Key: Yellow highlighting indicates a note to the copy editor.

Cross-language differences in acquisition

Jan R. Edwards, University of Wisconsin – Madison, Mary E. Beckman, Ohio State University, and Benjamin Munson, University of Minnesota

Key words: paidologos, phonological development, acquisition.

Pour l'étude du langage enfantin en général, une observation brute, même très complète, a encore un inconvénient. Elle ne suffit pas à distinguer clairement les particularités de l'enfant observé. L'individualité, chez l'adulte, n'affecte pas le système linguistique, qui est imposé socialement ; elle se réfugie dans l'équilibre particulier des ressource du lexique, la manière d'utiliser les possibilités de la phrase, le débit, la mimique, rarement dans certain détails de l'articulation ; elle se marque surtout au choix des choses dites. Chez l'enfant non encore adapté au langage normal, tout contient une part d'individuel, depuis l'articulation de chaque son du langage jusqu'au sens donné aux mots. Un fait observé chez l'enfant n'est bien utilisable pour le linguiste d'une part, le psychologue de l'autre, que si la part originale de l'individu peut y être délimitée, ce qui ne se réalise bien que par des comparaisons nombreuses. Ceci fait désirer, malgré toutes les difficultés, que les observations sur les enfants se multiplient en toutes langues, et doit encourager les linguists et les amateurs de linguistique à observer les enfants – particulier leurs enfants – même lorsqu'ils ne peuvent pas le faire complètement. (Cohen 1925: 111)

[For studying child language in general, a single set of raw observations, no matter how thorough, still has a drawback. It is not sufficient for clearly distinguishing the idiosyncracies of the child observed. Individuality, in an adult, does not affect the linguistic system, which is imposed socially; it retreats to the particular balance among lexical resources, the way in which one uses possible sentences, rate, gesture, and rarely in certain details of articulation; it is marked above all in the choice of what is said. In the child not yet adapted to language norms, everything contains an element of individuality, from the articulation of each sound of the language to the meaning assigned to words. A fact observed in a child can only be used by the linguist on the one hand or by the psychologist on the other, if the part that is unique to the individual can be delimited, which is something that can be accomplished only by making many comparisons. This makes it desirable, despite the difficulties, for observations of children to be multiplied in all languages, and it should encourage linguists and amateur linguists to observe children – particularly their own children – even if they cannot do so thoroughly.]

1. Introduction

Researchers have long been interested in comparing phonological acquisition across children learning different first languages (e.g. Cohen 1925, Locke 1983). Although children with typical phonological development can follow variable paths as their early vocalizations converge to the norms of their speech community (e.g. Ferguson 1979, Vihman 1993), they do begin learning to talk with the same constraints on production and perception, regardless of what language (or languages) to which they are exposed. In order to produce speech that is intelligible to other members of their speech community, children must acquire progressively more fine-grained phonetic control. Certain speech sound contrasts are more difficult than others (e.g. the /s/:/ʃ/ contrast requires more fine-grained motor control than the /p/:/m/ contrast), so researchers have hypothesized that early-acquired contrasts generally are ones that place relatively lesser demands on the talker/listener, while late-acquired contrasts place relatively greater ones.

Such reasoning has led researchers to propose that similarities in phonological acquisition across children learning different first languages should tell us something about pan-species constraints on what kinds of sound system are easiest to maintain in language transmission. In his influential monograph on child language, aphasia, and phonological universals, Jakobson (1941) proposed specific phonetic bases for generalizations that had been noted in the literature to that time (e.g. by Jespersen 1922). While our explanations have evolved considerably since then, many of the generalizations still stand. For example, across languages, children generally acquire stops and nasals before fricatives and liquids, and voiceless unaspirated stops before both voiceless aspirated stops and prevoiced stops. Moreover, these earlier acquired sound types also are ones that tend to occur in the phoneme inventories of more languages (e.g. Lindblom & Maddieson 1988).

More recently, a growing number of differences in phonological acquisition across languages have also been identified. For example, in a cross-language longitudinal study of consonants transcribed in babbling and early words, de Boysson-Bardies and Vihman (1991) found a dominance of labial sounds at the earliest recording sessions for the French- and English-learning children, but a dominance of lingual sounds for the Japanese- and Swedish-learning children. Moreover, there was a subsequent decrease in labials relative to linguals (particularly dentals) in the productions of the English-learning children, but not in the French-learning children. Examples of cross-language differences for later-acquired sounds include the earlier acquisition of the affricate /ts/ in Cantonese relative to Greek, the earlier acquisition of the non-sibilant fricative /θ/ in Greek relative to English (Edwards & Beckman 2008a), and the earlier acquisition of /s/ in English relative to Japanese (Li et al. 2009). Some of these differences are context specific. For example, /t/ is generally more accurate than /tʃ/ in both English- and Japanese-speaking toddlers, but for Japanese children, /t/ is less accurate than /tʃ/ in the context of /i/ (Edwards & Beckman 2008a).

Presumably, these cross-linguistic differences must be attributed to language-specific factors, such as differences in phoneme frequency in the earliest words that children learn or in the lexicon in general. Thus, Vihman et al. (1994) related the early dominance of labials in French-speaking children and of dentals in Swedish-speaking children to the differences in counts of labial- versus dental-initial content words in speech directed to the children. Similarly, Edwards and Beckman (2008a) related the earlier acquisition of /ts/ in Cantonese relative to Greek to the much higher type frequency of this affricate in the Cantonese lexicon as compared to Greek (/ts/ is almost as frequent as /t/ in Cantonese, while it is a very low-frequency sound in Greek). The same explanation holds for the earlier acquisition of /θ/ in Greek relative to English.

Differences in consonant-vowel (CV) sequence frequency can also help to explain differences in accuracy within and across languages. For example, Edwards and Beckman (2008a) relate the differences between /ti/ and /tʃi/ for Japanese- versus English-speaking children to the extremely low type frequency of /ti/ in Japanese. Similarly, Monnin et al. (2011) examined acquisition of /t/ and /k/ in different vowel contexts by children acquiring French and Drehu, an Austronesian language spoken in New Caledonia. They found that child speakers of both languages produced /k/ more accurately before /u/ than before /i/, in keeping with the predictions of the Frame/Content Theory developed by MacNeilage and colleagues (Davis, MacNeilage & Matyear 2002, other earlier literature reviewed in MacNeilage, this volume). However, contra the predictions of Frame/Content Theory, the French-speaking children (but not the Drehu-speaking children) also produced /t/ more accurately before /u/ than before /i/, a difference that Monnin and colleagues relate to the high frequency of /tu/ relative to /ti/ in French.

This approach of looking for differences in phoneme frequency and phoneme sequence frequency to explain exceptions to developmental universals was our primary focus when we began the παιδολογος (paidologos) project in 2003. This project was a systematic, large-scale cross-language comparison of accuracy in productions of lingual obstruents by children from 2 to 5 years (approximately 100 children in each language) elicited using a picture-prompted auditory word repetition task (Edwards & Beckman 2008b). The comparison began with recordings of children acquiring Cantonese, English, Greek, or Japanese and has been extended in collaboration with other researchers to Korean, two varieties of Mandarin Chinese, Taiwanese, two varieties of French, and Drehu. We were interested in identifying language-specific differences in phonological acquisition that were related to differences in phoneme frequency or phoneme sequence frequency across languages.

Differences in frequency, however, cannot explain all of the cross-linguistic differences that we investigated. For example, they cannot account for the later acquisition of /s/ relative to /ʃ/ in Japanese but not in English, as /s/ is a higher-frequency sound than /ʃ/ in both languages. A secondary focus, therefore, was to look also at finer-grained phonetic differences in how the sounds are produced by adult speakers and how children's immature productions are assimilated to the community norms for the phonetic cues for each relevant contrast (Beckman, Yoneyama & Edwards 2003). One of the most important lessons that we learned from the παιδολογος project is that we would have missed a great deal of what was interesting in our data if we had focused only on cross-linguistic differences in order of phoneme acquisition assessed using phonetic transcriptions, and explanations that focused solely on differences in phoneme frequency across languages.

In this chapter, we will present data from the παιδολογος project and from other cross-language studies to illustrate three important reasons why studies that rely solely on phonetic transcription as data and cross-linguistic frequency differences as explanations are overly simplistic. The first reason is that cross-linguistic differences in production begin very early in life, well before children produce speech sounds correctly. The second reason is that there is daunting cross-linguistic variation in what is ostensibly the "same" sound, which affects not just details of production but also the community-specific norms for what can be perceived as a "correct" production of the sound. This makes it difficult to generalize about speech sound acquisition across languages when studies use native speaker transcriptions as the sole measure of speech-sound acquisition. Finally, there is more to phonological development than learning to produce speech sounds that adults will recognize in terms of the lexical contrasts of the language. Children also need to acquire socio-phonetic competence if they are to be able to quickly and accurately parse the variation they hear, as well as to produce the subtle differences that convey different speaker attributes, such as gender, socioeconomic status, and ethnicity, among others.

