Journal of Phonetics 40 (2012) 725–744

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese

Eun Jong Kong [a,b,c,*], Mary E. Beckman [c], Jan Edwards [b,d]

[a] Department of English, Korea Aerospace University, 100 Hanggongdae gil, Hwajeon-dong, Deogyang-gu, Goyang-city, Gyeonggi-do 412-791, Korea
[b] Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705-2280, USA
[c] Department of Linguistics, The Ohio State University, Oxley Hall, 1712 Neil Avenue, Columbus, OH 43210-1298, USA
[d] Department of Communicative Disorders, University of Wisconsin-Madison, Goodnight Hall, 1975 Willow Dr. Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 3 November 2010
Received in revised form
12 July 2012
Accepted 24 July 2012
Available online 25 August 2012

ABSTRACT

The age at which children master adult-like voiced stops can generally be predicted by voice onset time (VOT): stops with optional short lag are early, those with obligatory lead are late. However, Japanese voiced stops are late despite having a short lag variant, whereas Greek voiced stops are early despite having consistent voicing lead. This cross-sectional study examines the acoustics of word-initial stops produced by English-, Japanese-, and Greek-speaking children aged 2–5, to investigate how these seemingly exceptional mastery patterns relate to use of other phonetic correlates. Productions were analyzed for VOT, f0 and spectral tilt (H1 − H2) in Japanese and English, and for amplitude trajectory in Greek and Japanese. Japanese voiceless stops have intermediate lag VOT values, so other "secondary" cues are needed to differentiate them from the voiced short lag VOT variant. Greek voiced stops are optionally prenasalized, and the amplitude trajectory for the voice bar during closure suggests that younger children use a greater degree of nasal venting to create the aerodynamic conditions necessary for voicing lead. Taken together, the findings suggest that VOT must be supplemented by measurements of other language-specific acoustic properties to explain the mastery pattern of voiced stops in some languages.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Cross-language differences in the age of children's mastery of adult-like voiced stops are typically explained in terms of the relative difficulty of the laryngeal gestures for the language's voice onset time distributions (VOT, Lisker & Abramson, 1964). There are three qualitatively different ranges of voice onset time – voicing lead, short lag, and long lag – and the voicing contrast in most languages uses two of these three ranges. The production of short lag VOT stops requires minimum temporal precision in aligning the oral opening gesture that produces the burst with the glottal gesture that postures the vocal folds for voicing given a sufficient trans-glottal air pressure difference. By contrast, voicing lead, in which the voicing onset occurs before the burst, is difficult to produce because the requisite aerodynamic condition for vocal fold vibration (i.e., lower supra-glottal air pressure than sub-glottal air pressure) is difficult to maintain when the oral cavity is sealed off by a stop closure gesture (Keating, 1983; Westbury, 1983; Westbury & Parris, 1970; Westbury & Keating, 1986). Across languages, then, short lag stops in word-initial position tend to be acquired earliest and voicing lead stops tend to be acquired last (Allen, 1985; Clumeck, Barton, Macken, & Huntington, 1981; Davis, 1995; Gandour, Petty, Dardarananda, Dechongkit, & Mukngoen, 1986; Macken & Barton, 1980a, 1980b; Pan, 1994). Thus, in languages such as English, which contrasts so-called "voiced stops" that often have short lag VOT values in word-initial position with voiceless stops that always have long lag VOT values, word-initial voiced stops are mastered at a relatively early age (by 2 years, Macken & Barton, 1980a; Kewley-Port & Preston, 1974; in babbling, Whalen, Levitt, & Goldstein, 2007), but in languages such as Spanish, which contrasts voiced stops that have voicing lead in word-initial position with voiceless stops that have short lag VOT values, voiced stops are mastered at a relatively late age (only after age 5, Macken & Barton, 1980b). Other languages in which stops with contrastive voicing lead have been found to be mastered relatively later than voiceless stops include French (Allen, 1985), Thai (Gandour et al., 1986), Taiwanese (Pan, 1994) and Hindi (Davis, 1995).

While language-specific VOT distributions correctly predict cross-language differences in mastery patterns of voiced stops

* Corresponding author at: Department of English, Korea Aerospace University, 100 Hanggongdae gil, Hwajeon-dong, Deogyang-gu, Goyang-city, Gyeonggi-do 412-791, Korea. Tel.: +82 2 300 0338; fax: +82 2 3158 5769.
E-mail addresses: ekong@kau.ac.kr (E.J. Kong), mbeckman@ling.osu.edu (M.E. Beckman), jedwards2@wisc.edu (J. Edwards).

0095-4470/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2012.07.002

for most languages, Greek and Japanese seem to be exceptions to the generalization, albeit in opposite ways. Specifically, both Japanese and Greek have been characterized as having a contrast between voiced and voiceless stops in word-initial as well as word-medial position (Fourakis, 1986; Homma, 1980; Kollia, 1993; Shimizu, 1989). This characterization predicts that voiced stops should be late in both languages, and the prediction is borne out in transcription studies of Japanese-acquiring children. For example, Yasuda (1970) found lower accuracy for /b/ and /d/ relative to /p/ and /t/ in the 100 three-year-olds she tested and Nakanishi, Owada, and Fujita (1972) found similar patterns in the much larger group of 4-year-olds that they tested. Both studies found equal error rates for /g/ and /k/, but these studies were conducted at a time when [ŋ] was in allophonic variation with [g] in the Tokyo dialect. In adult productions today, however, Japanese word-initial voiced stops often have short lag VOT values as in English (Riney, Takagi, Ota, & Uchida, 2007; Takada, 2011), but a recent transcription study of lingual obstruents (Beckman & Edwards, 2010) suggests that voiced stops are still late-acquired. Conversely, Greek word-initial voiced stops almost always have lead VOT values as in Spanish or French (Arvaniti, 2007; Botinis, Fourakis, & Prinou, 2000). However, Okalidou, Petinou, Theodorou, and Karasimou (2010) found that 2-year-old Standard Greek-speaking children could produce adult-like lead VOT. Perhaps, in these two languages, the age of mastery of voiced stops cannot be predicted by VOT alone. The current study examines the efficacy of VOT and other acoustic parameters for predicting the mastery pattern of voiced stops in Japanese and Greek.

In Tokyo Japanese today, VOT values overlap between voiced stops and voiceless stops in adult productions. Takada (2004a) found that voiced stops had short lag VOT values especially in younger adult speakers' productions. Takada (2011) reported analyses of VOT values in voiced stops measured in recordings of younger and older speakers made in the early 1990s from around the country and in another smaller-scale survey in 2005, suggesting that a sound change originating in the Touhoku dialects has been spreading south and west over the last half century. This VOT distribution of Tokyo Japanese voiced stops patterns similarly to that of English voiced stops in word-initial position. By contrast, voiceless stops of Japanese differ from those of English in that Japanese voiceless stops have VOT values intermediate between short lag VOT values and long lag VOT values (Riney et al., 2007; Takada, 2011), while English voiceless stops have long lag VOT values. This intermediate lag VOT was not documented in Lisker and Abramson (1964), but it has been reported in later studies of VOT distribution in at least two other languages with two-way distinctions – namely, Canadian French (Caramazza & Yeni-Komshian, 1974) and Hebrew (Raphael, Tobin, Faber, Most, Kollier, & Milstein, 1995) – and data reported in Cho and Ladefoged (1999) suggest that there may be other languages with similar patterns of overlap between a category with short lag and a category with intermediate lag VOT values.

The existence of this intermediate lag VOT category has, thus far, always involved a contrast with another category that has at least some short lag values, creating an overlap between two lag VOT ranges. Consistent with Canadian French and Hebrew, VOT alone cannot differentiate Japanese voiced and voiceless stops because there is overlap between the intermediate lag VOT values of voiceless stops and the short lag VOT values of voiced stops. This suggests that other acoustic differences must be needed to cue the voicing contrast in languages such as Japanese. For example, Sundara (2005) showed that the spectral characteristics of stop bursts play a role in disambiguating the voiced stop from the voiceless stop, despite the overlapping VOT in Canadian French. The burst intensity relative to the intensity of the following vowel was significantly higher in voiceless stops than in voiced stops in Canadian French. The questions we address for Japanese, then, are the following. Given that the voiced stops have this short lag variant, why are they not mastered as early as the voiced stops of English? Can the role of these other acoustic properties in differentiating two ostensibly short lag types explain why voiced stops are mastered so late in Japanese?

The questions for Greek are the opposite of the questions for Japanese. Greek-speaking children master voiced stops at very young ages (Beckman & Edwards, 2010; Kollia, 1987; Okalidou et al., 2010). We speculate that this early mastery pattern of voiced stops in Greek might be due to their amplitude characteristics, which cannot be captured by VOT. Despite the similarity of temporal properties between Greek and other languages that contrast voicing lead and short lag VOT stops, Greek voiced stops are differentiated from those of other languages in terms of the effects on voicing amplitude of the maneuvers used to produce prevoiced stops. For example, supra-glottal air pressure can be vented by making a "leaky" oral seal, as in the spirantized variant of Spanish intervocalic voiced stops (Cole, Hualde, & Iskarous, 1999; Lewis, 2002; Romero, 1995). It also can be vented by opening the velopharyngeal port to some degree and then closing it off just before the release of the oral closure (Iverson & Salmons, 1996; Yanagihara & Hyde, 1966). In Greek, voiced stops contrast with nasals in word-initial position, but they are optionally prenasalized in adult productions, in a pattern that has been linked to sociolinguistic variation across regional and class dialects, genders, and registers (Arvaniti, 2007; Arvaniti & Joseph, 2000, 2004; Charalambopoulos, Arapopoulou, Kokolakis, & Kiratzis, 1992). The existence of such voiced stop variants appears to influence young children's ability to maintain voicing during stop closure so as to produce truly voiced sounds. For example, the Spanish-learning children in Macken and Barton (1980b) consistently substituted the spirantized allophones of voiced stops for their word-initial voiced stops, suggesting that they had adopted a strategy of having a leaky oral seal to overcome the difficult aerodynamics of making a truly voiced stop. The questions we address for Greek, then, are the following. Does the prenasalized variant serve a similar function in allowing children to produce adult-like lead VOT values by opening the naso-pharyngeal port to vent air through the nose? Does the fact that Greek voiced stops are optionally prenasalized help to explain why Greek children can produce adult-like lead VOT values at an early age, as found in Okalidou et al. (2010)?

The purpose of this study was to explore these explanations for the apparently exceptional patterns of voiced stop mastery in Japanese and Greek. It reports results of two experiments, in which we first compared Japanese to English and then Greek to Japanese.

