Prosodic structure and consonant development across languages

Timothy Arbisi-Kelm[1], Mary E. Beckman[2]
[1] University of Wisconsin-Madison
[2] The Ohio State University

Abstract

This paper relates consonant development in first-language acquisition to the mastery of rhythmic structure, starting with the emergence of the "core syllable" in babbling. We first review results on very early phonetic development that suggest how a rich hierarchy of language-specific metrical structures might emerge from a universal developmental progression of basic utterance rhythms in interaction with ambient language input. We then describe salient differences in prosodic structures across the languages being studied in a cross-language investigation of phonological development, in which we are eliciting and analyzing recordings from hundreds of children aged two years through five years who are acquiring Cantonese, English, Greek, or Japanese. Finally, we present examples of how patterns of disfluent consonant production differ across children acquiring the different languages in this set, in ways that seem to be related to the differences in metrical organization across the languages.

1. Introduction

An enormous body of work over the past half century and more highlights the role of metrical structure in aligning different types of information from different parts of the grammar. Work on languages such as English, for example, supports the existence of a hierarchy of structures such as the syllable, the stress foot, and the intonational phrase, which is parsed, in part, from the role that these structural elements play in aligning intonation patterns with morphosyntactic constituents in the planning of attentional flow during an unfolding discourse (e.g., Gussenhoven, 1983; Selkirk, 1984; Welby, 2003). Just within the phonological component, lower-level metrical structures such as the syllable are evident from their role in aligning consonant gestures with vowel gestures in order to realize features that otherwise might not be audible (e.g., Mattingly, 1981; Browman & Goldstein, 2000). Moreover, studies such as Levelt and Cutler (1983) and Arbisi-Kelm (2006), among many others, provide striking evidence of the integral role also of higher-level metrical structures in facilitating the fluent production of consonants. In particular, these studies show that the most notable consonant disfluencies occur around structural positions such as the onsets of pitch-accented words, where entropy is maximal. Much of this evidence on adult disfluencies involves studies of speakers of English. However, there are a few studies of phenomena such as speech errors in other languages (e.g., Kubozono, 1989). These studies suggest that some metrical structures and/or the associated constraints on consonant alignment and production can differ from language to language. If metrical structures and/or consonant alignment and production constraints are to some extent language specific, how can they develop?

In this paper we will first clarify our assumptions about metrical structure and its role in organizing spoken discourse, and then review previous results on very early phonetic development that suggest how a rich hierarchy of language-specific metrical structures might emerge from a universal developmental progression of basic utterance rhythms in interaction with ambient language input. For example, while the basic motor rhythms of canonical babbling develop in a way that is fairly impervious to all but the most severe degradations in

input (see literature reviewed in Oller, 2000), there are recognizable differences in the "rhythmic feel" of babbling produced by infants acquiring different languages even during the first year or so of life (e.g., Whalen, Levitt, & Wang, 1991; Vihman, 1993).

After presenting our understanding of what this literature says about development, we will then review salient differences in metrical structures across the languages being studied in the παιδολογος project (http://ling.osu.edu/~edwards), a cross-language investigation of phonological development in which we are eliciting and analyzing recordings from hundreds of children aged two years through five years who are acquiring one of several rhythmically-diverse languages, including (so far) Cantonese, English, Greek, and Japanese. We will then briefly present examples of how patterns of disfluent consonant productions differ across children acquiring different languages in this set, in ways that seem to be related to the differences in metrical organization across the languages.

2. Ontology of metrical structure

The starting point for our paper is the observation that a primary function of prosody is to provide a rhythmic scaffolding that specifies designated temporal points of convergence and structural alignment among different components of the grammar. For instance, intonational phrase boundaries are points where talkers can align tones and gestures of consonants and vowels with respect to morphosyntactic clause boundaries. These alignment patterns are regular enough in read speech styles that some researchers (e.g., Selkirk, 1984; Nespor & Vogel, 1986; Gussenhoven, 1992) have hypothesized a deterministic mapping between prosodic constituency and morphosyntactic constituency. Observations of other speech styles do not support the strongest version of this hypothesis (see, e.g., Schafer, Speer, & Warren, 2004). At the same time, the alignment patterns are regular enough that listeners can use their expectations about the mapping between phonology and syntax to parse otherwise ambiguous strings, as illustrated in (1). In most lab-speech readings of the sentences in (1), the distribution of intonational phrase boundaries (marked by pause or final lengthening, as well as by the tonal cues to the ends of potentially stand-alone intonation contours) resolves the morphosyntactic ambiguity resulting from the homophony between the adjective and pronoun *her* and between the noun *pleas* and the adverb *please*. That is, the intonational phrase break following the word *pleas* in (1a) coincides with the end of a syntactic phrase, making *pleas* the object of *hear*. In (1b), by contrast, the intonational phrase boundary immediately following *her* disallows this interpretation. This kind of alignment between edges of tonally marked phrases and edges of syntactic constituents has been observed in many languages (e.g., Lehiste, 1973; Selkirk & Shen, 1990; Kang & Speer, 2003).

(1) a. [*When you hear her pleas*]$_{IP}$ [*don't respond without thinking.*]$_{IP}$
b. [*When you hear her*]$_{IP}$ [*please don't respond without thinking.*]$_{IP}$

In addition, metrical structure serves as the scaffolding for planning the pitch range relationships and the distribution of shorter-term melodic events that mark global and local discourse organization. For example, in English, pitch accents and phrase accents — the intonational morphemes that indicate local attentional foci within the discourse segment — are aligned to words and phrases by matching metrically strong positions to syntactically prominent constituents, such as lexical heads of the focus constituent, as illustrated in (2). The alignment of the pitch accent and phrase accent relative to the words in (2a) marks *the apples* as the focal constituent in the intonational phrase containing it, and suggests a discourse context in which the cook is about to start a round of pie-making and is considering different ingredients. By contrast, the distribution in (2b) makes *pies* the focal constituent of

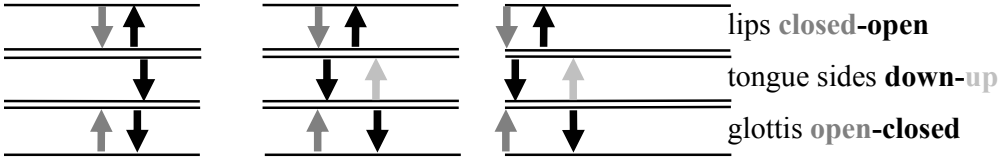
the phrase and suggests a context in which the cook has got a pile of apples and is considering various different desserts that could be made with them (see, e.g., Pierrehumbert & Hirschberg, 1990; Rooth, 1992). In Japanese, comparable differences in the relationship between the sentence and the larger discourse can be marked by differences in the intonational phrasing, the choice of boundary pitch movement, and the relative pitch ranges of successive intonational phrases (see, e.g., Venditti, Maekawa, & Beckman, 2008).

(2) a. Please only make pies with [the apples.]$_{FOC}$
[[σ]$_F$ [σ$_s$ σ$_w$]$_F$]$_{IP}$
L+H* L- L%

b. Please only make [pies]$_{FOC}$ with the apples.
[[σ]$_F$ [σ$_s$ σ$_w$]$_F$]$_{IP}$
L+H* L- L%

Another important point about metrical structure is that, just within the phonological component, we find evidence of alignment between different structural elements. In cases such as (2), for example, the foot structures that are parsed for monosyllabic *pies* versus trisyllabic *the apples* govern the alignment of the pitch accent and following phrase accent relative to the string of consonants and vowels in these words. Even more locally, consonant gestures coordinate with vowel postures, and differences in the coordination patterns reveal their alignment to larger metrical structures. For instance, the production of consonant sequences will vary depending on metrical position, as schematized in (3).

(3) [ʔ[æ]$_V$]$_σ$ [p [l]$_V$]$_σ$]$_F$]$_ω$ [ə]$_V$]$_σ$[p l [ɔ]$_V$ z]$_σ$]$_F$]$_ω$ [[p l[i:]$_V$ z]$_σ$]$_F$]$_ω$



When the sequence /pl/ is produced in foot-medial position (as in *apple*) the glottal opening gesture and lip closure for the /p/ are aligned to overlap considerably less with the lingual posture of the /l/ as compared to the degree of overlap for the cluster in foot-initial position (as in *applause*) and word-initial position (as in *pleas*). Evidence for this understanding of the role of metrical structure in facilitating consonant production comes from many sources, including studies of articulator kinematics in fluent read speech, such as Browman and Goldstein (1988), Krakow (1993; 1999), Byrd (1996), and Loevenbruck, et al., (1999). Some of the studies in this literature have shown layered effects of higher-level prosodic environment on consonant gestures at several levels, with longer constriction durations at intonational phrase edges compared to those at phrase-internal word edges as well as differences for different positions within a foot (Fougeron & Keating, 1997; Cho & Keating, 2001; Byrd, et al., 2005; Bombien, et al., 2006; Cho, 2006).

