NIH Public Access

Author Manuscript

Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development

Jan Edwards and
Department of Communicative Disorders, University of Wisconsin – Madison, Goodnight Hall, 1975 Willow Dr. Madison, WI 53706, PHONE: 608 262-6474, FAX: 608 262-6466

Mary E. Beckman
Department of Linguistics, Ohio State University, 1912 Neil Ave., 222 Oxley Hall, Columbus, OH 43220, PHONE: 614 292-9752, FAX: 614 292-8833

Jan Edwards: jedwards2@wisc.edu; Mary E. Beckman: mbeckman@ling.ohio-state.edu

Abstract

While broad-focus comparisons of consonant inventories across children acquiring different language can suggest that phonological development follows a universal sequence, finer-grained statistical comparisons can reveal systematic differences. This cross-linguistic study of word-initial lingual obstruents examined some effects of language-specific frequencies on consonant mastery. Repetitions of real words were elicited from 2- and 3-year-old children who were monolingual speakers of English, Cantonese, Greek, or Japanese. The repetitions were recorded and transcribed by an adult native speaker for each language. Results found support for both language-universal effects in phonological acquisition and for language-specific influences related to phoneme and phoneme sequence frequency. These results suggest that acquisition patterns that are common across languages arise in two ways. One influence is direct, via the universal constraints imposed by the physiology and physics of speech production and perception, and how these predict which contrasts will be easy and which will be difficult for the child to learn to control. The other influence is indirect, via the way universal principles of ease of perception and production tend to influence the lexicons of many languages through commonly attested sound changes.

In his influential monograph on child language, aphasia, and phonological universals, Jakobson (1941/1968) hypothesized that there are universal substantive principles — "implicational laws" —that structure the phoneme inventories of all spoken languages and that also determine how children acquire speech sounds. For example, one substantive principle that Jakobson noted, regarding phonation type contrasts, is that the presence of voiced and/or aspirated stops in an inventory necessarily implies the presence of voiceless unaspirated stops. In conjunction with this cross-language generalization, he predicted that young children, regardless of the language they are learning, will produce voiceless unaspirated stops before they produce either voiceless aspirated stops or voiced stops.

Jakobson's proposal of a half-century ago continues to influence our thinking about child phonology today, despite the fact that his two more specific claims about the ways that these principles operate in determining the course of phonological acquisition have been conclusively disproven. These claims were the following. First, there is a clear discontinuity between the large inventory of sounds that children produce in babbling and the more limited inventory of vowel and consonant phonemes that children produce in their first words. Second, all children acquire the same early vowels and consonants, regardless of the specific language to be learned, and each child expands this initial inventory in a rigid universal order.

Among subsequent studies countering the first claim are seminal papers by Vihman and colleagues showing that there is continuity — not discontinuity — between babbling and first words. For example, Vihman, Macken, Miller, Simpson, and Miller (1985) found that most children continue to babble for months after they begin to produce recognizable words. They also observed that the inventory of consonants and the relative frequencies of these consonants in a child's early words tend to be identical with the inventory and relative frequencies of consonants transcribed for the same child's babbling productions just before and during the months that the child is acquiring an initial lexicon of 25 to 50 words.

Among subsequent studies countering the second claim are the many studies and review articles showing that there is much variability in children's early phoneme inventories, both within and across languages. For example, Ingram (1999) compiled a table from reported results of studies of acquisition of English, Quiché, Turkish, and Dutch to show that typical consonant inventories differed across the four languages for children from 20 to 27 months. Vihman (1993) found that even within a single language, children begin to make recognizable productions of the different phonemes of the language in different orders.

The results showing continuity with babbling can be related directly to the results showing variation across children acquiring different languages if we consider the remarkable amount of language specificity that we see in infants' speech perception already by the end of the first year of life. Indeed, language-specific perceptual sensitivity can be seen already at birth, as evidenced by neonates' preference for listening to sentences of the mother's language as opposed to sentences of a rhythmically different language produced by a bilingual female speaker (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988). By 6 months, infants have started to tune their perceptual systems to the to-be-native vowel space, as evidenced by a lesser sensitivity to variation along a vowel formant continuum if the habituation stimulus is a prototypical vowel exemplar for the ambient language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). By about 10 to 12 months of age, they are already tuning their perceptual systems to the patterns of consonant allophony of the ambient language, as evidenced by their loss of ability to perceive non-native contrasts that they were able to perceive only a few months earlier (e.g., Werker & Tees, 1984).

Moreover, although language-specific tuning in production lags behind perception, influences of the ambient language are also evident in babbling during the first year. For example, de Boysson-Bardies, Hallé, Sagart, and Durand (1989) found that measurements of vowel formants in vocalizations produced by Arabic-, French-, English-, and Cantonese-acquiring 10-month-old infants differed in ways that reflected differences in vowel frequencies in the lexicons of the ambient languages. Similarly, de Boysson-Bardies and Vihman (1991) and de Boysson-Bardies, Vihman, Roug-Hellichius, Durand, Landberg, and Arao (1992) found that transcriptions of consonants in babbling in a longitudinal study of French-, English-, Swedish-, and Japanese-acquiring children reflected cross-language differences in the relative frequency of different consonants in the ambient adult languages. For example, the French-acquiring infants produced relatively more labial sounds than did the English-, Swedish-, and Japanese-acquiring infants, in keeping with the relatively greater frequency of labial consonants in French.

As children's phonological systems continue to develop in the second year, influences of the ambient language phonology emerge more clearly in the words that they produce. In the Vihman (1993) study, for example, variability across children acquiring the same language is drastically reduced between the babbling productions sampled just before the child's "first word" and the later babbling and word productions sampled at the child's 50-word stage. This reduction of variability across children acquiring the same language brings out more clearly the cross-language variability in the sample, leading to the conclusion that some sounds are

mastered earlier in some languages, as compared to analogous sounds in other languages. This conclusion is supported by cross-language comparisons of several "late" sounds. For example, /v/ seems to be mastered earlier in Swedish, Estonian, and Bulgarian, than it is in English (Ingram, 1988). Also, /l/ is mastered earlier by French-speaking children, as compared to English-speaking children, as shown by Chevrie-Muller and Lebreton's (1973) cross-sectional study as well as in the analyses of the longitudinal study reported in Vihman (1993).

Error patterns of young children also differ across languages. For example, Hua and Dodd (2000) found that in Putonghua (the standard variety of Mandarin Chinese spoken in the People's Republic of China), post-alveolar retroflex and alveolo-palatal affricates and fricatives (/ʂ, tʂ, tʂʰ, ɕ, tɕ, tɕʰ/) are among the earliest acquired sounds, while the similar post-alveolar sounds /ʃ, tʃ, dʒ/ of English are acquired relatively later by English-speaking children. The second most frequent error pattern in Putonghua is "backing", with 65 percent of the children substituting a more posterior constriction for a dental place of articulation (e.g., the apical post-alveolar [ʂ] for the initial dental /s/ in *suī* '*peepee*'). Similarly, in Japanese, dental or alveolar /s, ts, dz/ are mastered considerably later than post-alveolar /ʃ, tʃ, dʒ/, and the most common error pattern for /s/ is to substitute [ʃ] or [tʃ] (e.g., Nakanishi, Owada, & Fujita., 1972; Nishimura, 1980; Li & Edwards, 2006). These results for Putonghua and Japanese contrast with English where "fronting" of /ʃ/ to [s] (and /s/ to [θ]) is the typical error pattern (e.g., Weismer & Elbert, 1982; Baum & McNutt, 1990; Li & Edwards, 2006).

Given the continuity between babbling and early words and the evidence for language-specificity even within the first year of life, why do we continue to find Jakobson's general claim of phonological universals in development so appealing? Most likely, the idea of phonological universals is attractive because we know there are limits on what we can perceive and produce. Therefore, it would not be surprising if these constraints led — if not to universals — then at least to strong numerical trends in phonological acquisition. And, not surprisingly, these numerical tendencies have been observed, both within and across languages. For example, infant babbling contains many more stop consonants than fricatives, and most children master stop consonants before they master fricatives (e.g., Kent, 1992; Dinnsen, 1992; Smit, Hand, Freilinger, Bernthal, & Bird, 1990; Vihman et al., 1985). Kent (1992) suggests that stops are mastered earlier than fricatives for motoric reasons: stop consonants are relatively easy to produce because a complete closure of the vocal tract can be produced by a rapid ballistic gesture. By contrast, fricatives are relatively more difficult because the child must control the dimensions of the fricative constriction more exactly and coordinate this oral constriction gesture with a configuration of vocal tract postures behind the constriction to allow enough airflow through the fricative constriction to produced turbulence. This complex and precise coordination of gestures requires much more exact motor control than the simple ballistic movement for a stop closure.

Development of control over the voicing contrast for stop consonants provides another example of a phonological universal that has been well-studied across languages, using both transcription and measurement of voice onset time. These studies have shown that children produce voiceless unaspirated stops before they produce either aspirated stops or voiced stops (precisely as Jakobson predicted, in fact) in English (Macken & Barton, 1980a), French (Allen, 1985), Spanish (Macken & Barton, 1980b), Thai (Gandour, Holasuit Petty, Dardarananda, Dechongkit, & Munkongoen, 1986), Taiwanese (Pan, 1994), and Hindi (Davis, 1995). Kewley-Port and Preston (1974) suggest that this phonological universal of acquisition is phonetically grounded in the relative difficulty of satisfying aerodynamic requirements for the different stop types. The buildup of oral air pressure during stop closure inhibits voicing even when the vocal folds are adducted, so producing truly voiced stops (i.e., with audible voicing during the oral constriction) requires the child to perform other maneuvers, such as expanding the pharynx. The production of aspirated stops is not so complex, but it does require the child to keep the

glottis open long enough after the release of the oral closure to create an appropriately long interval of audible aspiration during the first part of the following vowel.

Although universal constraints on perception have been less well-studied than constraints on production, there is some evidence that perceptual constraints also influence the order of acquisition of consonants. "Strong" sibilant fricatives such as /s/ tend to be acquired earlier than "weak" non-sibilant fricatives such as /θ/. In English, /s/ is produced accurately by 75 percent of children by age 4;0 (years; months) while /θ/ is not produced accurately until about age 5;9 (Smit et al., 1990). This difference is most likely related to differences in perceptual saliency between the two classes of fricatives. Sibilant fricatives are easier to perceive than non-sibilant fricatives because place of articulation can be identified by the fricative noise alone for sibilant fricatives, while fricative noise and the CV transition are needed to identify place of articulation for non-sibilant fricatives (Harris, 1958; Jongman, Wang, & Sereno, 2000). This difference in perceptual salience leads to the Jakobsonian implicational universal of "strong fricatives before weak" which also describes the fact that [θ] is a very rare sound across languages. It occurs, for example, in only 18 of the 451 languages in the UPSID-PC database (Maddieson & Precoda, 1990), as compared to the 397 languages which are listed as having at least one voiceless sibilant fricative.