2. Experiment I: Japanese and English

2.1. Additional acoustic cues in Japanese voiced stops

As noted above, Japanese voiced stops today are similar to English voiced stops in often showing short lag values in word-initial position. Yet transcription studies suggest that Japanese voiced stops are mastered later than English voiced stops. Studies of laryngeal timing patterns show that the adduction gesture is consistently later for voiceless stops than for voiced stops in English (e.g., Flege, 1982), but not in Japanese (e.g., Takada, 2004b). The fact that Japanese voiced stops with short VOT values then overlap with voiceless stops with intermediate VOT values led us to speculate that Japanese-acquiring children need to learn to control other acoustic parameters in addition to VOT in order

to differentiate their voiced stops from voiceless stops. The goal of this first experiment, then, was to test the hypothesis that VOT alone differentiates the two voicing categories in English, but not in Japanese.

To test this hypothesis, we examined whether the acoustic properties of voice quality and fundamental frequency (f0) differentiated voiced from voiceless stops in English and Japanese. The f0 value at the onset of the following vowel has been shown to be a cue to the voicing contrast in many languages (Francis, Ciocca, Wong, & Chan, 2006; Haggard, Summerfield, & Roberts, 1981; Hombert, 1977; Hombert, Ohala, & Ewan, 1979; Ohde, 1984). For Japanese, too, Shimizu (1989) showed that voiceless stops tend to be followed by a significantly higher f0 than voiced stops. An acoustic measure of voice quality might also be useful, given the association between breathy voice and the lax stop series in Seoul Korean, another language that historically was characterized as having a contrast between short lag and intermediate lag VOT values (see, e.g., Holliday & Kong, 2011). We quantified the difference between modal and breathy voice quality by measuring the amplitude difference between the first harmonic and the second harmonic (H1 − H2). Breathy voice quality is associated with a greater H1 − H2 value (Esposito, 2004; Gordon & Ladefoged, 2001; Hanson, 1997; Hanson & Chuang, 1999; Holmberg, Hillman, & Perkell, 1988; Klatt & Klatt, 1990).

In this section, two cross-linguistic comparisons are made between Japanese and English stops. First, we compare the order of mastery of voiced stops and voiceless stops in English and Japanese based on native speaker transcription. Second, we examine the role of three acoustic parameters (VOT, f0, H1 − H2) in differentiating voiced from voiceless stops in the productions of English and Japanese adults and children.

2.2. Methods

2.2.1. Subjects

Japanese- and English-speaking children (2;0–5;11) and adults (18;0–30;0) were recruited and recorded in Tokyo, Japan and in Columbus, Ohio, USA, respectively. 51 Japanese children and 50 English children participated, along with 20 Japanese and 15 English adults. Table 1 shows the participants' age distribution. All of the child and adult subjects passed a hearing screening. Adult subjects reported no history of speech, language, or hearing problems. All children had age-appropriate speech and language, based on parent and teacher report.

2.2.2. Materials

Target consonants were word-initial lingual stops in three non-high vowel contexts. For Japanese, the three contexts were /e/, /a/ and /o/. For English, the vowel contexts grouped together vowels with comparable coarticulatory effects (for example, /ɛ/ and /e/ were grouped together into the "/e/" context for English; see Edwards and Beckman (2008) for the full list of groups). Three words were selected to elicit the target in each context. The target words were familiar to young children (see Appendix A for the lists of words). Culturally appropriate pictures of the target words were selected as stimuli to help prompt repetitions of the words.

2.2.3. Task and audio stimuli

Target consonants were elicited using a picture-prompted word-repetition task. On each trial, a computer program presented a picture stimulus on the computer monitor and played a recorded audio stimulus naming the picture through an external speaker. Subjects were asked to repeat each word immediately after the audio presentation. Their repetitions were recorded using a flash card recorder, and a cardioid pattern unidirectional microphone. (As noted in Edwards and Beckman (2008), we chose this method instead of a picture-naming task in order to be able to elicit word productions across a wide range of ages in a uniform way. In traditional picture-naming tasks, live voice prompts are used when the child does not know the word or fails to recognize the intended word from the picture. While there is no clear evidence that repetitions of live voice prompts differ from spontaneously named productions, the lack of uniformity in prompt is a potential confound in a cross-sectional study of age effects.)

Audio stimuli were made by recording multiple tokens of each target word produced in a child-directed style by an adult female native speaker of each language. In order to select tokens to be used in the auditory word repetition task, we presented multiple stimulus items in a shadowing task for repetition by five adult native-speakers. Only tokens that were repeated correctly by at least four of five adult speakers were included in the study. Three tokens of each word type were selected as stimuli. A Latin square design was used for selecting which tokens of each word type were included across participants. The adults used for stimuli selection were not included in the production tasks described below.

2.3. Analysis

2.3.1. Transcription analysis

A trained native-speaker transcriber coded all productions of word-initial stop consonants as correct or incorrect. Although the transcriber could listen to the entire word, she was instructed to focus on the initial consonant–vowel sequence to minimize lexical effects on accuracy judgments. When tokens were judged as incorrect, the transcriber provided an alphabetic transcription of those productions. While non-plosive productions were regarded as errors, place of articulation errors were ignored for the current analysis, as long as the voicing categories were correct. For example, [g] or [dʒ] for a target /d/ in English was counted as correct for voicing despite the error in the place of articulation. We excluded tokens that were deletions, distortions or non-plosive productions from the analysis. Table 2 summarizes the distribution of tokens that were selected for analysis.

2.3.2. Acoustic analysis

VOT was measured by subtracting the time of the burst from the time of the first indication of voicing, as evident from the voicing bar in the spectrogram, as well as a visible initiation of the first regular cycle of periodicity in the waveform. The f0 was

Table 1
Age distribution of the Japanese and English speaking participants.

Language	Gender	2; 0 − 2; 11	3; 0 − 3; 11	4; 0 − 4; 11	5; 0 − 5; 11	Adults	Total
Japanese	Male	7	5	8	7	10	37
	Female	6	7	6	5	10	34
English	Male	6	7	6	8	8	35
	Female	7	4	6	6	7	30

Table 2
The distribution of the consonant tokens produced by the Japanese or the English speaking children and adults that were used in the acoustic analysis.

Language	Consonant	2; 0–2; 11	3; 0–3; 11	4; 0–4; 11	5; 0–5; 11	Adult male	Adult female
Japanese	d	90	99	121	106	89	90
	t	97	96	121	108	89	90
	g	107	100	123	102	89	90
	k	104	91	118	104	89	90
English	d	69	91	106	118	63	72
	t	71	80	105	117	63	72
	g	34	32	32	40	21	24
	k	36	28	28	37	21	24

measured by taking the reciprocal of the interval between two neighboring pulses at 20 ms after the voicing onset. The glottal pulses were automatically detected using the pulse function in Praat (this is an autocorrelation-based periodicity detector). The analysis window was taken at 20 ms after the voicing onset instead of immediately at the voicing onset because it was observed that the function was more reliable (had a higher correlation coefficient) several glottal pulses after the exact onset of voicing. We used H1 − H2 as our spectral tilt measurement. It was measured by subtracting the amplitude (in dB) of the second harmonic from the amplitude of the first harmonic in the Fast Fourier Transform spectrum generated based on a 25 ms long Hamming window beginning at the voicing onset. The frequency location of the first harmonic was automatically detected by Praat, and then the researchers checked the precise frequency locations for the first and the second harmonic in a token-by-token manner so that DC noise and A1 (the amplitude of the first formant) would not be mistaken as H1 and H2, respectively. There were instances in which there was no clear second harmonic because the spectral roll-off was so extreme as to cause 0 dB SNR or even negative SNR at the frequency of that component. For these tokens, the frequency at twice the first harmonic was regarded as the frequency of the second harmonic, and the amplitude at that frequency location was taken as H2. (Including these tokens yields a more conservative estimate of the effect size by over-estimating the amplitude of the second harmonic for extremely breathy tokens.)

2.3.3. Statistical analysis

First, we used a mixed effects logistic regression model to predict the log odds of accuracy as a function of the child's age. The stop voicing categories (i.e., voiced vs. voiceless) were entered as another predictor. The model allowed the intercept to vary according to each speaker. We also evaluated the effects of each acoustic parameter using a mixed effects logistic regression model. The log of the odds of a production being a voiceless stop was predicted by the three acoustic parameters (VOT, H1 − H2 and f0). The slope of the VOT parameter was allowed to vary at the individual speaker level. The significance of the effect of each predictor was tested via the deviance test. The improvement of a model-fit can be evaluated through the change in deviance given the number of parameters added to the base model (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). To carry out the deviance test, three variations of the logistic regression models were created as expressed in Eq. (1a), (1b) and (1c). Each of the latter two models was compared with the base model (Eq. (1a)) to estimate the change of deviance resulting from the added parameter. In other words, the comparisons of each model fit to the base model, which has VOT as the only predictor tells us whether the added acoustic parameter (H1 − H2 or f0) is a redundant variable to the model in which the voicing contrast is explained. In addition, we calculated the percentage of correctly predicted categories by each model. If complex models with H1 − H2 or f0 do not yield higher percentages of correctly predicted categories than the base model, then this result suggests that the additional parameters are redundant to VOT.

The effect size of each explanatory variable was quantified by referring to the coefficients ($\beta_1, \beta_2 \ldots$) of the mixed effects logistic regression. The model estimated the fixed effect variables that were significant in the deviance test with random slopes of each variable at each speaker level. The greater the absolute value of the coefficient, the greater the role of the parameter in explaining the dependent variable. To remove the magnitude differences among measurement units (millisecond, decibel and hertz), the three acoustic parameters were standardized using a *z*-score transformation before fitting the regression models. The only difference in the models of the adults' productions and the children's productions was that the adult-speaker models used the target categories (e.g., /d, g/ versus /t, k/) as the dependent variables, whereas the child-speaker models used the transcribed categories (e.g., [d, g] versus [t, k]) as the dependent variables.

(a) Base model (VOT-only)

$$\log\left(\frac{/voiceless.stop/}{1-/voiceless.stop/}\right) = \beta_0 + \beta_1 VOT + \gamma Speaker \quad (1a)$$

(b) f0 model (f0 added to the base model)

$$\log\left(\frac{/voiceless.stop/}{1-/voiceless.stop/}\right) = \beta_0 + \beta_1 VOT + \beta_2 f0 + \gamma Speaker \quad (1b)$$

(c) H1 − H2 model (H1 − H2 added to the base model)

$$\log\left(\frac{/voiceless.stop/}{1-/voiceless.stop/}\right) = \beta_0 + \beta_1 VOT + \beta_2 H1H2 + \gamma Speaker \quad (1c)$$

2.4. Results

2.4.1. Transcribed accuracy

There was a difference in the production accuracy pattern of stops between Japanese and English, as shown in Fig. 1. Consistent with prior findings from longitudinal studies (Macken & Barton, 1980a; Zlatin & Koeigsknecht, 1975) and cross-sectional norming studies (Smit, Hand, Freilinger, Renthal, & Bird, 1990), both voiced and voiceless stops of English were produced by children with more than 75% voicing accuracy before 24 months. Japanese voiceless stops also have more than 75% voicing accuracy even before 24 months. However, Japanese voiced stops had a lower voicing accuracy than voiceless stops in the same age range as indicated by the regression curve from the mixed effects logistic regression model. This lower accuracy for Japanese voiced stops follows the earlier cross-linguistic findings of later mastery

Fig. 1. Transcribed accuracy of voicing categories for children's word-initial stops in English (left panel) and Japanese (right panel) as a function of the child's age in months. The curves are the inverse logit curves predicted by the mixed-effects logistic regression models.