The literature on speech errors provides a second source of valuable information regarding the role of metrical structure in consonant production. One key observation in this literature is that speech errors occur at different rates for consonants in different metrical positions (see, e.g., Dell, 1985; Shattuck-Hufnagel, 1987; Levelt, 1989). For example, English errors often move or exchange consonants at foot beginnings, as shown in the example in (4), from Fromkin (1973). This type of error almost never occurs in other

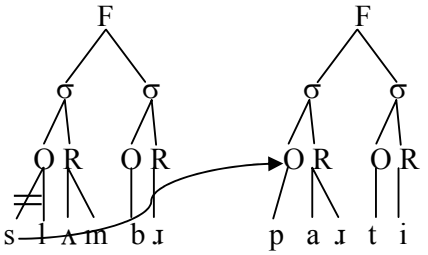
positions. That is, there are far fewer coda exchanges, and there are almost no exchanges between syllabic consonants and vowels (see, e.g., Fromkin, 1971, 1973; Shattuck-Hufnagel, 1979; MacNeilage, 1998). This asymmetry between onset and coda positions is often interpreted as evidence for differences in metrical affiliation at the level of the syllable, as schematized in (5) for two competing accounts of English syllable-internal organization.

(4) slumber party → lumber sparty

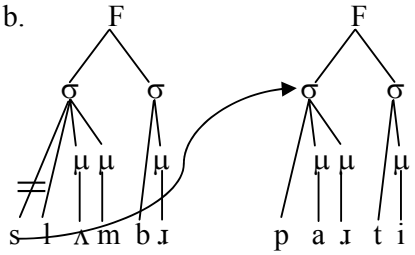
[[s l [ʌ]$_V$ m]$_σ$ [b [ɹ]$_V$]$_σ$]$_F$ [[p [a]$_V$ ɹ]$_σ$ [t [i]$_V$]$_σ$]$_F$ →

[[l [ʌ]$_V$ m]$_σ$ [b [ɹ]$_V$]$_σ$]$_F$ **[[sp** [a]$_V$ ɹ]$_σ$ [t [i]$_V$]$_σ$]$_F$

(5) a.



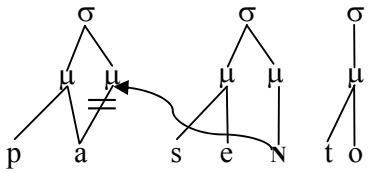
b.



Evidence that this asymmetry is an effect of metrical position per se (rather than a simpler sequential constraint on available cues in pre-vocalic versus post-vocalic position, as suggested implicitly by Steriade, 1999) comes from cross-language comparison. Specifically, in Japanese, errors can target gestures in analogous mora positions in different syllables, regardless of whether the gestures are vocalic or consonantal, as shown by the example in (6). Kubozono (1989) uses this difference in error patterns between the two languages to argue for a difference in syllable-internal metrical organization, as schematized in the tree diagram in (7), as contrasted to (5a), which is his account of English.

(6) paasento → pansento

(7)



Regardless of the structure that is posited, however, there is a notable fact about speech errors such as (4) and (6). When consonants move from one syllable to another in planning the utterance of a phrase, they are fluently aligned with the gestures of segments that are adjacent in the new environment, to make for the appropriate allophony. The glottal opening gesture of the /s/ in *lumber sparty* in (4), for example, merges with the glottal opening gesture of the following labial stop so that the /p/ is now unaspirated, as appropriate for its no-longer foot-initial position. The /l/ that is left behind, conversely, is now fully voiced, since its lingual posture no longer is aligned relative to the glottal opening gesture of a preceding foot-initial voiceless obstruent. And analogous fluent positionally appropriate allophonic variation characterizes the pronunciation of the segments around the intrusive moraic nasal in the Japanese example in (6), as well. In particular, the velic opening gesture for the displaced moraic nasal overlaps heavily with the lingual posture for the preceding /a/ to make for a heavily nasalized vowel. Also, the nasality cannot extend into the following /s/ constriction, so that the transcription of [n] for the moraic nasal is much less appropriate here than it is in the second syllable of this word. We interpret these observations about allophonic

appropriateness as evidence for the integral role of metrical structure in facilitating the fluent production of consonant gestures, particularly of gestures such as the labial release and glottal opening gestures of English /p/ which cannot even be heard unless they are timed so as to be co-produced appropriately with the gestures of other segments around a common metrical alignment point.

A third source of evidence for the integral role of metrical structure in consonant production comes from studies of disfluencies in the speech of stutterers. Following Levelt (1983), Shriberg (1999), and others, we understand disfluencies as arising from a speaker's detection, and attempted correction, of an error in language production. Different types of errors suggest errors at different stages of planning. In the speech of stutterers, the distribution of disfluencies suggests a breakdown at the stage of assembling an articulatory plan, particularly in the planning of gestures around those metrical positions that present the most challenging constraints on alignment. For example, in English-speaking stutterers, disfluencies occur much more frequently in word-initial position (Brown, 1938, 1945; Hahn, 1942; Soderberg, 1962; Taylor, 1966; Weiner, 1984; Hubbard, 1998; Natke, Grosser, Sandrieser, & Kalveram, 2002), and in stressed words more often than in unstressed words (Brown, 1938; Bergmann, 1986; Weiner, 1984; Wingate, 1988; Prins, et al., 1991; Natke, Grosser, Sandrieser, & Kalveram, 2002). Moreover, there are also metrically-conditioned differences between different types of stressed words. Specifically, there are more disfluencies on stressed words that are aligned to pitch accents in the intonation contour of the utterance (Arbisi-Kelm, 2006).

We attribute this last result to the exigencies of accent production in English. Consonant production in pitch-accented words is challenging for two reasons. The first is the set of linguistic conventions identified by researchers such as de Jong (1995) about the carefully "hyperarticulated" quality of the consonants and vowels associated to pitch-accented syllables in languages such as English. Even if there were no such conventions, however, we would expect accented words to be a locus of disfluency, because of the added burden of aligning the laryngeal and oral gestures for the consonants and vowels of the word together with the laryngeal and respiratory gestures for producing the tone pattern within the backdrop pitch range specified at that point in the discourse.

In summary, converging evidence from fluent and disfluent speech suggests to us that consonant production depends on a facility to coordinate gestures at multiple time scales, from the rapid sequence of raising and lowering movements that flick the tongue tip against the alveolar ridge to make the 20-30 ms closure for an alveolar tap to the gestures of the respiratory and laryngeal system that specify the pitch range over stretches of speech that can extend for 5-10 seconds. Our understanding of the hierarchy of metrical structures for any given language is that these structures provide a conventionalized schema for organizing the planning of speech production in real time. Like Ferreira (1993) and Keating & Shattuck-Hufnagel (2002), that is, we interpret speech error data as telling us that metrical structures for larger prosodic constituents, such as accented positions within intonation phrases, are assembled relatively early in the production process, before word-specific gestural ensembles are retrieved and aligned relative to the rhythmic frames specified for the larger constituents. Moreover, we understand from the comparison of intonation systems and word-level prosodic templates across languages that the metrical hierarchy is highly conventionalized and language-specific.

This language specificity raises an important question: how exactly can the higher-level metrical structures be acquired so that the child can begin to produce the consonants that distinguish words of the ambient language? In the next section, we suggest one route by which language-specific metrical organization can emerge from infants' developing motor

control over their respiratory and laryngeal systems, and over movements of their lips, tongue, and velum, in interaction with the voices and faces of speakers in the infants' environment.

3. Ontogeny of metrical structure

Evidence for the emergence of rudimentary metrical organization can be found very early in speech development, during the pre-babbling and canonical babbling stages as identified by researchers such as Oller (1980; 1986; 2000), Stark (1980), and Koopmans-van Beinum & van der Stelt (1986). In the "phonation stage" (using Oller's terms), infants practice their control of the most transparently autonomous articulatory gestures, exploring the motor space for different patterns of fundamental frequency or laryngeal source quality in conjunction with different vowel-like resonances. From around 2-3 months of age, during the "gooing stage", children advance to articulations of consonant-like sounds, including vaguely [k]-like releases that can be produced with a raised jaw and tongue-filled oral tract. At these early stages, the dominant audible rhythm is not internal to the infant's vocalization. Rather, it is the alternation of imitative turn-taking between the infant and the mother that can ensue if the mother responds to the baby's coos with contingent imitation of the baby's phonatory gesture and/or the baby's resonance gesture (see, e.g., Papoušek & Papoušek, 1989; Masataka, 1993, 2003). Papoušek, Papoušek, & Symmes (1991) even propose that there are cross-cultural commonalities in the melodic shapes that mothers use to engage an infant's attention and invite a bout of vocal turn-taking at this age. They cite in support of this proposal their results showing that Mandarin-Chinese speaking mothers will suppress lexical content of their utterances to their infants in order to enable the expression of these tunes.

