Preliminaries to Child Speech Analysis

Mary E. Beckman (Linguistics, Ohio State University)

Jakobson (1939) *The sound laws of child language and their place in general phonological theory*: What remains decisive in the correspondence between child language and the languages of the world is exclusively THE IDENTITY OF THE STRUCTURAL LAWS which underlie every modification of language. ... The phonological sequence of stages is rigorously consistent. It follows from the principle of MAXIMAL CONTRAST, and it proceeds, in the ordering of oppositions, from the SIMPLE and homogeneous to the COMPLEX and differentiated.

Jakobson, Fant, & Halle (1952) *Preliminaries to speech analysis*: But just as a musical scale cannot be grasped without reference to the sound matter, so in the analysis of the distinctive features such a reference is inevitable. ... A distinctive feature cannot be identified without recourse to its specific property.

Acknowledgements

Work supported by:

- NIDCD grant 02932 to Jan Edwards (Dept. Comm. Disorders & Waisman Center, University of Wisconsin – Madison)
- Ohio State University, Cognitive Science Interdisciplinary Summer Fellowships and/or Linguistics Department Targeted Investment Fellowship to Eunjong Kong, Fangfang Li, & Laura Slocum (Ohio State Univerity), Giorgos Tserdanelis (SUNY, Stony Brook), & Kiyoko Yoneyama (Daito Bunka University).

In collaboration also with:

- Tim Arbisi-Kelm, Hyunju Chung, Junko Davis, Sarah Schellinger, Asimina Syrika, & Peggy Wong (UW), Anne Hoffmann (OSU), Hélène Loevenbruck & Julia Monnin (ICP), Ben Munson (U Minnesota), Katerina Nicolaidis (Aristotle U), Areti Okalidou (U Macedonia), Kyuchul Yoon (Kyungnam U).

The παιδολογος project —target sounds and languages

- Comparing word-initial lingual obstruents elicited in a variety of following vowel contexts from children aged 2-5 years.
- Originally comparing Cantonese, English, Greek, Japanese, currently extending to Korean, Songyuan Mandarin, French.
  - All languages have /t/ and /k/ and Greek, and Japanese contrast /k/ with /k^{j}/ before /u, o, a/.
  - All have /s/ and all but Cantonese and Korean have at least 2 other voiceless fricatives from the set /θ, ʃ, ç, x/.
- Interesting differences in frequencies of target consonants and CV sequences across these languages. For example, …
  - /su/ is reasonably well attested in English and Japanese, but rare in Greek, and completely unattested in Cantonese.
  - /ti/ and /si/ are high-frequency sequences in Cantonese, but /ti/ is very rare and */si/ completely unattested in Japanese.

Phonological disorder (PD)

- A syndrome of habitual age-inappropriate mis-articulation in the absence of hearing impairment, cleft palate, or any other gross problems associated with delayed onset of speech.
- Assessed using tests such as the *Goldman-Fristoe Test of Articulation* (GFTA; Goldman & Fristoe, 1986), a picture-naming task that samples each of the consonants of English once in word-initial, medial, and final position.

Example misarticulations in PD (from Isermann, 2001)

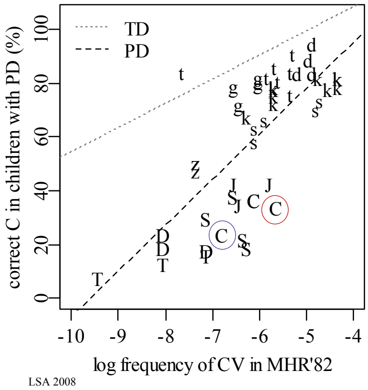
error type	target form		error	ID	sex (yr; mo)	
"stopping"	*socks*	/sɑks/	[datʰ]	p137	M	4;4
	sheep	/ʃip/	[ti]	p112	F	5;4
	cheeze	/tʃiz/	[ki]	p103	F	5;9
"fronting"	*cake*	/keɪ̯k/	[teɪ̯k]	p106	F	5;7 🔈
	brush	/bɹʌʃ/	[bwʌs]	p106		
	shoe	/ʃu/	[su]	p124	M	4;11