Our research program focuses on explaining cross-language differences in phonological development within a framework that recognizes that there must be substantive "phonetic universals" arising from constraints on production and perception. If these were the only constraints on speech development, there should not be cross-language differences between children acquiring languages with similar phoneme inventories. We suspect that at least some of the cross-linguistic differences we do observe are related to systematic variation in the frequencies of phonemes and phoneme sequences across the lexicons of different languages. Phoneme sequence frequency (also called phonotactic probability) has been shown to influence production accuracy of both real words and nonwords in English (Edwards, Beckman, & Munson, 2004; Munson, 2001; Zamuner, Gerken, & Hammond, 2004; Vodopivec, 2004). Therefore, it seems reasonable to hypothesize that at least some language-specific differences in phonological development might be related to differences in phoneme and phoneme sequence frequency across languages. For example, Ingram (1988) cites the low type frequency of /v/ in English as a reason for the later mastery of this sound by English-acquiring children relative children acquiring Swedish, Estonian, or Bulgarian. Similarly, the earlier mastery of /l/ in French as compared to English might be related to the much greater token frequency of /l/ in French than in English, in combination with the fact that this high token frequency is due primarily to its use in the pre-vocalic allomorph of the definite article, which occurs as a proclitic before nearly every vowel of the language. The different orders of mastery of /s/ relative to the post-alveolar fricatives in Putonghua and English also may be related to the differences in type frequency, since Putonghua /s/ occurs in only a third as many words either /ʂ/ and /ɕ/, whereas English /s/ occurs in six times as many words as /ʃ/.

The aim of this paper is to begin to explore how such language-specific facts about the distribution of phonemes and phoneme sequences in the lexicon might affect the child's acquisition of the phonology of the ambient language. That is, we are trying to devise methods for identifying effects of language-specific distributional asymmetries such as the especially low frequency of [v] relative to [f] in English, and for teasing these apart both from the direct effects of universal constraints and from any indirect effects of universal constraints that might lead to commonly attested distributional asymmetries. To examine these potential language-specific factors in phonological acquisition, we have designed a study to compare the accuracy of production of word-initial consonants across four languages: Cantonese, English, Greek, and Japanese. This set of languages affords various cross-language pairings of phonetically similar consonants which differ in initial consonant frequency and/or in initial CV sequence

frequency. When data collection is entirely complete, we will be able to look at productions of the target consonants and target CV sequences in real words and nonwords elicited from a hundred 2-, 3-, 4-, and 5-year-old children for each language. Here we report some of the results from a smaller feasibility test in which we elicited productions of the real words and of a few nonwords from about twenty 2- and 3-year old children for each language, prior to commencing the larger-scale study.

The analyses that we will report are three-fold. First, we correlated CV sequence frequencies for all of the consonants in each language with the five cardinal vowels (/a, e, e, o, u/) against the average frequency ranks for the most comparable CV sequences in the other three languages, and we also correlated these frequencies between every pair of languages for the consonant-vowel sequences that the paired languages have in common. These analyses were designed to examine whether phonotactic probabilities, in general, are rooted in universal constraints on perception and production. That is, if phonetic constraints such as the aerodynamic difficult of making voiced stops give rise to "markedness" universals, then phoneme frequencies should be correlated across languages. Similarly, if difficulty of perceptual parsing or difficulty of gestural coordination between the gestures for a consonant and a following vowel give rise to commonly observed phonotactic constraints on which vowels can follow which consonants, and if these universals are the primary factor determining differential mastery of consonants in different prevocalic contexts, then the correlations of CV frequencies across the pairs of languages should be significant.

Second, we also correlated the children's consonant accuracy against phoneme-sequence frequency within each language, for every CV target that we elicited in real words. This second analysis asked if children's accuracy can be partially explained by frequency of CV sequences. If there is such an explanation — that is, if the effects of universal constraints on which vowels can follow which consonants are modulated by specific-language experience — then there should be significant within-language correlations between frequency and accuracy. In this analysis, we included as a second independent variable, a measure of markedness developed in the correlation of frequencies across languages. Our reasoning was that, if universals are the primary factor determining the order of mastery of different consonants, and if these universals can be uncovered by examining correlations in the distributions of consonants across languages, then the measures of shared frequency variation across languages should be more predictive of the accuracy of consonants produced by young children than the language-specific measures of frequency variation alone. That is, a significant correlation among among CV sequence frequencies across languages in the first set of analyses might be taken as evidence for universal constraints on segment perception and production. A significant correlation between children's production and within-language frequency in the second set of analyses, with cross-language effects partialled out, might then be interpreted as evidence that frequency in children's experience contributes to production accuracy over and above any effect of universal constraints on the distribution of consonants in the language's lexicon.

Finally, we chose three pairs of consonant contrasts to study in more detail. The contrasts we chose were ones where the predictions from the phonetics were quite clear, but there were strikingly different frequency relationships for the analogous contrasts across at least two of the languages, so that we could juxtapose the predictions from independently well-motivated implicational universals against the predictions from language-specific opportunities for practicing the contrast. For each of these contrasts, we compared phoneme frequency in the adult lexicon and accuracy of young children's productions across two languages. The contrasts that we chose were phonemes /s/ versus /θ/ for English and Greek, phonemes /t/ versus /ts/ for Cantonese and Greek, and phonemes /t/ versus /tʃ/ for English and Japanese. For all three of the comparisons, we found evidence both for phonological universals such as "stops before affricates" and "strong fricatives before weak fricatives," and we also found evidence for

language-specific differences in acquisition that could be related to phoneme and phoneme-sequence frequencies within a particular language.

Methods

Participants

The participants were 41 2-year-olds and 44 3-year-olds. All participants were monolingual native speakers of the languages under examination and the data were collected in their native countries (Columbus, OH; Thessaloniki, Greece; Tokyo and Hamamatsu, Japan; and Hong Kong) by native language testers. Table 1 gives the number of children and mean age for the two age groups for each of the four languages. All children were typically developing, based on parent and teacher report, and had passed a hearing screening (either pure tone audiometry or otoacoustic emissions).

Stimuli

We chose to focus on word-initial lingual obstruents in four languages, Cantonese, English, Greek and Japanese. We chose word-initial position because this is where all obstruents can occur in all four languages whereas the distribution of obstruents in other word positions is much less comparable. For example, Cantonese has only /p, t, k/ at the ends of words, and Japanese has no word-final obstruents at all. We chose to focus on lingual obstruents because we wanted to avoid sounds that are very easy to produce (such as /b/ or /n/) and which are mastered before the age of 2;0 in most languages. We also wanted to avoid sounds that are very difficult to produce (such as the English and Mandarin approximant /ɹ/ and the Japanese or Spanish trilled /r/) which are mastered after the age of 6;0 in most languages.

We chose the four languages to be studied for three reasons. First, online lexicons are available for all four languages, with both segmental transcriptions and information about word frequency and/or word familiarity. Second, all four languages have a rich inventory of lingual obstruents, and although the exact inventory of sounds differs across languages (see Table 2), they all have a good number of sounds that can be compared to similar sounds in at least one other language in CV sequences that differ in frequency across the languages. For example, Greek dental /θ/ can be compared to the analogous fricative in English, and Greek alveolar /ts/ can be compared to the analogous affricates in both Cantonese and Japanese. Finally, we knew from our experience with these languages that some phonemes and phoneme sequence frequencies for some shared sounds differ across these languages. For example, /ti/ is a relatively high-frequency sequence in Cantonese and Greek, whereas in Japanese, it occurs in only a handful of recent loan words such as /tiʃʃu/ 'tissue'.

We elicited the target consonants in word-initial position before the vowels /i, e, a, o, u/, which appear in all four languages. For English and Cantonese, which both have more than the canonical five vowels of Greek and Japanese, we collapsed together vowels that have similar coarticulatory effects. For example, for English, we included both lax and tense vowels in each vowel category where the tense/lax contrast is relevant (for example, both /i/ and /ɪ/ were included in the /i/ category) and we included all three low back vowels /ɑ,ʌ, ɔ/ in the /a/ category. Similarly, for Cantonese, we included both long and short vowels for each of the /i, a, o/ categories. Note that while these decisions were motivated primarily by the need to set up comparable coarticulatory environments when comparing consonants across the languages, they also are in accord with what is known about the vowel systems of these four languages. For example, although the Greek mid front vowel is written with the same alphabetic symbol as the English tense /e/, the region that it occupies in the vowel space covers both English /e/ and English /ɛ/ (see, e.g., Hawks & Fourakis, 1995), as would be predicted by work on

implicational universals for vowel systems such as Liljencrants and Lindblom (1972) and Schwartz et al. (1997a, 1997b).

For each target CV sequence, we tried to choose about three pictureable words that we expected would be familiar to children. This resulted in some empty cells. For example, /θu/ is a permissible sequence of English, but it does not begin any words that young children can be expected to recognize. In order to find enough familiar and pictureable real words for each CV sequence, it turned out that we could not control lexical frequency within and across cells (for example, *coat* is more frequent than *cone* in English). However, Edwards and Beckman (2006) found that these differences in lexical frequency did not result in significant differences in CV accuracy for any of Cantonese, English, and Greek (the three languages where frequency information was available in the online lexicons that we used). CV accuracy was also not correlated with word familiarity in either English or Japanese (the two languages for which word familiarity ratings were available).

Each word was paired with a color photograph. The pictures were chosen to be culturally appropriate for the language and country in question. For instance, the English word *cake* and the Greek word *torta* are translation equivalents, but we chose different pictures because an American *cake* does not look like a Greek *torta*. All pictures were edited to fit on a fixed-size window on a laptop computer screen.

For each language, the stimulus items were spoken by an adult female native speaker and digitally recorded at a sampling rate of 22,500 Hz. For each word type, two tokens that were perceived with at least 80 percent accuracy by five adult native speakers were selected for use. Once the two tokens for each word type were chosen, two master lists were created for each language so that all participants heard either one token or the other of each word type.

Frequency

The frequency information was taken from adult online lexicons. For English, we used the *Hoosier Mental Lexicon* (HML, Pisoni, Nusbaum, Luce, & Slowiacek, 1985), which is a list of 19,321 wordform types created by collapsing homophones in the *Webster's Pocket Dictionary.* For Cantonese, we used the Cantonese language portion of the *Segmentation Corpus* (Chan & Tang, 1999; Wong, Brew, Beckman, & Chang, 2002), which is a list of 33,000 words extracted from transliterated newspaper texts. For Greek, we used the ILSP database (Gavrilidou, Labropoulou, Mantzari, & Roussou, 1999), which is a list of the 18,853 most frequent word types from a large corpus of newspaper texts. For Japanese, we used a subset of words from the NTT database (Amano & Kondo, 1999), which is based on the third edition of the Sanseido Shinmeikai Dictionary (Kenbou, Kindaichi, Shibata, Yamada, & Kindaichi). We used the subset of 78,801 words from this list that was also used by Yoneyama (2002) to calculate neighborhood densities for Japanese. While, unfortunately, these online lexicons were not uniform in size or type across languages, this is what is available at the present time. Another consideration is whether frequencies should be taken from adult lexicons or from lexicons based on child-directed speech (see, for example, Zamuner et al., 2004). A few studies suggest that there may be small but significant differences in some phoneme frequencies or some phoneme-sequence frequencies between some registers of adult-directed speech and some styles of child-directed speech in some languages (e.g., Tserdanelis, 2005), we know of no studies showing large differences affecting most or all lingual obstruents in any of the four languages that we are comparing. While we would prefer to have frequencies based on child-directed speech rather than the adult lexicon, we are currently only part-way through the long and laborious process of building lexicons that are phonetically transcribed and segmented into prosodic words based on comparable corpora of child-directed speech for the four languages (Beckman & Edwards, 2007).