Fig. 2. Histograms of VOT, f0 and H1 − H2 measured in voiced and voiceless stops produced by English-speaking adult males and females. Vertical lines are the medians.

of 'true' voiced stops in Spanish (Macken & Barton, 1980b), French (Allen, 1985) and Thai (Gandour et al., 1986).

2.4.2. Acoustic patterns

2.4.2.1. Adults' productions. Figs. 2 and 3 show the distributions of the VOT values for English and Japanese adult speakers separated by gender. In English, VOT values of word-initial voiceless stops were in the long lag range. The voiced stops in English were realized predominantly with short lag VOT values, with less than a quarter of the tokens showing lead VOT. That is, only 23% (73 out of 314) were 'true' voiced stops. This bimodal distribution of the voiced stops was found in both genders. The VOT values for voiced vs. voiceless stops show almost no overlap, confirming that VOT successfully differentiates voiced and voiceless stops in English.

Similar to English, Japanese voiced stops have two variants, a voiceless unaspirated variant (with short lag VOT) in addition to 'true' voiced stops with lead VOT values (the right panels of Fig. 3). 79.92% (309/418) of voiced stops had short lag VOT values so that again, fewer than a quarter were 'true' voiced stops. The pattern is consistent with the findings in Takada (2004a, 2011). This variation in Japanese voiced stops appears to be associated with gender, in that 11 prevoiced tokens (out of 13 in total) in the females' productions were made by only three female speakers, whereas only two male speakers produced a majority of voiced targets with short lag VOT values.

In contrast to English, Japanese voiceless stops have intermediate lag VOT values that are between those of English voiceless stops and voiced stops. No clear-cut boundary was found between voiced stops and voiceless stops in Japanese along the VOT continuum, when productions of all speakers were combined. The wide range of the interval of overlap between voiced and voiceless stops in Japanese reflects both the intermediate VOT values of voiceless stops (3–94 ms) and the presence of voiceless variants of voiced stops (0–56 ms).

The middle panels of Figs. 2 and 3 are histograms of the f0 values 20 ms after vowel onset for the two stop voicing categories in English and Japanese. A common pattern was that voiceless stops in both languages tended to have higher f0 values than voiced stops. This pattern was observed for stops produced by both males and females.

The rightmost panels of Figs. 2 and 3 show the distributions of the breathiness measure for English and Japanese speakers. While English H1 − H2 values for female speakers were higher than those for male speakers, H1 − H2 values for voiceless stops for both males and females were greater than those for voiced stops. Similarly, in Fig. 3 for Japanese, H1 − H2 values for voiceless stops

Fig. 3. Histograms of VOT, f0 and H1 − H2 measured in voiced and voiceless stops produced by Japanese-speaking adult males and females. Vertical lines are the medians.

Table 3
The log-likelihood ratio tests (the deviance test) between the models of adult productions.

Group	Models (parameter)	Df	LLK[a]	χ^2	χ df	Pr (> χ^2)
Japanese						
Female	Base (VOT)	3	−145			
	H1 − H2 (VOT + H1 − H2)	4	−135	20.23	1	< 0.001
	F0 (VOT + f0)	4	−116	58.46	1	< 0.001
Male	Base (VOT)	3	−94			
	H1 − H2 (VOT + H1 − H2)	4	−84	19.96	1	< 0.001
	F0 (VOT + f0)	4	−94	1.25	1	0.26

[a] Log-likelihood.

Table 4
The parameter estimation (i.e., fixed effect coefficients) of the optimal models in each group.

Parameter estimation (fixed effect coefficients)	Parameters	Estimate	Std. error
English	Predictors: VOT		
	VOT	31.27	8.25
Japanese (female)	Predictors: VOT + f0 + H1 − H2		
	VOT	7.75	1.35
	f0	2.35	0.45
	H1 − H2	1.31	0.28
Japanese (male)	Predictors: VOT + H1 − H2		
	VOT	14.68	1.96
	H1 − H2	1.21	0.30

were generally higher than those for voiced stops across both male and female speakers.

Table 3 summarizes the results of the deviance tests, in which each of two different mixed effects logistic regression models with multiple parameters (Eq. (1b) and (1c)) was compared to a simpler model, which had only the VOT parameter (Eq. (1a)) to test the efficiency of added acoustic parameters. For the English adult productions, it was not meaningful to carry out the log-likelihood ratio test between the base model (VOT-only) and the models that also contained either f0 or H1 − H2, because the base model with VOT alone could yield an almost perfect fit. That is, the base model predicted the target categories with an accuracy of 100%. Therefore, VOT was sufficient to separate the voiceless stops from voiced stops in English.

In contrast, the base model in Japanese did not predict the target categories with perfect accuracy, and there was an even lower accuracy rate for females (80.6%) relative to males (88.5%). There was a significant contribution of the two acoustic parameters (H1 − H2 and f0) to the prediction of stop voicing categories in Japanese. The addition of f0 or H1 − H2 to the base model for Japanese-speaking females' productions yielded significant log-likelihood ratio differences, increasing the accuracy of model prediction of the target categories to 90% and 84%, respectively. For Japanese-speaking males' productions, the addition of H1 − H2 (but not f0) to the VOT model resulted in a significant change in the log-likelihood ratio. This model correctly predicted 91.3% of the two voicing categories. These results indicate that H1 − H2 and f0 were not redundant factors in the model that differentiates voiceless stops from voiced stops in Japanese. It is noted that even higher percentages of correctly predicted categories were obtained when logistic regression models were made separately for individual speakers. When the model just had a VOT predictor, there was only one individual's model (out of 10 female adults' models) that had more than 90% accuracy of model prediction. The addition of f0 and H1 − H2 to this VOT model increased the model prediction accuracy in nine female individual models. As a result, eight individual female models out of 10 females had at least 92% accuracy of model prediction (100% in three individual models). There were five male speakers' individual models with more than 92% prediction accuracy with VOT predictor alone (100% in four individuals' models). When VOT, f0 and H1–H2 were predictors, the increased accuracy of model predicted categories were observed in five male individuals' models. In the end, there were seven male individuals' models of more than 92% prediction accuracy. This suggests that there are substantial inter-individual differences in how Japanese adults carve up the acoustic dimensions for the stop voicing contrast.

The coefficients for each of the acoustic parameters from the mixed effects models in English and Japanese base models are presented in Table 4 and Fig. 4. It can be observed that the VOT

coefficient was significantly greater in English than in Japanese, as shown by the steeper shape of the fitted regression curve for English in Fig. 4. Table 4 also presents the coefficients of H1 − H2 and f0 parameters as well as VOT estimated in Japanese males' and females' models. While the coefficients of VOT were greatest in both models, the coefficients of f0 and H1 − H2 in females' model were relatively higher than in males' model. The relative effects of these parameters are illustrated by steepness of slopes in Fig. 5.

2.4.2.2. Children's productions. Fig. 6 shows histograms of the VOT values for target voiced and voiceless stops produced by English- and Japanese-speaking children, separated into four different age groups. In English (the left panels), the voiced stops were clustered in the short lag VOT range and the voiceless in the long lag VOT range. The youngest children showed more variability. That is, in the graphs for the English-speaking 2-year-olds, some voiced stops are realized with long lag VOT values and some voiceless stops are realized with very long lag VOT values, resulting in a less peaky distribution, and a somewhat greater median VOT value for the voiceless stops.

Fig. 4. The fitted curves of the logistic regression models in which the voiceless stops produced by English-speaking and Japanese-speaking adults are predicted by the VOT parameter alone.

The VOT values of voiced stops in children's productions patterned similarly between English and Japanese. As in the English voiced stops, a majority of the Japanese voiced stop targets were realized with short lag VOT values. The one difference to the adult pattern was that there was no gender-related pattern of boys producing more tokens of the lead VOT variant for the voiced stops. The voiceless stops had only slightly longer VOT values overall, but there were different degrees of overlap between the target voiced and voiceless stops across the four age groups. Older children appeared to make a better separation between target voiced stops and voiceless stops than younger children.

Table 5 summarizes the output of the deviance tests for the Japanese-speaking children's productions. The complex models with H1 − H2 or/and f0 included additional predictors had significantly different log-likelihood ratios from that of the base model (VOT-only). The base model predicted 82.9% of the transcription categories correctly. When H1 − H2 or f0 or both were added as a fixed variable, the prediction accuracy improved: 84.7% (H1 − H2 only), 85.2% (f0 only) and 87% (both). Similar to the findings for the Japanese-speaking adults' models, logistic regression models for individual children had higher prediction accuracy. There were 12 individual models with more than 90% of prediction accuracy when the predictor was VOT only. When H1 − H2 or/and f0 were added to each individual's models, 26 individual models had at least 90% prediction accuracy. This pattern contrasts strikingly with the pattern for the English-speaking children, where the base model alone predicted 93.7% of the transcription categories, and adding f0 and/or H1 − H2 did not improve the prediction accuracy rate.

Fig. 7 further illustrates the relative effects of VOT, f0 and H1 − H2 in differentiating transcribed voiceless stops from voiced stops in both languages. These effects are represented by the fixed effect coefficients of the mixed effects logistic regressions in Japanese- and English-speaking children's productions. Those regression models included all three acoustic parameters (VOT, f0 and H1 − H2) as fixed effects. While the coefficients for VOT

Fig. 5. The estimated probability of the stops being voiceless or voiced with respect to VOT, f0 and H1 − H2. The overlaid curves were the model fits of the mixed-effects logistic regression models of Japanese-speaking females' and males' productions. Each data point indicates the proportion of voiceless stops for a given acoustic value produced by individual speakers. The size of data points indicates how many tokens by a speaker belong to the probability. The estimated probability of '1' indicates voiceless stops, whereas '0' indicates voiced stops. The exact coefficient values are shown in Table 4.

Fig. 6. Histograms of English and Japanese stop VOT values separated into the four different age groups (2;0–2;11, 3;0–3;11, 4;0–4;11 and 5;0–5;11). Vertical lines indicate the median VOT for voiced/voiceless stops in each age group.