2. When do cross-language differences in production begin?

It is by now well established that there are extensive cross-linguistic differences in fine phonetic detail in adult speech. That is, sounds that are transcribed with the same phonetic symbol in cross-language comparisons of phoneme inventories and phonotactics are not the same when examined using finer-grained analysis tools than IPA transcription. For example, the vowel transcribed as /i/ is not as high and front in American English or Dutch as it is in German, Swedish, or Danish, and the vowel transcribed as /u/ is not as back in American English as it in German or French (e.g. Disner 1983, Flege 1987). The voicing contrast in Hungarian, Dutch, Polish, French, and Swedish differentiates stops with short-lag voice onset time (VOT) from stops with voicing lead, even in word-initial position, whereas the "same" contrast in German and English differentiates stops with long-lag VOT from stops with short-lag VOT (e.g. Lisker & Abramson 1964, Keating 1984, Flege 1987, Stoel-Gammon et al. 1994). The consonants transcribed as /t/ and /d/ are dental and typically laminal in French (Dart 1998) or Swedish (Stoel-Gammon et al. 1994), but alveolar and typically apical in American English (Dart 1998), and the consonants transcribed as /ʈ/ and /ɖ/ in Hindi are not nearly so retroflex as the "same" sounds in Tamil or Telegu (Ladefoged & Bhaskararao 1983). Similarly, although both English and Japanese have a sibilant fricative contrast, /s/ in English is alveolar and often apical (Dart 1998) and /ʃ/ also is apical and typically rounded, while /s/ in Japanese is dental and typically laminal and /ʃ/ is an alveolopalatal produced with spread lips (Toda & Honda 2003).

<u>Early differences</u>. These cross-linguistic differences are known to influence early infant speech perception (e.g. Kuhl et al. 1992 for Swedish versus English /i/, Mattock et al. 2010 for French versus English VOT), and a growing literature suggests that language-specific phonetic detail is acquired quite early in production as well. For example, Stoel-Gammon, Williams and Buder (1994) found systematic cross-linguistic differences in fine phonetic detail of productions of /t/ by a sample of 10 Swedish- and 10 American-English-speaking 30-month-old children. They found that the children already were producing language-specific differences in VOT, burst intensity, and burst spectral diffuseness in their /t/ productions, measures that systematically differentiated between the unaspirated laminal dental /t/ of Swedish productions and the aspirated apical alveolar /t/ of English productions by a control group of twenty adults (10 per language). We found similar results when comparing peak frequencies for stop burst spectra in productions of /k/ (and /k^j/) across languages. In adult productions, the dorsal stops of English are less palatalized (less "acute") before front vowels or /j/ and less rounded and backed (less "grave") before /u/ and /o/ by comparison to the "same"

sounds of Greek, and we observed this cross-language difference in productions by English- and Greek-speaking children as young as 24 months (Arbisi-Kelm et al., 2009).

Stoel-Gammon and colleagues looked at 30-month-old children because /t/ is acquired by that age by the majority of typically developing children in both languages, and the question that they were asking was, "Do children begin by 'hitting the right target' for their language, or do they share some default place of articulation and then acquire the language-specific target with increased exposure to the language and practice" (p. 150). The results for dorsal stops in the παιδολογος project are noteworthy, then, because /k/ is typically acquired somewhat later than /t/ by English-speaking children, and "fronting" errors (transcribed [t] for /k/ substitutions) are not unusual in the speech of typically developing 24-month-olds. Moreover, while /k/ before back vowels is error free in productions by Greek-speaking children at an age when some English-speaking ones make [t] for /k/ substitutions in that context, "fronting" errors for the palatalized allophone are not uncommon in either group at this age. In other words, the evidence is against any "unmarked" universal default place for these "young" consonants. Even before the variable productions of children within a speech community converge on a pattern that reliably differentiates /k/ from /t/, the "undifferentiated lingual gesture" (Gibbon 1999) is one that is aimed toward "the right target" for the ambient speech community.

Fricative place contrasts. For "older" consonants that tend to be acquired late across languages as well, we often found language-specific patterns in production before children had reliably mastered a particular contrast – i.e. before productions were identified consistently as accurate by native speaker/transcribers. For example, Li (2008) observed systematic cross-linguistic differences between English and Japanese /ʃ/ and /s/ productions, even by the two-year-olds. At this age, transcribed accuracy rates were below 50% for both fricatives for both languages, and they were especially low for Japanese, where fewer than 10% of /s/ targets and only 20% of /ʃ/ targets were transcribed as correct. Figure 1 shows mean values for three acoustic measures that differentiate /s/ from /ʃ/ in adult productions plotted against age for English and Japanese-acquiring children. The distribution of means for /s/ and /ʃ/ for most of the youngest children show considerable overlap in both English and Japanese. Note, however, that the region of overlap differs between the two languages; the means for the Japanese-speaking children have lower (more /ʃ/-like) centroid frequencies whereas means for the English-speaking children have higher (more /s/-like) values. That is, we see differences between 2-year-old English and Japanese speakers for these two sibilant fricatives, even though the majority of productions of both sounds were transcribed as incorrect by the native-speaker transcribers.

This difference in fine phonetic detail is reflected also in the transcribed substitution patterns (Tables 6.2 and 6.3 in Li 2008). For English-learning children, the most frequently transcribed substitutions were "fronting" errors – i.e. [s] for /ʃ/ and [θ] for /s/. For Japanese-learning children, they were "palatalization" and "stopping" errors – i.e. [ɕ] for /s/ and [tɕ] for /ʃ/. The only one of these patterns that is predicted by a frequency difference between the target consonant and the transcribed substitution is the [s] for /ʃ/ substitution transcribed for the English-learning children. A question that naturally arises, then, is how to explain such differences.

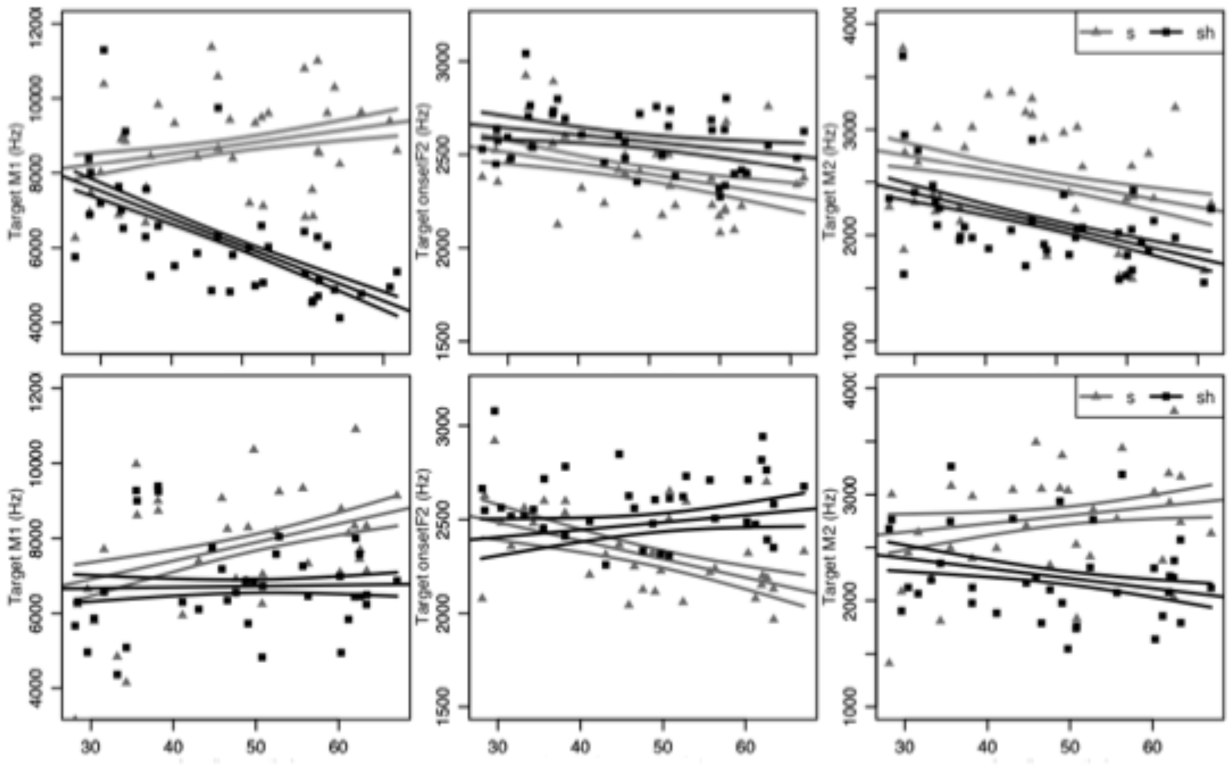

Figure 1. Child-by-child mean values for centroid frequency in a spectrum taken from the middle of the frication interval, F2 frequency at the onset of the vowel, and standard deviation for the fricative spectrum in productions of /s/ (triangles and gray regression lines) and of /ʃ/ (squares and black regression lines) plotted against age in months for English- (top plots) and Japanese- (bottom plots) speaking children. (Figure adapted from Figure 6.3 in Li 2008, used with permission.)

