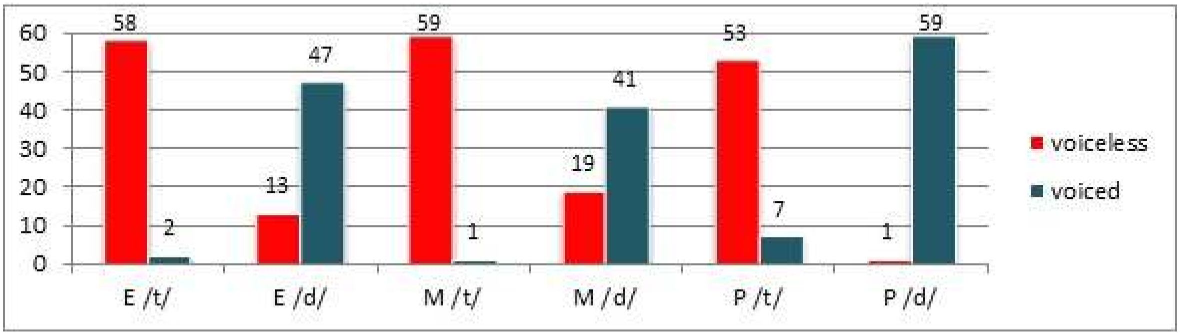
Statistical analysis showed that there were significant between-group differences in the perception of the English /b/ (Chi-square=19.564; p=.000) and the Mandarin /b/ (Chi-square=19.091; p=.000), including:

(1) differences between the two groups of adults (ALs and ELs) (Chi-square=7.059; p=.008 for English, and Chi-square=11.613; p=.001) for Mandarin);

(2) differences between children (BL) and adults (ELs) for the English /b/ (Chi-square=4.912; p=.027). There were no statistically significant differences between the two groups in the perception of the Mandarin /b/ (Chi-square=1.616; p=.204).

4.1.2. Apical stops

Nearly all subjects labelled the English /t/ and the Mandarin /t/ as fortis (58 and 59 subjects respectively, 96.7–98.3%), and the Polish /d/ as lenis (59 subjects, 98.3%). Surprisingly, there was more variation in the perception of the Polish /t/: 7 subjects (11.7%) classified it as lenis. Views on whether the English /d/ was lenis or fortis were even more divided, and there were 13 subjects (21.7%) who identified it as being fortis. In the case of the Mandarin /d/, 41 subjects (68.3%) opted for its being lenis (Figure 5).



F i g u r e 5. **Perception of apical stops (overall results)**

Statistically significant between-group differences were found in the perception of the English /d/ (Chi-square=20.229; p=.000) and the Mandarin /d/ (Chi-square=16.791; p=.000), including:

(1) differences between the two groups of adults (AL and EL) in the perception of the Mandarin /d/ (Chi-square=8.485; p=.004). There were no significant differences between the groups in the perception of the English /d/ (Chi-square=2.105; p=.147).

(2) differences between children (BL) and adults (EL) for the English /d/ (Chi-square=9.231; p=.002). There were no statistically significant differences between the two groups in the perception of the Mandarin /d/ (Chi-square=2.506; p=.113).