A bit later, starting at about 6 months, the rudimentary consonant-like releases of the "goo" stage become coupled to more rhythmically consistent mandibular oscillation patterns. At this stage of "canonical babble", parents in homes where the babies are acquiring a spoken language are much more likely to describe their children's vocalizations in terms of the consonants and vowels of the ambient language, as in Darwin's (1877: 292) description of one infant's progression from the earlier stages into the canonical babbling stage:

> At 46 days old, he first made little noises without any meaning to please himself, and these soon became varied. An incipient laugh was observed on the 113th day, but much earlier in another infant. At this date I thought, as already remarked, that he began to try to imitate sounds, as he certainly did at a considerably later period. When five and a half months old, he uttered an articulate sound "da" but without any meaning attached to it.

Researchers have long noted that the onset of this canonical babble (or "reduplicative babble") comes when the baby is maximally engaged in a regular rhythmic exploration in general, waving hands and feet, shaking rattles, and so on (e.g., Thelen, 1979, 1991; see review in Ejiiri & Masataka, 2001). While the timing of the mandibular cycle makes the infant's vocalizations at this stage sound very much like consonant-vowel alternations, there are strong constraints on what consonants can combine with what vowels. These constraints suggest that the infant is still controlling tongue posture only at the whole-utterance level, in what Davis and MacNeilage (1995) refer to as "frame dominance" — a precursor to the "variegated babble" seen in many infants, when shorter term rhythms allow the infant to begin to control sequences of labial and lingual postures on a "syllable-by-syllable" basis.

Summarizing this work, we can say that observations of these stages of pre-linguistic babbling strongly support the idea that the intonation phrase and the core syllable are universal units of spoken language simply because they harness rhythmic structures that

emerge universally in normal development. The intonation phrase (or "breath group" as Lieberman, 1967, calls it) emerges as a very young infant explores the auditory consequences of coordinating the respiratory cycle with oral gestures for sustained egressive phonation. The core syllable, similarly, emerges as the somewhat older infant explores the auditory consequences of coordinating a basic mandibular oscillation with lingual and labial constrictions for a fluent sequence of consonants and vowels.

While these structures have a universal ontogenetic basis, however, the metrical structures that eventuate do differ across languages. The tone shapes that mark the "breath group" are specific to each language variety, and so are the dominant syllable structures. An obvious question that arises, then, is the following. How and when are these universal rhythms tuned to become the metrical structures specific to the phonological grammar of the ambient speech community? We believe that the universal rhythms are entrained to ambient language structures very early, because of the critical role of auditory feedback in normal development of motor control for speech. As Kent (1984: R890) asserts in his seminal paper on the "psychobiology of speech development":

> *Production and perception capabilities that ultimately lead to speech are initially largely separate, but they begin to be coordinated (integrated) within the first few months of life.* The integration of the two systems also interacts with the child's linguistic background, such that the child's exposure to the sounds around him/her eventually influences the child's own pattern of vocalization. [emphasis in original]

Much research on early infant vocalization highlights the importance of auditory feedback. For example, Langlois, Baken, and Wilder (1980) attributed the onset of a sustained exhalation phase of the respiratory cycle (and consequently longer vocalizations) entirely to anatomical factors — especially to the growth of the rib cage. However, the crucial role of audition becomes apparent when examining how utterance duration changes when there is little or no auditory feedback. Clement, Koopmans-van Beinum, and Pols (1996) found that while both typically-developing and hearing-impaired children showed the expected increase in mean utterance duration around 3-4 months, the hearing-impaired children showed a significantly smaller increase. The authors ascribed this difference to a lack of auditory feedback, which is consistent with Lieberman's (1986) suggestion that insufficient laryngeal muscle exercise during this period could result in reduced ability to manipulate sub-glottal air pressure. In other words, normal development of the universal basis of the "breath group" depends on auditory feedback at 3-4 months.

Analogous effects of auditory experience are also observed for the later development of canonical babbling rhythms. For example, in a study of two monozygotic twins — one profoundly hearing-impaired and the other with normal hearing — Kent, Osberger, Netsell, and Goldschmidt Hustedde (1987) found key production differences at 8, 12, and 15 months. In addition to producing a smaller range of vowel formant frequencies at each of the three recording sessions, the hearing-impaired twin showed an overall later onset of canonical babbling. A larger study by Oller and Eilers (1988) of 21 infants with normal hearing and 9 deaf infants replicates this effect of input. All of the infants with normal hearing began canonical babbling between 6-10 months of age, while none of the deaf infants began this stage before 11 months.

Given this critical role of auditory input for the onset of canonical babbling, we might expect to see entrainment to ambient language speech patterns before the onset of language. And, indeed, influence of language-specific autosegmental content on infants' vocalization patterns is found as early as 10 months. In their study of 10-month-old infants growing up in Arabic-, Cantonese-, English-, or French-speaking homes, de Boysson-Bardies, Hallé, Sagart,

and Durand (1984) found that the distribution of formant frequencies measured in vowel-like intervals in each infant's canonical babbling reflected the frequencies of different vowels in the lexicon of the ambient language. Transcribed consonant place frequencies in babbling reflect a similarly early influence of adult language input. In their cross-linguistic study of infant vocalization patterns, de Boysson-Bardies and Vihman (1991) found that 9-10-month-old infants growing up in English-, French-, Japanese-, or Swedish-speaking homes produced different proportions of labial consonants relative to lingual consonants, in keeping with the different distributions of labials in the adult lexicons of the four ambient languages.

Analogous results have been found for the time course of language-specific effects in the prosodic domain, as well. In a longitudinal study of babies growing up in French- or English-speaking homes, Levitt and Wang (1991) found no cross-language differences in "syllable" durations during the pre-canonical stages at 4-6 months. During the later reduplicative babbling stage, however, the English-acquiring children produced significantly shorter final syllables than did the French-acquiring infants, reflecting the predominantly trochaic pattern of English words, as opposed to the predominantly iambic rhythms of French. A companion study of the fundamental frequency patterns produced by these infants lends further support to this interpretation of the duration patterns in terms of the language-specific rhythmic patterns. Specifically, Whalen, Levitt, and Wang (1991) report consistent differences in the pitch contours for two- and three-syllable reduplicative babbling trains, such that the French-acquiring babies produced a mix of both sequence-final rises and sequence-final falls, while the English-acquiring infants produced almost exclusively falling melodic contours.

Such cross-language prosodic differences in reduplicative babbling are important because canonical babbling provides the basic "vocal motor schemes" of the first words (Vihman et al., 1985; Davis, et al., 2000; McCune & Vihman, 2001). That is, the ability of infants to integrate the longer-term coordination of subglottal pressure modulation and laryngeal control for the global and local pitch patterns of intonation phrases together with the shorter-term control of jaw, tongue, and lip movements for vowels and consonants is prerequisite to the later fluent production of one-word utterances. Therefore, the fact that there are measurable effects of ambient language prosody on the tunes and rhythms of reduplicative babbling months before the infant makes the connection to meaning strongly supports the primacy of prosodic structure in constraining the development of motor control for later word production. Therefore, we are not surprised to find that the first recognizable words that young toddlers produce also reflect the prosodic structures of the languages that they are acquiring, even as they reflect more universal constraints on the motoric complexity of early utterances. For example, Vihman, DePaolis, and Davis (1998) note differences in the prosodic structures in early multi-syllabic word productions by French- and English-acquiring toddlers at 13-20 months. Mirroring the stress patterns of the ambient language, the English-acquiring babies produced only recognizable trochees, while the French-speaking babies produced recognizable iambs. The difference was especially striking in cases where a child truncated a long word or re-configured a word's underlying syllabic structure to match the desired stress pattern, as in the French-acquiring baby's production of [pɑˈpji] for the adult target *papillon* 'butterfly' or the English-acquiring baby's production of [ˈwɑnə] for the adult target *around*. While older toddlers tend not to show such dramatic prosodic reorganization, we might still expect to see constraints from the prosodic organizations that the child has mastered interacting with and affecting the child's productions of one-word utterances. In section 5, we will look in greater detail at several examples of disfluency patterns produced by young children acquiring English or one of three other languages, particularly with respect to how these error patterns are shaped and constrained by both language-specific and language-general aspects of metrical organization. In the next section, we describe the larger database of productions from which these examples are drawn.