Tasks used in project on phonological disorder

- Non-word repetition task compares accuracy for high-frequency sequence /twɛkɪt/ /mæbɛp/ 🔈 vs low /pwagəb/ /moɪpəd/ 🔈 (No difference in size of frequency effect between children with phonological disorder and children with typical development.)
- Phonetic inventory assessed with a picture naming task, that probed each consonant in each position in at least 3 words:

initial /ʃ/ in *shovel*, *sheep*, *shoe*, *shaving*

Vodopivec (2004) Consonant and vowel-context frequency



key

d: *deer*, *dishes*, *doll*, *duck*
t: *teeth*, *top*, *toys*, …
k: *camel*, *cat*, *coffee*, …
g: *goat*, *gum*, *game*, …
s: *scissors*, *sun*, *soap*, …
z: *zipper*, *zoo*, *zebra*, *zoo*
J: *jelly*, *juice*, *jeep*
S: *sheep*, *shoe*, *shovel*, …
C: *chicken* (22), *chair*, *cherries*, *cheese* (7)
D: *those*, *that*, *those*
T: *thorn*, *thumb*, *thigh*

The chicken (adult state) or egg (child state) question

- English-acquiring children tend to be less accurate in producing consonants like /z/, /tʃ/, and /θ/, that occur in fewer words.
- These are also the consonants that occur in the inventories of fewer languages. For example, 438 of the 451 UPSID languages have /k/ or /g/, and while fewer (286) have /t/ or /d/, only 188 have /tʃ/ or /dʒ/. Similarly, 234 UPSID languages have /s/, but only 79 have /z/ and even fewer (18) have /θ/.
- Are children less accurate at producing infrequent consonants because they get less practice at parsing, remembering, and reproducing new words that contain these consonants?
- Or do these consonants occur in fewer words of English because they are combinations of distinctive features that are difficult to parse and/or reproduce accurately (perhaps the same reason they occur in fewer languages)?

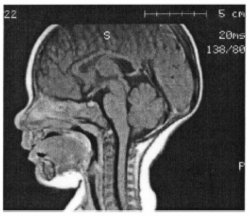
Jakobson (1939) lexical contrast drives acquisition

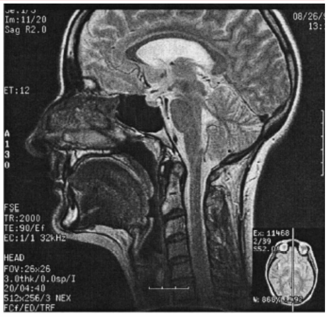
"It is these first distinctions, aiming at becoming significant, which necessitate simple, clear-cut, and stable sound oppositions, capable of being engraved in the memory and implemented at will. The PHONETIC RICHNESS of the babbling period thus gives way to a PHONOLOGICAL LIMITATION."

Argument:

- Babbling infants make all the sounds of the world's languages, so it's not that they can't make the sounds.
- Therefore it must be the fact of lexical contrast that drives simplification, so principle of maximal contrast should be paramount.
- Lexical frequency should be not be directly relevant.
- Simplifications should be the same for children acquiring all languages.

"The PHONETIC RICHNESS of the babbling period"





Midsagittal MRIs of 7-mo-old girl (above) and woman (right). (Vorperian, Kent, Lindstrom, Gentry, Yandell, 2005)

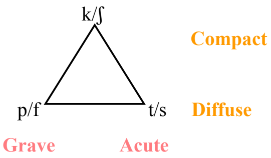
24-mo-old 🔈 🔈 adult 🔈 🔈

[from the παιδολογος database]

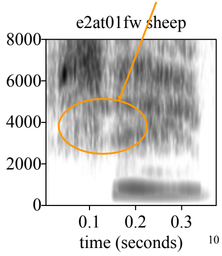
9-mo-old 🔈 [©Damon Hart-Davis/DHD Multimedia Gallery]

The specific property, according to Jakobson, Fant, & Halle

English-speaking transcriber transcribes this *sheep* 🔈 as [sep].



"Compact phonemes are characterized by the relative predominance of one centrally located formant region."



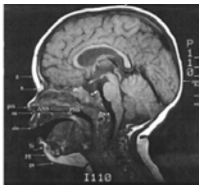
Jakobson (1939): compact consonants "carry a higher degree of specific intensity" and hence are less distinct from vowels, less good consonants.

Kent (1984) on "the psychobiology of speech development"

1) *Vocal tract anatomy changes markedly in the first year of life and in fact continues to be gradually remodeled over the first few years of life. …*

These differences give "a set of articulatory capabilities and propensities different from those of the adult speaker."

For example, the "common preference for front vowels … can be explained anatomically with respect to the infant's tendency toward a frontal carriage of the tongue."



MRI of 15-mo-old boy (Vorperian et al., 2005).

Sorting out the different possible explanations

English-speaking transcriber transcribes /ʃ/ of *sheep* 🔈 as [s].