Table 5
The output of deviance tests between the models and parameter estimation (the fixed effect coefficients) of the mixed effects logistic regression models for children's productions.

(a) Deviance test						(b) Parameter estimation		
Group	Parameters	Df	LLK	χ^2	χ df	Parameters	Estimate	Std. error
						Predictors: VOT+f0+H1−H2		
English	Base (VOT)	3	−276			VOT	6.89	0.42
	H1−H2 (VOT+H1−H2)	4	−272	**10.9**	1	f0	0.45	0.12
	F0 (VOT+f0)	4	−279	**6.9**	1	H1−H2	0.53	0.12
Japanese	Base (base)	3	−738			VOT	4.25	0.32
	H1−H2 (VOT+H1−H2)	4	−707	**62.7**	1	f0	0.81	0.11
	F0 (VOT+f0)	4	−691	**93.9**	1	H1−H2	0.72	0.12

were greatest among the three acoustic parameters in both Japanese and English, the ratio of the VOT coefficient to the f0 and H1−H2 coefficients was greater in English than in Japanese. This is illustrated in Fig. 7.

2.5. Discussion

In Experiment I, we examined the acoustic characteristics that differentiate voiced and voiceless stops produced by English- and Japanese-speaking adults and children. Adult productions of Japanese voiced stops differed from English ones in that the VOT values for the many short lag variants overlapped with the intermediate VOT values of Japanese voiceless stops. This greater overlap between the two categories left considerably more room for H1−H2 and f0 to play a role. The VOT realizations for voiced and voiceless stops produced by Japanese-speaking children patterned similarly to those produced by adult speakers of the language, except that they lacked any gender effect. Also, whereas the English native speaker's voicing judgments for the English-speaking children's stops were predicted very accurately by their VOT values alone, with no role for the other parameters, the Japanese native speaker's transcriptions of the Japanese children's stops were predicted less accurately by VOT values, and prediction accuracy improved to be nearly as high as for the English speaking children's productions only when f0 and H1−H2 values were added to the models.

These differences suggest that stops produced by an English-acquiring child will be judged as accurate in voicing as soon as the

Fig. 7. The estimated probability of the transcribed voiceless stops vs. voiced stops with respect to VOT, f0 and H1–H2. The overlaid curves were the model fits of the mixed-effects logistic regression models of English- and Japanese-speaking children's productions. Each data point of the scatter plots indicates the proportion of voiceless stops for a given acoustic values produced by individual speakers. The estimated probability of '1' indicates the production was transcribed as voiceless, whereas '0' indicates that it was transcribed as voiced. The exact coefficient values are shown in Table 5.

child has achieved a sufficiently adult-like differentiation in VOT values. Stops produced by a Japanese-acquiring child, one the other hand, will be judged as correct only when the child has achieved sufficiently adult-like differentiation of secondary cues such as H1 − H2 and f0 as well. Thus, the late mastery of Japanese voiced stops, despite their short lag VOT values, might be explained, at least in part, by this language-specific use of other acoustic characteristics in differentiating voiced stops from voiceless stops.

It should be noted that it is also possible that the greater variability in the input (with some adults producing voiced stops produced as prevoiced and others as short lag) also contributes to the later mastery of voiced stops in Japanese relative to English. However, phoneticians have only recently noted this sound change in progress (see Takada, 2011), which may be related to the rapid urbanization since the 1950s and massive in-migration to the Tokyo area of workers from the Touhoku region. That is, the shift in VOT values for the initial voiced stops seems to be the spread of a pattern that was attested first in dialects to the north of Tokyo. The shift in the Tokyo dialects may be related to the much better documented change in the sociolinguistically marked position-specific nasal realization of /g/ in urban Tokyo dialects (see, e.g., Hibiya, 1995). That is, in the 1980s, older speakers and speakers from Shitamachi (the working class and lower-middle class wards in the eastern part of the Tokyo metropolitan area) produced a nasal variant of the velar stop in word-medial position, and produced the "hard" stop variant only in word-initial positions. The second author's own casual observations in the early 1980s suggested to her then that some speakers (particularly Shitamachi speakers) had a comparable allophonic variation for /d/ between a weak fricative variant in word-medial position and a "hard" voiced stop in initial position. The age-grading noted in Takada (2004a, 2011) suggests that the lag VOT variant today is a further extension of this "strengthening" of the word-initial allophones reinforcing a "dialect borrowing" of the Touhoku pattern. However, until we have more information about the social demographics of this more recent change, it is premature to speculate further about the effect of this variation on children's acquisition of the voiced stops today.

3. Experiment II: Greek and Japanese

3.1. Measuring nasality in Greek voiced stops

While VOT plays an important role in differentiating voiced stops with lead VOT from voiceless stops with short lag VOT in Greek, the temporal measure of VOT cannot be used to describe prenasalization. Burton, Blumstein, and Stevens (1992) found that phonemically contrastive voiced stops, prenasalized stops and nasals in Moru were distinguished by their amplitude characteristics. When the amplitude trajectories during the voicing bar of voiced and prenasalized stops and the nasal murmur of nasals were compared, the two types of stops were distinguished from nasals by lower amplitude energy either throughout the closure or in an interval of some duration just before the burst, and prenasalized stops were distinguished from oral voiced stops by the higher energy of the nasal murmur during at least the earlier part of the closure. In other words, prenasalized stops are characterized as having higher (nasal-like) amplitude at the beginning and lower (voiced stop-like) amplitude at the end. As Chan and Ren (1987), Gordon and Maddieson (1999), Demolin, Haude, and Storto (2006), and others have pointed out, prenasalized stops may vary in the extent to which amplitude is high at the beginning and/or lowered at the end, resulting in a continuum of types from "hyper-voiced stops" (Iverson & Salmons, 1996) to "post-stopped nasals" (Chan & Ren, 1987). Arvaniti and Joseph (2000) also note considerable variation in the realization of voiced stops in Greek, with the ratio of the nasal part to the oral part at the end varying between and even within speakers.

In the second experiment, we aimed to uncover acoustic evidence of some degree of prenasalization of Greek voiced stops by making a within-language comparison of voiced stops to nasal consonants in Greek and then by making a cross-linguistic comparison to Japanese voiced stops and nasals. First, we hypothesized that the high incidence of prevoicing will be related to the option of nasal venting in prenasalized stops. If Greek voiced stops are prenasalized, we predict that Greek-speaking children will produce many voiced stop productions with lead VOT values because nasal venting during the closure provides a favorable aerodynamic

condition for prevoicing. By contrast, the low frequency of prevoiced tokens in Japanese is most likely related to Japanese children and adults not using nasal venting when producing prevoicing in voiced stops. This direct comparison between voiced stops and nasals assesses whether Greek voiced stops acoustically resemble nasals more than Japanese voiced stops do.

3.2. Methods

3.2.1. Subjects

Child (2;0–5;11) and adult (18;0–30;0) subjects were recruited in Tokyo, Japan, and Thessaloniki, Greece. 93 Greek and 80 Japanese children and 6 Greek (three male and three female) and 20 Japanese (10 male and 10 female) adults participated in the task. The Japanese adult participants were the same speakers as in Experiment I; the Japanese child speakers were the 51 speakers from Experiment I along with an additional 29 Japanese children. Table 6 shows the child subjects' age distribution. Adult subjects reported no history of speech, language, or hearing problems. Both adult and child subjects passed a hearing screening. All children had age-appropriate speech and language, based on parent and teacher report.

3.2.2. Materials and tasks

Voiced stops and nasals were elicited in word-initial position in real words. Target consonants included /b/, /d/, /m/ and /n/ (only /d/ and /n/ were elicited from Japanese adults). The sets of words for children and adults were not completely identical, although the target words overlapped between these two participant groups in Greek because the set of words for children was a subset of the words elicited from adults (see Appendix B).

The children's productions were elicited using a picture-naming task. If a given picture failed to cue the target word, the researcher directly asked the children to repeat after the researcher's voice prompt. The adults did a slightly different task from the children. Greek target consonants were collected from a word-reading task presented on a computer screen, whereas Japanese target consonants were collected from the word-repetition task described in Experiment I. The productions by children and adults were digitally recorded using a unidirectional microphone. Experimenters were native speakers of the target languages. Tables 7 and 8 provide the distributions of tokens collected from adults and children.

Table 6
The age distribution of Greek and Japanese child subjects.

Language	Gender	2; 0–2; 11	3; 0–3; 11	4; 0–4; 11	5; 0–5; 11	Total
Greek	Boys	12	12	11	11	46
	Girls	10	12	13	12	47
Japanese	Boys	10	13	10	8	41
	Girls	9	5	15	10	39

3.2.3. Acoustic analysis

The duration and amplitude characteristics were measured in the voicing lead interval for the voiced stops and in the nasal murmur for nasals. The duration was measured based on the interval between the two acoustic landmarks of prevoicing onset and burst. The prevoicing onset was determined based on the first evidence of periodicity in the waveform, and burst was defined as the first spike of aperiodic energy. For nasal consonants, the beginning of the nasal murmur was considered the onset of the nasal (analogous to the onset of prevoicing in a voiced stop) and the boundary between the nasal and the following vowel was considered the offset of the nasal (analogous to the burst of a voiced stop). We will refer to this offset location as the burst in discussing both cases for expository convenience. Therefore, the duration of the voice bar in voiced stops (i.e., the absolute value of the VOT) corresponds to the duration of the nasal murmur in the nasals.

Amplitude values over the voice bar and the nasal murmur were extracted to describe the energy trajectory. Replicating Burton et al. (1992), there were as many amplitude values as periods in the interval starting at the burst and extending to the onset of prevoicing. Each of the amplitude values was measured by taking the first peak amplitude in the FFT spectrum made from a 6 ms Hamming window centered at the glottal pulse for that period. The vowel amplitude was measured to normalize the amplitudes measured in the nasal murmur or voicing lead. The amplitude of the vowel was obtained by measuring the amplitude of the first harmonic in the spectrum of a 25 ms analysis window, which started at the third pulse after the burst, as suggested by Burton et al. (1992).

3.2.4. Statistical analysis

These time-series amplitude values were entered into a mixed effects regression model as the dependent variables. Because absolute duration and f0 (and hence the number of periods) were not constant across productions, we used the amplitudes at five proportional locations—80%, 60%, 40% and 20% of the whole

Table 7
The distribution of acoustically analyzed Greek voiced stops and nasals collected from child and adult participants. Numbers in parentheses indicate the number of tokens used for the amplitude analysis.