4. The παιδολογος project — cross-linguistic research on phonological acquisition

The examples that we will discuss were recorded as part of a larger project in which we are comparing accuracy rates for productions of word-initial lingual obstruents in a range of following vowel contexts across a variety of languages. For each language, the subjects we record are twenty young adults (aged between 18 and 30 years) and 100 children aged from 2 years through 5 years. The target consonants are elicited using a picture-name repetition task that allows us to record them in real words that are not likely to be familiar to the youngest children, as well as in the highly familiar picturable real words used in most word-naming tasks. In this way, we can sample the consonants evenly across following vowel environments even when one coarticulatory context is relatively rare. The task even lets us elicit the target consonants in nonsense words so that we can compare the children's accuracy rates across pairs of languages in all coarticulatory conditions of interest, whether the consonant vowel context is phonotactically licit in both languages or just in one.

Although we have recently begun to record productions by children acquiring Korean, Mandarin Chinese, and French, the first languages that we chose to study were Cantonese, English, Greek, and Japanese. These are the languages for which we have the most complete analyses to date, and therefore they are the languages from which we will draw all of our examples for this paper. These initial four languages (and the subsequent three) were chosen because we have access to children acquiring them and because each has a rich inventory of lingual obstruents which can be compared to identically transcribed consonants with different contrastive properties, different allophonic patterns, or different phonotactic frequencies in one or more of the other languages. For example, all of the languages have at least one sibilant fricative that is transcribed as /s/, and in all of the languages but Cantonese and Korean, this alveolar fricative contrasts with at least one other voiceless lingual fricative at a different place of articulation. In Greek and English, /s/ contrasts with a more anterior dental /θ/ as well as with a more posterior fricative, which is the dorso-palatal /ç/ in Greek and the coronal post-alveolar /ʃ/ in English, whereas in Japanese /s/ contrasts only with two more posterior fricatives – an alveolopalatal /ɕ/ as well as the dorso-palatal /ç/. Moreover, in English, both /s/ and /ʃ/ occur readily in all following vowel environments, whereas in Japanese, /s/ does not occur before /i/ and /ɕ/ is only marginally attested before /e/. All the languages also have voiceless dental or alveolar and "plain" dorsal stops that can be transcribed as /t/ and /k/, as well as at least one other voiceless dorsal stop with a contrastive "secondary" labial or palatal articulation.

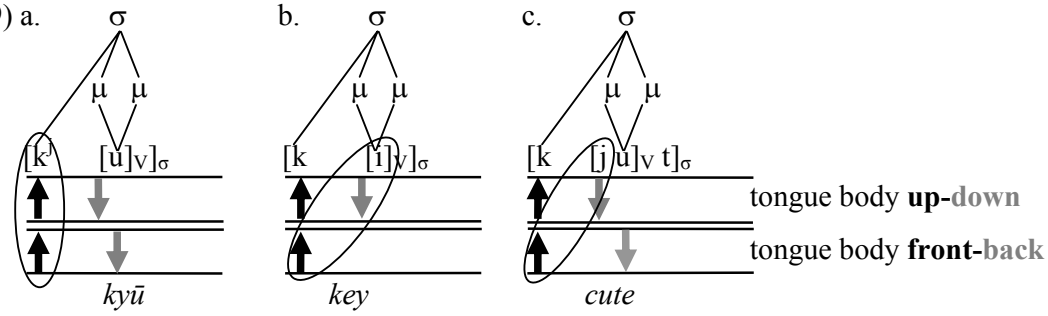
Our initial analysis is a categorical judgment (by a native speaker of that language who is a trained phonetician) of the accuracy of each target word-initial consonant and of the following vowel. That is, the transcriber opens the audio file for a recording session in Praat, along with a TextGrid file in which the interval for each word has been marked off and tagged with an initial broad phonemic transcription. The transcriber zooms in on each target word in turn, listens to the word production, and compares the word-initial consonant and the following vowel to their target transcriptions. If a sound is judged to be correct, the phonemic transcription is accepted. Otherwise, the transcriber analyzes the error to provide one of the types of tags listed in (8), the first three of which are types of "substitution" that require a transcription of the substituted sound. Both the automatically provided initial broad phonemic transcription of the target sounds and the transcribers' analyses of the substitution errors use the WorldBet ASCII encoding of the IPA devised by Hieronymous (1994), and both adhere to conventions that sometimes are at odds with the usual phonemic analysis for the language. For example, the Greek front dorsal stop that occurs before /i/ and /e/ and that

contrasts with "plain" /k/ before the other three vowels of the language in words such as /ˈkʲoskʲi/ 'kiosk' in (8) is transcribed as a palatalized stop rather than as a palatal stop, as in the usual phonemicization of Greek provided by Arvaniti (1999) and others. The English initial stop in words such as *cute* and *key* also is transcribed as a palatalized velar, and the English initial stop in words such as *quake* is transcribed as a labialized velar.

(8) `$` plus a broad transcription, for a within-inventory substitution error
e.g., `$ts` for a /ts/ for /kʲ/ substitution for Greek /ˈkʲoskʲi/ 'kiosk'
`+` plus a narrow transcription, if the substitution goes outside the language's inventory
e.g., `+tc}` for a [tɕ] for /kʲ/ substitution for Greek /ˈkʲoskʲi/ 'kiosk'
`:` separating some pair of IPA tags
e.g., `+tc}:$ts` for a production intermediate between the above two
`deletion` for a segment that is simply not pronounced
`distortion` for an error that cannot be captured by any IPA symbol

This transcription of labialized or palatalized stops bears more explanation. By adopting these transcription conventions, we are equating the configuration of lingual and labial gestures at the beginning of English *quake* with the configuration of lingual and labial gestures at the beginning of Cantonese /kʷa:^{55}/ 'melon' and we are equating the configuration of tongue body gestures at the beginning of Greek /ˈkʲoskʲi/ 'kiosk' and of English *key* and *cute* with the configuration of tongue body gestures at the beginning of Japanese *kippu* /kʲip:u/ 'ticket' and *kyū* /kʲu:/ 'nine'. That is, for the sake of cross-language comparison, we are analyzing the CV sequence in words such as *cute* as having the prosodic structure and alignment constraints in (9a). In this analysis, the front place of constriction of the dorsal stop in English *cute* is parsed as a property inherent to the consonant, just as in the contrastively palatalized dorsal stop in Japanese *kyū*. A more traditional analysis is shown in (9c). Here, the tongue fronting gesture is parsed as a feature of the first target in a following diphthong, which is co-produced with the consonant, yielding the "front allophone of /k/" that is also seen in words such as *key*, as in (9b). The schemata in (9) are intended to highlight the idea that the difference between English *cute* and Japanese *kyū* is a prosodic difference rather than a difference in intrinsic autosegmental content. Comparing these "palatalized velars" across the two languages gives us a way of evaluating the consequences of such cross-language differences in prosodic organization at the level of the syllable.

(9) a. b. c.



The languages examined also have other, more obviously prosodic differences involving structures above the syllable. For instance, Greek and English both have intonational morphemes (pitch accents) that are constrained to anchor to syllables that are rhythmically prominent (i.e., have "lexical stress"). However, the word-level rhythms differ. English

words are predominantly one or two syllables long, and words longer than one syllable show a trochaic bias, grouping alternating strong and weak syllables into bimoraic feet (Hayes 1980; Halle & Vergnaud 1987; Kager 1989). As a consequence, the most typical word shapes are a stressed syllable alone or a disyllable with initial stress, so that the consonant marks not just the beginning of the intonation phrase (in the citation form utterances we elicited) but also the beginning of a stress foot, with all of the concomitant constraints on precise alignment to coordinate the constriction gesture for the consonant with the lingual and labial postures specified for the following full (and typically pitch-accented) vowel.

This contrasts with Greek, which has almost no monosyllabic words, and which has a trisyllabic stress window aligned to the end of the word, with no phonetic evidence for stress alternations elsewhere (see Joseph & Philippaki-Warburton, 1987, and other references cited in Arvaniti & Baltazani, 2005). Many more words of Greek, therefore, have unstressed initial syllables. Furthermore, while in English only /ə/ and its variants are unstressed and subject to deletion, any vowel in Greek may be unstressed, and both /i/ and /u/ may be lenited in unstressed environments (Arvaniti, 1994). This is illustrated in Figure 1. The word /ˈçiʲlʲa/ 'lips' in the spectrogram on the left has a stressed initial syllable, with a fully realized vowel /i/, whereas the word /çiˈmonas/ 'winter' in the spectrogram on the right has stress on the second syllable, and the unstressed vowel /i/ is significantly reduced.