In an earlier paper (Beckman, Yoneyama & Edwards 2003), we suggested that the different accuracy rates and stereotypical error patterns might be related to two other differences between the two languages. First, there is a difference in lingual fricative phoneme inventory. In English, /s/ and /ʃ/ contrast with dental /θ/, whereas in Japanese, they contrast with palatal /ç/. Second, there are different sequential constraints. In particular, although /s/ in Japanese is more frequent than /ʃ/ overall, it is restricted to the contexts of following back vowels /a/, /o/, /u/, and also /e/ (which has the lowest type frequency of the five vowels). In contrast, /ʃ/ occurs very frequently before /i/, less frequently before /a/, /o/, and /u/, and before /e/ in only a small number of words such as /ʃeri:/ 'sherry' which are unlikely to be among the words that young children learn. In English, on the other hand, /s/ and /ʃ/ have no such dependencies. Rather, both sibilants are attested before all of the many more vowels of

English. We further noted that there are many more English words beginning with a lingual obstruent followed by a front vowel than a lingual obstruent before a back vowel. We speculated (Beckman et al. 2003, p. 26) that these facts "might conspire to induce a difference in 'basis of articulation' (Heffner, 1950) between the two languages."

More recently, the advent of easier-to-obtain articulatory measures has begun to provide instrumental support for this long-standing idea of a "basis of articulation" or "articulatory setting" (Honikman 1964) specific to each target language (see review of this concept in Laver 1978). Wilson and colleagues (Wilson 2006, Wilson, Horiguchi & Gick 2007) use Ultrasound in combination with Optotrack to examine tongue and lip postures during inter-sentence pauses. They note consistently higher tongue tip postures for English speakers relative to both French and Japanese speakers. Following our earlier speculation, we wonder whether this "high front" articulatory setting for English might be related to the very high (type and token) frequencies for front vowels in the language. In an early cross-language comparison of formant frequencies measured in babbling productions by 10-month-olds, de Boysson-Bardies et al. (1998) found a concentration of values in the high-front region for English-learning infants, by comparison to Arabic-, French-, and Cantonese-learning infants. They related the differences in formant distributions to cross-language differences in vowel token frequencies in running speech. The difference between English- and French-learning 10-month-olds was replicated in a cross-sectional study by Rvachew et al. (2006). That is, Rvachew and colleagues found that 10-month-olds had a smaller vowel space than older infants for both languages, but centered differently. Plots of mean values across ages showed a developmental expansion of the vowel space in both languages, but in different dimensions, in keeping with the different starting points. If the fine-grained phonetic differences for sibilant fricative place in English- versus Japanese-learning children are related to a difference in habitual tongue posture that begins to be set in place in preverbal babbling, we might expect to see cross-language differences in babbling for features that define other late contrasts as well.

Stop voicing contrasts. Indeed, other researchers have observed language-specific fine phonetic detail for some consonant sounds much earlier than 24 months, in vocalizations of infants before they have begun to produce any words at all. For example, Whalen, Levitt, and Goldstein (2007) examined voice onset time (VOT) in initial stops in babbled utterances of French- and English-acquiring 9- and 12-month-old infants. They found that the French-learning infants produced a much higher proportion of initial stops with voicing lead than did the English-learning infants. This is despite the fact that French word-initial prevoiced stop consonants are not transcribed as being produced correctly until much later in word productions. That is, at 30 months (an age when English-speaking children begin to have good control of the contrast between aspirated /p, t, k/ and unaspirated /b, d, g/ in their language), French-speaking children are transcribed as either substituting the voiceless stop or as producing "filler" syllables – preceding homorganic nasals or vowels "that are easily perceived as one of the indefinite articles *un* or *une*" by the transcriber (Allen 1985, p. 41).

Kewley-Port and Preston (1974) also found very few prevoiced stops in their longitudinal studies of English-learning children (see also Macken & Barton 1980), and their explanation for this pattern predicts the early "substitution" of voiceless unaspirated stops for prevoiced stops in French. 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 or making a "leaky" naso-pharyngeal closure to allow the oral air pressure to vent. The French children's production of filler syllables, then, seems to be a reflex of the latter maneuver, which is then interpreted (and reinforced) as a meaningful morpheme where appropriate, thus explaining the early mastery of determiners by French-learning children (see Demuth & Tremblay 2008 and literature reviewed there).

In the παιδολογος data, we found two more patterns in the acquisition of the voicing contrast in languages that have been described as being like French in contrasting "true" voiced stops with voiceless unaspirated stops. Word-initial voiced stops produced by the Greek-speaking children systematically showed pronounced prevoicing, and unlike in French, even the 24-month-old children were transcribed as making virtually no voicing errors. By contrast, very few of the voiced stops produced by the Japanese-speaking children showed any prevoicing, and there were many instances of transcribed substitutions of [t] for target /d/ and [k] for /g/. Kong et al. (2012) explain the difference between the Greek and Japanese patterns in terms of the different community norms, which are complicated because of sound changes in progress in both communities.

Specifically, in native Greek words, the voiced stops developed fairly recently from nasal-plosive clusters (see, e.g. Arvaniti & Joseph 2004), and in the standard variety today, voiced stops show voicing lead, as in French, but they differ from French in that they can be pronounced with more or less strong prenasalization (Arvaniti & Joseph 2000). Kong and colleagues developed a measure of degree of nasal venting, based on a study by Burton, Blumstein, and Stevens (1992). By this measure, many adult productions showed evidence of some nasal venting, and the children's productions showed even more clear signs of nasal venting to produce long voicing lead. Because prenasalized stops are an accepted allophonic variant, when Greek-learning children use nasal venting, they are not transcribed as producing a preceding indefinite article as the French-acquiring children are.

In Tokyo Japanese also, there are at least two variant realizations of the contrast, due to a sound change in progress (Takada 2011). For some speakers, particularly male and older adult speakers, it is a contrast between prevoiced and short lag stops. However, for younger female adult speakers, many voiced stops have short lag VOT values and the voiceless stops have VOT values intermediate between short and long lag VOT values (as noted also by Riney et al. 2007). Note that this sound change is distinct from the better-known older sound change, whereby [ŋ] is no longer produced as an allophone of /g/ by most Tokyo Japanese speakers today. The older [ŋ]~[g] alternation may explain why Yasuda (1970) noted virtually no errors for /g/ in the three-year-old children she studied, in marked contrast to the high error rates for /g/ and /d/ in Kong et al. (2012). Presumably, children must be learning to control other cues in addition to VOT in order to make the voicing contrast today. Kong and colleagues suggest that this presumed greater complexity might explain why the Japanese-learning children in the παιδολογος database were transcribed as making more voicing errors than either the Greek- or the English-learning children.

In Seoul Korean, another language that had traditionally been described as contrasting short-lag and intermediate-lag VOT values, one cue to this contrast is a difference between modal and breathy voice, as noted by Kong and colleagues (Kong et al. 2011, Holliday & Kong 2011) among others. One measure of breathy voice quality is the difference in power between the first and second harmonic (H1-H2), a measure that has been correlated with contrastive breathy voice in many languages (e.g. Miller 2007 for Ju|'oansi vowels and consonants, Gordon & Ladefoged 2000 for a review of earlier studies for other languages with contrastive breathy voice on vowels or consonants). Kong et al. (2012) found that H1-H2 values were systematically higher for voiceless than for voiced stops produced by adult Japanese speakers. However, they also found this to be true for productions by adult English speakers. Therefore, in order to assess whether there are fine-grained phonetic differences between Japanese and English in the use of voice quality, and to see whether these differences might explain the later acquisition of the voicing contrast in Japanese-learning children relative to English-learning children, it is necessary to adopt more sensitive behavioral measures of the community norms. We describe these measures in the next section.