4.1.3. Velar stops

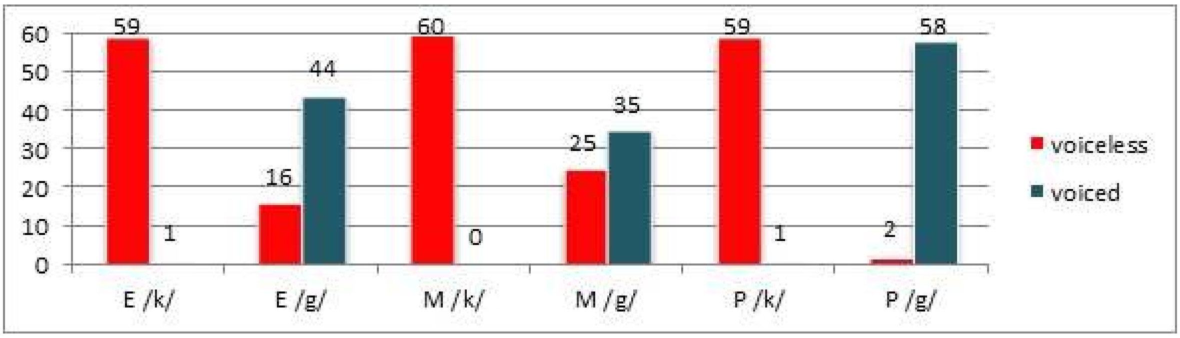
The English, Mandarin and Polish /k/ were perceived as fortis by 59–60 participants of the study (98.3–100%), and the Polish /g/ was classified as lenis by 58 subjects (96.7%). Once again, voicing judgments were divided in the case of two sounds: the English and Mandarin lenis velars. The English /g/ was labelled as lenis by 44 subjects (73.3%), and in the case of the Mandarin /g/, 35 subjects opted for its being lenis (58.3%) (Figure 6).

Statistical analysis showed significant between-group differences in the perception of the Mandarin /g/ (Chi-square=31.406; p=.000), including:

(1) differences between the two groups of adults (AL and EL) (Chi-square=4.329; p=.037);

(2) differences between children (BL) and adults (EL) (Chi-square=15.000; p=.000).

There were no statistically significant between-group differences in the perception of the English /g/ (Chi-square=4.773; p=.092).



F i g u r e 6. **Perception of velar stops (overall results)**

4.2. Perception of English and Mandarin lenis stops

4.2.1. Individual results

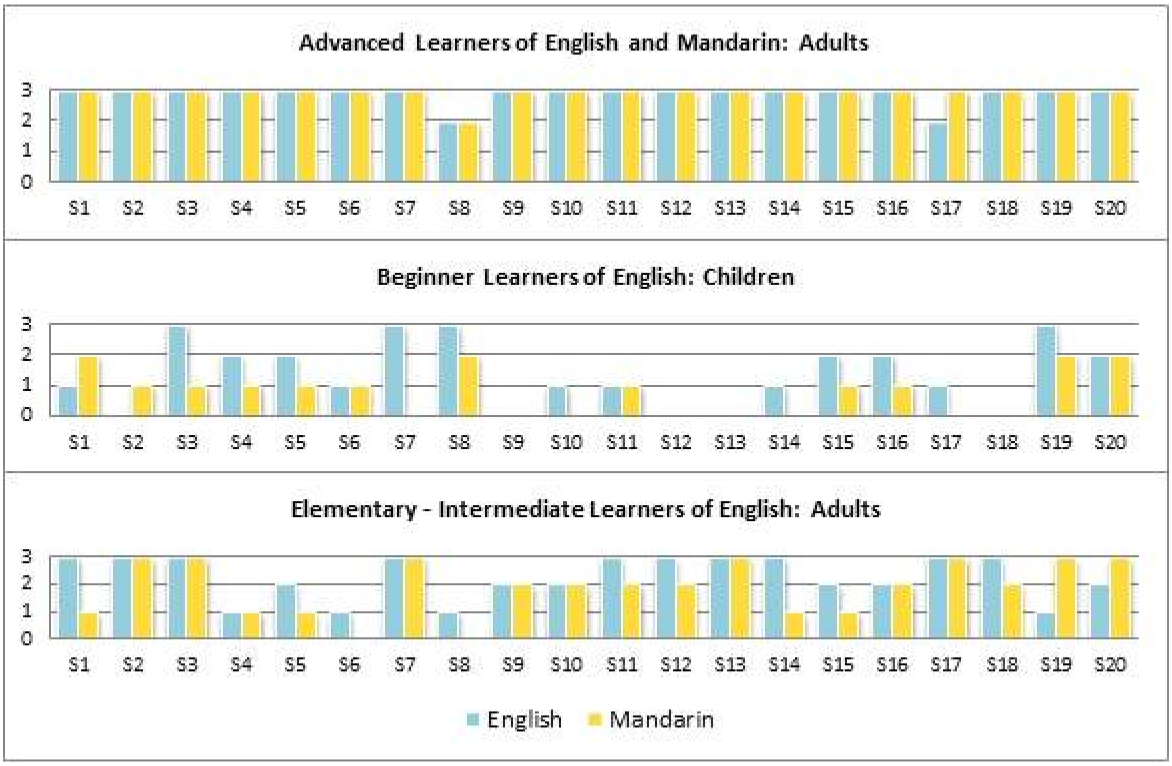
As shown in Figure 7, ALs showed almost no within-group variation in the perception of English and Mandarin lenis stops, and 18 out of 20 subjects perceived all 6 (three English and three Mandarin) stops as lenis. Mean number of lenis stops perceived as lenis was 5.85 (Std.Dev.=0.49): 2.9 for English (Std. Dev.= 0.31) and 2.95 for Mandarin (Std. Dev.=0.22).