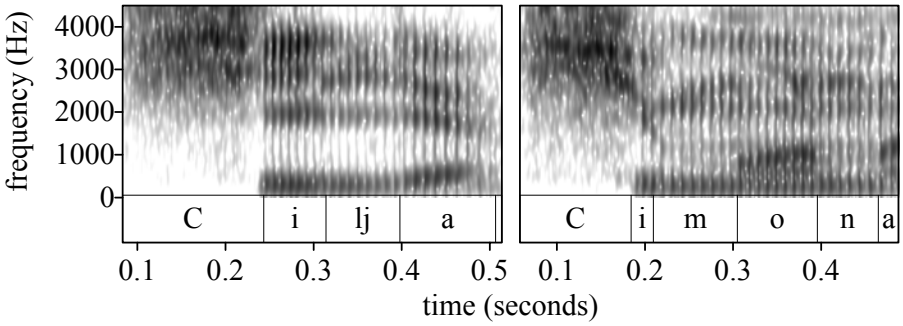


Figure 1. Spectrograms and WorldBet transcriptions of (the first 430 ms of) a Greek adult speaker's productions of a word beginning with stressed initial target /çi/ (left) versus one beginning with an unstressed and lenited initial target /çi/ (right).

The consonantal contexts that favor this lenition of high vowels in Greek are very similar to those that promote "high vowel devoicing" in Tokyo Japanese, and the acoustic patterns that result are similar. The Japanese phenomenon differs prosodically, however, because Japanese does not have anything like stress in Greek. There are localized pitch events that Greek- and English-speaking second language learners of Japanese assimilate to the "pitch accents" in their native intonation systems. However, the pitch accents of Japanese are simply lexically specified tone patterns that are not associated to prominent syllables and which interact only very indirectly with the prosodic mechanisms for focus-marking (see Venditti, Maekawa, & Beckman, 2008, for a review). It is not at all uncommon for high vowels in lexically accented syllables to be deleted, as is clearly evident in the numbers reported for the Corpus of Spontaneous Japanese by Maekawa and Kikuchi (2005) as well as in many earlier studies of lab speech. Moreover, there is a rising initial boundary pitch movement covering the first one or two moras of every well-formed accentual phrase, and hence of every utterance, whether or not there is a lexical accent anywhere in the utterance. Therefore, the Japanese word-initial consonants in our database are not in a position of relaxed prosodic alignment constraints even in syllables with devoiced or deleted vowels.

Finally, Cantonese also has segmental lenition and even whole vowel deletion, in a phenomenon called "syllable fusion" (Wong, 2004, 2006). This reduction phenomenon is not constrained by word-level prosodic prominence, since Cantonese does not have anything like the contrast between stressed and unstressed syllables of English and Greek. This description may make Cantonese syllable fusion seem like Japanese vowel devoicing. However, the two phenomena are quite different, as are the two prosodic systems more generally. Every syllable of Cantonese is prosodically strong in the sense that every syllable bears a full lexical tone. The lack of tonally unspecified syllables differentiates Cantonese not just from Japanese but also from most varieties of Mandarin Chinese as well as from Wu dialects such as Shanghainese. In these other Chinese varieties, prosodic words typically are at least two syllables, and high vowel deletion like that seen in Greek occurs on weak syllables that have "neutral tone" — i.e., syllables that either are intrinsically not specified for tone, as in the second syllables of Mandarin *dōufu* 'tofu' and *yìsi* 'meaning', or that "lose" their tone specification in the word-formation process, as in the middle syllable of *yào bu yao* 'want?' (examples from Peng et al., 2005). Cantonese differs from these other varieties of Chinese in having far fewer disyllabic (and longer) words as well as in not having neutral tone syllables. Syllable fusion is much more strictly a post-lexical process, but even the most extreme cases of reduction and vowel deletion typically occur without tone loss. That is, even in cases where the medial consonants have been completely deleted and the two vowels merged into a single sandhi form, both syllables' lexical tone specifications are preserved, so that the syllable count is effectively unchanged (see examples and description in Wong, Chan, & Beckman, 2005).

Our original motivation for eliciting several productions of each word-initial lingual obstruent in each of several vocalic contexts was to be able to compare accuracy rates across the different languages. By comparing accuracy rates for ostensibly the same CV sequence across languages that have different phonotactics or different frequencies for the sequence, we can begin to tease apart the effects of language-specific phonotactic probabilities from any language-universal effects of the intrinsic difficulty of producing that particular constellation of consonant and vowel gestures (see Edwards & Beckman, 2008). Given the differences in prosodic structure outlined in this section, we wonder whether the error patterns might also be informative. Even if children acquiring different languages make the same number of errors for a particular CV sequence, might they show differences in the types of errors that they make if the same autosegmental content is parsed differently by the prosodic organization of the language? For example, given the different prosodic analyses in (9), we might expect Japanese-acquiring children and English-acquiring children to latch onto different strategies for producing this difficult assemblage of lingual gestures. Moreover, even when there are no obvious differences at the level of the syllable, different demands for more or less precise coordination of consonant, vowel, and tone might lead to different patterns of disfluency. In order to be able to explore these possibilities, we encourage the transcribers to note examples of different types of disfluent productions on a "notes" tier, and we are developing consensus analyses and conventions for tagging recurring types. We have culled interesting examples of these recurring types which we present in the next section.

5. Disfluencies and deletions in the παιδολογος recordings

One of the disfluency tags on the "notes" tier is the tag "E" for an initially incorrect or very "effortful" production of a target fricative or affricate that eventually homes in on an acceptable target constriction. This is the disfluency type that seems most similar to the stereotype of adult stuttering. In the example in Figure 2, an English-speaking child produces

an "E"-tagged disfluency on the initial /s/ in the trisyllabic nonword /ˈsɛvɨˌfɹæʃ/. The "effortful" nature of the production is clearly evident in the acoustic pattern of an initially less strident interval followed by a momentary dip in amplitude as the child apparently repositions the tongue body or tongue blade to better direct the airstream to hit the edge of the incisors downstream of the constriction.

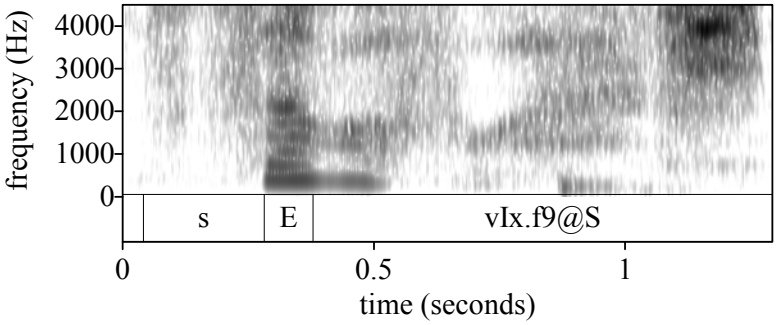


Figure 2. An "E" tagged disfluency of initial /s/ in the English non-word /ˈsɛvɨˌfɹæʃ/.

We originally devised the "E" label so that transcribers could analyze cases such as the /s/ in Figure 2 as different from fluent productions of target fricatives and affricates, while still recognizing that the child eventually achieves some configuration of lingual and laryngeal postures that is recognized as the target strident sound. The "E" tag, therefore, is reserved for target fricatives and affricates, where visible (and audible) changes in amplitude and spectral shape over the course of the turbulent interval can reveal the child's ongoing struggle with the difficult aerodynamic requirements of these sounds.

Figure 3 is an example of a different analysis category — the "split-CV" — which we now think reflects a similarly effortful struggle to coordinate the lingual and laryngeal gestures for a target stop with the tongue posture for the following vowel. In contrast to the rapid succession of more or less fluent stop releases that define the stereotypical English-speaking adult stuttering pattern, the children's disfluent stop productions seem much more effortful and less organized rhythmically. In Figure 3, for example, the child successfully releases the dorsal stop closure into the lingual posture for the target /ɑ/, but fails to initiate voicing. There is a nearly 300-ms period of "aspiration" which would be interpreted, if this were Japanese, as a fluent devoiced vowel. The child then overshoots the laryngeal adduction target for voicing, so that there is a 180-ms pause before another short interval of "aspiration" and then finally the full stressed target vowel. This creates the percept of an epenthetic vowel followed by a glottal stop, with a correspondingly increased syllable count.

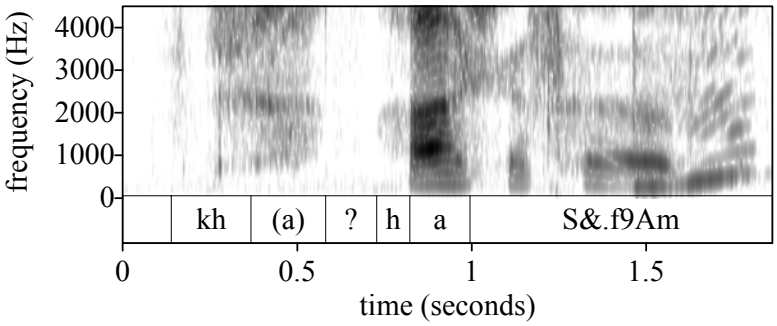


Figure 3. A "split-CV" production of /kʰa/ in the English non-word /ˈkʰɑʃəˌfɹɑm/.