To calculate frequency for each consonant in each of the five coarticulatory environments, we first counted the number of times the CV sequence occurred in word-initial position. Since we were specifically concerned with the coarticulatory effects of the following vowel (rather than with evaluating the sequence as a potential prosodic unit, as in Treiman, Kessler, Knewasser, & Bowman, 2000), we decided to ignore rhyme structure, syllable weight, and most other language-specific word-internal prosodic properties in these counts. That is, for example, when counting the frequency of /d/ in the context of following /a/ in Japanese, we included both words beginning with light syllables (such as *dashi* 'soup stock' and *dabudabu* 'baggy') and words beginning with various types of heavy syllables (such as *dakko* 'hug', *daikon* 'daikon radish', and *dango* 'dumpling cake').

Once we had the type count for each for each CV, we divided this number by the total number of words in the database. Using this ratio allows us to partially correct for the different sizes of the lexicons across languages. We then took the log of the ratio, which effectively weights a percentage change at the low-frequency end of the distribution more heavily than the same percentage change at the high-frequency end. This is a commonly-used transform for frequency data, corresponding to the intuition that changes at low frequencies are relatively more important than changes at high frequencies. For example, if new words coming into a lexicon result in a particular CV sequence being attested in 20 words when it was previously attested only in 10 words, this is a more substantial change than if the CV sequence comes to be attested in 210 words rather than in 200 words.

Procedure

The testing took place in a quiet room at one or more preschools in each of the four countries. Children were tested one at a time in one to three testing sessions (i.e., blocks of trials were either presented on the same day or on successive days, depending on the child's age and attention span). Each trial consisted of a picture and the associated sound file, which were presented simultaneously to the participant over a laptop with a 14-inch screen using a program written specifically for our purposes. The computer program included an on-screen VU meter to help the children monitor their volume and a picture of an animal (duck or frog or koala bear) walking up a ladder on the left side of the screen to provide visual feedback to the children about how close they were to completing the task. The children were instructed to repeat each word exactly as they heard it. Children were asked to repeat responses in the following cases: (1) if the response was different from the prompted word (e.g., said *duck* when prompted with *goose*) or (2) if the tester thought the target sequence would be impossible to transcribe because the response was spoken very softly, or overlapped with the prompt or with excessive background noise (e.g., a door slam or another child screaming during the response). The children's responses were recorded directly onto a CD or a digital audiotape, using a high-quality head-mounted microphone. The first audible response to each prompt was transcribed and included in the statistical analyses. We chose to use real words rather than nonwords because we thought it would be much easier for young children (especially 2-year-olds) to repeat a large number of familiar real words that were paired with a picture, rather than nonwords. (This pilot experiment also included between 15 and 20 nonwords for each language paired with pictures of unfamiliar objects in order to test whether the youngest children would repeat nonwords paired with novel objects. However, the results for the nonwords were not included in the analyses reported here.) We chose to use a repetition paradigm rather than a picture-naming paradigm because this was a pilot experiment for a much larger study that would include children ranging in age from 2;0 to 5;11. We wanted to make sure that all children had the same stimulus presentation conditions, in spite of the fact that 5-year-olds can name more pictures than 2-year-olds. In many picture-naming studies of phonological development, immediate or delayed repetition is used to elicit responses if the child does not spontaneously name the picture (e.g., Smit et al., 1990).

Transcription

A native speaker who is also a trained phonetician listened to the response and examined the acoustic waveform for each repetition. The target consonants were coded as *correct*, as *incorrect*, or a *voicing error only* (for those responses in which the phonation type was incorrect but the place and manner of the response was correct as, in *gate* perceived as [k^{h}et]). The *voicing error only* category accounted for 9% of the responses not judged to be *correct*. There are small differences in the coeffecients returned by the statistical analyses depending on how this category is treated, but these differences do not change the results in any substantive way, so we will report only results that collapse together the *incorrect* and *voicing error only* categories as equally not *correct*.

For each language, a second native speaker and trained phonetician, blindly re-transcribed 20 percent of the data (repetitions of two two-year-olds and two three-year-olds). Phoneme-by-phoneme inter-transcriber reliability for accuracy was at or above 89% for all four languages (90% for English, 96% for Cantonese, 94% for Greek, and 89% for Japanese).

Analysis

We did two types of regression analysis on the frequencies and the transcribed accuracy rates for the different consonants and different CV frequencies within and across the four languages.

First, as noted earlier, we correlated CV frequencies between every pair of languages for the consonant-vowel sequences that the paired languages have in common, and we correlated the children's consonant accuracy against phoneme-sequence frequency within each language, for every CV target that we elicited in real words. For each of these two analyses, we used a simple linear model with log frequencies of the target CV sequences in one of the four languages as the independent variable. In the first type of analysis, the dependent variable also was log frequencies of the target CV sequences, but for the other language in the pair. In the second type of analysis, the dependent variable was the percent correct consonant productions for each target CV in the language averaged first over the productions of the three words by each child (to adjust for missing tokens) before averaging over all the children recorded for the language.

Finally, in analyzing the results of each of the three specific consonant pair by language pair comparisons, we used a multilevel logistic regression model (Raudenbush & Bryk, 2002). We chose this type of analysis because of the nested structure of the data. That is, multiple responses were collected from each child, so a multilevel model was considered appropriate to account for the dependence among responses coming from the same child. Multilevel models also allow variables to be linked to different levels of a hierarchy (i.e., response-level variables versus subject-level variables), thus permitting hypotheses to be tested with respect to the appropriate units of analysis. The specific models used for these analyses are described in more detail in the Appendix. The factors included in each of the models were the following: language, consonant, and following vowel, including a language-by-consonant interaction factor, and the continuous variable age (in months). All multilevel analyses were performed using the statistical software HLM 6.0 (Raudenbush, Bryk, Cheong, & Congdon, 2004).

Results

Frequency relationships

For each of the six language pairs (e.g., Cantonese versus English, Cantonese versus Greek, etc.), we did a regression analysis to see whether the relative CV frequencies in one language could be predicted from the relative CV frequencies in the other language. Only the CV pairs which both languages had in common were included in the analyses. These results are shown in Figure 1. There was a significant correlation for two of the language pairs ($r^2 = 0.15$, *p* =

0.045, for English/Greek; $r^2 = 0.17$, $p = 0.01$ for Greek/Japanese). A third pair (Cantonese/English) showed a small (albeit insignificant) shared variance ($r^2 = 0.1$, $p = 0.1$). For the other three of the six pairs (Cantonese/Greek, Cantonese/Japanese, and English/Japanese), there was essentially no correlation ($r^2 \leq 0.01$).

These results are inconsistent with the hypothesis that there are universal constraints on perception and production that govern the frequency of segment sequences across languages. However, as noted above, these twelve correlation analyses only included CV sequences that are shared between the paired languages, so that any CV sequence that has a particularly low frequency in one language but is completely unattested in the paired language could not contribute to the correlation, reducing the value of these correlations as a measure of universal "markedness" tendencies. To overcome this limitation of the pairwise correlations, we also did a second set of regression analyses for each language, to see whether the log frequencies of all of the CV sequences in the language can be predicted from the mean frequency ranks in all three other languages, as a somewhat more global measure of the universal "markedness" constraints.

Figure 2 shows the results of this analysis. Although all of these regression models showed a significant relationship, the proportion of the variance that was accounted for is generally small. These small r^2 values reflect the existence of outlier sequences, such as /si/ in Greek and /ti/ in Japanese, where the frequency of the consonant in context is much higher or much lower than expected given the frequency of the consonant overall. While the two analyses of frequency relationships across languages do not support the claim that universal constraints on production and perception influence the frequency of phoneme sequences across languages, the results of the analyses of accuracy-frequency relationships discussed below suggests that there is some evidence for a role of universal constraints in acquisition.

Accuracy-frequency relationships

For each language, we then performed a regression analysis with the markedness measures and log frequencies plotted in Figure 2 as the independent variables and percent correct consonant production as the dependent variable. Figure 3 shows the analyses for overall consonant accuracy and Table 3 shows the associated multiple regression analyses. If both language-specific experience and universal constraints affects accuracy in the predicted ways, the coefficient for log frequency in the target language should be positive (the more words that begin with the CV sequence, the more accurate the children's productions of it) and the coefficient for the average frequency rank should be negative (the closer to being highest-ranked for frequency across different lexicons, the more accurate the consonant). The regression models behaved in this way. The resulting R^2 values covered a wide range, with marginally significant ($p<0.05$) low values of 0.13 and 0.26 in Japanese and Greek, respectively, contrasting with clearly significant high values of 0.41 and 0.47 in Cantonese and English, respectively. However, it was only for English that the language-specific log frequencies contributed significantly to the model, accounting for 37% of the variance in consonant accuracy even after the correlation of these frequencies with the more general pattern of frequency rankings was partialled out. On the face of it, these results provide some support for the claim that language-universal factors influence production accuracy, given the significant contribution of the markedness factor in Cantonese and Greek, and some support for the claim that language-specific phonotactic frequencies influence production accuracy, given the results for English. The equivocal outcome of these analyses is perhaps due to differences across the adult lexicons that we used to compute the frequencies of the CV sequences. One difference in particular that we suspect may be at play involves the relationship between sequence frequencies in the adult lexicon and sequence frequencies in the words that a child is likely to hear. In Japanese, for example, Hayashi, Yoshida, & Mazuka (1998) showed

that there are prosodic differences between adult words and child-directed words and our own research in progress (Beckman & Edwards, 2007) suggests that there are differences in distributions of dental versus post-alveolar sounds in the two registers.

Three contrast pairs

In the remainder of the Results section, we focus on the three comparisons that we had chosen to examine in detail. The first comparison involved the relationship between frequency and accuracy of /s/ versus /θ/ in English versus Greek. Given the greater perceptual salience of /s/ relative to /θ/, we might expect /s/ to be both more accurate and more frequent in both languages. Figure 3 shows this to be true. In both the panel for English and the panel for Greek, the datapoint point for /s/ is above and to the right of the datapoint for /θ/ (symbolized by "T" in the figure). This relationship is confirmed in Figure 4. The upper panel of the figure graphs the relative frequencies of the two fricatives for the two languages in each of the five vowel contexts, as well as the mean overall frequency across these five vowel contexts. As the graph shows, in English, dental /θ/ is much less frequent than alveolar /s/ across the board. In Greek, the difference in frequency between the two fricatives is smaller and in the /e/ vowel context, /θ/ is slightly more frequent than /s/. The lower panel of the figure shows the accuracy rates for the same consonants. In both languages, /s/ has a similarly high rate of accuracy, while /θ/ has a much lower accuracy rate in English than in Greek.

Table 4 shows the results of the multilevel model for this first comparison. The first part of the table reports the estimated coefficients for the child's age, for the three sets of dummy variables Language=English, Consonant=/θ/, and Vowel=/e/,/i/,/o/,/u/, and for the interaction variable Language=English&Consonant=/θ/. Since the model is a logistic model, these coefficient estimates do not reference input units such as months (for the age variable), but instead need to be interpreted with respect to their effects on the logit, which will not be linear with respect to the probability of a correct production. For example, based on the estimates in Table 4, the expected probability of a correct production of the target consonant /θ/ in the context of the vowel /e/ by a hypothetical 30-month-old English-speaking child would be $P = 0.05$, derived from the logit value:

$$\log[P/(1 - P)] = -2.902$$

which can be calculated from the coefficient estimates given in the table. (See the Appendix for details of the calculation.) This logit is negative, reflecting the large negative contributions of the coefficients for the interaction variable and for the dummy variables Consonant=/θ/ and Vowel=/e/. Although these coefficients cannot be interpreted directly in terms of such real world units as age in month, however, since the logit has a strictly increasing relationship with the probability, the sign and significance of each coefficient estimate indicates the consistent direction of effect. Consequently, from Table 4, we can see that the target consonant /θ/ and context vowel /e/ both have significant main effects, each reducing the probability of a correct production (in comparison to the reference categories /s/ and /a/). In addition, the negative sign associated with the significant interaction implies an even larger decrease in the probability of correct response when the target consonant /θ/ is presented to the English-speaking children as compared to the Greek children.