In the group of BLs, there were 4 subjects who categorised all three English stops as lenis, and none of the children perceived all Mandarin lenis stops as lenis. However, 5 subjects classified all three English stops, and 8 perceived all Mandarin stops, as fortis. Two subjects perceived 5 or more stops as lenis, whereas 8 as fortis. Mean number of lenis stops perceived as lenis was 2.2 (Std. Dev.=1.64): 1.4 for English (Std. Dev.= 1.09) and 0.8 for Mandarin (Std. Dev.=0.76).

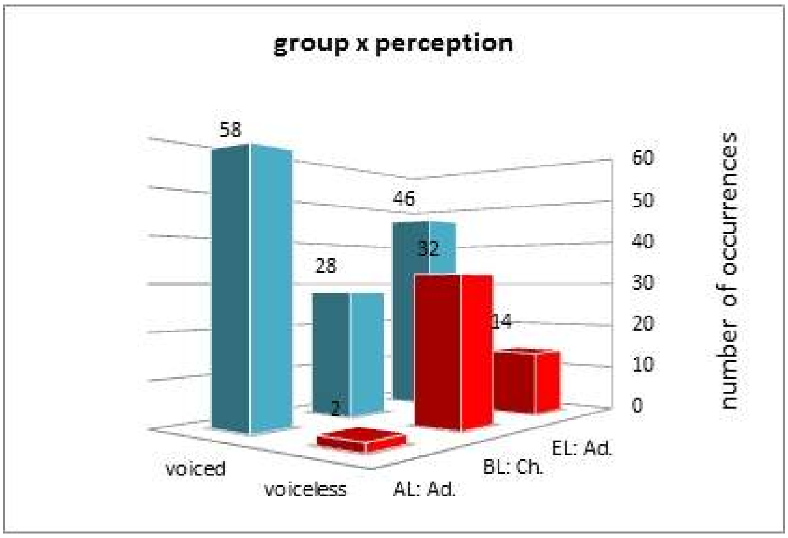
All English lenis stops were perceived as lenis by 10 ELs, and 7 subjects perceived all Mandarin lenis stops as lenis. There were 9 subjects who perceived 5 or more stops as lenis, and two who perceived 5 out of 6 stops as fortis. Mean number of lenis stops perceived as lenis was 4.2 (Std. Dev.=1.58): 2.3 for English (Std. Dev.= 0.8) and 1.9 for Mandarin (Std. Dev.=1.02).

4.2.2. Overall results

Between-group analysis of the perception of English and Mandarin lenis stops showed that ALs classified English /b d g/ as lenis 58 times, i.e. 96.7%. In the other group of adults, EL, English /b d g/ were recognised as lenis in 46 instances (76.7%), whereas in the group of children (BL) they were perceived as lenis only 28 times (46.7%) (Figure 8). The difference in the perception of English /b d g/ between the three groups was statistically significant (Chi-square=38.864; p=0.000).