In English, such "split-CV" cases occur especially frequently when the child attempts the especially demanding coordination involved in sequencing the lingual and laryngeal gestures for a fluent aspirated palatalized velar stop before the labial and lingual gestures for a back vowel, either in the legal sequence /k^{jh}u/ — traditionally analyzed as /k^{h}/ followed by /ju/ as in (9c) — or in the "illegal" sequence */k^{jh}o/ which we could elicit only in nonwords. Figure 4 shows such a case. There is the same apparent epenthetic syllable as in Figure 3, and also a perceived fronting substitution of /t^{h}/ for the target dorsal.

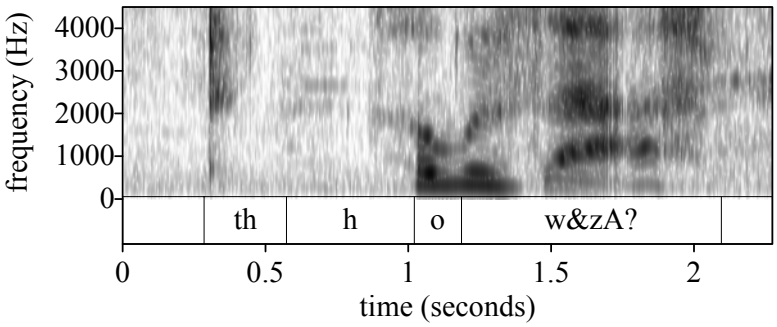


Figure 4. A "split-CV" production of */k^{jh}o/ in the English non-word /ˈk^{jh}oˌzɑm/.

The error patterns that Japanese-acquiring children show for the analogous sequences in their language show two interesting differences from the English error patterns in Figure 4, both of which are illustrated in Figure 5. First, the more common place error corresponding to the fronting to [t^{h}] in Figure 4 is the substitution of an alveolopalatal affricate, as in the production shown in the left panel of Figure 5. Tsurutani (2004) also notes this substitution pattern, which mirrors a frequent substitution pattern for the palatalized velar in the less challenging sequence /k^{j}i/ in studies such as Nakanishi, Owada, and Fujita (1972). Second, in cases that are analogous to the "split-CV" aspect of the example in Figure 4, the mistiming does not involve the laryngeal gesture. Instead, the palatal place-of-articulation gesture is extended into the voiced vocalic interval, creating the percept of an [i] vowel before the transition to the unrounded back vowel, as in the production shown in the right panel of Figure 5. Although this mistiming increases the perceived syllable count, the prosodic reorganization does not sound nearly as disfluent as the voiceless vowel followed by glottal stop that the English-native-speaker transcriber tagged with the "split-CV" label in Figure 4. These differences between the two languages suggest that it is partly the added requirement of controlling the alignment of the laryngeal gesture for the English aspiration contrast in word-initial stressed syllables that makes the English sequence more challenging.

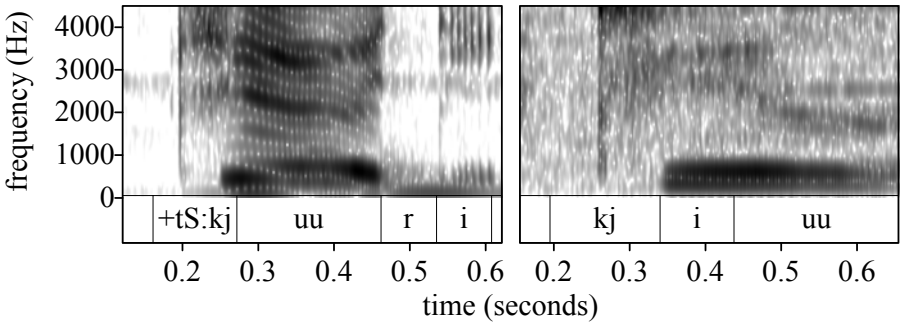


Figure 5. Examples of two different error patterns for the target sequence [k^{j}uː] in the Japanese words *kyūri* 'cucumber' (left) and *kyū* 'nine' (right).

Further support for this suggestion comes from Cantonese disfluencies that are analogous to the English cases in Figures 3 and 4. Figure 6 gives an especially illuminating example. The child backs the initial /t/ to [k] and then seems to hesitate momentarily before initiating voicing half-way through the second target in the following diphthong. Although there is a 130-ms interval where /a/-like formants are excited by soft [h]-like turbulence, followed by a 200-ms interval where /u/-like formants are excited by even softer turbulence, the Cantonese transcriber does not perceive this as an aspiration error. Instead, she tags it as a deletion of the first (short /a/) target in the diphthong. That is, since there is voicing to carry the tone only during (the second half of) the interval where the tongue body and lips are postured for /u/, this is the only part that can be counted as a vowel. Or, to put it another way, because Cantonese, unlike English, does not have unstressed syllables that license vowel devoicing, the hesitation after the release of the stop cannot be perceived in terms of a prosodic reorganization that increases the syllable count.

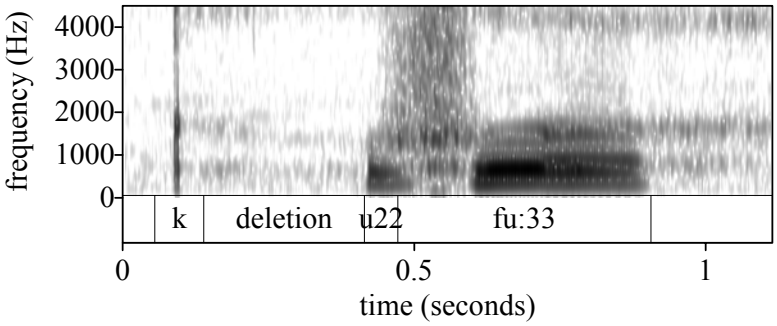


Figure 6. Partial vowel "deletion" in a disfluent production of Cantonese /tau^{22}fu:^{33}/ 'tofu'. Tone is preserved and realized on the later voiced portion of the vowel.

Such differences in interpretation of ostensibly identical misalignment patterns across languages highlights the close dependency between fluent production of segments and mastery of the language-specific prosodic constraints on gestural coordination. They also raise an important question about the act of transcription. As suggested earlier in our discussion of Japanese *kyū* and English *cute* in (9), consonant and vowel gestures do not carry their prosodic affiliations on their sleeves. Assigning a particular segmental transcription to a child's production implies a particular prosodic analysis, and different transcribers can disagree on the prosodic analysis. The production in Figure 7 is a case in point.

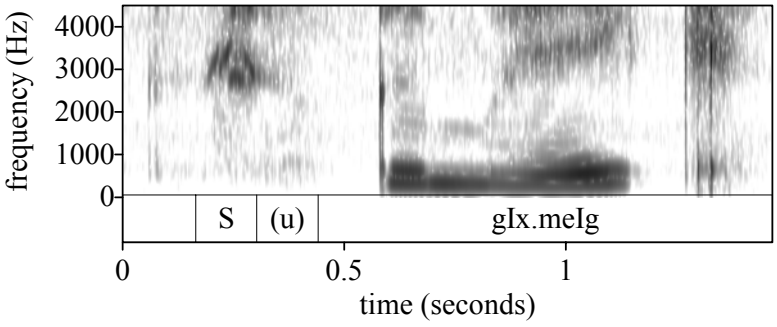


Figure 7. Vowel devoicing in a production of English nonword /ˈʃuˌgɨmeg/.

This production of the English nonword /ˈʃugɨˌmeg/ is another case of an "E" disfluency. Unlike the example in Figure 2, however, here the effortful struggle to produce the aerodynamic conditions necessary for strident turbulence results in a failure to initiate voicing after the (eventually successful) configuration of gestures for the /ʃ/ constriction is released into the wider oral passage for the following vowel. One English speaker (the second author) interpreted this failure to voice the /u/ in terms of a prosodic reorganization that displaced the stress to the second syllable. Another English speaker (the first author) picked up on residual cues to the stress pattern, such as the high intensity during the fricative in alternation with the low intensity on the weak vowel in the second syllable, to perceive this disfluency instead as an illegal vowel lenition in a stressed syllable.

We might be tempted to ascribe such disagreements to the different backgrounds that any two transcribers necessarily bring to the task of transcription, even when they share a common native language. However, the examples in Figures 8 and 9 show that there can be truly ambiguous cases, where there is no good basis for choosing between two different prosodic analyses for two different transcriptions. In our Greek data, cases of "effortful" productions of fricatives were most frequent in four-syllabic non-words, which necessarily have initial unstressed syllables, given the three-syllable window for stress. Such four-syllable forms with antepenultimate stress are very rare (although non-initial stress is not rare), and the children find the nonwords with this shape particularly challenging, as illustrated in these two figures.