- Could this be because /ʃ/ is less frequent than /s/ in English, particularly before /i/ and /ɪ/, giving the child less opportunity to make robust generalizations about the CV segmentation and the locus of the contrast between /ʃi/ or /ʃu/ and /si/ or /su/?
- Jakobson's principle of MAXIMAL CONTRAST, suggests that when fricatives are differentiated from stops, the "best" fricative (i.e., the one most different from vowels) will emerge first; diffuse [s] substitutes for compact /ʃ/ (just as diffuse [t] substitutes for compact /k/ in children with PD, e.g. *cake* 🔈).
- Anatomical/physiological constraints identified by Kent (1984) suggests that younger children might be likely to "front" /ʃ/ to [s] in *sheep* (and /k/ to [t] in *cake*) because of the "frontal carriage of the tongue" (cf. Locke, 1983).

Mastering the dynamics

Slowed down cineradiographic clip of *It's 10 below outside* (from Queens University/ATR database, downsampled and at half speed)

Kent (1984) on motor constraints close to the start state

5) Rhythmic or cyclic patterns are a natural basis for the organization of movement systems and may contribute to the acquisition of skilled, coordinated movements in speech.

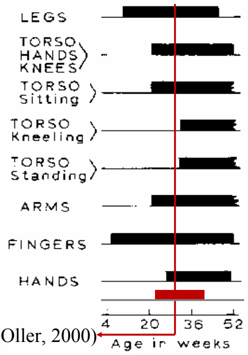
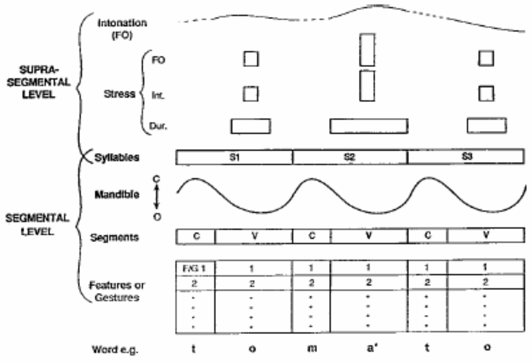


FIG 2. Age ranges of rhythmic stereotypic activity for various parts of the body. [Adapted from Thelen (1981).]

onset of canonical babbling (Oller, 2000)

MacNeilage & Davis (1990) on the jaw rhythm of canonical babble



(Fig. 1 in MacNeilage, 1998)

Kent (1984) on coordinated articulator movements

7) Acquisition of phonology interacts with the acquisition of motor control for speech. … Certainly the child's acquisition of phonology depends on the capacity to shape the vocal tract as required for individual sounds and to orchestrate the musculature to produce sequences of sounds. The reverse is also true: the child's motor facility depends in part on the emergence of phonetic contrasts and the need to utter long sequences of sound contrast.

Even after a child has accomplished the basic concatenation of the phonetic segments of speech, he or she continues to improve in the speed and efficiency of motor control, so that the individual movements of articulation increasingly overlap one another without compromising intelligibility.

cf. Maye, Daland, & Goldrick and Goffman papers yesterday.

Motor control and the time course of consonant mastery

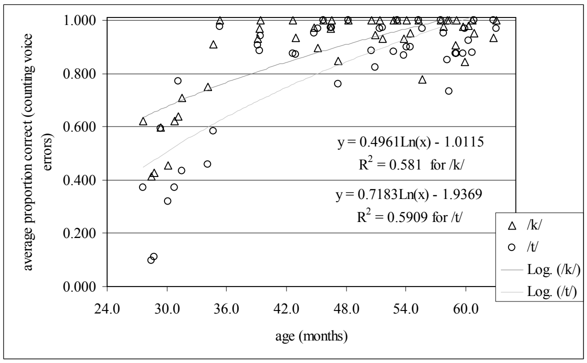
- Simple stops such as [t] and [k], which can be made by a ballistic gesture that throws the tongue against the roof of the mouth during up cycle of jaw, generally first sounds mastered.
- Stops which require coordination of two different constrictions, such as [k^w], and also stop-vowel sequences which require a careful sequencing of different tongue body postures, such as [k^{j}u], generally mastered later.
- Fricatives, such as [θ], [s], and [ʃ], which require a more precise posturing of the tongue tip and/or tongue body generally are later than stops.
- Affricates, such as [ts] and [tʃ], which require a careful sequencing of stop followed this more precise posturing of the tongue tip and/or tongue body also are generally even later.

These are similar to implicational laws for consonant inventories.