Thus, the table shows that there was a significant effect detected for consonant, with fewer correct productions observed for /θ/ than for /s/ across both languages. There was also a small but significant effect of vowel, reflecting the surprisingly low accuracy of the /s/ in the context of /e/ in English. Finally, and most importantly, there was a significant interaction between language and consonant. A word-initial /θ/ elicited from an English-acquiring child was significantly less likely to be transcribed as an accurate production than predicted from the

other characteristics of that word token. We suggest that the difference between English and Greek /θ/ production is due to within-language frequency effects overlaid on a universal perceptual salience effect that makes this consonant infrequent in both languages as well as unattested in the other two target languages.

For the second comparison, we looked at /t/ versus /ts/ in Cantonese and Greek. Given the greater articulatory demands for an affricate relative to a stop, we might expect the affricate to be both lower in frequency and less accurately produced in both languages. However, as Figure 3 showed, the affricate is substantially less frequent only in Greek. The upper panel of Figure 5 confirms the difference in frequency relationships between the two languages. In Greek, the affricate is substantially less frequent than the homorganic stop in all following vowel environments, but in Cantonese, /ts/ is surprisingly frequent, beginning at least as many words as /t/ in all four vowel contexts where dentals can occur in the language. The lower panel of Figure 5 shows the accuracy comparisons for /t/ versus /ts/ for the two languages. There was a considerably smaller disadvantage for the affricate in Cantonese as compared to Greek.

Table 5 shows the results of the multilevel model applied to these data. There was a significant effect of the child's age in months, with older children producing more accurate productions of target consonants than younger children. There was also a significant effect of consonant, with more incorrect productions observed for /ts/ than for /t/. Finally, there was again a significant interaction between language and consonant. A word-initial /ts/ elicited from a Greek-acquiring child was significantly less likely to be transcribed as an accurate production than predicted from the other characteristics of that word token. We suggest that the difference between Cantonese and Greek /ts/ production is due to a within-language frequency effect overlaid on a universal articulatory ease effect that makes affricates in general less frequent than stops.

The third comparison was the contrast between /t/ versus /tʃ/ in English and Japanese. The predictions about relative frequencies and relative accuracy are the same as for the second comparison, and this time the overall frequencies do show the expected relationship in both languages. However, as the top panel in Figure 6 shows, in English, /t/ begins more words than /tʃ/ regardless of vowel, while in Japanese, the pattern is dependent on vowel context. That is, in Japanese, */tu/ is not attested, and /tʃi/ is more frequent than /ti/, so that /t/ is more frequent than /tʃ/ only in the /e, a, o/ contexts. (The difference is especially large in the /e/ context, where /tʃ/ primarily occurs in recent loanwords such as /tʃekku/ 'check'.) The accuracy patterns for the same CV contexts are shown in the lower panel of Figure 6. In English, /t/ is produced much more accurately than /tʃ/ in all vowel contexts. In Japanese, by contrast, /t/ is not substantially more accurate than /tʃ/ overall, which is surprisingly accurate for an affricate. Moreover, looking more closely at the individual vowel contexts, we can see that the surprisingly high accuracy for the affricate in Japanese is completely due to the /i/ context, the one vowel context where /t/ is substantially less frequent than /tʃ/. (Note that we are comparing the always unaspirated and typically voiceless stop and affricate in English words such as *dish*, *doctor*, *jeans*, and *joker* to the sometimes slightly aspirated and always voiceless stop and affricate in Japanese words such as /tiʃʃu/ 'tissue', /tamago/ 'egg', /tʃizu/ 'map', and /tʃakku/ 'zipper', but the frequency and accuracy relationships were qualitatively the same when we compared the aspirated stop and affricate in English words such as *tissue*, *tomato*, *cheeze*, and *chalk*.)

Table 6 shows the results of the multilevel model. Again, there was a significant effect of age, with older children more accurate than younger ones. There was also a significant effect of consonant, with a lower mean accuracy rate for /tʃ/ than for /t/. Finally, there was again a significant interaction between language and consonant. A word-initial /tʃ/ elicited from a Japanese-acquiring child was significantly more likely to be transcribed as an accurate

production than predicted from the other characteristics of that word token. We suggest that the difference between Japanese and English /t/ production is due to a within-language frequency effect overlaid on a universal articulatory ease effect that makes voiceless unaspirated [t] the second most frequent consonant on average across the four languages included in this study.

Discussion

This study examined the relationships between phoneme frequency and consonant production accuracy in young children's productions of lingual obstruents across four languages. In an initial set of regression analyses, we compared CV sequence frequencies in the adult lexicon for shared consonant-vowel sequences between pairs of languages. If phonotactic probabilities within a language are determined primarily by universal principles of ease of perception and production, then these CV frequencies should be highly correlated between each pair of languages. However, only two of the six possible pairings showed a significant correlation for the paired languages' CV sequence frequencies. We hypothesized that these essentially negative results could be an artifact of having to exclude low-frequency sequences that are unattested in one language of each pairing.

We therefore developed a measure of "markedness" that averaged over the frequency ranks of the CV sequences in the four online lexicons that we used in this study. This allowed us to include in the evaluation of language-specific frequencies, those low-frequency sequences that are not attested in all or even in any of the other languages compared. In a second set of regression analyses, we regressed each language's CV sequence frequencies against the average frequency ranks for the other three languages. All four of these models showed a significant relationship, although the proportion of variance accounted for was relatively low. It is interesting to note that the relationship is strongest for English and Greek, the two languages for which we had the smallest lexicons, and hence the largest variation in log frequencies along the y-axis. The results of the first two analyses do not provide strong support for the claim that universal constraints on production and perception influence the frequencies of consonant-vowel sequences across languages. However, the results of the frequency-accuracy analyses suggest that there is a role for universal constraints on production and perception in phonological acquisition.

In the frequency-accuracy regression analyses, we used our markedness measure to try to numerically disentangle the effects of "markedness" universals from the effects of language-specific frequencies. In these analyses, we correlated children's consonant accuracy against CV frequency within each of the four target languages. All four of the models showed a significant relationship, although the effect was not as robust as we might have guessed. In all languages except English, the effect of language-specific frequencies disappeared when the expectations from the frequency relationships in other languages was partialled out, suggesting a substantial contribution from markedness universals. In English, CV frequency accounted for 37% of the variance in consonant accuracy even after the effects of the average CV frequency ranks in the other three languages were partialled out, suggesting large effects of language-specific frequencies.

This range of predictive values is perhaps due to differences across the adult lexicons that we used to compute the frequencies of the CV sequences. One difference in particular that we suspect may be at play involves the relationship between sequence frequencies in the adult lexicon and sequence frequencies in the words that a child is likely to hear. In Cantonese, for example, there is a sound change in progress that almost certainly makes /k^wo/ and /k^{wh}o/ less frequent in child-directed speech than in the *Segmentation Corpus* wordlist. Such a discrepancy would help explain why these two sequences fall well below the regression curve in Figure 3,

and also why the most frequent "error" for these targets was, in fact, a "substitution" of [k] or [kʰ]. In Japanese, too, Hayashi, Yoshida, & Mazuka (1998) showed that there are prosodic differences between adult words and child-directed words and our own research in progress (Beckman & Edwards, 2007) suggests that there are significant differences in the distributions of dental versus post-alveolar sounds in the two registers (see also, Fischer (1970), among others). The unexpectedly high frequency of the Greek sequence /si/ in Figure 3, similarly, is primarily due to the high incidence of the morpheme /sin/ συν 'with' which occurs in many compound verbs in the newspaper texts from which the ISLP wordlist was created. This morpheme is rare in words that young children would know, a discrepancy that may help explain why the datapoint for /si/ falls below the regression curve in the panel for Greek in Figure 3 (and in fact, is at about the same height as the datapoints for /sa/, /su/, and /so/ which cluster around the regression curve in the middle of the panel to its left).

The results of the three specific comparisons also suggest that language-specific distributions influence mastery. At the same time, they show that these language-specific factors in phonological acquisition must be evaluated against the role of implicational universals. For all three comparisons, implicational universals based on ease of production or perception predicted which consonant would be produced more accurately. For two of the three comparisons, implicational universals also predicted which consonant would be more frequent. Two of the contrasts compared a stop to an affricate. Stop consonants are easier to produce than affricates and, not surprisingly, we found that stop consonants were produced more accurately than fricatives by young children across languages. This was true even in Cantonese, where the affricate /ts/ is *more* frequent than the stop /t/. The only context in which an affricate was produced more accurately than a stop was before the vowel /i/ in Japanese, where /ti/ is only marginally acceptable and /tʃi/ has an especially high type frequency.

Thus effects of phoneme frequency and phoneme sequence frequency must be interpreted against the backdrop of expectations related to the inherent phonetic difficulty of the target consonants and sequences. For example, the first of the three more specific comparisons contrasted the accuracy of children's productions of /s/, which is a "strong" sibilant fricative, with the accuracy of their productions of /θ/, a "weak" non-sibilant fricative. As noted above, sibilant fricatives are more perceptually salient than non-sibilant fricatives (Harris, 1958; Jongman et al., 2000) and are mastered earlier in English (Smit et al., 1990). Across languages, sibilant fricatives are more frequent than non-sibilant fricatives (Maddieson & Precoda, 1990), as noted in the Jakobsonian implication universal of "strong fricatives before weak." Within languages, sibilant fricatives are more frequent than non-sibilant fricatives in the languages that have both. In both Greek and English, for example, there are many more words beginning with /s/ than with /θ/, although the difference in relative frequency between /s/ and /θ/ is considerably smaller in Greek than in English.

The second and third comparisons contrasted a stop with an affricate. Stop consonants should be produced more accurately than affricates by young children because they are motorically easier. The stop closure simply requires a rapid ballistic gesture, while fine temporal coordination and articulatory precision are required to produce a strong sibilant release after a plosive burst. Even a cursory glance at older studies shows that, as predicted, stop consonants are mastered before affricates in English, with all stops being produced accurately by 75 percent of children by 3;0, while the two affricates are not produced accurately until 4;6 (Smit et al., 1990). We chose the comparison of a homorganic stop to an affricate because one language-specific difference that has already been noted in the earlier literature is that affricates are produced earlier in languages where they are relatively frequent, such as Quiché (Pye, Ingram, & List, 1987) and Cantonese (So & Dodd, 1995), than they are in languages such as English. Our comparison of /t/ and /ts/ in Cantonese versus Greek supports a frequency-based interpretation of earlier results for Cantonese. Both groups of children were less accurate in

producing /ts/ than in producing /t/, as predicted by the principle of "stops before fricatives". However, the difference in accuracy was smaller for the children who are acquiring Cantonese, the language where the language-specific phoneme frequency patterns work against the effects of inherent motor difficulty. Our comparison of /t/ and /tʃ/ in Japanese and English showed similar evidence for the universal principle being modulated by language-specific distributions, but this time by the distribution of the two consonants across the different vowel contexts. That is, the Japanese children showed much less of a difference in accuracy between the "harder" and the "easier" sound in just the context where the affricate is considerably more frequent than the stop.