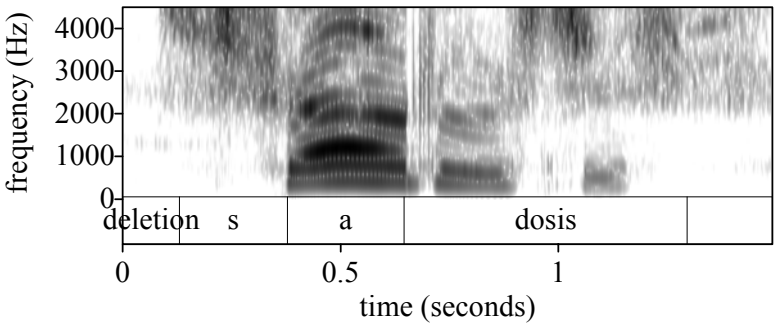


Figure 8. Deletion of initial unstressed syllable in Greek nonword /diˈsamonis/.

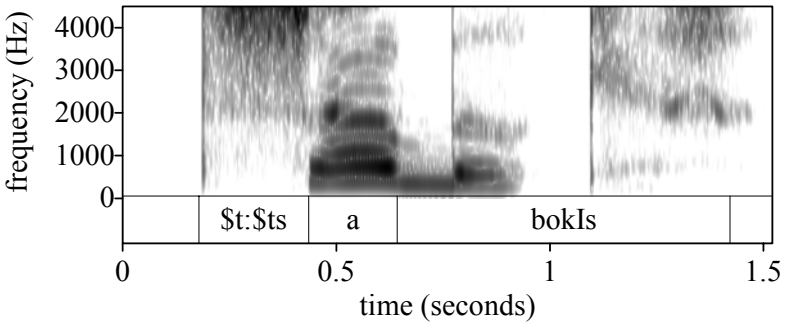


Figure 9. Ambiguous case that can be analyzed either as complete deletion of the initial unstressed syllable with a stopping error on the initial /s/ of the second syllable, or as deletion of the vowel only, with a possible merger of the features of the resulting /ds/ in a disfluent production of the Greek nonword /daˈsamonis/.

In the example in Figure 8, the child appears to have deleted the entire first syllable of the target form /di'samonis/, and then produced a slightly distorted /s/ from the second syllable. By contrast, in the example in Figure 9, since there is clear evidence of an initial stop burst, the interpretation of the disfluency is ambiguous. Does this [ts] sequence represent a complete initial CV deletion, and the substitution of the affricate [ts] for the now word-initial target /s/? Or does the [ts] instead come about via a less substantial reduction of the first syllable — i.e., a "devoicing" of the vowel as well as of the initial /d/? And if so, is the original syllable count preserved so that the apparent [ts] sequence is interpreted as a sequence of two syllable onsets, as it would be if the "devoiced" vowel were /i/ or /u/? Or does the unexpected lenition of a target low vowel (observed sometimes in Japanese but never reported in Greek) effectively change the syllable count, so that the resulting [ts] is now a merging of the onset consonants into the affricate [ts]? Our Greek transcriber entertained all three possibilities, but did not find any one of them more compelling than the others.

6. Summary and conclusion

A key insight of Autosegmental-Metrical theory is that metrical structures are parsed from the way in which they license the choice of autosegmental content from the language-specific inventory of paradigmatic contrasts, and govern the syntagmatic alignment of different autosegmental content specifications with each other. As children acquire the ambient spoken language, they must learn the metrical structures as well as the inventory of autosegmental content specifications specific to their target language. In this paper, we have reviewed a few examples of young children's disfluent productions of consonants in positions where the gestural alignment patterns or the prosodic interpretation of the gestural alignments is particularly challenging. We have tried to show how such a comparison of young children's mis-productions of consonants cross-linguistically can illuminate the interaction between ostensibly universal constraints imposed by immature motor control systems and the language-specific metrical structures and autosegmental content inventories. In future work, we hope to further uncover these processes and interactions, in order to gain a greater understanding of the mechanisms available to a young speaker in planning and generating utterances.

Acknowledgements

This work was supported by an NIH traineeship to the first author and NIDCD grant 02932 to Jan Edwards, Principal Investigator of the παιδολογος project. We thank the adult and child participants who produced the utterances that were recorded on this project and the native speaker transcribers (Wai-Yi Peggy Wong, Sarah Schellinger, Asimina Syrika, and Junko Davis) who transcribed the utterances and noted the disfluencies. We thank these transcribers, Jan Edwards, Eunjong Kong, and Fangfang Li for discussion of the prosodic analyses.

References

Arbisi-Kelm, T. (2006). Intonation structure and disfluency detection in stuttering. Paper presented at LabPhon10.

Arvaniti, Amalia (1994). Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, 22, 239-268.

Arvaniti, Amalia (1999). Illustrations of the IPA: Modern Greek. *Journal of the International Phonetic Association*, 19, 167-172.

Arvaniti, Amalia, & Baltazani, Mary (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In Sun-Ah Jun (ed.) *Prosodic typology: The phonology of intonation and phrasing* (pp. 84-117). Oxford: Oxford University Press.

Bergmann, G. (1986). Studies in stuttering as a prosodic disturbance. *Journal of Speech and Hearing Research*, 29, 290-300.

Bombien, L., Mooshammer, C., Hoole, P., Kühnert, B. & Schneeberg, J. (2006). An EPG study of initial /kl/ clusters in varying prosodic conditions in German. In *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, Brasil.

de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11, 1-15.

de Boysson-Bardies, B., & Vihman, M.M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297-319.

Browman, C.P., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica*, 45, 140-155.

Browman, C.P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Cahiers de la Institut de la Communication Parlée: Bulletin de la Communication*, 5, 25-34.

Brown, S.F. (1938). Stuttering with relation to word accent and word position. *Journal of Abnormal & Social Psychology*. Volume 33, 112-120.

Brown, S.F. (1945). The loci of stutterings in the speech sequence. *Journal of Speech Disorders* 10, pp. 181–192.

Byrd, Dani (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209-244.

Byrd, D., Lee, S., Riggs, D., & Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America*, 118(6), 3860-3873.

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In: Louis Goldstein, Doug H. Whalen, and Catherine T. Best (Eds.), *Laboratory Phonology 8: Varieties of Phonological Competence*, Mouton de Gruyter, Berlin and New York, pp. 519-548.

Cho, T., & Keating, P.A. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155-190.

Clement, C.J., Koopmans-van Beinum, F.J. & Pols, L.C.W. (1996). Acoustical characteristics of sound production of deaf and normally hearing infants. *Proceedings of the Fourth International Conference on Spoken Language Processing*, Philadelphia, Volume 3: 1549-552.

Darwin, C.R. (1877). Biographical sketch of an infant. *Mind*, 2, 285-294.

Davis, B.L. & MacNeilage, P.F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38: 1199-211.

Davis, B.L., MacNeilage, P.F., Matyear, C.L., Powell, J.K. (2000). Prosodic correlates of stress in babbling: An acoustical study. *Child Development*, 71, 1258-1270.

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491-504.

Dell, G.S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science*, 9, 3-23.

Edwards, Jan, & Beckman, Mary E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. In press in *Language Learning and Development*.

Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 41, 40-48.

Ferreira, Fernanda (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 2, 233-253.

Fougeron, C., & Keating, P.A. (1997). Articulatory strensgthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728–3740.

Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.

Fromkin, V.A. (1973). *Speech errors as linguistic evidence.* The Hague, The Netherlands: Mouton.

Gussenhoven, Carlos (1983). Testing the reality of focus domains. *Language and Speech*, 26, 61-80.

Gussenhoven, Carlos (1992). Intonational phrasing and the prosodic hierarchy. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, & J. R. Rennison (Eds.) *Phonologica 1988. Proceedings of the 6th International Phonology Meeting*, pp. 89-99. Cambridge, UK: Cambridge University Press.

Hahn, E.F. (1942). A study of the relationship of stuttering occurrence and phonetic factors in oral reading. *Journal of Speech Disorders*. 7:143-151.

Halle, M. & Vergnaud, J.R. (1987). *An essay on stress*. Cambridge, MA: MIT Press.

Hayes, B. (1980). *A metrical theory of stress rules*. Doctoral dissertation, MIT, Cambridge, MA. Published 1985, New York: Garland.

Hieronymous, J.L. (1994). *ASCII phonetic symbols for the world's languages: Worldbet.* AT&T Technical Report. Distributed at http://www.cslu.ogi.edu/publications/

Hubbard, C. (1998). Stuttering, stressed syllables, and onsets. *Journal of Speech, Language, and Hearing Research*, 41, 4, 802-807.