Consonant	2; 0–2; 11	3; 0–3; 11	4; 0–4; 11	5; 0–5; 11	Adult male	Adult female
b	20 (18)	24 (21)	24 (21)	22 (19)	57	61
m	12 (12)	20 (19)	21 (20)	21 (20)	54 (26)	60 (60)
d	15 (10)	24 (16)	24 (21)	22 (13)	52 (27)	58 (56)
n	25 (22)	27 (26)	22 (20)	23 (23)	53 (35)	54 (59)

Table 8
The distribution of acoustically analyzed Japanese voiced stops and nasals collected from child and adult participants. Numbers in parentheses indicate the number of tokens used for the amplitude analysis.

Consonant	2; 0–2; 11	3; 0–3; 11	4; 0–4; 11	5; 0–5; 11	Adult male	Adult female
b	34 (6)	31 (2)	45 (4)	34 (0)	0	0
m	29 (24)	32 (29)	60 (52)	45 (42)	0	0
d	18 (8)	17 (4)	23 (3)	17 (2)	210 (37)	210
n	34 (31)	38 (34)	60 (53)	49 (48)	137 (44)	150

duration, and the burst location. (We also measured the amplitude at the beginning of the voice bar or nasal murmur, and included that measurement point in figures. We excluded the beginning value in the statistical analysis, so that the abrupt amplitude rise observed in utterance initial position for both types does not dominate the curve fit.) In the regression model, this dependent variable of normalized amplitude was explained by the factor of time (i.e., the five distinct proportional locations). Since the amplitude trajectories turned out to have curvilinear shapes, it was necessary to include orthogonal polynomial parameters up to the third order (i.e., linear, quadrant and cubic terms) as fixed effects (Barr, 2007; Raudenbush & Bryk, 2002; Singer & Willett, 2003; Snijders & Bosker, 1999; Mirman, Dixon, & Magnuson, 2008). Each term of the orthogonal polynomial time parameters captures a different component of the curve shape. The linear term describes the slope of the curve, and the higher order polynomials describe the symmetric (quadratic or second order) and asymmetric (cubic or third order) peaks/valleys of the curve. The regression model had random effects for both tokens and speakers.

The effects of each orthogonal polynomial term were assessed by the deviance test as described in experiment I. We used the fixed intercept model as a base and then we added each polynomial term to the set of fixed effect variables in order to estimate the amount of improvement for each additional parameter relative to the base model. The sign and magnitude of the coefficients of the three polynomial parameters indicate the direction and rate of excursion of the curve.

3.3. Results

3.3.1. Greek and Japanese adult productions

3.3.1.1. Duration. Greek voiced stops produced by adults had voicing lead in most cases regardless of gender. 94.3% (215 out of 228) of Greek voiced stops were produced with a voicing bar in the adult productions. Fig. 8 (top panels) shows the histograms of stop voice bar duration separated by gender. The mean durations of the voice bars of voiced stops (/b/ and /d/) produced by female speakers were 102 ms and those by males were 57 ms.

In contrast, Japanese voiced stops were realized with voiceless variants (short lag VOT values) as well as with prevoiced variants (lead VOT values), as reported in Riney et al. (2007) and Takada (2004a). When separated by gender, 10.4% (22 out of 210) of females' voiced stops and 43.8% (92 out of 210) of males' voiced stops had lead VOT values. Fig. 8 (bottom panels) shows the duration distributions of Japanese voiced stops separated by gender. The mean duration of the voice bars of /d/ produced by male speakers was 57 ms, and the mean duration of /d/ by females was 37 ms. On average, the voice bars of Greek voiced stops were longer in duration than those of Japanese voiced stops.

3.3.1.2. Amplitude trajectory. Figs. 9 and 10 plot medians and confidence intervals of successive amplitudes measured over the voicing bar of voiced stops and the nasal murmur of nasals separated by speaker. Across all of the Greek adult speakers' amplitude curve patterns, nasals were consistently different from voiced stops at the burst region ($x=0$, rightmost side of the curve) by having greater amplitude than voiced stops. The nasal murmur amplitude trajectories have a gradual increase over time without abrupt amplitude fluctuations toward the end. The mean amplitude just before the burst with reference to the amplitude of the following vowel was -5.29 dB (standard deviation$=\pm 3.52$) for nasals as compared to a mean of only -22.2 dB (standard deviation$=\pm 6.08$) for voiced stops when averaged across speakers.

In contrast to the consistent shape of the nasal murmur amplitude trajectory, the prevoicing of Greek voiced stops showed several different patterns of amplitude trajectory across the six Greek adult speakers. The difference was primarily at the beginning of the voice bar. Speaker gtf07 in the top center panel of Fig. 9, for example, began her voiced stops with energy as high as her nasals followed by an energy decrease over time toward the burst. For speaker gtm08 in the bottom right panel of Fig. 9, on the other hand, voiced stops began with energy lower than in his nasals, and he maintained this same amplitude difference throughout the voicing bar. The other Greek adult speakers showed amplitude trajectories similar either to that of speaker

Fig. 8. Histograms showing the distributions of the VOT values of voiced stops produced by Greek- and Japanese-speaking adults separated by gender.

Fig. 9. The amplitude trajectories of Greek voiced stop lead and nasal murmur elicited at word initial position produced by six Greek-speaking adults (voiced stops (/d/) vs. nasals (/m/, /n/)). The abscissa represents the six proportional locations arranged chronologically (*burst* at the rightmost edge), and the ordinate represents the sound pressure normalized with reference to the following vowel amplitude. Zero at the ordinate corresponds to an amplitude equal to that of the following vowel. Each solid trajectory line shows the median values of the amplitude measured in the voiced stops and nasals across the glottal pulses prior to the burst. The dotted lines above and below the solid lines are 95% confidence interval lines.

Fig. 10. The amplitude trajectories of Japanese voiced stop leads of /d/ and nasal murmurs of /n/ in word initial position produced by selected Japanese-speaking male adults.

gtf07 or of speaker gtm08, with minor variations in the relative amplitude at the initial region of the voicing lead. Despite the variation in the amplitude at the initial portion of the voiced stops, the amplitude of the voicing lead in Greek voiced stops near the burst was consistently distinctively lower than that of the nasal murmur in the nasal consonants.

From their match to the description of amplitude changes over the voicing lead and the nasal murmur in Burton et al. (1992), it

appears that Greek voiced stops have the amplitude characteristics of prenasalized stops; they have high amplitude at the beginning, as in an interval of nasal murmur, but a lower amplitude in some interval before the burst. The varying degrees of similarity of the initial portion of the voicing lead of Greek voiced stops to nasal murmurs suggests that Greek voiced stops are prenasalized to varying degrees depending on the speaker.

The amplitude trajectories of the voice bars in the Japanese voiced stops and of the nasal murmurs in the nasal consonants are shown in Fig. 10. Only male productions were analyzed since females produced few voiced tokens with prevoicing. The Japanese nasal amplitude characteristics were congruent with those of the Greek nasals in that the amplitude of the Japanese nasal murmurs gradually increased over the murmur. However, compared to amplitude characteristics of Greek nasals, the overall energy of individual speakers' nasals was more variable in Japanese. The normalized nasal murmur amplitude just before the burst was, on average, −4.3 dB (standard deviation= ±2.3 dB), which was consistently higher than the amplitude of the voice bar just before the voiced stop burst (mean= −20.2 dB, standard deviation= ±5.2 dB).

The amplitude curves of the Japanese voiced stops patterned slightly differently from those of the Greek voiced stops for these adult speakers in that no individual speaker showed steep amplitude falls. Instead, the amplitude curves of voice bars in the Japanese voiced stops either decreased gradually or were at a sustained level parallel to the amplitude trajectories of the nasal murmurs. In either case, the initial amplitude values in the voiced stops were consistently lower than in the nasal murmurs.

3.3.1.3. Mixed effects models of amplitude trajectory. The amplitude trajectories of the Japanese and Greek voiced stops were quantified by a mixed effects regression model. As described in Section 3.2.4, the orthogonal polynomial time parameters were the explanatory variables of the regression and we used only the final 80% of the curve duration. That is, we removed the initial 20% of the amplitude signal (the data points aligned with '0' in Figs. 9 and 10) so that the overall shape of the amplitude trajectory was not mischaracterized by the abrupt amplitude rise, which presumably occurs in utterance initial position. Fig. 11 shows both the group (fixed effect) and individual-level (random effect) model fits for the amplitude trajectories of voiced stops and nasals produced by Greek- and Japanese-speaking adults. The effects and the coefficients of the model parameters are summarized in Table 9.

Fig. 11. Model-fits for the amplitude trajectories of voiced stops and nasals produced by Greek-speaking (left) and Japanese-speaking (right) adults. Time (proportional values of the total duration) is shown on the x-axis and sound pressure (normalized with respect to the amplitude of the following vowel) is shown on the *y*-axis. The averaged linear slopes are drawn in thick lines, and the speaker-level variations of the linear slopes were overlaid to show the individual differences observed in Figs. 9 and 10.

Table 9
The output of deviance tests and the parameter estimations: adult productions of nasals and voiced stops in Greek and Japanese (bold indicates $p<0.05$).

Group	(a) Deviance test				(2) Parameter estimation		
	Models (predictors)	LLK	χ df	χ^2		Estimate	Std. error
Voiced stops							
Greek	Intercept (intercept)	−1065			Intercept	−17.94	1.04
	Linear (intercept+linear)	−1063	1	**4.04**	Linear	−4.43	1.99
	Quadratic (intercept+linear+quadratic)	−992	1	**141.79**	Quadratic	−3.37	0.23
	Cubic (intercept+quadratic+cubic)	−987	1	**9.73**	Cubic	0.74	0.23
Japanese	Intercept	−307			Intercept	−17.53	1.25
	Linear	−298	1	**17.01**	Linear	−4.09	1.16
	Quadratic	−277	1	**41.97**	Quadratic	−2.70	0.46
	Cubic	−277	1	0.96	Cubic	0.43	0.46
Nasals							
Greek	Intercept	−1918			Intercept	−6.65	0.74
	Linear	−1909	1	**17.90**	Linear	3.63	0.31
	Quadratic	−1851	1	**115.11**	Quadratic	−1.19	0.10
	Cubic	−1851	1	0.61	Cubic	0.08	0.10
Japanese	Intercept	−307			Intercept	−6.05	1.03
	Linear	−298	1	**17.01**	Linear	4.87	0.53
	Quadratic	−277	1	**41.97**	Quadratic	−1.96	0.27
	Cubic	−277	1	0.96	Cubic	0.26	0.27

The amplitude curves for productions of voiced stops for both Greek and Japanese had significant linear and quadratic components. The negative signs of the coefficients for the first order term capture the gradual overall energy decrease with time. Despite the similarity in the overall curve shapes of the averaged model-fits, there were differences between the Greek and Japanese voiced stops in terms of the degree of amplitude fall in individual speakers' productions, as shown in Fig. 11. Specifically, the Greek-speaking adults had much greater individual differences in the linear slope coefficients than did the Japanese-speaking adults. The estimated slopes for the Greek speakers ranged from −8.4 (very steep downtrend) to 1.92 (almost flat or gradual increase) with an estimated σ of 4.65, whereas the estimated slopes for the Japanese speakers indicated more or less shallow downtrends ranging from −4.75 to −2.73, with an estimated σ of only 1.23. Also there was a subtle difference between the Greek and Japanese voiced stops in terms of the effect of a higher order polynomial term in the trajectory models. The important thing to consider in Table 9 is whether the cubic component of the model was significant for the voiced stops. The cubic component was significant for Greek voiced stops, but not for Japanese voiced stops. A significant cubic component indicates an asymmetric curve and a positive sign of the estimated coefficient indicates that the curve was skewed during the onset portion of the amplitude trajectory (i.e., a curve beginning with a skewed rise followed by a fall). This positive cubic component of the curve indicates that there was additional amplitude during the first part of the voicing bar where there was nasal energy. This is observed for the Greek, but not the Japanese, voiced stops.