Taken together, the results of the different types of analyses suggest a picture of phonological acquisition whereby developmental patterns that are common across languages arise in two ways. One influence is direct, via the universal constraints imposed by the physiology and physics of speech production and perception, and how these predict which contrasts will be easy and which will be difficult for the child to learn to control. The other influence is indirect, via the way universal principles of ease of perception and production tend to influence the lexicons of many languages through commonly attested sound changes.

This is a more nuanced picture of the relationship between Jakobsonian universals and language-specific factors than we see in two other approaches to phonological acquisition which Vihman and Velleman (2000) describe as "phonology all the way down" accounts versus "phonetics all the way up" accounts. The first approach includes much work done in phonological frameworks such as Optimality Theory. In this approach, Jakobson's implicational universals are understood to be grammatical constraints (i.e., "markedness constraints") fully on par with the morpheme structure constraints of a language (i.e., "faithfulness constraints"). The second approach includes such work as Locke (1983), Lindblom (1992), and MacNeilage, Davis, Kinney, and Matyear (2000). In this approach, Jakobson's "implicational laws" are understood to be "phonetic universals" (Maddison, 1997) rather than grammatical constraints. However, as in the first approach, these accounts do not usually tease apart the direct effects of the phonetic universals from the indirect ones, so that there seems to be no way to account for why the effects of the phonetic universals are modulated by the contingencies of the language's history.

By either of these approaches, we would expect to see strongly predictive relationships among frequently observed patterns across languages, frequently observed patterns within a language, and frequently observed patterns in phonological development. For example, we would expect that consonants and consonant-vowel sequences which occur in many languages also should occur in many words of any given language, and these consonants should be the ones that are mastered early by all children in all vowel contexts. Conversely, we would expect that consonants which occur in few phoneme inventories and consonant-vowel sequences which are prohibited in many languages also should occur in few words in any language in which they do occur, and these should be mastered late by children. In the context of the four particular languages in this study, then, we would expect to see strong correlations between CV frequencies across languages, and these correlations should mirror the correlations between CV frequencies and production accuracy within a language. However, this was not the case. CV frequencies across languages were not highly correlated, although consonant accuracy was correlated with CV frequency in two of the languages. Furthermore, in the consonant-specific comparisons, the four languages differed in terms of which phonemes and phoneme sequences were relatively more frequent and more accurate.

Before concluding, we should enumerate a number of ways in which the particular data we examined may have influenced the findings reported here. The CV sequence frequencies that we used were derived from adult lexicons, rather than child-directed speech lexicons, which

would be more reflective of the input the children are actually receiving. We are in the process of developing child-directed speech lexicons in all four languages, but this is a very labor-intensive process (Beckman & Edwards, 2007). In addition, the sizes of the adult lexicons we used varied widely from a high of 70,000 words in the Japanese NTT database to a low of less than 20,000 words in the HML and ILSP (the English and Greek lexicons). Pierrehumbert (2001) and others have pointed out that the kinds of phonological generalizations speakers and listeners can make depend very much on the size of the lexicon over which they are generalizing. Smaller lexicons support only coarser-grained generalizations. An inspection of the data along the x-axes in Figure 2 suggests that the CV frequency relationships that are observed in smaller (more child-like) lexicons will also be "coarser-grained" in the sense of being more similar to the average patterns that are expected from the universal constraints on perception and production. We also examined only one aspect of phonological acquisition – namely, accuracy of word-initial obstruents produced by 2- and 3-year-olds.

With the caveats just expressed, these results suggest that our understanding of phonological acquisition needs to take into account both phonological universals that are grounded in perception and production constraints and also language-specific differences in phoneme and phoneme sequence frequency. That is, we need an account of acquisition in which the direct influences of the Jakobsonian implicational universals are modulated by the contingencies of the language's history. By such an account, universals shape phonological acquisition because phonological structure is grounded in the natural world; consonant sounds and consonant-vowel sequences that are easy to say and easy to recognize will be used to make word forms in many languages, and acquiring an ambient spoken language includes the process of phonetic mastery of ambient language word forms. However, acquiring phonological structure involves more than phonetic mastery of word forms. Children make generalizations about sublexical patterns that can be reused in perceiving and saying new word forms. These generalizations will be more robust for patterns that must be reused often to acquire many different word forms. Universals can influence children acquiring the "same" consonants and consonant-vowel sequences differently, if the children are acquiring different languages, because phonological acquisition is a process mediated by the lexicon, which is the language learner's source of information about phoneme and phoneme sequence frequency.

Acknowledgments

References

Allen GD. How the young French child avoids the pre-voicing problem for word-initial voiced stops. Journal of Child Language 1985;12:37–46. [PubMed: 3980606]

Amano, S.; Kondo, T. Lexical properties of Japanese. Sanseido; Tokyo: 1999.

Bates, E.; Bretherton, I.; Snyder, L. From first words to grammar: Individual differences and dissociable mechanisms. New York: Cambridge University Press; 1988.

Baum SR, McNutt JC. An acoustic analysis of frontal misarticulation of/s/in children. Journal of Phonetics 1990;18:51–63.

Beckman ME, Yoneyama K, Edwards J. Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. Journal of the Phonetic Society of Japan 2003;7:18–28.

Beckman ME, Edwards J. Phonological characteristics of child-directed speech in Japanese and Greek. 2007 Manuscript in preparation.

de Boysson-Bardies B, Hallé P, Sagart L, Durand C. A cross-linguistic investigation of vowel formants in babbling. Journal of Child Language 1989;16:1–17. [PubMed: 2925806]

de, Boysson-Bardies; Vihman, MM. Adaptation to language: Evidence from babbling and first words in four language. Language 1991;67:297–319.

de Boysson-Bardies, B.; Vihman, MM.; Roug-Hellichius, L.; Durand, C.; Landberg, I.; Arao, F. Material evidence for infant selection from the target language: A crosslinguistic phonetic study. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological development: Models, research implications. Parkton, MD: York Press; 1992. p. 369-392.

Chan SD, Tang ZX. Quantitative analysis of lexical distribution in different Chinese communities in the 1990's. Yuyan Wenzi Yingyong[Applied Linguistics] 1999;3:10–18.

Chevrie-Muller C, Lebreton MT. Étude de la réalization des consonnes au cours d'une épreuve de mots, sur des groupes d'enfants de 3 ans et 5 ans ½. Revue de layngologie 1973;94:109–152.

Davis, Katharine. Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. Journal of Child Language 1995;22:275–305. [PubMed: 8550724]

Dinnsen, DA. Variation in developing and fully-developed phonetic inventories. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological development: Models, research, implications. Timonium, MD: York Press; 1992. p. 191-210.

Edwards J, Beckman ME. Subject-, stimulus-, and task-related factors influencing the accuracy of children's imitative productions of word-initial lingual obstruents. Manuscript submitted for review. 2006

Edwards J, Beckman ME, Munson B. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. Journal of Speech, Language, and Hearing Research 2004;47:421–436.

Gandour J, Holasuit Petty S, Dardaranandana R, Dechongkit S, Mukongoen S. The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. Journal of Child Language 1986;13:561–572. [PubMed: 3793815]

Gavrilidou, M.; Labropoulou, P.; Mantzari, E.; Roussou, S. Prodiagrafes gia ena ipologistiko morphologiko lexiko tis Neas Ellininkis [Specifications for a computational morphological lexicon of Modern Greek]. In: Mozer, A., editor. Greek Linguistics 97, Proceedings of the 3rd International Conference on the Greek Language; Athens: Ellinika Grammata; 1999. p. 929-936.

Harris KS. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech 1958;1:1–7.

Hawks JW, Fourakis MS. The perceptual vowel spaces of American English and Modern Greek. Language and Speech 1995;38:237–252. [PubMed: 8816085]

Hayashi, A.; Yoshida, K.; Mazuka, R. Baby-word rhythm preferences of Japanese infants. Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America; 1998. p. 2067-2068.

Hua Z, Dodd B. The phonological acquisition of Putonghua (Modern Standard Chinese). Journal of Child Language 2000;27:3–42. [PubMed: 10740966]

Ingram D. The acquisition of word-initial [v]. Language and Speech 1988;31:77–85. [PubMed: 3256768]

Ingram, D. Phonological acquisition. In: Barrett, M., editor. The development of language. Hove, UK: Psychology Press; 1999. p. 73-97.

Jakobson, R. Kindersprache, aphasie und allgemeine lautgestze. [Published originally in 1941; citations are from the 1968. In: Keiler, AR., editor. Child language, aphasia, and phonological universals. The Hague: Mouton; 19411968.

Jongman, A.; Wang, Y.; Sereno, J. Acoustic and perceptual properties of English fricatives. Proceedings of the 6th International Conference on Spoken Language Processing; 2000. p. 536-539.

Kenbou, H.; Kindaichi, H.; Shibata, T.; Yamada, T.; Kindaichi, K. Sanseido Shinmeikai Dictionary. 3. Tokyo, Japan: Sanseido; 1981.

Kent, RD. The biology of phonological development. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological development: Models, research, implications. Timonium, MD: York Press; 1992. p. 65-90.

Kewley-Port D, Preston MS. Early apical stop production: A voice onset time analysis. Journal of Phonetics 1974;2:195–210.

Kong, E.; Beckman, ME. Fine-grained phonetics and developmental universals for glottal features. Poster presented at the 10th Conference on Laboratory Phonology; Paris, France. June 27–30; 2006.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 1992;255:606–608. [PubMed: 1736364]

Li, F.; Edwards, J. Contrast and covert contrast in the acquisition of /s/ and /ʃ/ in English and Japanese. Poster presented at the 10th Conference on Laboratory Phonology; Paris, France. June 27–30; 2006.

Liljencrants L, Lindblom B. Numerical simulations of vowel quality systems: The role of perceptual contrast. Language 1972;48:839–862.

Macken MA, Barton D. The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. Journal of Child Language 1980a;7:41–74. [PubMed: 7372738]

Macken MA, Barton D. The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. Journal of Child Language 1980b;7:433–458. [PubMed: 6969264]

Maddieson, I.; Precoda, K. UPSID-PC. The UCLA Phonological Segment Inventory Database. 1990. [MS-DOS package available online from the UCLA Phonetics Laboratory at http://www.linguistics.ucla.edu/faciliti/sales/software.htm]

Mehler J, Jusczyk PW, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. A precursor to language acquisition in young infants. Cognition 1988;29:144–178.

Munson B. Phonological pattern frequency and speech production in adults and children. Journal of Speech, Language, and Hearing Research 2001;44:778–792.

Nakanishi Y, Owada K, Fujita N. Kōon kensa to sono kekka ni kansuru kōsatu. Tokushū kyōiku kenkyū shisetu hōkoku (Bulletin of the Tokyo Gakugei University Special Education Research Group) 1972;1:1–41.

Nishimura, B. Emergence of functional articulation disorders. Proceedings of the First International Congress for the Study of Child Language; Lanham, Md: University Press of America; 1980.

Pan, Ho-Hsien. Ph.D. thesis. Ohio State University; 1994. The voicing contrasts of Taiwanese (Amoy) initial stops: Data from adults and children.

Pierrehumbert J. Why phonological constraints are so coarse-grained. Language and Cognitive Processes 2001;16:691–698.

Pisoni D, Nusbaum H, Luce P, Slowiacek L. Speech perception, word recognition, and the structure of the lexicon. Speech Communication 1985;4:75–95.