Joseph, B., & Philippaki-Warburton, I. (1987). *Modern Greek*. London: Croom Helm.

Kager, R. (1989). *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris

Kang, S., & Speer, S.R. (2003). Prosodic disambiguation of syntactic clause boundaries in Korean. In Gina Garding & Mimu Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp.259-272). Somerville, MA: Cascadilla Press.

Keating, Patricia & Shattuck-Hufnagel, Stephanie (2002). A Prosodic View of Word Form Encoding for Speech Production. *UCLA Working Papers in Phonetics*, 101, 112-156.

Kent, R.D. (1984). Psychobiology of speech development: Coemergence of language and a movement system. *American Journal of Physiology*, 246, R888–R894.

Kent, R.D., Osberger, M. J., Netsell, R., & Goldschmidt Hustedde, C. (1987). Phonetic Development in identical twins differing in auditory function. *Journal of Speech and Hearing Disorders*, 52, 64-75.

Koopmans-van Beinum, F.J., & van der Stelt, J.M. (1986). Early stages in the development of speech movements. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 37–50). Basingstoke, Hampshire: MacMillan Press.

Krakow, R.A.(1993). Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In Marie A. Huffman & Rena A. Krakow (Eds.), *Nasals*, *nasalization*, *and the velum* (pp. 87-116). New York: Academic Press.

Krakow, R.A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, 27, 23-54.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech*, 32(3), 249-278.

Langlois, A., Baken, R.J., & Wilder, D.N. (1980). Pre-speech respiratory behaviour during the first year of life. In T. Murray and J. Murray (Eds.), *Infant communication: cry and early speech*. Houston, Texas: College Hill Press.

Lehiste, I. (1973) Phonetic disambiguation of syntactic ambiguity. *Glossa* 7(2): 107-121.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.

Levelt, W.J.M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, 2(2), 205–217.

Levitt, Andrea G., & Wang, Qi (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech*, 34, 235-249.

Lieberman, Philip (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.

Lieberman, Philip (1986). The acquisition of intonation by infants: physiology and neural control. Catherine Johns-Lewis (ed.). *Intonation in Discourse*, p. 239-257.

Lœvenbruck, H., Collins, M.J., Beckman, M.E., Krishnamurthy, A.K., & Ahalt, S.C. (1999). Temporal coordination of articulatory gestures in consonant clusters and sequences of consonants. In Osamu Fujimura, Brian D. Joseph, & B. Palek (Eds.) *Proceedings of LP'98, Item Order in Language and Speech* (Vol II, pp. 547-573). Prague: The Karolinum Press.

MacNeilage P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21: 499–511.

Maekawa, Kikuo, & Kikuchi, Hideaki (2005). Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In Jeroen Maarten van de Weijer, Kensuke Nanjo, & Tetsuo Nishihara (Eds.) *Voicing in Japanese* (pp. 205-228). Berlin: Walter de Gruyter.

Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *Journal of Child Language*, 20, 303-312.

Masataka, N. (2003). *The onset of language*. Cambridge, UK: Cambridge University Press.

Mattingly, I.G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.) *The cognitive representation of speech*, pp. 415-420. Amsterdam: North-Holland.

McCune, L., & Vihman, M.M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44, 670-694.

Nakanishi, Y., Owada, K., & Fujita, N. (1972). Kō'on kensa to sono kekka no kōsatsu [Results and interpretation of an articulation test]. *RIEEC Report [Annual Report of Research Inst. Education of Exceptional Children, Tokyo Gakugei University]*, 1, 1–41.

Natke, U., Grosser, J., Sandrieser, P., and Kalveram, K.T. (2002). The duration component of the stress effect in stuttering. *Journal of Fluency Disorders*, Volume 27, Issue 4, 305-318.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Oller, D.K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.). *Child phonology*: Vol. 1. Production (pp. 93-112). New York: Academic Press.

Oller, D.K. (1986). Metaphonology and infant vocalizations. In B. Lindblom & R. Zetterstrom (Eds.), Precursors of early speech (pp. 21-36). Basingstroke, Hampshire: Macmillan.

Oller, D.K. (2000). *The Emergence of the speech capacity*. Mawah, NJ: Lawrence Erlbaum and Associates, Inc.

Oller, D.K., & Eilers, R. (1988). The role of audition in infant babbling. *Child Development*, Vol. 59, No. 2, 441-449.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, 9, 137-158.

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, 14, 415-440.

Peng, Shu-hui, Chan, Marjorie K.M., Tseng, Chiu-yu, Huang, Tsan, Lee Ok Joo, & Beckman, Mary E. (2005). Towards a pan-Mandarin system for prosodic transcription. In Sun-Ah Jun (ed.) *Prosodic typology: The phonology of intonation and phrasing* (pp. 230-270). Oxford: Oxford University Press.

Pierrehumbert, J.B., & Hirschberg, J. (1990). The meaning of intonation contours in the interpretation of discourse. In Philip R. Cohen, Jerry L. Morgan, & Martha E. Pollack (Eds.) *Intentions in communication* (pp. 271-311). Cambridge, MA: MIT Press.

Prins, D., Hubbard, C., Krause, M. (1991). Syllabic Stress and the Occurrence of Stuttering. *Journal of Speech, Language, and Hearing Research*, 34, 5, 1011-1016.

Rooth, Mats (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, 75-116.

Schafer, A.J., Speer, S.R., & Warren, P. (2004). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In John C. Trueswell & Michael K. Tanenhaus & (Eds.) *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 209-226). Cambridge: MIT Press.

Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, E.O., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas & D. Zec (Eds.) *The Phonology-Syntax Connection* (pp. 313-337). Stanford, CA: Center for the Study of Language and Information.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale, NJ: Lawrence Erlbaum.

Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning: New evidence form speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processing in language* (pp. 17-51). Hillsdale, NJ: Erlbaum.

Shriberg, Elizabeth (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences* (ICPhS-99), San Francisco, Vol. I, 619-622.

Soderberg, G.A. (1962). Phonetic influences upon stuttering. *Journal of Speech and Hearing Research*, 5, 315-320.

Stark, Rachel E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, Volume 1, Production* (pp. 73–90). New York, NY: Academic Press.

Steriade, Donca (1999). Alternatives to the syllabic interpretation of consonantal phonotactics. In Osamu Fujimura, Brian D. Joseph, & B. Palek (Eds.) *Proceedings of LP'98, Item Order in Language and Speech* (Vol I., pp205-242). Prague: The Karolinum Press.

Taylor, I.K. (1966). The properties of stuttered words. *Journal of Verbal Learning and Verbal Behavior* 5, 112-118.

Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Animal Behaviour*, 27, 699-715.

Thelen, E. (1991). Motor aspects of emergent speech: A dynamic approach. In N. A. Krasneger, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.)

Biological and behavioral determinants of language development (pp. 339-362). Hillsdale, NJ: Erlbaum.

Tsurutani, C. (2004). Acquisition of Yo-on (Japanese contracted sounds) in L1 and L2 phonology in Japanese Second Language Acquisition. *Journal of Second Language*. l(3), 27-48.

Venditti, J.J., Maekawa, K., & Beckman, M.E. (2008). Prominence marking in the Japanese intonation system. To appear in S. Miyagawa & M. Saito (Eds.) *Handbook of Japanese linguistics*. Oxford: Oxford University Press.

Vihman, M.M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21, 61-82.

Vihman, M.M, DePaolis, R.A., & Davis, B.L. (1998). Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development*, 69, 4, 935-949.

Vihman, M.M., Macken, M.A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: a re-assessment of the continuity issue. *Language*, 61, 2, 397-445.

Weiner, A.E. (1984). Stuttering and syllable stress. *Journal of Fluency Disorders*, 9, 4, 301-305.

Welby, P. (2003). Effects of pitch accent type and status on focus projection. *Language and Speech*, 46, 53-81.

Whalen, D.H., Levitt, A.G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*, 18, 501-516.

Wingate, M.E. (1988). *The structure of stuttering: A psycholinguistic approach*. New York: Springer-Verlag.

Wong, Wai-Yi P. (2004). Syllable fusion and speech rate in Hong Kong Cantonese. In: *Proceedings of Speech Prosody*, 255-258. 22-26 March , Nara, Japan.

Wong, Wai-Yi P. (2006). *Syllable fusion in Hong Kong Cantonese connected speech.* Doctoral dissertation. Department of Linguistics, Ohio State University.

Wong, Wai-Yi P., Chan, Marjorie K.M., & Beckman, Mary E. (2005). An autosegmental analysis and prosodic annotation conventions for Cantonese. In Sun-Ah Jun (ed.) *Prosodic typology: The phonology of intonation and phrasing* (pp. 271-300). Oxford: Oxford University Press.