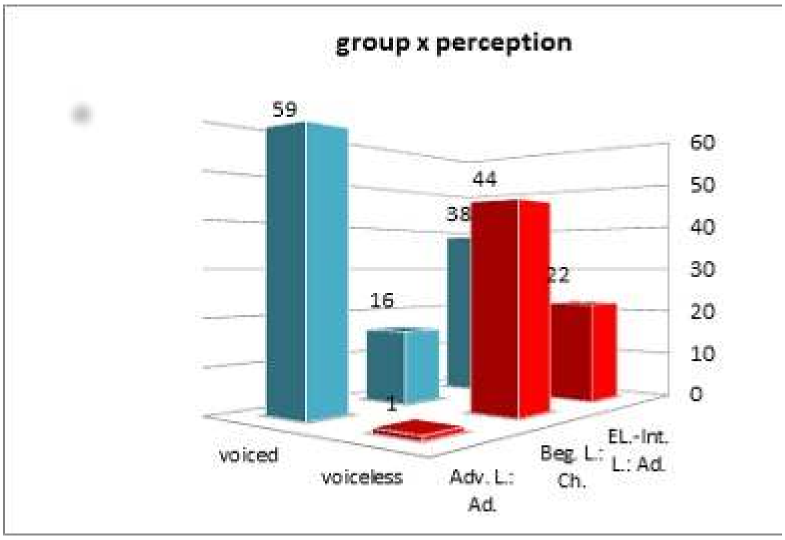


F i g u r e 7. **The number of English and Mandarin /b d g/ perceived as lenis by each subject**



F i g u r e 8. **Recognition of English lenis stops in three groups**

Children recognised Mandarin /b d g/ as lenis 16 times (26.7%), whereas ELs 38 times (63.3%). In the group of ALs, they were perceived as lenis in as many as 59 out of 60 instances (98.3%) (Figure 9). The difference in the perception of Mandarin /b d g/ between the three groups was statistically significant (Chi-square=71.492; p=0.000).

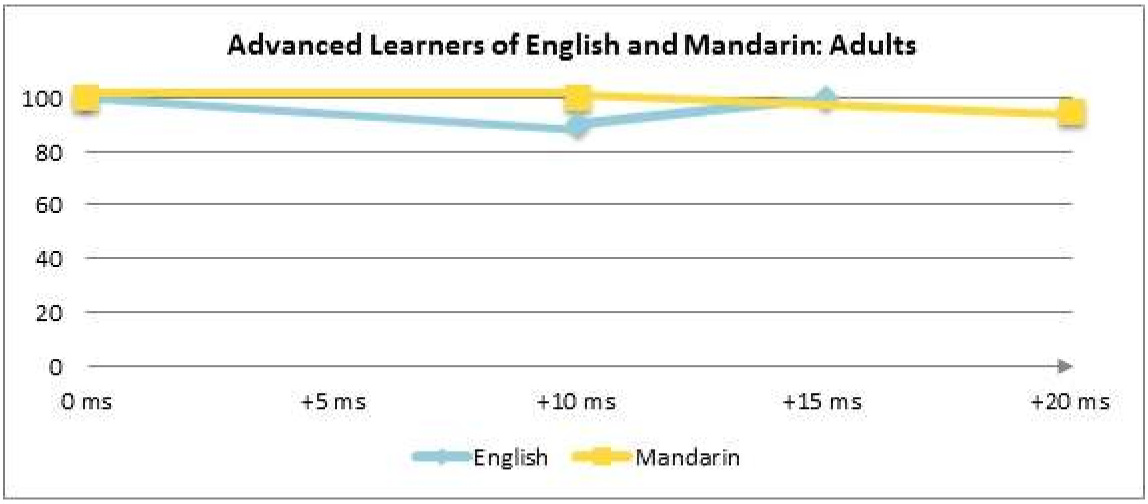


F i g u r e 9. **Recognition of Mandarin lenis stops in three groups**

As shown in Figure 10, in the group of ALs, the recognition of each English and Mandarin stop ranging between 0 and 20 ms as lenis was very high (90%–100%). ELs, however, perceived English /b d g/ as lenis 76.7% of the time, with the highest lenis judgements for the +15 ms VOT /d/ (90%), and Mandarin /b d g/ were categorised as lenis 73.3% of the time, with the highest lenis judgements for the +20 ms VOT /g/ (70%) and the lowest for +10 ms VOT /b/ (55%) (Figure 11). The difference between the two groups in the perception of both English and Mandarin short-lag stops was found to be statistically significant (Chi-square=10.385; p=0.001 for English, and Chi-square=23.720; p=0.000 for Mandarin). Further analyses showed that in the case of English stops, statistically significant differences occurred in the perception of /b/ exclusively (p=0.007), whereas in the case of Mandarin stops, there were high differences in the perception of all three stops (see Table 4).