Another difference emerges when the models for the voiced stops are contrasted with the models for the nasal consonants. Whereas the voiced stops showed more individual variability for the Greek speakers, the nasals showed more individual variability in the Japanese speakers. Specifically, there is a great deal of variability in the overall level of energy in the Japanese nasals as contrasted to consistently high energy in the Greek nasals in Fig. 11. However, despite the difference in the speaker-level variations, the averaged models for the nasal consonants were highly similar between Greek and Japanese, in that for both languages, the effects of the linear and quadratic terms were significant in explaining the amplitude trajectories during nasal murmurs. The positive signs of the coefficients of the linear terms for the two languages suggest that the amplitude curves of nasals linearly increase as they approach the burst.

3.3.2. Greek and Japanese children's productions

3.3.2.1. Duration. In Greek, 84.7% (72 out of 85) of girls' voiced stops and 82.2% (74 out of 90) of boys' voiced stops were produced with prevoicing in word initial position. In each age group, the occurrence of prevoiced voiced stops was generally high (85.7%, 79.1%, 87.5% and 81.8% in the 2∼5-year-old groups, respectively). The left panels of Fig. 12 show the distributions of the durations of the voicing interval for voiced stops (i.e., VOT) and of the nasal murmurs of nasals in Greek separately by group. The right panels of Fig. 12 show the histograms of the durations of the Japanese voiced stops and nasals. Unlike the voiced stops produced by Greek children, very few voiced stops produced by Japanese children were produced with prevoicing in word-initial position. Only 29.3% (32/109) of girls' voiced stops and 24.5% (27/110) of boys' voiced stops were prevoiced. The voiced stops

Fig. 12. The VOT values of the voice bar in word initial position by Greek-speaking (left panels) and Japanese-speaking children (right panels) separated by age group (2;0–5;11).

and nasals produced by children in both languages mimic the adult patterns in that more Greek voiced stops were prevoiced than Japanese voiced stops. Japanese children rarely produced voice bars in their voiced stops, whereas Greek children almost always produced voiced stops with voicing leads.

3.3.2.2. Amplitude trajectory. Fig. 13 shows the trajectories of the median amplitude values of the voice bar of voiced stops and of the nasal murmur of nasals produced by Greek children (left panels) and Japanese children (right panels). Similar to Greek- and Japanese-speaking adult productions of nasals, the amplitude curve of nasal murmur in nasals produced by Greek children lacked abrupt amplitude changes during the murmur. This was true for all four age groups. The amplitude of nasals immediately before the burst location ($x=0$) was higher than the amplitude of voiced stops, although the difference between the two consonant types was smaller than the difference in adult productions. The mean normalized amplitude just before the burst location was -8.6 dB (standard deviation$=\pm 5.1$ dB) for nasals and -13.6 dB (standard deviation$=\pm 6.1$ dB) for voiced stops in Greek children's productions.

The amplitude trajectories of Greek children's voiced stops differed from those of the Greek adults' in that the amplitude stayed high or showed only a very gradual decrease in amplitude over time. That is, even for the oldest children, the amount by which the amplitude dropped as the burst approached was smaller, resulting in less differentiation between the amplitude of voiced stops and nasals at the burst region, as noted above. This amplitude pattern over time suggests that many of the Greek children's voiced stops have a strong nasal quality throughout the voicing lead interval. The duration of nasality during the voicing lead varied such that some voiced stops had a prenasalized property at the initial portion of the voicing lead, and many others had a fully nasal-like quality over the entire duration of the voicing lead. Therefore, Greek children's voiced stops are characterized either as strongly prenasalized voiced stops or as post-stopped nasals. Note that even the most nasal-like productions of /d/ had clear bursts, which distinguished them from /n/, and the Greek-speaking phoneticians who reviewed the stops were consistent in identifying them as stops rather than nasals.

The right panels in Fig. 13 show the amplitude trajectories of Japanese children's nasals and voiced stops. Across all four age

Fig. 13. The amplitude trajectories of voiced stop leads and nasal murmurs in word initial position produced by Greek-speaking (left panels) and Japanese-speaking children (right panels). The *x*-axis represents the 6 proportional locations arranged chronologically, and the *y*-axis represents the sound pressure normalized with reference to the amplitude of the following vowel. Each solid trajectory line shows the median values of the amplitude and the dotted lines below and above the solid line are 95% confidence interval lines.

groups, the amplitude of the nasal murmurs gradually increased just as in the nasals produced by Japanese and Greek adults. Immediately before the burst, the amplitude of the nasals was consistently higher than that of voiced stops where the amplitude means were −9.01 dB (standard deviation= ±4.7 dB) for nasals and −16.6 dB (standard deviation= ±6.5 dB) for voiced stops.

The amplitude trajectory patterns of the Japanese children's voiced stops was different from that of the Greek children's voiced stops in that the amplitude values of the Japanese children's voiced stops were lower than those of the nasals over all frames. Unlike Greek children's voiced stops, the amplitude rarely rose in the middle of the voicing bar nor did it decrease as the burst approached. The amplitude of the Japanese children's voiced stops remained relatively level over the duration of the voicing lead, and this level was lower than the amplitude of the nasals from the beginning of the voiced stops. Although the initial amplitude of the Japanese children's voiced stops was closer to the amplitude of their nasals, the amplitude of their voiced stops was clearly lower than that of their nasals immediately before the burst. This difference in amplitude between voiced stops and nasals increased with age for the Japanese-speaking children.

3.3.2.3. Mixed effects model of amplitude trajectory. The amplitude trajectories of the children's nasals and voiced stops were described by the orthogonal polynomial factors in the mixed effects regression model as summarized in Tables 10–12. The model-fits for the amplitude trajectories are drawn in Fig. 14. As in the adult production models, we analyzed only 80% of the curve duration so as not to include the initial amplitude rise.

In both Japanese and Greek nasals, there were significant effects of linear, quadratic and cubic terms. As shown in Table 10, the coefficients of the intercept and the three polynomial parameters were very similar between the two languages. This similarity suggests that the overall nasal amplitude was comparable in Japanese and Greek, and the amplitude increased at a similar rate till the end of the murmur in the two languages.

In the models for the children's voiced stops, there was no effect of linear slope for any of the age groups except the 2-year old group, which also had significant higher-order polynomials in the amplitude curves. This suggests that the curve shapes of the Japanese and Greek voiced stops were similar, in that the amplitude slope during the voicing bar was more or less flat in the 3-, 4- and 5-year-olds' voiced stop productions, while the amplitude significantly increased over time in the voiced stops of the 2-year-old children.

The most interesting difference between Greek and Japanese voiced stops was found in the intercept coefficients of the models across the age groups (i.e., the overall amplitude level produced by the different age groups). The intercept coefficient for the

Table 10
The output of deviance tests and parameter estimation for children's production of Greek and Japanese nasals (bold indicates $p<0.05$).

Nasals							
(a) Deviance test					(b) Parameter estimation		
	Models	LLK	χ df	χ^2		Estimate	Std. error
Greek	Intercept (intercept)	−3789			Intercept	−8.61	0.30
	Linear (intercept+linear)	−3660	1	**259.21**	Linear	4.61	0.36
	Quadratic (intercept+linear+quadratic)	−3521	1	**278.0**	Quadratic	−1.50	0.15
	Cubic (intercept+linear+quadratic+cubic)	−3514	1	**14.2**	Cubic	0.31	0.15
Japanese	Intercept	−2137			Intercept	−9.01	0.20
	Linear	−2082	1	**109.43**	Linear	4.68	0.23
	Quadratic	−2038	1	**89.31**	Quadratic	−1.80	0.09
	Cubic	−2036	1	**4.26**	Cubic	0.37	0.09

Table 11
The output of deviance tests and the estimated coefficients of parameters for children's production of Greek voiced stops (bold indicates $p<0.05$).

Greek voiced stops							
(a) Deviance test					(b) Parameter estimation		
	Models	χ df	LLK	χ^2		Estimate	Std. error
2-year-olds	Intercept		−381		Intercept	−14.11	0.75
	Linear	1	−370	**22.22**	Linear	5.20	0.90
	Quadratic	1	−370	0.75	Quadratic	0.36	0.41
	Cubic	1	−367	**6.40**	Cubic	1.05	0.41
3-year-olds	Intercept		−496		Intercept	−13.63	0.84
	Linear	1	−496	0.01	Linear	0.17	1.23
	Quadratic	1	−496	0.15	Quadratic	−0.16	0.41
	Cubic	1	−493	**6.37**	Cubic	1.05	0.41
4-year-olds	Intercept		−582		Intercept	−13.59	0.73
	Linear	1	−581	0.78	Linear	0.90	1.02
	Quadratic	1	−575	**13.11**	Quadratic	−1.27	0.33
	Cubic	1	−572	**6.62**	Cubic	0.87	0.33
5-year-olds	Intercept		−443		Intercept	−11.57	1.07
	Linear	1	−443	1.64	Linear	1.43	1.12
	Quadratic	1	−436	**13.49**	Quadratic	−1.36	0.36
	Cubic	1	−436	0.46	Cubic	0.24	0.36

Table 12
The output of deviance tests and the estimated coefficients of parameters: children's production of Japanese voiced stops (bold indicates $p < 0.05$).