Pye, C.; Ingram, D.; List, H. A comparison of initial consonant acquisition in English and Quiché. In: Nelson, KE.; Van Kleeck, A., editors. Children's Language. Vol. 6. Hillsdale, NJ: Lawrence Erlbaum; 1987. p. 175-190.

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2. Thousand Oaks, CA: Sage; 2002.

Raudenbush, SW.; Bryk, AS.; Cheong, YF.; Congdon, R. HLM 6: Hierarchical linear and nonlinear modeling. Lincolnwood, IL: Scientific Software International; 2004.

Schwartz L, Boë LJ, Vallée N, Abry C. Major trends in vowel system inventories. Journal of Phonetics 1997a;25:233–253.

Schwartz L, Boë LJ, Vallée N, Abry C. The Dispersion-Focalization Theory of vowel systems. Journal of Phonetics 1997b;25:255–286.

Smit AB, Hand L, Freilinger JJ, Bernthal JE, Bird A. The Iowa articulation norms project and its Nebraska replication. Journal of Speech and Hearing Disorders 1990;55:779–798. [PubMed: 2232757]

So LKH, Dodd B. The acquisition of phonology by Cantonese-speaking children. Journal of Child Language 1995;22:473–495. [PubMed: 8789511]

Treiman, R.; Kessler, B.; Knewasser, S.; Tinco, R.; Bowman, M. English speakers' sensitivity to phonotactic patterns. In: Broe, MB.; Pierrehumbert, JB., editors. Papers in Laboratory Phonology V: Acquisition and the lexicon. Cambridge, UK: Cambridge University Press; 2000. p. 269-282.

Tserdanelis, G. The role of segmental sandhi in the parsing of speech. Ohio State University; Columbus: 2005. Unpublished doctoral dissertation

Vihman MM. Variable paths to early word production. Journal of Phonetics 1993;21:61–82.

Vihman MM, Macken MA, Miller R, Simmons H, Miller J. From babbling to speech: A reassessment of the continuity issue. Language 1985;61:397–445.

Vodopivec, S. The influence of phonotactic probability on consonant acquisition. Ohio State University; 2004. Unpublished senior honor's thesis

Weismer G, Elbert M. Temporal characteristics of "functionally" misarticulated /s/ in 4- to 6- year old children. Journal of Speech and Hearing Research 1982;25:275–286. [PubMed: 7120967]

Werker J, Tees R. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav Dev 1984;7:49–63.

Wong, W-YP.; Brew, C.; Beckman, ME.; Chan, SD. Using the segmentation corpus to define an inventory of concatenative units for Cantonese speech synthesis. Proceedings of the First SIGHAN Workshop on Chinese Language Processing; Taipei, Taiwan: Academia Sinica; 2002. p. 119-123.

Yoneyama, K. Phonological neighborhoods and phonetic similarity in Japanese word recognition. Ohio State University; Columbus: 2002. Unpublished doctoral dissertation

Yoneyama K, Beckman ME, Edwards J. Phoneme frequencies and acquisition of lingual stops in Japanese. 2002 Manuscript in preparation.

Zamuner TS, Gerken LA, Hammond M. Phonotactic probabilities in young children's production of coda consonants. Journal of Child Language 2004;31:515–536. [PubMed: 15612388]

Appendix

The results in Table 4 to 6 are based on application of the same two-level generalized linear model in which stimulus responses (t) were nested within subjects (i). At Level 1 is a logistic regression model that can be written as:

$$\log[P_{it}/(1 - P_{it})]$$
$$= \beta_{0i} + \beta_{1i}CONS_t$$
$$+ \beta_{2i}VOWEL\_e_t$$
$$+ \beta_{3i}VOWEL\_i_t$$
$$+ \beta_{4i}VOWEL\_o_t$$
$$+ \beta_{5i}VOWEL\_u_t$$

where

P_{it} = Probability of a positive response to stimulus t from person i

$CONS_t$ = Dummy coded (0/1) variable identifying consonant type for stimulus t

$VOWEL\_e_t$ = Dummy coded (0/1) variable indicating a vowel of 'e' for stimulus t

$VOWEL\_i_t$ = Dummy coded (0/1) variable indicating a vowel of 'i' for stimulus t

$VOWEL\_o_t$ = Dummy coded (0/1) variable indicating a vowel of 'o' for stimulus t

$VOWEL\_u_t$ = Dummy coded (0/1) variable indicating a vowel of 'u' for stimulus t

At Level 2, the model expresses the level 1 coefficients as a function of subject-level variables and residuals:

$$\beta_{0i}=\gamma_{00}+\gamma_{01}AGE_i+\gamma_{02}LANGUAGE_i+U_{0i}$$
$$\beta_{1i}=\gamma_{10}+\gamma_{11}LANGUAGE_i+U_{1i}$$
$$\beta_{2i}=\gamma_{20}$$
$$\beta_{3i}=\gamma_{30}$$
$$\beta_{4i}=\gamma_{40}$$
$$\beta_{5i}=\gamma_{50}$$

where

AGE_i= age of person i

$LANGUAGE_i$= Dummy coded (0/1) variable indicating language of subject i and U_{0i}, U_{1i} are subject-level residuals assumed normally distributed with means of 0 and estimated variances and covariance.

For each analysis, the levels of the CONS and LANGUAGE variables that were assigned a level of '1' are identified in parentheses in Tables 4 to 6.

Because the model applied is a logistic model, the coefficient estimates reported in any of the tables should be interpreted with respect to their effects on the logit, which will not be linear with respect to the probability of a positive response. The coefficient estimates are also in reference to a unit-specific model representation, so their effects are best interpreted in reference to hypothetical individuals (see Raudenbush et al. 2004 for details). For example, based on the estimates in Table 4, the expected probability of a correct production of the target consonant /θ/ in the context of the vowel /e/ by a hypothetical 30-month-old English-speaking child would be $P = 0.05$, derived from a logit value of:

$$\log[P_{it}/(1 - P_{it})]=\gamma_{00}+\gamma_{01}{}^*30+\gamma_{02}{}^*1+(\gamma_{10}+\gamma_{11}{}^*1)^*1+\gamma_{20}{}^*1=$$
$$-.946+0.60^*30+.165^*1+(-1.454+(-1.806)^*1)+(-.661)^*1= - 2.902$$

where −0.946 is the intercept, 0.060 is the coefficient for age, 0.165 is the coefficient for the dummy variable Language=English, −1.454 is the coefficient for the dummy variable Consonant=/θ/, and −0.661 is the coefficient for the dummy variable Vowel=/e/. If the context vowel were instead /a/ (treated here as the reference category for Vowel), the coefficient for vowel would drop out, increasing the logit to 2.241, returning a probability estimate of approximately 0.10.

Other effects in Table 4 can be interpreted similarly. It should be noted that the effect of any predictor on the logit is the same for all hypothetical individuals, and since the logit has a strictly increasing relationship with the probability, the sign and significance of each coefficient estimate indicates the consistent direction of effect. Consequently, from Table 4 we can see that the consonant /θ/ and vowel /e/ both have significant main effects, each reducing the probability of a correct response (in comparison to the reference categories /s/ and /a/). In addition, the negative sign associated with the significant interaction implies an even larger decrease in the probability of positive response when target consonant /T/ is presented to English-speaking children as compared to Greek-speaking children. It is only in translating these effects to probabilities that it becomes necessary to interpret them with respect to hypothetical individuals.

The variance estimates reported in each table indicate the residual variability in subject-specific intercept term and consonant effects, each of which are being predicted by age (as well as language, in the case of the consonant effect). They can also be incorporated in the above