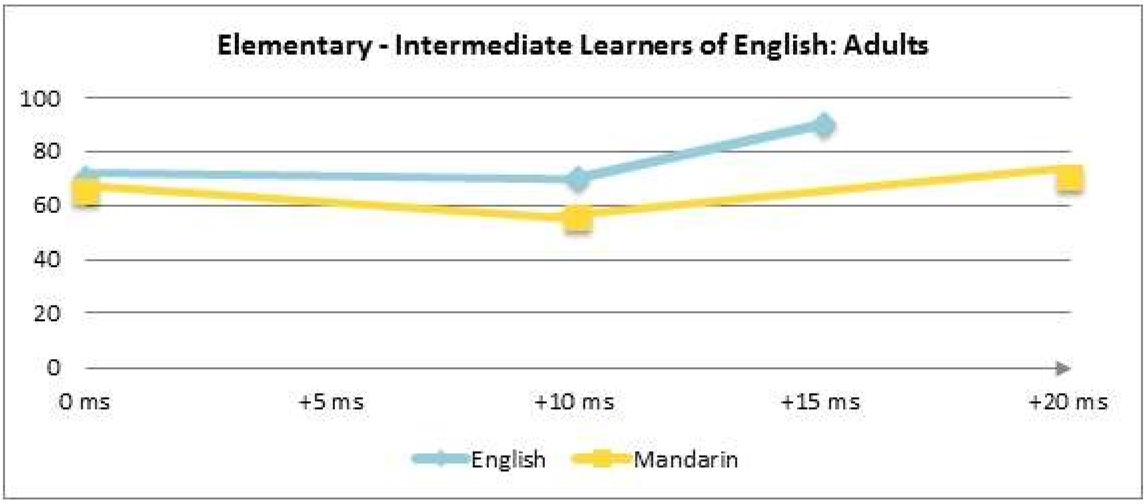
T a b l e 4. **Differences in the perception of lenis stops between ALs and ELs (p-values)**

	0 ms	+10 ms	+15 ms	+20 ms
English	**p=0.007**	p=0.5	p=0.147	-
Mandarin	p=0.113	**p=0.001**	-	**p=0.037**



F i g u r e 10. **Recognition of the English and Mandarin word-initial sounds as lenis across the VOT continuum [in percentage]**

It can be seen that ALs had a strong categorisation effect along the VOT continuum, and all the values between 0 and +20 ms VOT – both in English and Mandarin – were consistently categorised as lenis (mean number of English and Mandarin lenis stops recognised as lenis was 5.85, Std. Dev.=0.49). This was not the case for ELs, although they tended to perceive English lenis stops as lenis and a half of the group categorised all the three of them as such (mean=2.3; Std. Dev.=0.8). Mandarin stops were less frequently recognised as lenis: there were 7 subjects who recognised all the three stops as lenis, and two who perceived them as fortis (mean=1.9; Std. Dev.=1.02) (Table 5).



F i g u r e 11. **Recognition of the English and Mandarin word-initial sounds as lenis across the VOT continuum [in percentage]**