Japanese voiced stops							
(a) Deviance test					(b) Parameter estimation		
	Models	χ df	LLK	χ^2		Estimate	Std. error
2-year-olds	Intercept		−163		Intercept	−16.50	1.62
	Linear	1	−160	**5.81**	Linear	3.30	1.26
	Quadratic	1	−153	**14.02**	Quadratic	−1.43	0.32
	Cubic	1	−149	**6.73**	Cubic	0.86	0.32
3-year-olds	Intercept		−77		Intercept	−14.48	0.96
	Linear	1	−77	0.44	Linear	2.09	3.23
	Quadratic	1	−74	**5.90**	Quadratic	−2.19	0.86
	Cubic	1	−73	1.07	Cubic	0.83	0.86
4.5-year-olds	Intercept		−122		Intercept	−18.54	2.50
	Linear	1	−121	0.11	Linear	0.96	2.70
	Quadratic	1	−120	3.40	Quadratic	−2.00	1.01
	Cubic	1	−118	**4.00**	Cubic	2.02	1.01

Fig. 14. Model-fits for the amplitude trajectories of voiced stops and nasals produced by Greek-speaking (left) and Japanese-speaking (right) children. Time (proportional values of the total duration) is shown on the *x*-axis and sound pressure (normalized with respect to the amplitude of the following vowel) is shown on the *y*-axis.

amplitude of the voiced stops produced by 2-year-olds was higher than that of those produced by 4- and 5-year-olds in Japanese, whereas the intercept coefficient of Greek 2-year-olds' voiced stops was lower than that of older age groups of Greek. The higher level amplitude trajectories of Greek-speaking 5-year-olds' voiced stops indicates that their voiced stops were more nasal-like than the voiced stops of younger age groups of Greek. In the same vein, the lower level amplitude trajectories in Japanese 5-year-olds' voiced stops suggest that their voiced stops were the most voiced-stop-like among the age groups in Japanese. To summarize, as age increased, the differences between the amplitude trajectories of voiced stops for the two languages also increased. This can be observed both with respect to the shape of the curves and the overall amplitude level during the prevoicing.

3.4. Discussion

The high amplitude and (for the adults, especially) the amplitude trajectory of the voicing lead captured the prenasalized quality of voiced stops of Greek fairly well. The onset of the amplitude trajectories of the voicing lead in Greek stops was as high as in nasal murmurs, followed (in the adults at least) by a pronounced amplitude drop toward the burst. However, although the degree of similarity between the voiced stops and the nasals varied depending on the speaker, the amplitude of Greek voiced stops immediately before the burst was consistently lower than the nasal amplitude across speakers. Note that while previous research has found that prenasalization of voiced stops is much more common in word-medial than word-initial position, we found a substantial amount of prenasalization even in word-initial position. This might be attributed to a dialect characteristic of the Greek spoken in Thessaloniki or perhaps to the word-repetition elicitation method used in the current study (see Arvaniti and Joseph (2000) for Standard Modern Athenian Greek voiced stops and Charalambopoulos et al. (1992) for Thessaloniki Greek voiced stops). In any case, this prenasalized quality was a language-specific characteristic of Greek voiced stops and was not observed in Japanese voiced stops. That is, all of the Tokyo Japanese speakers in our study were too young to produce even a mildly nasalized variant even for /g/.

A possibly related difference is that nasal consonants in Japanese exhibited varying degrees of overall amplitude levels, whereas the energy in Greek nasals were consistently higher. We speculate that the energy in the nasals can be more variable in Japanese, because voiced stops in Japanese are never strongly prenasalized. In Greek, by contrast, the energy in nasal consonants has to be consistently higher, because the Greek nasals are in contrast with the more variable and often prenasalized voiced stops.

The voiced stops produced by children acquiring Greek or Japanese also showed language-specific characteristics. The Greek-speaking children, like the Greek-speaking adults, produced pre-voiced stops with varying degrees of nasality. In addition, the Greek children showed even more nasalization than the adults, with the nasality extending over the entire duration of the voicing lead. That is, the amplitude decline in the child speakers was often minimal,

making their voiced stops look more like post-stopped nasals than like prenasalized stops. Thus, Greek children's voiced stops might be characterized as strongly prenasalized. By contrast, the prevoiced stops produced by the Japanese-speaking children were not prenasalized, similar to those of the Japanese-speaking adults.

This supports the idea that the earlier acquisition of voiced stops by Greek-speaking children is related to a cross-linguistic difference in the degree of nasal venting during voiced stop production. As we hypothesized earlier, the use of the naso-pharyngeal port opening should help children overcome the difficulty of maintaining a high enough supra-glottal pressure throughout the closure, facilitating the production of prevoicing for voiced stops. The evidence of a prenasalized or nasalized quality shown in the voicing lead of Greek-speaking children's productions suggests that they are taking advantage of the prenasalized nature of the voiced stops of their native language and thus achieve an early production proficiency for voiced stops.

4. Conclusion

This study explored how otherwise puzzling cases of cross-linguistic differences in the mastery patterns for the stop voicing contrast can be explained by a better understanding of the language-specific acoustic characteristics of the phonetic categories. For many languages, the adult VOT norms are a useful acoustic measure for predicting the relative order in which children master stop phonation types. In general, across languages, a common observation has been that the short lag VOT category appears first in children's productions, the long lag VOT category appears next, and the lead VOT category appears last, often after the age of 5 years. We investigated two specific cases in which predictions based on this common pattern failed.

The first case was the late mastery of Japanese voiced stops, many of which are transcribed as incorrect even when they are produced as short lag stops that are well within the adult VOT range of Japanese voiced stops. The second case was the early mastery of Greek voiced stops, which even very young children produced with highly adult-like lead VOT values. In order to understand these two puzzles, we had to look more carefully at the VOT patterns, and we had to look beyond VOT to understand the role of other acoustic parameters.

To resolve the Japanese puzzle, we examined the roles of three acoustic parameters, VOT, H1 − H2 and f0, in a regression model that related these parameters to native speaker transcriptions of Japanese- and English-speaking children's productions. Although VOT was a very important predictor for both languages, H1 − H2 and f0 were more important in Japanese than in English in predicting the transcription categories. These findings suggest that the cue-weightings of the contrast between voiced and voiceless stops are different for Japanese and English. We hypothesized that voiced stops are mastered later in Japanese relative to English, because Japanese-acquiring children need to learn how to produce differences in several acoustic parameters, while English-acquiring children need to learn to how to produce differences in only one.

To resolve the Greek puzzle, we developed an acoustic measure of prenasalization, as Greek voiced stops are optionally prenasalized. We hypothesized that this prenasalization would facilitate the production of prevoicing of voiced stops, since an open naso-pharyngeal channel should help children meet the aerodynamic conditions for initiating and maintaining voicing during closure. By comparing the spectral characteristics of the voicing lead in voiced stops and of the nasal murmur in nasals, we showed that the amplitude characteristics of the voicing bar at the beginning of voiced stops was highly similar to those of nasals in Greek. The Greek-acquiring children were mimicking this partially nasalized property of voiced stops, and further extended the nasal quality to the entire duration of the voicing lead. We inferred that Greek children overcome the difficulty of prevoicing by using the nasal venting of intraoral air pressure during closure.

A consistent finding across these two cases was that even the cross-linguistically common phonological stop voicing contrast is realized in the acoustic space in a language-specific way. Although Japanese voiced stops are becoming like English voiced stops in allowing a short lag variant, Japanese voiceless stops are different from English voiceless stops with respect to VOT values. The English voiceless stops in our study were consistently realized in the long lag VOT range whereas Japanese voiceless stops were realized in an intermediate lag VOT range. The intermediate lag VOT might be understood as a fourth type of VOT type – a "partially" aspirated stop – which was previously noted in Canadian French (Caramazza et al., 1973; Caramazza and Yeni-Komshian, 1974), in Hebrew (Raphael et al., 1995) and in Korean as described in older studies predating a recent sound change (e.g., Han and Weitzman, 1970; Lisker & Abramson, 1964). The Greek voicing contrast also taps language-specific categories. Greek voiced stops differ from Japanese and English voiced stops in having a partially nasalized variant rather than a short lag variant. VOT alone was not adequate to capture the difference between Greek and Japanese voiced stops.

These subtle acoustic differences have implications for acquisition. Children learning voiced stops or voiceless stops have to learn the specific acoustic parametric values appropriate for the ambient language. The findings of this study are consistent with earlier research showing that the language-specific phonetics of the stop voicing contrast predicts the order of mastery within the language. This study adds to that literature in showing that some seemingly exceptional patterns of earlier-than-usual or later-than-usual mastery of some speech categories do not deviate from our expectations when we make an appropriately fine-grained acoustic characterization of the category.

Acknowledgments

We thank the children who participated in the study, the parents who gave their consent, and the child-care centers at which the data were collected. Portions of this research were conducted as part of the first author's Ph.D. thesis from the Ohio State University, Department of Linguistics, completed in June 2009. This research was supported by NIDCD Grant 02932 to Jan Edwards.

Appendix A

Wordlists for Experiment I:

(1) Japanese word-initial voiced stops and voiceless stops

Words	Gloss	Words	Gloss
daikoN	daikon radish	tako	octopus
daNɡo	round rice cake	tatamaɡo	egg
daruma	Dharma doll	taNpopo	dandalion
dekoboko	jagged	tebɯkɯro	gloves
deNkʲi	light	teɡami	letter
depaːto	department store	tempɯra	tempura
doa	door	tiːkappɯ	tea cup
doNɡɯri	acorn	tiːʃatsɯ	t-shirt
doːnatsɯ	donut	tiʃʃɯ	tissue
ɡakkoː	school	tomato	tomato
ɡarasɯ	glass	toːɸɯ	tofu
ɡasɯ	gas	tora	tiger

ɡeːmɯ	game	kaba	hippo
ɡeNkʲi	vigor	kame	turtle
ɡeta	clogs	karasɯ	crow
ɡoma	sesami	keːkʲi	cake
ɡomi	trash	kemɯri	smoke
ɡorira	gorilla	keʃiɡomɯ	eraser
ɡumi	gummy candy	koara	koala
ɡɯrɯɡɯrɯ	spinning	kodomo	child
ɡɯː	rock/fist (in rock, paper, scissors game)	koppɯ	cup
		kɯma	bear
		kɯri	chestnut
		kɯrɯma	car

(2) English word-initial voiced stops and voiceless stops

Words	Gloss	Words	Gloss
dʌv	dove	kʰʌləɹ	color
dɑl.fn	dolphin	kʰʌtɪŋ	cutting
dɑŋki	donkey	kʰɑɹ	car
dɛ.zəɹt	desert	kʰɛtʃəp	ketchup
dɛk	deck	kʰek	cake
dendʒəɹ	danger	kʰev	cave
dɪɡɪŋ	digging	kʰi	key
diɹ	deer	kʰɪkɪŋ	kicking
ditʃ	ditch	kʰɪtʃɪn	kitchen
donət	donut	kʰoko	cocoa
doɹ	door	kʰon	cone
domz	domes	kʰot	coat
dud	dude	kʰuɡəɹ	cougar
duk	duke	kʰʊki	cookie
dul	duel	kʰʊkɪŋ	cooking
ɡʌmdɹɑps	gumdrops	tʰʌŋ	tongue
ɡɑɹdɪn	garden	tʰɑko	taco
ɡɑlf	golf	tʰɑl	tall
ɡɪɡəlz	giggles	tʰeəl	tail
ɡɪvɪŋ	giving	tʰɛnt	tent
ɡɪft	gift	tʰest	taste
		tʰɪkəl	tickle
		tʰipi	tepee
		tʰitʃəɹ	teacher
		tʰoɹn	torn
		tʰod	toad
		tʰost	toast
		tʰunə	tuna
		tʰub	tube
		tʰuθ	tooth