3. Using perception tasks to assess differences in community norms

As the results reviewed in the previous section make clear, when children's productions are examined using the same acoustic measures that have been applied in documenting the extent of differences in adult productions, no exact comparisons are possible. Even when the two languages being compared have ostensibly the "same" inventory – e.g. a two-way contrast in sibilant fricative place or a two-way contrast between prevoiced and voiceless stops – the children are never really acquiring the "same" sounds. This means that, in order to assess perceived production accuracy relative to the community norms, we also need to supplement accuracy as gauged by the phonetic symbol assigned by a transcriber with finer-grained measures. Recently child language researchers have begun to develop such measures by designing perception tasks to elicit accuracy judgments for children's productions of target sounds from samples of naïve listeners from each target community.

For example, Li, Munson, Edwards, Yoneyama, and Hall (2011) extracted CV stimuli from a large subset of the same productions of English and Japanese words beginning with /ʃ/ and /s/ examined in Li et al. (2009). Li and colleagues presented these stimuli (N=200 for each language) twice, in two different blocks, to 19 English-speaking listeners (tested in Minneapolis) and 20 Japanese-speaking listeners (tested in Tokyo). In one block, listeners said whether a stimulus was an acceptable production of /ʃ/ and, in the other, whether it was an acceptable production of /s/. The target consonant was defined in the instructions at the beginning of the block in terms of orthographic categories appropriate for the language and illustrated with sample words, such as *shape* for the "sh" category for the English-speaking listeners and さる /saru/ 'monkey' for the "さ行" ("s") category for the Japanese. The data were analyzed separately by both listener and speaker language. A token was judged to be classified as /ʃ/ if 70% or more of the listeners answered "yes" in the "sh" block and "no" to the "s" block. Conversely, it was judged to be /s/ if 70% of the listeners answered "yes" in the "s" block and "no" in the "sh" block. The stimuli were plotted in a two-dimensional space defined by the centroid frequency of a spectrum taken over the middle 40 ms of the fricative (as in the left panels in Figure 1) and the F2 frequency of the following vowel at its onset (as in the middle panels of Figure 1). Li, Edwards, and Beckman (2009) had shown earlier that the first of these measures discriminates between the two fricatives in productions by those English-speaking children who were transcribed by the English-speaker phonetician as having a contrast, and that the two measures together discriminate between the two fricatives in productions by those Japanese-speaking children who were transcribed by the Japanese phonetician as having a contrast. In the naïve English-speaking listeners' classifications of the children's productions in the Li et al. (2011) study, the /s/ stimuli occupied a larger area in this space than the /ʃ/ stimuli. For the Japanese-speaking listeners, the relationship was the opposite: the /ʃ/ space was larger than the /s/ space. Li and colleagues argue that these cross-linguistic differences in perception might be part of the explanation for the cross-linguistic differences in acquisition. When faced with a sound intermediate between /s/ and /ʃ/, Japanese listeners are more apt to call it "sh", while English-speaking listeners are more apt to call it "s".

An even more sensitive measure can be obtained by asking listeners to provide a rating along a continuous visual analog scale (VAS, Massaro & Cohen, 1983), instead of a simple yes/no response. Kong et al. (2012) used this method to elicit goodness ratings for CV stimuli extracted from English- and Japanese-speaking children's productions of words beginning with /t/, /k/ versus /d/, /g/. They analyzed the responses by building models that regressed the ratings against VOT alone, or against VOT and either fundamental frequency or H1-H2 (after transforming the acoustic measures into z-scores to be able to compare the regression coefficients directly). Adding H1-H2 values to a model that differentiated between voiced and voiceless stops significantly improved the model fit for the productions of Japanese- but not English-speaking adults. Adding H1-H2 values to the model significantly improved the model fit for the productions of Japanese- but not English-speaking children. These results suggest that children learning Japanese must learn to control voice quality in addition to VOT in order to be recognized as making a distinction between voiced and voiceless stops by adults in the ambient speech community.

An added advantage of the VAS design is that it elicits a continuous response that is simultaneously listener-specific and stimulus-specific. This means that it can be used in combination with other responses, to begin to understand the social dynamics of the input that children receive during acquisition. For example, Julien and Munson (2012) elicited productions of target English words beginning with /s/ versus /ʃ/ in both a casual and a clear speech style from a 22 English speakers, to provide a baseline measure of each speaker's range for the hypoarticulation/hyperarticulation continuum. The same subjects then participated in a "listen-rate-say" task, in which they listened to CV stimuli extracted from English-speaking children's productions of these words, rated the initial sound of each stimulus on a VAS scale from "s" to "sh", and then said the target word as a model for the child. The results suggest that English-speaking adults in this particular speech community, at least, hyperarticulate speech in response to listening to a child's production that they perceive to be inaccurate.

More recently, we have begun to use VAS ratings also to explore cross-linguistic differences in vowel perception, as a first step in building models of how pre-verbal infants might use responses from adults in the ambient community in the "imitation game" (de Boer 2000, Plummer, 2014). Figure 2 illustrates the kinds of difference that we are finding, showing a small part of the results from an experiment in which adult native speakers of Cantonese (n=15), English (n=21), Greek (n=20), Japanese (n=21), and Korean (n=20) categorized 6 sets of 38 synthetic vowels created with an articulatory synthesizer (Boë & Maeda 1997) to simulate the vocal tract and voice source of a very young infant (in the block shown in the figure), or of a two-year-old, four-year-old, or ten-year-old child, or of a 16-year-old or 21-year-old man (see Plummer et al. 2013 for more complete results). Cantonese- and English-speaking listeners categorized each stimulus by clicking on any of 11 keywords representing the monophthongal vowels in each language (e.g. Cantonese 歡 /fun/ and English *soup* /sup/ for the "shared" phoneme /u/). Listeners for the other languages categorized by clicking on a symbol or symbol string that unambiguously represented a (short monophthongal) vowel in isolation (e.g. Korean 우, Japanese う, Greek ου for the "shared" phoneme /u/), choosing among 7 vowels (Korean-speaking listeners) or among 5 vowels (Greek- and Japanese-speaking listeners).

The first panel of the figure shows the location in the vowel formants space of the stimuli simulating infant productions (see Ménard et al. 2009 for further details). The remaining panels of the figure show some of the categorization results, separately by adult listener language. In each, the size of the symbol is proportional to the percentage of listeners for that language who categorized the stimulus as /u/. Consider first the results for the adult Japanese listeners. Stimuli in the high-mid portion of the vowel space were just as likely to be identified as う as stimuli in the high-back portion. This is consistent with the description of the Japanese /u/ as an unrounded [ɯ]. Contrast this pattern with the stimuli labeled *soup* in English and ου in Greek. Both English and Greek have only two high vowels, /i/ and /u/, and adult listeners in both of these languages labeled a large set of non-front high vowels as /u/, albeit not as many as the Japanese did. (The same pattern held for the Texas dialect English-speaking listeners' responses to these stimuli in Ménard et al. 2009.) Finally, contrast both patterns to the very few stimuli that were labeled as 우 by the Korean- or as 歡 by the adult Cantonese-speaking listeners. Korean has three high vowels, /i/, /ɨ/, and /u/. Many of the stimuli that English, Greek, and Japanese listeners identified as "/u/" (i.e. responding with *soup*, ου, or う) were identified as /ɨ/ by adult speakers of Korean. Cantonese has a different set of three high vowels, /i/, /y/ and /u/. Moreover, the type frequency of /u/ in Cantonese is lower than that of many of the other languages because of a phonotactic prohibition against /u/ after any of the dental consonants /t, t^h, s, ts, ts^h, n, l/. Presumably, these two factors are responsible for Cantonese speakers' unwillingness to identify many synthesized vowels as /u/. (The same pattern holds for the adult French-speaking listeners' responses to the similar synthetic stimuli in Ménard et al. 2004, and it mirrors the differences between the English- versus French-dominant listeners' categorization of the natural vowel tokens in Figure 3 of Rvachew et al. 2008.) These findings emphasize that categories such as /u/ can be associated with very different patterns of perception across languages that ostensibly share this phoneme.