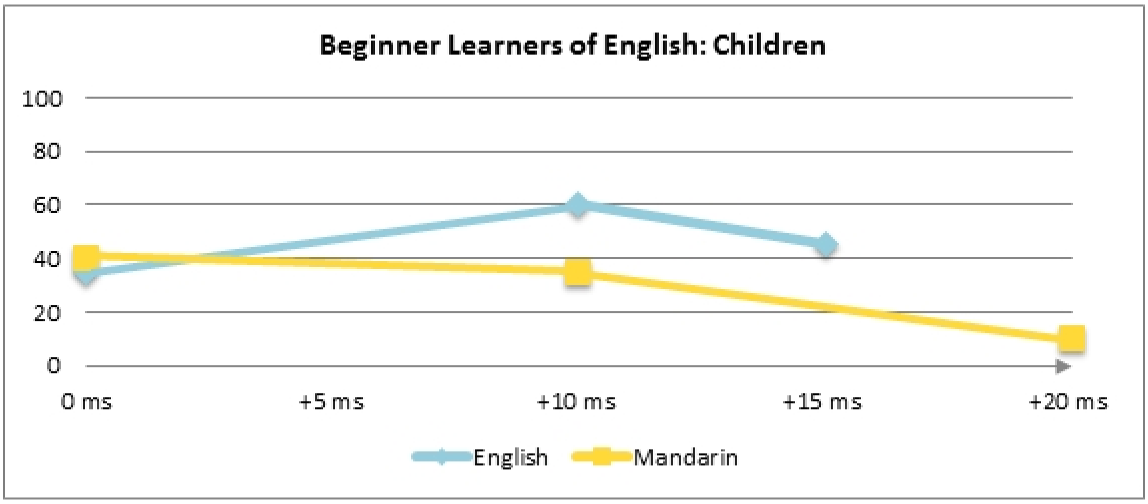
T a b l e 5. **Mean number of English and Mandarin lenis stops perceived as lenis**

Advanced Learners of English and Mandarin: Adults			
	English	**Mandarin**	**All**
Mean	2.9	2.95	5.85
Std. Dev.	0.31	0.22	0.49
Elementary-Intermediate Learners of English: Adults			
	English	**Mandarin**	**All**
Mean	2.3	1.9	4.2
Std. Dev.	0.8	1.02	1.58

As compared to the group of adults who generally perceived English and Mandarin lenis stops as lenis (although the adults did not have a strong categorisation effect along the VOT continuum, especially for Mandarin stops), children were inconsistent and perceived English /b d g/ as lenis 46.7% of the time, with the highest lenis judgements for the +10 ms VOT /d/ (60%) and the lowest for 0 ms /b/ (35%), and Mandarin /b d g/ were categorised as lenis only 26.7% of the time, with the highest lenis judgements for the 0 ms VOT /d/ (40%) and the lowest for +20 ms VOT /b/ (10%) (Figure 12). The difference between the two groups in the perception of both the English and Mandarin short-lag stops was found to be statistically significant (Chi-square=11.422; p=0.001 for English, and Chi-square=16.298; p=0.000 for Mandarin). Further analyses showed that only differences in the perception of the English +10 ms VOT /g/ and the Mandarin +10 ms VOT /b/ were not statistically significant (p>0.05) (Table 6).

T a b l e 6. **Differences in the perception of lenis stops between ELs and BLs (p-values)**

	0 ms	+10 ms	+15 ms	+20 ms
English	**p=0.027**	p=0.114	**p=0.002**	-
Mandarin	**p=0.004**	p=0.204	-	**p=0.000**



F i g u r e 12. **Recognition of the English and Mandarin word-initial sounds as lenis across the VOT continuum [in percentage]**

As shown in Table 7, in the group of children, the mean number of lenis tokens perceived as lenis was almost two times lower than in the group of adults (Mean=2.2; Std. Dev.=1.64). Less than a half of the subjects categorised two or three English stops as lenis, and the mean number of English stops perceived as lenis was 1.4 (Std. Dev.=1.09). None of the children perceived all Mandarin lenis stops as lenis, and the mean number of lenis judgements for these sounds was very low. i.e. 0.8 (Std. Dev.=0.76). Differences between the perception of English and Mandarin lenis stops met the criteria of statistical significance (Q=5.143, $p<0.023$), and within-group analyses showed that in the case of Mandarin, the stimulus effect was statistically significant (Q=7.167, $p < 0.028$). In the case of English, however, no stimulus effect was found (Q = 3,455, $p < 0.178$).