Appendix B

Wordlists for Experiment II

(1) Greek word-initial voiced stops and nasals for children

Words	Gloss	Words	Gloss
bala	ball	mixani	motorbike
dulapa	wardrobe	nero	water

(2) Greek word-initial voiced stops and nasals for adults

Words	Gloss	Words	Gloss
bala	ball	maɡʲes	macho guys
beno	I enter	meno	stay
bira	beer	minima	message
botes	boots	monimos	permament
buti	thigh	mumja	mummy
baloni	ballon	maɡʲes	macho-like action
berdevo	confuse	meta	afterwards
biskoto	biscuit	mixani	motorcycle
boreso	I will be able to	modelo	model
bukali	bottle	muɡos	dumb
dama	queen of hearts	nani	sleeping
defi	tambourine	nevro	nerve
dini	he/she is dressing	nikʲi	victory
dopjos	local	notos	south
duku	to pay in cash	numero	number
dalika	truck	narkono	drug
debuto	debut	nero	water
divani	divan	nistazo	drowsey
domata	tomato	nomizo	I think
dulapa	wardrobe	nuvela	short story

(3) Japanese word-initial voiced stops and nasals for children

Words	Gloss	Words	Gloss
basɯ	bus	mame	beans
budoː	grape	megane	glasses
deNɥa	telephone	mikaN	orange
		naiteirɯ	crying
		neko	cat
		niNɡʲoː	doll

(4) Japanese word-initial voiced stops and nasals for adults

Words	Gloss	Words	Gloss
darɯma	Dharma doll	namida	tear
deNkʲ	electricity	nomimono	drink
doːnatsɯ	donut	nemɯri	sleep
naa	dinner	nɯma	swamp
disɯkɯ	disk	nihoN	Japan
depaːto	department store	nattoː	fermented beans
daikoN	radish	nikoniko	smiling
doa	door	nedaN	price
disɯpɯreː	display	noriba	platform
doNɡɯri	acorn	nɯimono	needlework
dekoboko	ragged	nodo	throat
daNɡo	rice cake	neko	cat
		nɯkɯnɯkɯ	snugly
		naʃi	pear
		niNɡʲoː	doll

References

Allen, G. (1985). How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language*, *12*, 37–46.

Arvaniti, A., & Joseph, B. D. (2000). Variation in voiced stop prenasalization in Greek. *Glossologia*, *11–12*, 131–166.

Arvaniti, A., & Joseph, B. D. (2004). Early modern Greek /b d ɡ/: Evidence from Rebétika and Folk. *Journal of Modern Greek Studies*, *22*(1), 73–94.

Arvaniti, A. (2007). Greek phonetics: The state of the art. *Journal of Greek Linguistics*, *8*, 97–208.

Barr, D. J. (2007). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Beckman, M. E., & Edwards, J. (2010). Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology*, *1*(2), 319–343.

Botinis, A., Fourakis, M., & Prinou, I. (2000). Acoustic structure of the Greek stop consonant. *Glossologia*, *11-12*, 167–199.

Burton, M., Blumstein, S., & Stevens, K. (1992). A phonetic analysis of prenazalized stops in Moru. *Journal of Phonetics*, *20*, 127–142.

Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *Journal of the Acoustical Society of America*, *54*, 421–428.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, *2*, 239–245.

Chan, M.K-M., & Ren, H. (1987). Post-stopped nasals: An acoustic investigation. *University of California Working Papers in Phonetics*, *68*, 121–131.

Charalambopoulos, A., Arapopoulou, M., Kokolakis, A., & Kiratzis, A. (1992). Phonologial variation: Voicing-prenasalization. *Studies in Greek Linguistics*, *13*, 289–303.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.

Clumeck, H., Barton, D., Macken, M., & Huntington, D. (1981). The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics*, *9*, 210–224.

Cole, J., Hualde, J. I., & Iskarous, K. (1999). Effects of prosodic and segmental context on /g/-lenition in Spanish. in *Proceedings of the fourth international linguistics and phonetics conference*, pp. 575–589.

Davis, K. (1995). Phonetics and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language*, *22*, 275–305.

Demolin, D., Haude, K., & Storto, L. (2006). Aerodynamic and acoustic evidence for the articulation of complex nasal consonants. *Revue Parole*, *39/40*, 177–206.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying phonological acquisition. *Clinical Linguistics and Phonetics*, *22*(12), 939–958.

Esposito, C. (2004). Santa Ana del Valle Zapotec phonation. *UCLA working papers*, Vol. 103, pp. 71–105.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance–initial English stops. *Journal of Phonetics*, *10*, 177–192.

Fourakis, M. (1986). A timing model for word-initial CV syllables in Modern Greek. *Journal of the Acoustical Society of America*, *79*, 1982–1986.

Francis, A., Ciocca, V., Wong, V., & Chan, J. (2006). Is fundamental frequency a cue to aspiration in initial stops?. *Journal of the Acoustical Society of America*, *120*, 2884–2896.

Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., & Mukngoen, S. (1986). The acquisition of the voicing contrast in Thai: A study of Voice Onset Time in word-initial stop consonants. *Journal of Child Language*, *13*, 561–572.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.

Gordon, M., & Maddieson, I. (1999). The phonetics of Ndumbea. *Oceanic Linguistics*, *38*(1), 66–90.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading f0 cues in the voiced–voiceless distinction. *Journal of Phonetics*, *9*, 49–62.

Han, M. S., & Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p,t, k/ and /ph, th, kh/. *Phonetica*, *22*, 112–128.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of Acoustical Society of America*, *101*, 466–481.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, *106*, 1064–1077.

Hibiya, J. (1995). The velar nasal in Tokyo Japanese: A case of diffusion from above. *Language Variation and Change*, *7*, 139–152.

Holliday, J. J., & Kong, E.J. (2011). Dialectal variation in the acoustic correlates of Korean stops. in *Proceedings of the XVIIth international congress of phonetic sciences*, pp. 878–881.

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. *Journal of the Acoustical Society of America*, *84*, 511–529.

Hombert, J. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, *120*, 173–190.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*, 37–58.

Homma, Y. (1980). Voice Onset Time in Japanese. *Onseigakkai kaihao*, *163*, 7–9.

Iverson, G. K., & Salmons, J. C. (1996). Mixtec prenasalization as hypervoicing. *International Journal of American Linguistics*, *62*(2), 165–175.

Kollia, H. (1987). *An instrumental analysis of the voicing contrast in word-initial stops in young greek-speaking children*. M.A. Thesis.

Kollia, L. B. (1993). Segmental duration changes due to variation in stress, vowel, place of articulation and voicing of stop consonants in Greek. *Journal of the Acoustical Society of America*, *42*, 22–98.

Keating, P. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, *11*, 277–290.

Kewley-Port, D., & Preston, M. S. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, *2*, 195–210.

Klatt, D., & Klatt, L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male speakers. *Journal of the Acoustical Society of America*, *87*, 820–857.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Words*, *20*, 384–442.

Lewis, A. M. (2002). *Contrast maintenance and intervocalic stop lenition in Spanish and Portuguese: When is it alright to lenite?*. John Benjamins pp. 159–171.

Macken, M. A., & Barton, D. (1980a). The acquisition of the voicing contrast in English: A study of Voice Onset Time in word-initial stop consonants. *Journal of Child Language*, *7*, 41–74.

Macken, M. A., & Barton, D. (1980b). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, *7*, 433–458.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*, 475–494.

Nakanishi, Y., Owada, K., & Fujita, N. (1972). Kou'on kensa to sono kekka ni kansuru kousatsu [Results and interpretation of articulation tests for children]. *Tokushuu kyouiku kenkyuu shisetu houkoku [Bulletin of the Tokyo Gakugei University Special Education Research Group]*, *1*, 1–41.

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, *75*, 224–230.

Okalidou, A., Petinou, P., Theodorou, E., & Karasimou, E. (2010). Development of voiced onset time in standard-Greek and Cypriot-Greek-speaking preschoolers. *Clinical Linguistics & Phonetics*, *24*(7), 503–519.

Pan, H. (1994). *The voicing contrasts of Taiwanese (Amoy) initial stops: Data from adults and children*. Ph.D. Thesis, Ohio State University.

Raphael, L., Tobin, Y., Faber, A., Most, T., Kollier, H., & Milstein, D. (1995). Intermediate values of voice onset time. *Producing Speech: Contemporary Issues*, 117–127.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* ((2nd ed.). Thousand Oaks, California: Sage.

Riney, T., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, *35*, 439–443.

Romero, J. (1995). *Gestural organization in Spanish. An experimental study of spirantization and aspiration*. Ph.D. Thesis, University of Connecticut.

Shimizu, K. (1989). A cross-language study of voicing contrasts of stops. *Studia Phonologica*, *23*, 1–12.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Smit, A. B., Hand, L., Freilinger, J., Renthal, J., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*, 779–798.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. London: Sage.

Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America*, *118*, 1026–1037.

Takada, M. (2004a). Nihongo no gotou no yuuseion /d/ ni okeru +VOT-kato sedaisa. [+VOT tendency in the initial voiced alveolar plosive /d/ in Japanese and the speaker's.]. *Journal of the Phonetic Society of Japan*, *8*(3), 57–66.

Takada, M. (2004b). Nenrei no kotonaru washa ni yoru haretsuon no seiriteki ichi kousa [A physiological investigation of age-related differences in Japanese stops.]. In *Proceedings of the 18th pan-national meeting of the phonetic society of Japan*, pp. 149–154.

Takada, M. (2011). *Nihongo no gotou heisa'on no kenkyuu: VOT no kyoujiteki bunpu to tsuujiteki henka [Research on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT]*. Tokyo: Kurosio.

Westbury, J., & Parris, P. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America*, *47*, 625–633.

Westbury, J. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, *73*, 1322–1336.

Westbury, J., & Keating, P. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*, 145–166.

Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*, *35*, 341–352.

Yasuda, A. (1970). Articulatory skills in three-year-old children. *Studia Phonologica*, *5*, 52–71.

Yanagihara, N, & Hyde, C. (1966). An aerodynamic study of the articulatory mechanism in the production of bilabial stops consonants. *Studia Phonologica*, *4*, 70–80.

Zlatin, M., & Koeigsknecht, R. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Reseach*, *18*, 541–553.