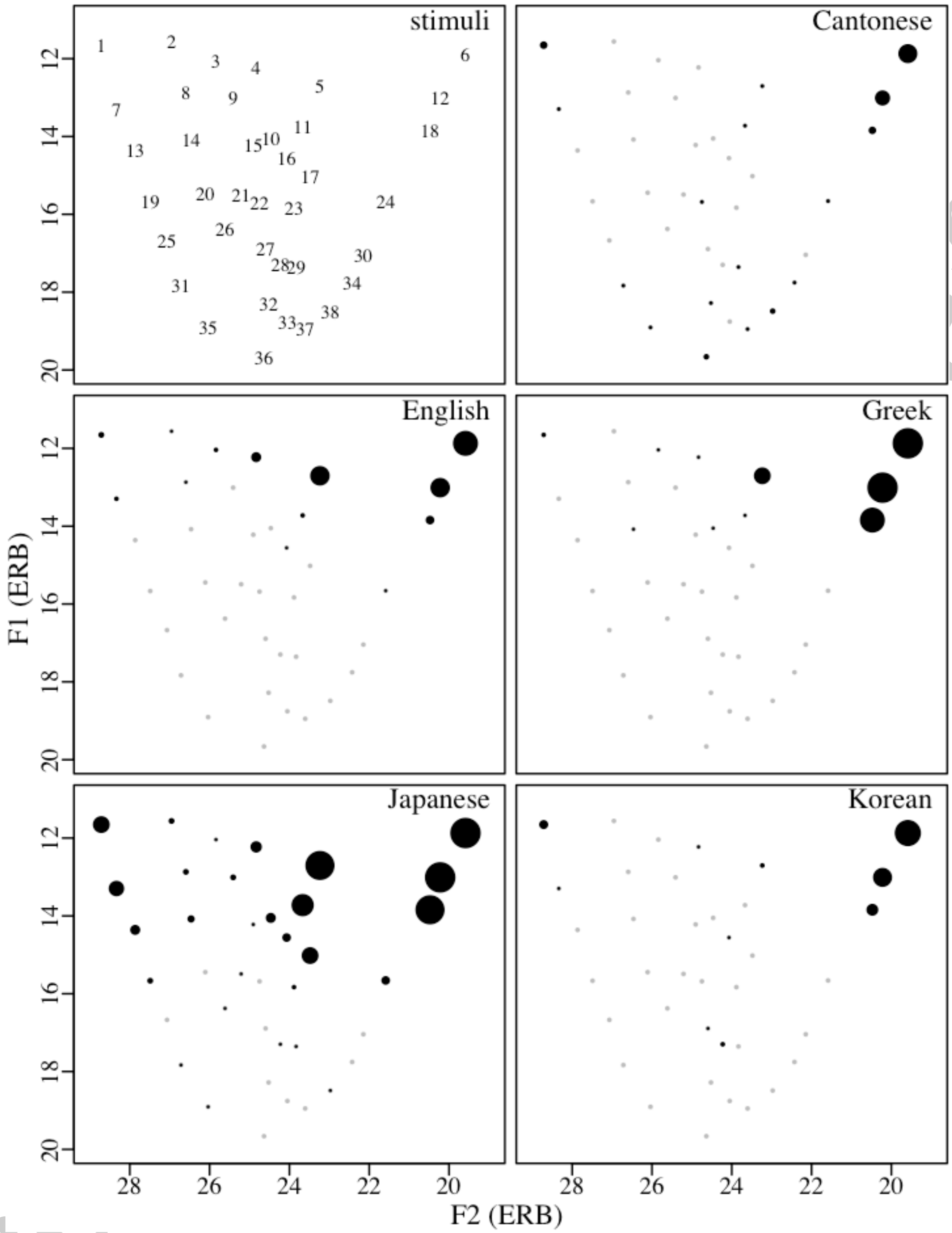

Figure 2. Stimuli synthesized with an articulatory synthesizer scaled to represent an infant's vocal tract (top left) and proportion of listeners who identified each one as /u/ in a cross-language vowel perception experiment.

Cross-linguistic differences in perception like those described in the previous paragraph may underlie some cross-linguistic differences in the acquisition of speech sounds. For a hypothetical illustration of how cross-linguistic differences in perception might influence acquisition, consider children's acquisition of the vowel /u/. Imagine that a child were saying a word such as Cantonese /k^{h}u:^{55}ŋa:^{21}/ 'braces', English *cougar*, French /kuto/ 'knife', Greek /kukla/ 'doll', Japanese /kutsu/ 'shoes', or Korean /kutu/ 'shoes', and produced a sound in the high central region of the vowel space – i.e.

something near to stimulus number 5 – for the vowel in the first syllable of the word. If this child were acquiring English, Greek, or Japanese, the token might well be recognized as a correct token of /u/. If this same child were acquiring Cantonese, French, or Korean, the very same token would not be recognized as correct. Rvachew et al. (2008) present acoustic and perceptual evidence that /u/ is acquired earlier in English than in French. If we were to find, similarly, that /u/ is acquired earlier in Greek and Japanese than it is in Cantonese and Korean, then we should consider the possibility that these apparent cross-language differences in production are due, at least in part, to cross-language differences in adults' perception of children's productions. That is, different speech community norms for carving up the acoustic-phonetic space map differently onto the distribution of sounds that children are producing.

4. Variation conditioned by position and by social categories

In other future applications of these methods, we plan to begin to explore the effects of cross-language differences in positional variants of more challenging contrasts, to see how these affect acquisition. For example, children who are acquiring Tokyo Japanese must not only learn to produce the contrast between /ʃi/ and /ʃu/, but they must also learn to recognize and reproduce this contrast even before voiceless stops, in words such as /ʃika/ 'deer' and /ʃukudai/ 'homework', where these CV sequences are often produced with no voiced interval to carry the distinct resonances of the [i] versus [ɯ] contrast, so that the vowel posture must be deduced from the effects on the frication spectrum (Beckman & Shoji 1984). Imaizumi, Hayashi and Deguchi (1995) show that at least some adults adjust their productions of vowels in syllables such as these in talking to children, particularly to children with hearing impairment, so that there is a voiced interval to carry the vowel formants. We could use the "listen-rate-say" task to see whether comparable adjustments can be induced in adults just in case they rate a child's production as being a less acceptable production of the syllable.

Another such positional effect involves word-medial /d/ for children who are learning English or Japanese. In American English, /d/ and /t/ contrast with each other but not with [ɾ]. However, children who are learning American English master both stops in word-medial position relatively late, because most longer words that they are learning are trochees, and foot-medial position is a prosodic environment where both /t/ and /d/ are typically produced as [ɾ] (see Klein & Altman 2002 and literature reviewed there). This flap consonant is very similar to the difficult /r/ phoneme of Japanese, which is one of the last sounds to be mastered by Japanese-learning children. In word-medial position, the Japanese /r/:/d/ contrast is perceptually difficult as well, and /r/:/d/ confusions are attested even in school-age children (Otuka 2005). Children who are learning American English do not need to differentiate [ɾ] from [d] as phonemes, but they do need to learn whether words such as *Daddy* and *water* have medial /d/ or /t/, in order to be able to map from the usual variant to the "correct" hyperarticulate variant as they begin to command the distinction between casual and careful speech styles. Again, we could use the "listen-rate-say" task to see whether adult speakers in the two different speech communities behave differently in accommodating to young children's difficulties with [d] and [ɾ] in word-medial position.

The example of position-specific flapping (and neutralization of the /d/:/t/ contrast) in English drives home the point that children must amass a substantial body of knowledge about systematic sources of variation in production, and they must exploit this knowledge in real-time speech perception. One substantial source of variation relates to social categories: At least some of the variability in speech sounds occurs because the phonetic characteristics of sounds are manipulated in a way that conveys attributes about speakers. These attributes can be highly individual (i.e. distinctive pronunciations that allow familiar people to identify a talker), or they can relate to group-level characteristics, be they macro-sociological categories such as race, age, and gender, or local structures like social cliques in schools.

Researchers have begun to document how children learn to produce socially meaningful variation in language, such as the difference between the careful-speech register of American English in which medial /t/ and /d/ are differentiated and the casual-speech register in which they are neutralized to a flap. Redford and Gildersleeve-Newman (2009), examined the production and perception of these two registers in typically developing 3- to 5-year old children. They found that even the youngest children produced speech that adults perceived as more accurate in careful speech than in conversational speech (a result that is replicated in a perception study and accompanying acoustic analyses by Syrett and Kawahara 2013). However, the difference in rated accuracy of words produced in careful versus conversational styles was larger for adults than for 5-year olds, and larger for 5-year-olds than for younger children. Redford and Gildersleeve-Newman's results show that careful and casual speech styles become more distinct between ages 3 and 5, and control of the distinction continues to develop as children become adults.