T a b l e 7. **Mean number of English and Mandarin lenis stops perceived as lenis**

Beginner Learners of English: Children			
	English	**Mandarin**	**All**
Mean	1.4	0.8	2.2
Std. Dev.	1.09	0.76	1.64
Elementary-Intermediate Learners of English: Adults			
	English	**Mandarin**	**All**
Mean	2.3	1.9	4.2
Std. Dev.	0.8	1.02	1.58

5. General discussion

Several conclusions can be drawn from the present study.

Most importantly, it has been shown that for Polish speakers, the perception of voicing contrasts in English and Mandarin differs with their language experience and age, whereas in the case of Polish, it remains practically unchanged.

Having both theoretical and practical knowledge of English, Mandarin and, naturally, Polish phonetics, advanced learners consistently divide word-initial stops into fortis and lenis: Polish short-lag stops and English and Mandarin long-lag stops are categorised as belonging to fortis category, whereas Polish stops with voicing lead and English and Mandarin short-lag stops are perceived as lenis. On the other hand, adults with little or no knowledge of the two FLs do not have such a strong categorisation effect, and whereas they generally do categorise the English lenis stops as lenis, their voicing judgments on the Mandarin stops are more ambiguous. These between-group differences appear to be directly related to the differences in FL proficiency.

It can also be seen that children show a chaotic pattern in reacting to English lenis stops, but as faced with Mandarin lenis stops, they have a tendency to categorise them as fortis. As their judgments are significantly different from those of adults at similar proficiency level, it is possible to conclude that the perception of VOT continuum is age- or experience-dependent.

In the group of children and adults with little or no FL experience, English short-lag stops, as compared to Mandarin short-lag stops, are more frequently recognised as lenis, although they occupy the same place on the VOT continuum and range between 0 and +20 ms VOT. This might be due to the fact that subjects had received, probably subconsciously, relatively high input from native English speakers and they are familiarised, at least to some extent, with the pronunciation of English. The sounds of Mandarin, on the other hand, are rather novel for Polish speakers as the probability of exposure to this Asian language is relatively low. Also, the results suggest that subjects with no knowledge of Mandarin categorise Mandarin lenis stops basing on their native language and experience in FL, i.e. in English.

It appears that sensitivity to English and Mandarin boundaries with the positive VOT values can be acquired in the process of learning, and advanced learners can obtain a complete shift into lenis category for English and Mandarin VOT values of +20 ms and lower. Children, whose FL experience is the smallest, however, have their critical point at +20 ms VOT, at which we can observe almost a complete shift into the fortis category – at least in the case of Mandarin stops. For VOT values lower than +20 ms, there is a gradual shift into lenis category, although the number of fortis judgments is still significantly higher. In the case of English stops, no critical point can be observed in the present study and children are generally inconsistent in their judgments.

Regarding the perception of fortis stops, it can be observed that although highly aspirated Mandarin stops are easily recognisable as fortis, subjects who are not familiarised with this language frequently have difficulties in specifying their place of articulation correctly. In the case of advanced learners, thanks to the high level of input they received from native Mandarin Chinese speakers, mistakes are relatively rare.