analysis in discussing hypothetical individuals, although play no role in understanding the predictor effects.
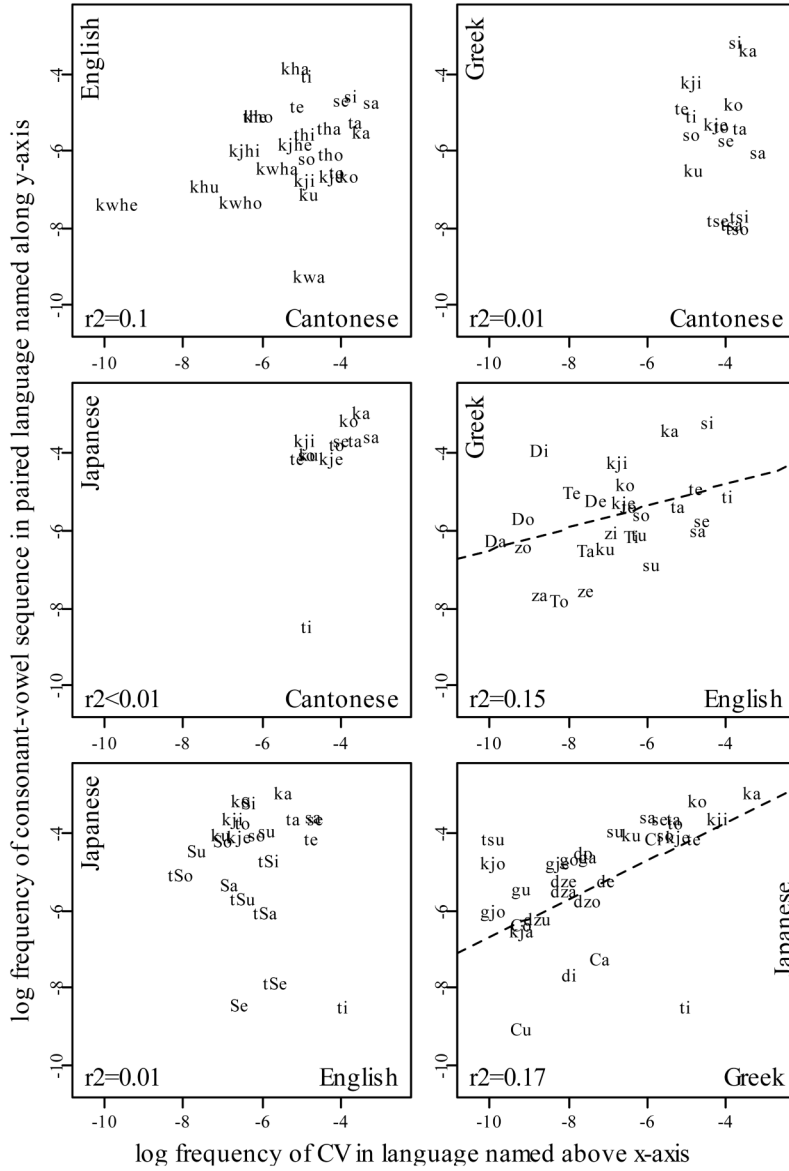
Figure 1.
Relationship between log relative frequencies of CV sequences shared by each pair of languages, with regression curves shown for the two language pairs where the relationship is significant at the *p*<0.05 level.
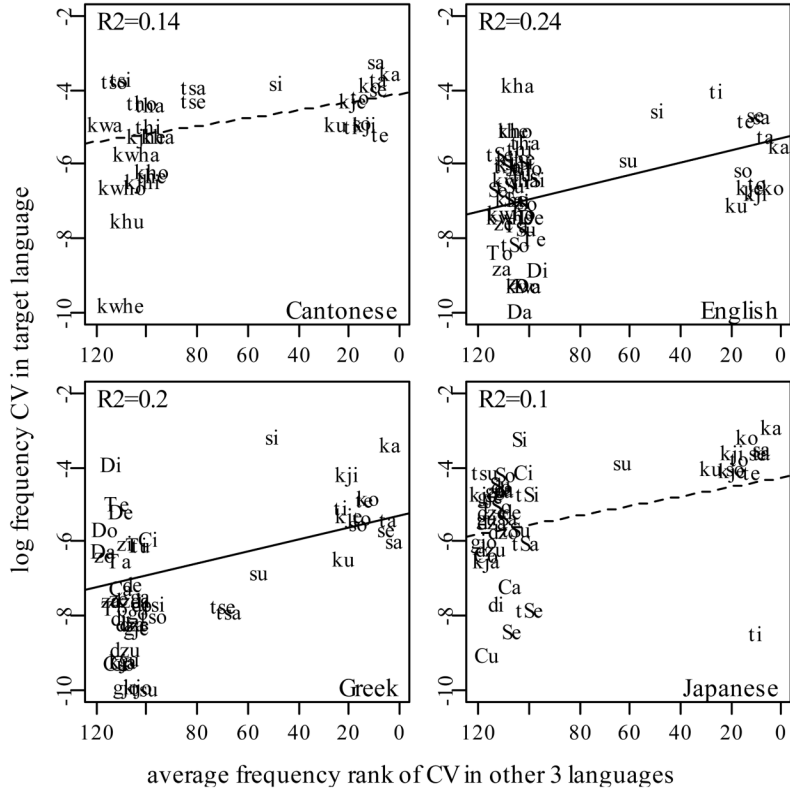
Figure 2.
Relationship between log relative frequency of consonant-vowel sequences in each language and mean rank frequency of the analogous sequences in the adult lexicons of the other three languages, with regression curves showing significance at the *p*<0.05 level (dashed line) or at the *p*<0.01 level (solid line).
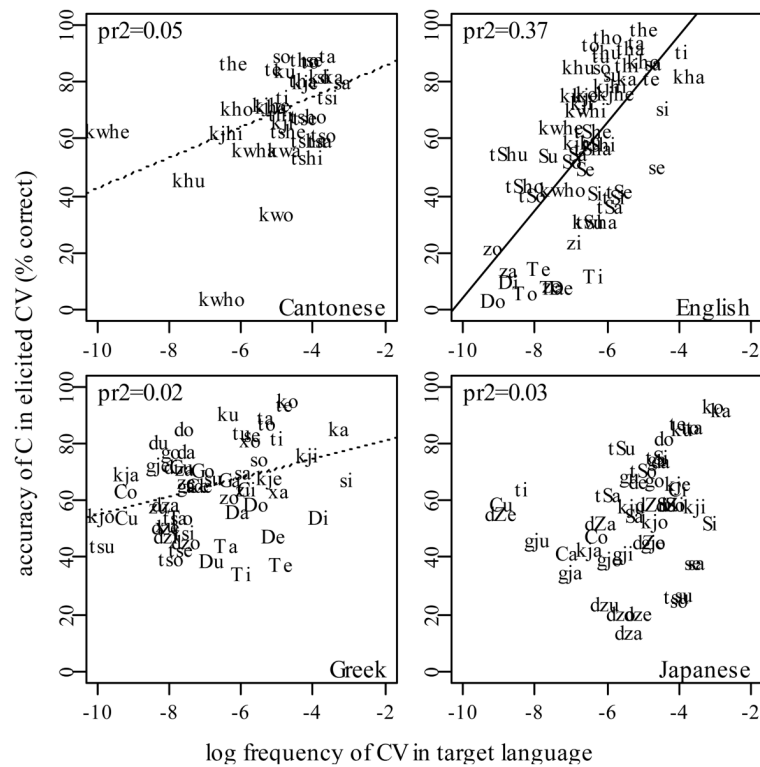
Figure 3.
Relationship between accuracy of children's productions of each consonant in each of the vocalic environments in which it was elicited as a function of the log relative frequency of the CV sequence in the language. The regression curves show significance at the *p*<0.01 level, which holds only when the mean frequency rank in the other languages is not partialled out (dotted line) or even when the mean frequency rank in the other languages is partialled out (solid line).

Figure 4.
Log relative frequency (top plot) and percent correct (bottom) for English and Greek /s/ and /θ/ in different vowel contexts.
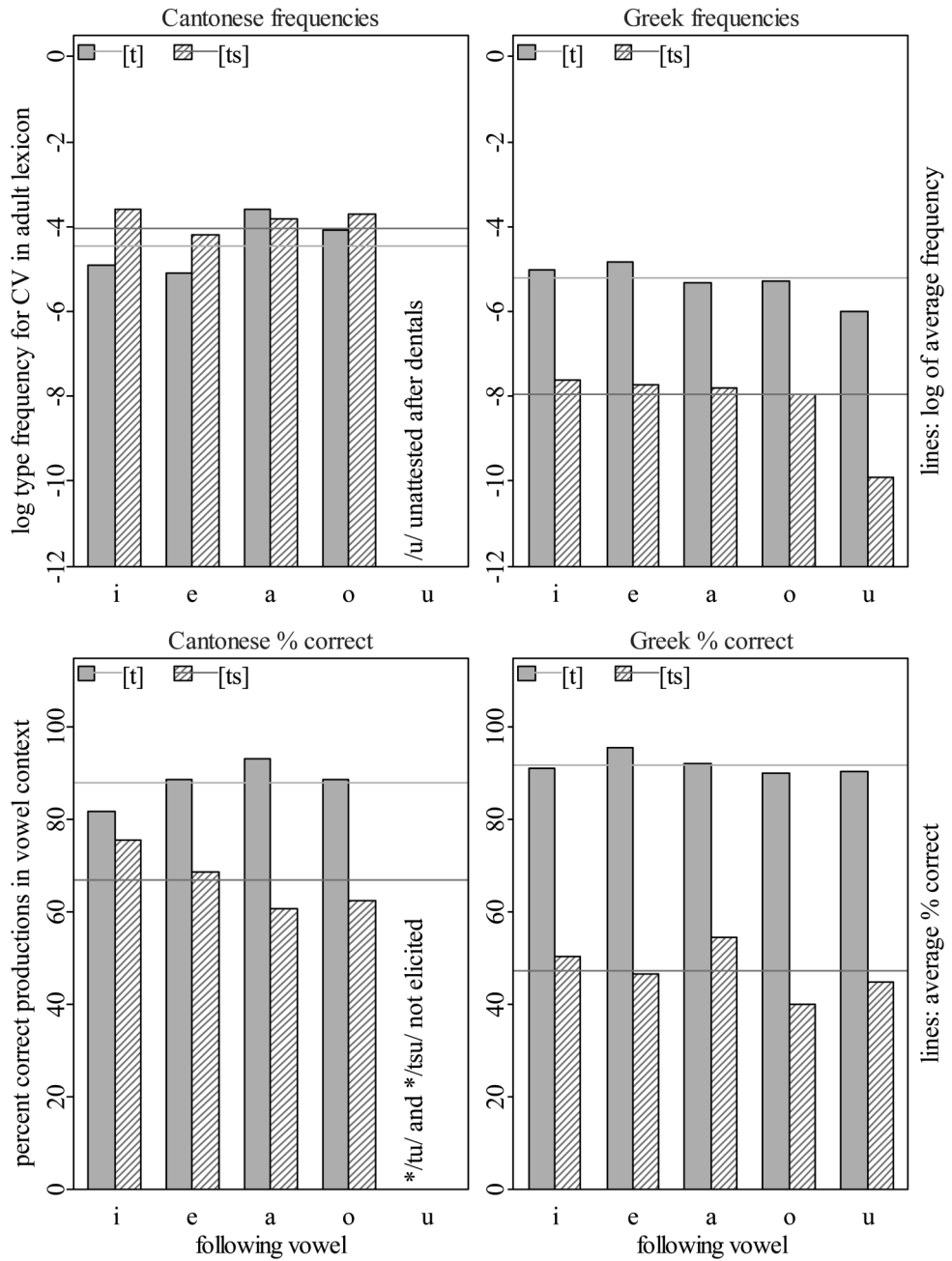
Figure 5.
Log relative frequency (top plot) and percent correct (bottom) for Cantonese and Greek /t/ and /ts/ in different vowel contexts.