Consider next the acquisition of one of the most widely studied social categories, gender. Docherty et al. (2006) show that girls acquiring the variety of English spoken in the Tyneside region of Northeast England begin to produce a gendered variant, preaspiration in medial stop consonants, sometime between 42 and 48 months of age. They also show that this gender variation is correlated with differences in the early input; mothers of girls produce the preaspirated variant more in talking to their babies relative to mothers of boys.

Another case of early socially relevant gender differentiation comes from the παιδολογος database. Li and colleagues (Li 2008, Li et al. 2008) examined the development of the three-way sibilant fricative contrast in a dialect of Dongbei (northeast) Mandarin spoken in Songyuan City. In the speech of young adults, /s/ has a higher centroid frequency than both /ɕ/ and /ʂ/, while /ʂ/ has a higher F2 onset frequency than both /s/ and /ɕ/. Sociolinguistic studies of Mandarin dialects spoken in Beijing and further north have documented the emergence of a so-called "feminine accent" variant of alveolopalatals, a systematically higher centroid frequency for /ɕ/ and /tɕ/ relative to /ʂ/ and /tʂ/ for younger women relative to older women and men (see Hu 1991, Li 2005, and even earlier literature reviewed in Li 2005). Presumably, this mimics the effect of having a smaller vocal tract and thus sounding more "feminine" (or child-like). Li examined spectral characteristics of the three Mandarin fricatives by two-, three-, four-, and five-year old boys and girls in Songyuan. She found that the centroid frequencies of 2- and 3-year-old boys' and girls' productions of /ɕ/ and /ʂ/ were not different. However, there was a significant gender by fricative interaction for the centroids of 4- and 5-year-old children's productions, such that girls produced a larger difference between /ɕ/ and /ʂ/ than did boys. Li interprets this finding as potential evidence for acquisition of the feminine-accented variant of /ɕ/ by the age of 4.

By contrast, Kong et al. (2012) found no gender differentiation for voiced stops in Japanese. As described in section 2, Japanese voiced stops are generally produced by adult males and older speakers with prevoicing but by adult females and younger speakers with short lag VOT. Kong et al. (2014) show that lack of prevoicing in adult males is associated with less masculine-sounding voices. However, Kong and colleagues found very high rates of the short lag variant among 2- to 5-year-old boys as well as girls. This is likely due to the difficult intrinsic aerodynamic demands of producing prevoicing. That is, the short-lag variant that is associated with female speakers who are leading this sound change in progress is also an "easier" sound that occurs "naturally" as a characteristic early misarticulation of voiced stops in languages such as French. As Li, Kong, Beckman and Edwards (2008) note, these findings suggest that socially meaningful phonetic variation could be acquired early, but evidence for or against acquisition must be interpreted carefully, taking more general constraints into account.

Roberts (1994) makes a similar point in her study of the acquisition of phonological and morphological constraints on socially meaningful variation in the production of coda /t/ and /d/ in American English. Roberts studied preschool children in Philadelphia and found that the children deleted coda /t/ and /d/ less often before a vowel than before a following consonant, a pattern that could be due simply to the easier perceptual parsing of the stop closure in prevocalic positions. However, Roberts also found that the children deleted /t/ and /d/ less often before a pause, a pattern that is characteristic of Philadelphia, which differentiates it from the New York pattern. Of course, Roberts's argument would have been even stronger if she had been able to compare productions by preschool children acquiring the New York dialect.

Imaizumi et al. (1999) is a good example of how cross-dialect comparison (like cross-language comparison) can help differentiate developmental differences from sociophonetic differences. Imaizumi and colleagues looked at vowel devoicing in productions of words such as /kitsutsuki/ 'woodpecker' elicited from three groups of speakers of the Tokyo dialect and of the Osaka dialect. There were age differences in both dialects, with the adults producing more devoiced vowels than the 5-year-olds and the 5-year-olds producing more devoiced vowels than the 4-year-olds. These age differences could reflect differences in the input, if speech directed to younger children differs from speech directed to older children in the direction expected from the results of Imaizumi et al. (1995). Also, the Osaka speakers showed much lower rates of vowel devoicing than the Tokyo speakers, confirming earlier research. Moreover, the cross-dialect difference was attested for all three age groups.

In other related work, Roberts and Labov (1995) suggest a "critical period" for the acquisition of regional variation in the contexts for a split of /æ/ into a "tensed" low-mid falling diphthong contrasting with the original "lax" monophthong that characterizes many dialects of northeastern U.S. Young children of parents who had moved to Philadelphia as adults acquired the Philadelphia pattern, rather than the pattern of their parents' native dialect. The older siblings of these children either showed a mixed pattern or their parents' pattern. This finding is in accord with the much older observation that children of immigrants generally grow up producing the community language with the appropriate regional accent, rather than the foreign accent of their parents, particularly if they are very young when the parents immigrated or are born after the immigration.

Within-child (as opposed to within-family) comparisons provide evidence that children of immigrants may be learning to command socially meaningful variation when they learn to speak like their peers. In a study of vowels produced by children of immigrants and close neighboring friends, Khattab (2007) found that the children produced variants more like the Arabic-accented immigrant parents' speech when talking with the parents or when caricaturing their parents' speech as compared to when talking with the non-immigrant neighbors.

Baron-Cohen and Staunton (1994) uncovered an interesting exception to the general rule that children learn to control the accent of their peers and not (just) that of their parents. They compared the speech of children with autism to that of their siblings with typical language development. Half of the children had mothers who were non-native speakers of English, while half had native-English speaking mothers. All children had grown up in England and attended school with native English-speaking peers. Based on ratings by naïve listeners, most of the children with autism (83%) who had non-native English-speaking mothers were judged to speak "like their mother," while most of their siblings with typical language development (88.5%) were judged to speak "like their peer group." In this case, the comparison between neurotypical children and children with autism provides the control for interpreting the general rule as evidence of the early development of socially meaningful variation in regional accent or foreign accent.

Comparisons across groups within a language community also can help in interpreting evidence of gender differentiation. In addition to control of specific gendered sounds, children also learn control of more global aspects of speech production that let them sound progressively more like the adults in the ambient speech community who share their gender identity. Naïve adults can reliably differentiate between the speech of boys and girls in their own language community, even for children as young as 4 years (Perry, Ohde & Ashmead 2001), and evidence is beginning to emerge suggesting that this differentiation results from learned control rather than from the subtle differences in vocal tract size and shape recently documented by Vorperian et al. (2011). For example, Munson and Baylis (2007) found that 3- to 7-year-old boys with phonological disorder were rated to sound less "boy-like" (more "girl-like") than their typically developing age peers. Furthermore, boys with Gender Identity Disorder or GID (a clinical label sometimes given when individuals display behavior that is not expected for their sex, such as having opposite-sex peer preferences, preferences for opposite-sex typed toys, and, in some cases, overt gender dysphoria) have less masculine-sounding speech than age-matched boys whose gender development was deemed to meet cultural expectations. Crocker and Munson (2006) examined the characteristics of the speech of 5- to 13-year-old boys clinically identified as having GID. A perception test with content-neutral speech samples showed that even the youngest boys with GID were rated to sound less masculine relative to boys with typical gender development. Acoustic analysis further suggested that this difference was due to the production of specific gender-marked variants of sounds, rather than to overall characteristics such as average fundamental frequency or the spacing of the ensemble of vowels in the F1/F2 space. These findings suggest that boys with GID learn specific gendered speech variants early in life, perhaps as the result of selective attention to specific adult models.

5. Summary and conclusion

In this chapter we have reviewed studies comparing children's productions across languages and across varieties within a language. These studies illustrate the three points with which we began. First, even before they begin to produce vocalizations that are reliably recognized as words by the ambient speech community, children's productions reflect language-specific norms. Second, adults perceive children's productions in terms of language-specific perceptual norms. Finally, children must learn to produce socio-indexical characteristics that let them signal their identities and their social affiliations.

In the last half century, there has been much attention to the interplay between universal constraints and language-specificity in regards to the first two points – i.e., the production and perception of lexical contrasts. The fact that human language is also used to signal group affiliation may be deeply embedded in the evolution of the species. Fitch (2004) reviews research on cases of non-human animal communication where subgroups of species produce distinctive vocalizations to mark themselves as kin. For example, in large breeding colonies, seal pups produce vocalizations that are sufficiently distinct that when their mothers return from hunting for food they can recognize and locate their kin even in very large groups. Socially-indexed phonetic variation serves an analogous purpose, as anyone who has discerned a familiar regional accent among a large crowd of people can attest.