* This article was written on the basis of the author's unpublished MA thesis

References

1. Boersma, P. Praat: doing phonetics by computer [Computer program]. Version 5.4.08 / P. Boersma, D. Weenink. – Retrieved 24 March 2015 from http://www.praat.org/.
2. Chao, K.-Y. A cross-linguistic study of voice onset time in stop consonant productions [Text] / K.-Y. Chao, L.-M. Chen // Computational Linguistics and Chinese Language Processing. – 2008.– Vol. 13. – N 2. – P. 215–232.
3. Chao, K.-Y. Comparison of VOT patterns in Mandarin Chinese and in English [Text] / K.-Y Chao, G. Khattab, L.-M. Chen // Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities. – 2006.– P. 840–859.
4. Cho, T. Variations and universals in VOT: Evidence from 18 languages [Text] / P. Ladefoged, T. Cho // Journal of Phonetics. – 1999. – Vol. 27. – P. 207–222.
5. Dell, G. S. Effects of frequency and vocabulary type on phonological speech errors [Text] / G. S. Dell // Language and Cognitive Speech Processes. – 1990. – Vol. 5. – P. 313–349.
6. Docherty, G. J. The Timing of Voicing in British English Obstruents [Text] / G. J. Docherty. – New York: Foris, 1992.
7. Keating, P. A. Phonetic and phonological representation of stop consonant voicing [Text] / P. A. Keating // Language. – 1984. – Vol. 60. – P. 286–319.
8. Keating, P. A. Patterns in allophone distribution for voiced and voiceless stops [Text] / P. A. Keating, W. Linker, M. Huffman // Journal of Phonetics. – 1983. – Vol. 11. – P. 277–290.
9. Keating, P. A. A cross-language study of range of voice onset time in the perception of initial stop voicing [Text] / P. A. Keating, M. J. Mikoś, W. F. Ganong III // Journal of the Acoustical Society of America.– 1981. – Vol. 70. – P. 1261–1271.
10. Klatt, D. H. Voice-onset time, frication, and aspiration in word-initial consonant clusters [Text] / D. H. Klatt // Journal of Speech and Hearing Research. – 1975. – N 18. – P. 687–703.
11. Kopczyński, A. Polish and American English Consonant Phonemes: A Contrastive Study [Text] / A. Kopczyński. – Warszawa: Państwowe Wydawnictwo Naukowe, 1977.
12. Lisker, L. A cross language study of voicing in initial stops: Acoustic measurements [Text] / L. Lisker, A. S. Abramson // Word. – 1964. – Vol. 20. P. 384–422.
13. Luce, P. A. Recognizing spoken words: The Neighborhood Activation Model [Text] / P. A. Luce, D. B. Pisoni // Ear and Hearing. – 1998. – Vol. 20. – P. 19–36.
14. Peng, J.-F. Factors for Voice Onset Time: Stops in Mandarin and Hakka: Master thesis [Text] / J.-F. Peng. – National Cheng Kung University, 2009.

15. Port, R. F. Relation between voice-onset time and vowel duration [Text] / R. F. Port, R. Rotunno // Journal of the Acoustical Society of America. – 1979. – Vol. 66. – P. 654–662.
16. Ran, Q.-B. Jiyu putonghua de hanyu zuse fuyin shiyanjiu (Experimental studies on Chinese obstruent consonants: with the emphasis on standard Chinese): Ph. D. Dissertation [Text] / Q.-B. Ran. – Nankai University, 2005.
17. Rochet, B. L. Effect of Consonant and Vowel Context on Mandarin Chinese VOT: Production and Perception [Text] / B. L. Rochet, Y. Fei // Canadian Acoustics. – 1991. – Vol. 19. –N 4. – P. 105–106.
18. Shi, F. Zhongmei xuesheng hanyu seyin shizhi duibi fenxi (Contrastive analysis of the length of stops in Mandarin between Chinese and American students) [Text] / F. Shi, R.-R. Liao // Language Teaching and Linguistic Studies. – 1986. – N 4. – P. 67–83.
19. Sobkowiak, W. English Phonetics for Poles: A Resource Book for Learners and Teachers [Text] / W. Sobkowiak. – Poznań: Wydawnictwo Poznańskie, 2001.
20. Summerfield, A. Q. Aerodynamics vs. mechanics in the control of voicing onset in consonant-vowel syllables [Text] / A. Q. Summerfield // Speech Perception. – 1975. – N 4. – P. 61–72.
21. Theodore, R. M. Individual talker differences in voice-onset-time: Contextual influences [Text] / R. M. Theodore, J. L. Miller, D. DeSteno // Journal of the Acoustical Society of America. – 2009. – N 125. – P. 3974–3982.
22. Weismer, G. Sensitivity of voice onset measures to certain segmental features in speech production [Text] / G. Weismer // Journal of Phonetics. – 1979. – N 7. – P. 194–204.
23. Wu, Z.-J. Shiyan Yuyinxue Gaiyao (=An Outline of Experimental Phonetics) [Text] / Z.-J. Wu, M.-C. Lin. – Beijing: Higher Education Press, 1989.