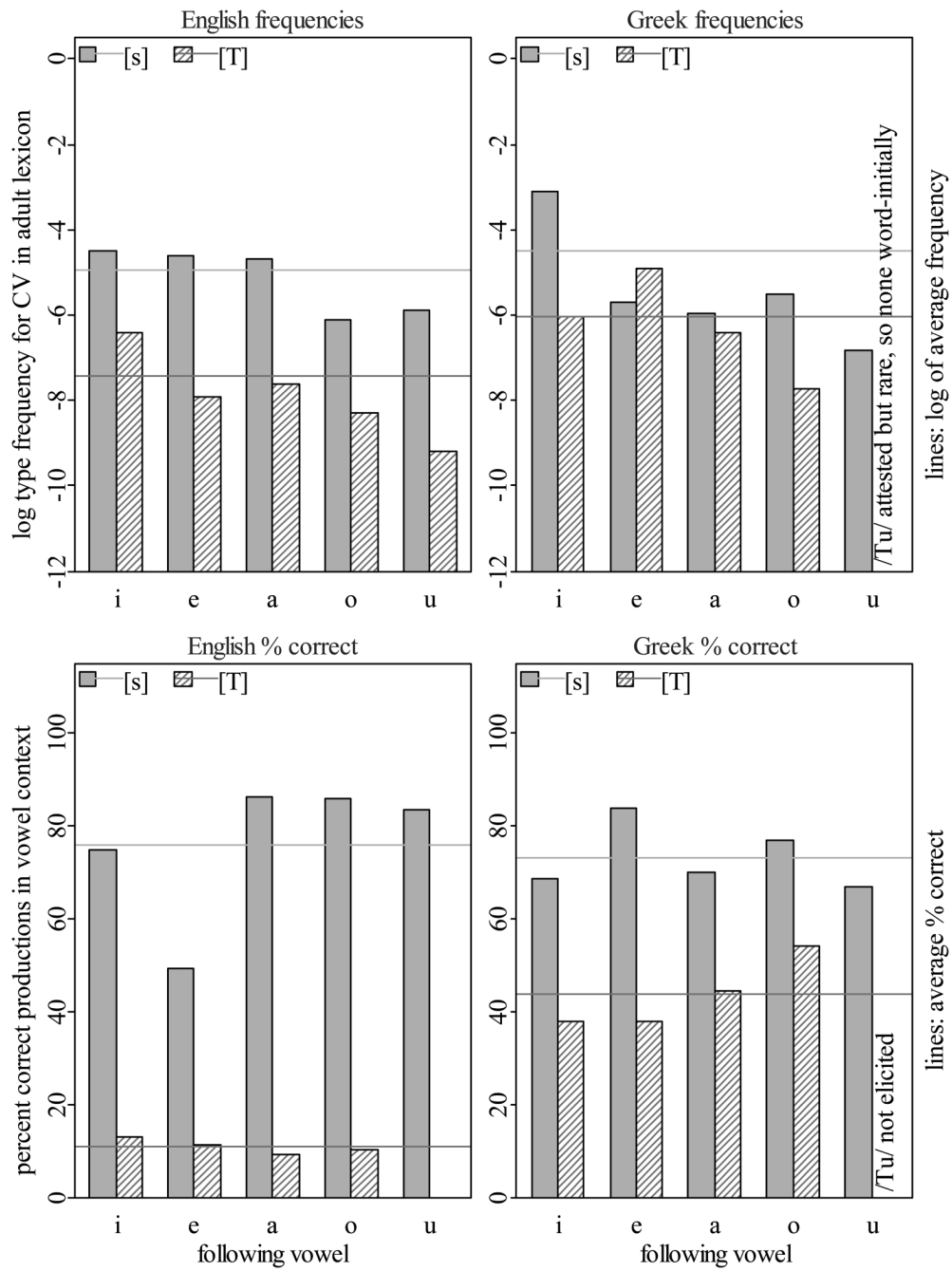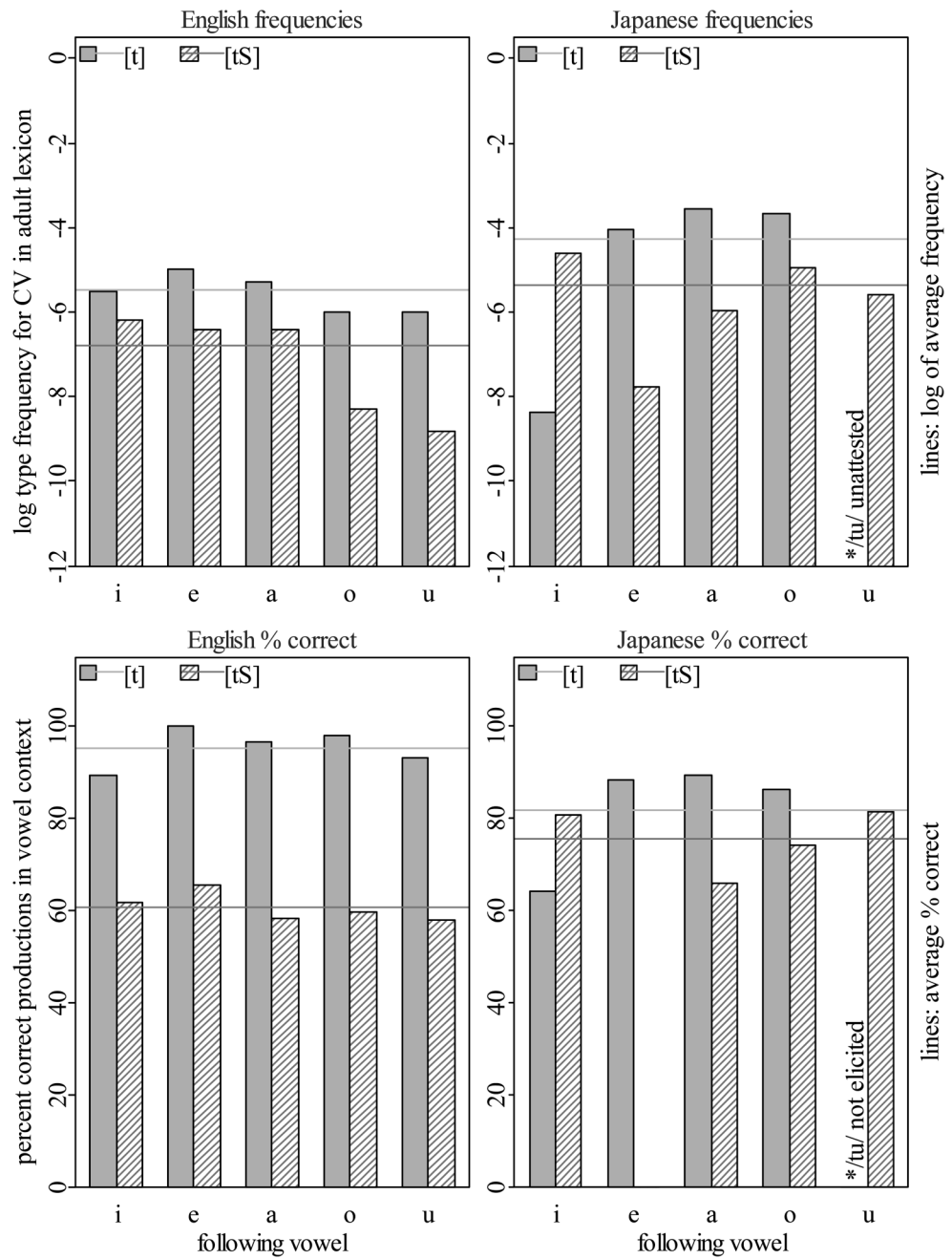**Figure 6.**
Log relative frequency (top plot) and percent correct (bottom) for English and Japanese /s/
and /tʃ/ in different vowel contexts.

Table 1

Mean age in months (standard deviation in parentheses) and number of subjects for participant groups for each of the four languages.

Age groups:	Language	Cantonese	English	Greek	Japanese
2-year-olds	age	31 (3.6)	31 (3.4)	33 (1.3)	32 (1.8)
	N	10	10	9	12
3-year-olds	ages	43 (2.4)	39 (2.6)	42 (3.5)	44 (2.4)
	N	12	13	11	8

Table 2

Distribution of lingual obstruents in the four languages.

consonant type:	Cantonese[1]	English[2]	Greek[3]	Japanese[4]
stops	/t, tʰ, k, kʰ/	/d, tʰ, g, kʰ/	/d, t, gʲ, kʲ, g, k/	/d, t, gʲ, kʲ, g, k/
fricatives	/s/	ð, θ, z, ʒ, s, ʃ/	/ð, θ, z, s, ʐ, ç, ɣ, x/	/s, ʃ, ç/
affricates	/ts, tsʰ/	/dʒ, tʃ/	/dz, ts/	/dz, ts, dʒ, tʃ/

[1]Notes: <u>Cantonese:</u> The phonation type contrast is between voiceless unaspirated plosives and aspirated plosives, and the coronal stops and fricative are more dental than alveolar. Dentals are not attested before /u/.

[2]<u>English:</u> The phonation type contrast in word-initial position is between "voiced" plosives with a short lag VOT (or, sometimes, voicing lead) and voiceless aspirated plosives, and the coronal stops and fricatives are alveolar.

[3]<u>Greek:</u> The voicing contrast is between a voiced (or sometimes prenasalized) stop with a voicing lead and a voiceless unaspirated stop. Dorsal stops and fricatives are palatalized before front vowels.

[4]<u>Japanese:</u> The phonation type contrast is between voiced plosives with voicing lead (or, sometimes, a short lag VOT) and mildly aspirated plosives with VOT values intermediate between those for the short lag unaspirated plosives and those for the long lag aspirated plosives of Cantonese and English. Dorsal stops are palatalized before front vowels. The voiced affricates [dz] and [dʒ] are word-initial allophones of phonemes which are typically realized as fricatives [z] and [ʒ] in word-medial position. The sequences */tu/ and */si/ are unattested and /ti/, /ʃe/ and /tʃe/ are attested only marginally, primarily in recent loan words from languages such as English. The voiceless alveolar affricate is attested only in the sequence /tsu/.

Table 3

Multiple regression models for consonant accuracy values in different vocalic environments plotted on y-axis in Figure 3. Significant p-values are in bold.

1) Model for Cantonese:	$F_{(2,32)} = 11.01$, $R^2 = 0.41$, p < 0.001			
Coefficient for:	**Estimate**	**Std. Error**	**t-value**	**p-value**
Intercept	99.8	9.28	10.76	**<.001**
Log freq of CV in Cantonese	3.16	1.95	1.623	0.1
Ave fr rank, other 3 lgs.	−0.18	0.05	−3.66	**<.001**
2) Model for English:	$F_{(2,53)} = 23.56$, $R^2 = 0.58$, **p < 0.001**			
Coefficient for:	Estimate	Std. Error	t-value	p-value
Intercept	158.54	15.15	10.45	**< 0.001**
Log freq of CV in English	14.31	2.47	5.79	**< 0.001**
Ave fr rank, other 3 lgs.	−0.09	0.07	−1.22	0.226
3) Model for Greek:	$F_{(2,55)} = 9.87$, $R^2 = 0.26$, p < 0.001			
Coefficient for:	Estimate	Std. Error	t-value	p-value
Intercept	86.23	7.83	11.01	**< 0.001**
Log freq of CV in Greek	0.91	1.31	0.69	0.49
Ave fr rank, other 3 lgs.	−0.16	0.04	−3.44	**0.001**
4) Model for Japanese:	$F_{(2,46)} = 3.37$, $R^2 = 0.13$, **p = 0.04**			
Coefficient for:	Estimate	Std. Error	t-value	p-value
Intercept	76.93	10.20	7.54	**< 0.001**
Log freq of CV in Japanese	1.86	2.21	0.84	0.40
Ave fr rank, other 3 lgs.	−0.12	0.07	−1.73	0.09

Table 4

Hierarchical linear model for English and Greek /s/ and /θ/ consonant accuracy values in lower panel of Figure 4. Significant p-values are in bold.

Coefficient for:	Estimate	Std. Error	t-value	df	p-value
Intercept	−.946	1.214	−.78	40	.441
Language(English)	.165	.364	.45	40	.652
Consonant(θ)	−1.454	.308	−4.715	41	**<.001**
English x Consonant(θ)	−1.806	.465	−3.880	41	**.001**
Vowel(e)	−.661	.214	−3.088	1125	**.003**
Vowel(i)	−.421	.216	−1.951	1125	.051
Vowel(o)	.239	.235	1.013	1125	.312
Vowel(u)	−.087	.282	−.309	1125	.758
Age	.060	.031	1.929	40	.060
Variances	Estimate	df	chi-square		p-value
Intercept	.967	40	145.050		**<.001**
Consonant	.984	41	73.018		**.002**

Table 5

Hierarchical linear model for Cantonese and Greek /t/ and /ts/ consonant accuracy values in lower panel of Figure 5. Significant p-values are in bold.

Coefficient for:	Estimate	Std. Error	t-value	df	p-value
Intercept	−.879	1.043	−.843	39	.405
Language(Greek)	.223	.340	.657	39	.515
Consonant(ts)	−1.023	.394	−2.597	40	**.013**
Greek x Consonant(ts)	−1.373	.558	−2.459	40	**.019**
Vowel(e)	−.068	.261	−.261	1113	.795
Vowel(i)	−.196	.258	−.758	1113	.448
Vowel(o)	−.350	.255	−1.371	1113	.171
Vowel(u)	−.409	.331	−1.235	1113	.218
Age	.080	.028	2.847	39	**.007**
Variances	Estimate	df	chi-square		p-value
Intercept	.395	39	65.653		**.005**
Consonant	1.844	40	97.305		**<.001**

Table 6

Hierarchical linear model for English and Japanese /t/ and /t∫/ consonant accuracy values in lower panel of Figure 6. Significant p-values are in bold.

Coefficient for:	Estimate	Std. Error	t-value	df	p-value
Intercept	−.745	.992	−.751	40	.457
Language(Japanese)	−.599	.340	−1.763	40	.085
Consonant(t∫)	−3.008	.317	−9.499	41	**<.001**
Japanese x Consonant(t∫)	2.320	.476	4.873	41	**<.001**
Vowel(e)	−.008	.268	−.031	1100	.975
Vowel(i)	.066	.249	.265	1100	.791
Vowel(o)	.120	.250	.481	1100	.630
Vowel(u)	−.007	.258	−.029	1100	.977
Age	.087	.028	3.178	40	**.003**
Variances	Estimate	df	chi-square		p-value
Intercept	.343	40	55.834		**.049**
Consonant	.962	41	70.566		**.003**