The data that we have discussed drive home two important methodological points that we are far from the first to make. The first point concerns the measures that we use to study children's production. Far more phonetic variation exists, both across and within languages, than simple IPA-style transcriptions would suggest. In some cases, the cross-language variation explains cross-linguistic differences in patterns of acquisition that would be difficult to understand if only IPA transcriptions were used. In particular, the studies described in Section 2 show that there are cross-linguistic differences in children's earliest productions, even before they have mastered particular contrasts. Moreover, the studies described in Section 3 show that adults in different speech communities may interpret children's productions differently even when they are phonetically the same. For example, when French children and Greek children both use nasal venting to produce strong prevoicing, their productions are perceived as following different developmental trajectories on the way to mastering the "same" difficult voicing contrast.

The second methodological point concerns the appropriate control comparisons. Taken together, the results in Sections 2 and 3 make it clear why we need to compare across languages in order to make sensible models of the development course that children follow as they learn to produce speech. The results also make it clear why we need to exercise caution in interpreting evidence that very young children have acquired socially meaningful variation. That is, it is not enough simply to show that children's productions vary in ways that reflect the variation within the society into which they are born. If Northeastern Mandarin-speaking girls' productions of /ɕ/ and /tɕ/ differ from boys' productions of these sounds, it does not necessarily mean that they have chosen to mark themselves as little women. It could simply mean that their mothers and other caretakers use the "feminine accent" more in talking to them. Similarly, it is not surprising that children growing up in Tokyo show higher devoicing rates than children growing up in Osaka. Children acquire the production patterns that let them match the sound patterns that they hear. The studies described in Section 4 that compare productions by neurotypical children to productions by children with autism, or that compare productions by a single child when talking to different addressees or in different styles, allow us to see this point especially clearly.

Taken together, then, the studies reviewed in all three of these sections help us to appreciate better the true complexities of speech production and the importance of the social group at all stages of language acquisition. The results in section 4, especially, suggest that the acquisition of socially meaningful phonetic variation cannot be taken for granted – children gradually learn to vary their productions in ways that let them control how they mark their identities as members of a particular speech community in terms of gender, social class, and regional accent. In contrast to the relatively large literature on stylistic variation in other linguistic variables, such as word choice (Andersen, 1992), we know remarkably little about children's acquisition of socio-indexical phonetic variation. In fact, we do not even know at what age children's voices are as recognizable as adult voices. We might ask whether the fact that all mothers turn to look for their child on the playground when they hear a child cry "mommy" is because they are such good caregivers, or is it simply because young children's voices are not individually recognizable? When we contrast this to what the seal pups (do not) need to learn to be individually identifiable, it helps us begin to understand how kin selection may have acted as a ratchet in driving the evolution of greater and greater complexity and cultural diversity that spoken language enabled.

Acknowledgements

This work was supported in part by NIH grant R01 DC02932 and NSF grant BCS0729140 to Jan Edwards, by NSF grant BCS0729277 to Benjamin Munson, and by NSF grant BCS0729306 to Mary E. Beckman. We also gratefully acknowledge the contributions of Tim Arbisi-Kelm, Eunjong Kong, and Fangfang Li.

References

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

Cohen, Marcel 1925. Sur les langages successifs de l'enfant. *Mélanges linguistiques offerts à M. J. Vendryes*, pp. 109-127. Paris: Champion.

Arbisi-Kelm, Timothy, Mary E. Beckman, Eun Jong Kong & Jan Edwards. 2009. Production of dorsal places of articulation by children and adult speakers of four languages. Annual Meeting of the Linguistics Society of America, January 8-11, San Francisco, CA.

Arvaniti, Arvaniti & Brian D. Joseph. 2000. Variation in voiced stop prenasalisation in Greek. *Glossologia* 11/12. 131-166.

Arvaniti, Arvaniti & Brian D. Joseph. 2004. Early Modern Greek /b d g/: Evidence from rebétika and folk songs. *Journal of Modern Greek Studies* 22. 73-94.

Baron-Cohen, Simon & Ruth Staunton. 1994. Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language* 14. 241-248.
Beckman, Mary E. & Atsuko Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41. 61-71.
Beckman, Mary E., Kiyoko Yoneyama & Jan Edwards. 2003. Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7(2). 18-28.
Boë, Louis-Jean & Shinji Maeda. (1997). Modélization de la croissance du conduit vocal. Éspace vocalique des nouveaux-nés et des adultes. Conséquences pour l'ontegenèse et la phylogenèse. *Journée d'Études Linguistiques* (La Voyelle dans Tous ces États) 98-105.
de Boer, Bart. 2000. Self organization in vowel systems. *Journal of Phonetics* 28(4). 441-465.
de Boysson-Bardies, Bénédicte, Pierre Hallé, Laurent Sagart & Catherine Durand 1989. A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language* 161. 1-17.
de Boysson-Bardies, Bénédicte & Marilyn May Vihman. 1991. Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67(2). 297-319.
Burton, Martha W., Sheila E. Blumstein & Kenneth N. Stevens. 1992. A phonetic analysis of prenazalized stops in Moru. *Journal of Phonetics* 20(1). 127-142.
Cohen, Marcel. 1925. Sur les langages successifs de l'enfant. *Mélanges linguistiques offerts à M. J. Vendryes*, 109-127. Paris: Champion.
Crocker, Laura & Benjamin Munson. 2006. *Speech characteristics of gender-nonconforming boys*. Conference on New Ways of Analyzing Variation in Language, Columbus, OH.
Dart, Sarah N. 1998. Comparing French and English coronal consonant articulation. *Journal of Phonetics* 26. 71-94.
Davis, Barbara L., Peter F. MacNeilage & Christine L. Matyear. 2002. Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches. *Phonetica* 59.75-107.
Demuth, Katherine & Annie Tremblay. 2008. Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language* 35. 99-127.
Disner, Sandra Ferrari. 1983. Vowel quality: *The relation between universal and language-specific factors*. Los Angeles, CA: University of California dissertation. (*UCLA Working Papers in Phonetics* 58.)
Docherty, Gerard, Paul Foulkes, Jenny Tillotson & Dominic Watt. 2006. On the scope of phonological learning: Issues arising from socially-structured variation. In Louis Goldstein, D. H. Whalen & Catherine T. Best (eds.) *Laboratory phonology 8: Varieties of phonological competence*, 393-421. Berlin: Mouton de Gruyter.
Edwards, Jan & Mary E. Beckman. 2008a. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development*, 42: 122-156.
Edwards, Jan & Mary E. Beckman. 2008b. Methodological questions in studying consonant acquisition. *Clinical Linguistics and Phonetics* 22(12). 939-958.

Ferguson, Charles A. 1979. Phonology as an individual access system: Some data from language acquisition. In Charles J. Fillmore, Daniel Kempler & William S-Y. Wang (eds.) *Individual differences in language ability and language behavior*, 189-201. New York: Academic Press.

Fitch, W. Tecumseh. 2004. Kin Selection and "mother tongues": A neglected component in language evolution. In D. Kimbrough Oller & Ulrike Griebel (eds.) *Evolution of communication systems: A comparative approach*, 275-296. Cambridge, MA: MIT Press.

Flege, James Emil. 1987. The production of "new" and "similar" phonemes in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15.47-65.

Gibbon, Fiona E. 1999. Undifferentiated lingual gestures in children with articulation / phonological disorders. *Journal of Speech, Language and Hearing Research* 42. 382-397.

Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29. 383-406.

Heffner, Roe-Merrill Secrist. 1950. *General Phonetics*. Madison, WI: University of Wisconsin Press.

Holliday, Jeffrey J. & Eun Jong Kong. 2011. Dialectal variation in the acoustic correlates of Korean stops. *XVIIth International Congress of Phonetic Sciences,* 878-881.

Honikman, Beatrice. 1964. Articulatory settings. In David Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott & J. L. M. Trim (eds.) *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday*, 73-84. London: Longman.

Hu, Mingyang. 1991. Feminine accent in the Beijing vernacular: a sociolinguistic investigation. *Journal of the Chinese Language Teachers Association* 26. 49-54.

Imaizumi, Satoshi, Akiko Hayashi & Toshisada Deguchi. 1995. Listener adaptive characteristics of vowel devoicing in Japanese dialogue. *Journal of the Acoustical Society of America* 98(2). 768-778.

Imaizumi, Satoshi, Kiyoko Fuwa & Hiroshi Hosoi. 1999. Development of adaptive phonetic gestures in children: Evidence from vowel devoicing in two different dialects of Japanese. *Journal of the Acoustical Society of America* 106(2). 1033-1044.

Jakobson, Roman. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksell. Translated from the German by Allan R. Keiler for publication as Jakobson 1968. *Child language. Aphasia and phonological universals Janua linguarum. Series minor 72*. The Hague: Mouton.

Jesperson, Otto. 1922. *Language: Its nature, development, and origin*. New York: Henry Holt and Company

Julien, Hannah & Benjamin Munson. 2012. Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech, Language, and Hearing Research* 55(6). 1836-1849.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60(2). 285-319.

Kewley-Port, Diane & Malcolm S. Preston. 1974. Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2. 195-210.

Khattab, Ghada. 2007. Variation in vowel production by English-Arabic bilinguals. In Jennifer Cole & José Ignacio Hualde (eds.) *Laboratory phonology 9*, 383-410. Berlin: Mouton de Gruyter.
Klein, Harriet B. & Elaine K. Altman. 2002. The acquisition of medial /t, d/ allophones in bisyllabic contexts. *Clinical Linguistics and Phonetics* 16(3). 215-232.
Kong, Eun Jong, Mary E. Beckman & Jan Edwards. 2011. Why are Korean tense stops acquired so early? The role of acoustic properties. *Journal of Phonetics* 39. 196-211.
Kong, Eunjong, Mary E. Beckman & Jan Edwards. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40. 725-744.
Kong, Eunjong, Kiyoko Yoneyama, & Mary E. Beckman. 2014. Effects of a sound change in progress on gender-marking cues in Japanese. 14th Conference on Laboratory Phonology, Tokyo, Japan, July 25, 2014.
Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044). 606-608.
Ladefoged, Peter & Peri Bhaskararao. 1983. Non-quantal aspects of consonant production: A study of retroflex consonants. *Journal of Phonetics* 11(3). 291-302.
Laver, John. 1978. The concept of articulatory settings: An historical survey. *Historiographia Linguistica* 5(1/2). 1-14.
Li, Fangfang. 2005. An acoustic study on feminine accent in the Beijing Dialect. In Qian Gao (ed.) *North American Conference on Chinese Linguistics 17*, 219-224. Los Angeles: GSIL Publications.
Li, Fangfang. 2008. *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese*. Columbus, OH: Ohio State University dissertation.
Li, Fangfang, Eunjong, Kong, Mary E. Beckman & Jan Edwards. 2008. Adult acoustics and developmental patterns for gender-marked phonetic variants in Mandarin fricatives and Japanese stops. 11th Conference in Laboratory Phonology, Wellington, New Zealand, June 30, 2008.
Li, Fangfang, Jan Edwards & Mary E. Beckman. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics* 37. 111-124.
Li, Fangfang, Benjamin Munson, Jan Edwards, Kiyiko Yoneyama & Kathleen Hall. 2011. Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. *Journal of the Acoustical Society of America* 1292. 999-1011.
Lindblom, Björn & Ian Maddieson. 1988. Phonetic universals in consonant systems. In Larry M. Hyman & Charles N. Li (eds.) *Language, speech, and mind: Studies in honor of Victoria A. Fromkin*, 62-78. New York: Routledge.
Lisker, Leigh & Arthur S. Abramson 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 203. 384-422.
Locke, John L. 1983. *Phonological acquisition and change*. New York: Academic Press.
Macken, Marlys A. & David Barton. 1980. 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.

MacNeilage, Peter F. 2013. Frame/Content Theory. This volume.
Massaro, Dominic W. & Michael M. Cohen. 1983. Categorical or continuous speech perception: A new test. *Speech Communication* 2. 15-35.
Mattock, Karen, Linda Polka, Susan Rvachew & Madelaine Krehm. 2010. The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science* 13(1). 229–243.
Ménard, Lucie, Jean-Luc Schwartz & Louis-Jean Boë. 2004. Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research* 475. 1059-1080.
Ménard, Lucie, Barbara L. Davis, Louis-Jean Boë & Johanna-Pascale Roy. 2009. Producing American English vowels during vocal tract growth: A perceptual categorization study of synthesized vowels. *Journal of Speech, Language, and Hearing Research* 525. 1268-1285.
Miller, Amanda L. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35(1). 56-84.
Monnin, Julia, Hélène Loevenbruck, Mary E. Beckman & Jan Edwards. 2011. Vowel context and frequency effects in dorsal and coronal acquisition in Drehu and French. 9th International Seminar on Speech Production. Montreal, 20-23 June.
Munson, Benjamin & Adriane L. Baylis. 2007. *Gender typicality in the speech of children with phonological disorder*. Symposium for Research on Child Language Disorders, Madison, WI.
Munson, Benjamin, Jan Edwards, Sarah Schellinger, Mary E. Beckman & Marie Meyer. 2010. Deconstructing phonetic transcription: language-specificity, covert contrast, perceptual bias, and an extraterrestrial view of *vox humana*. *Clinical Linguistics and Phonetics* 24. 245-260.
Otuka, Noboru. 2005. Articulation development and phonological perception. *Nihon University Graduate School of Social and Cultural Studies Journal* 6. 150-160. [In Japanese.]
Perry, Theodore L., Ralph N. Ohde & Daniel H. Ashmead. 2001. The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America* 109(6). 2988-2998.
Plummer, Andrew R., Mary E. Beckman, Mikhail Belkin, Eric Fosler-Lussier & Benjamin Munson. 2010. Learning speaker normalization using semisupervised manifold alignment. *InterSpeech2010* 2918-2921.
Plummer, Andrew R., Lucie Ménard, Benjamin Munson, and Mary E. Beckman. 2013. Comparing vowel category response surfaces over age-varying maximal vowel spaces within and across language communities. *InterSpeech2013* 421-425.
Plummer, Andrew R. 2014. *The acquisition of vowel normalization during early infancy: Theory and computational framework*. Columbus, OH: Ohio State University dissertation.
Redford, Melissa A. & Christina E. Gildersleeve-Neumann. 2009. The development of distinct speaking styles in preschool children. *Journal of Speech, Language, and Hearing Research* 52. 1434-1448.
Riney, Timoth James, Naoyuki Takagi, Kaori Ota & Yoko Uchida. 2007. The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics* 35. 439-443.

Roberts, Julia Lee. 1994. *Acquisition of variable rules: (-t, d) deletion and (ing) production in preschool children*. Philadelphia, PA: University of Pennsylvania dissertation.

Roberts, Julia Lee & William Labov. 1995. Learning to talk Philadelphian: Acquisition of short *a* by preschool children. *Language Variation and Change* 7. 101-112.

Rvachew, Susan, Karen Mattock, Linda Polka & Lucie Ménard. 2006. Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *Journal of the Acoustical Society of America* 1204. 2250-2259.

Rvachew, Susan, Abdulsalam Alhaidary, Karen Mattock & Linda Polka. 2008. Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French. *Journal of Phonetics* 36. 564-577.

Stoel-Gammon, Carol, Karen Williams & Eugene Buder 1994. Cross-language differences in phonological acquisition: Swedish and English /t/. *Phonetica* 51. 146-158.

Takada, Mieko. 2011. *Nihongo no goto heisaon no kenkyu: VOT no kyojiteki bunpu to tsujiteki henka* [Research on the word-initial stops of Japanese: Synchronic distribution and diachronic change in VOT]. Tokyo: Kurosio.

Toda, Martine & Kiyoshi Honda 2003. An MRI-based cross-linguistic study of sibilant fricatives. *5th International Seminar on Speech Production*.

Vihman, Marilyn May. 1993. Variable paths to early word production. *Journal of Phonetics* 21. 61-82.

Vihman, Marilyn May, Edwin Kay, Bénédicte de Boysson-Bardies, Catherine Durand & Ulla Sundberg. 1994. External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to 1-year-old children. *Developmental Psychology* 30(5). 651-662.

Vorperian, Houri K., Shubing Wang, E. Michael Schimek, Reid B. Durtschi, Ray D. Kent, Lindell R. Gentry & Moo K. Chung. 2011. Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research* 54(4). 995-1010.

Whalen, Douglas H., Andrea G. Levitt & Louis M. Goldstein 2007. VOT in the babbling of French- and English-learning infants. *Journal of Phonetics* 35. 341-352.

Wilson, Ian Lewis. 2006. *Articulatory settings of French and English monolingual and bilingual speakers*. Vancouver: University of British Columbia dissertation.

Wilson, Ian, Naoya Horiguchi & Brian Gick. 2007. Japanese articulatory setting: The tongue, lips and jaw. Ultrafest IV, New York University. https://files.nyu.edu/ld43/public/PEPLab/presentations/WilsonHoriGickUF4.pdf

Yasuda, Akiko. 1970. Articulatory skills in three-year-old children. *Studia Phonologica,* 5: 52-71. [In Japanese.]