Different typical consonant errors for different languages

- For example, /ʃ/ →[s] is stereotypical error in English-acquiring children, but /s/→[tʃ] or [ʃ] much more typical of Japanese-acquiring children (see Munson, Li, et al., tomorrow at 10:30).
- Also, Möhring (1938), Ingram (1974), and many others, document "velar fronting" in German- and English-acquiring children —e.g., P106 (female, at 5;7) in Isermann (2001) produces these tokens for targets *cake* 🔈 and *goat* 🔈.
- Ushijima & Hayashi (1943), Nakanishi, Owada, & Fujita (1972) report "dental backing" in Japanese-acquiring children — e.g., T17 (female, 3;7) in Yoneyama, Beckman, & Edwards (2003) produces these 4 targets for /tora/ 'tiger'. 🔈 🔈 🔈 🔈

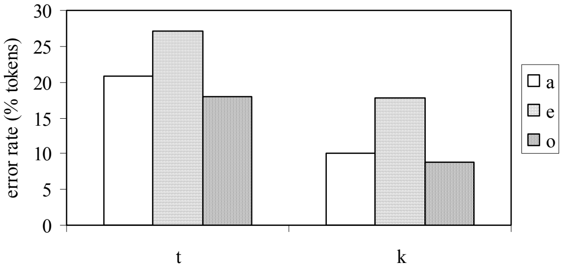
Japanese children’s accuracy rate for /k/ and /t/ by age



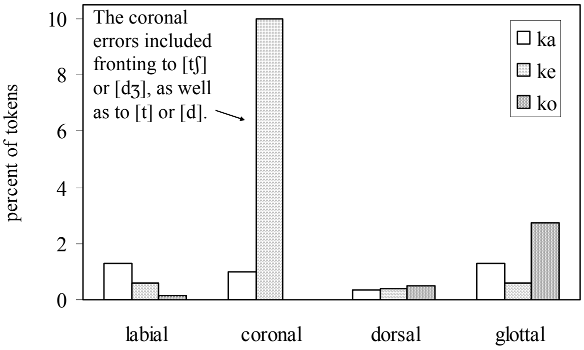
(from Yoneyama, Beckman, & Edwards, 2003)

Yoneyama et al. (2003) — error rates by vowel context

- Three Japanese-speaking phoneticians independently did a narrow transcription of initial sound in each production. (86-96% agreement between each pair of transcribers).
- Used only transcriptions where at least 2 transcribers agreed.



Yoneyama et al. (2003) — error patterns for /k/



The coronal errors included fronting to [tʃ] or [dʒ], as well as to [t] or [d].

The chicken or egg nature of the question, again

- Japanese-acquiring children tend to be less accurate in producing dorsal stops in the /e/ vowel context, which is also the context that occurs in fewer words of the language.
- Their error patterns are somewhat different from those of English-acquiring children, but …
- These error patterns match processes of “velar softening” that have been attested diachronically in many languages of the world, so ....
- Are Japanese children less accurate at producing /k/ before /e/ because they get less practice at parsing, reproducing, and remembering new words that contain this consonant-vowel sequence?
- Or is it because this sequence (and [kʲ] generally?) is universally difficult to parse and/or reproduce accurately?

The παιδολογος project (2) — elicitation methods

Munson, Edwards, & Beckman (2005), repetition task comparing high-frequency /twɛkɪt/ /mæbɛp/ to low /pwagəb/ /moɪpəd/

Yoneyama, Beckman, & Edwards (2003), picture naming task:

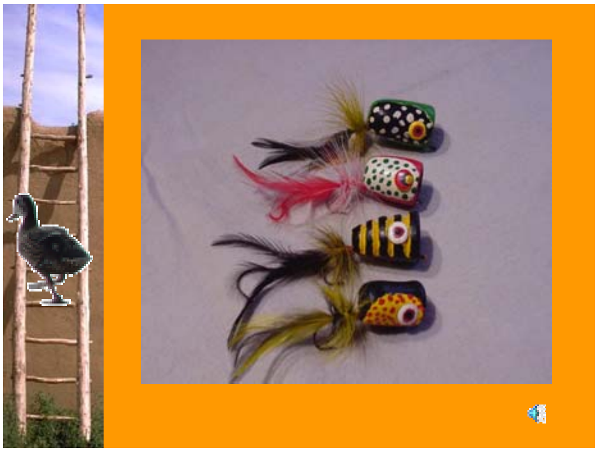
target /to/ in *tora, tomato,*

target /ko/ in *koara*

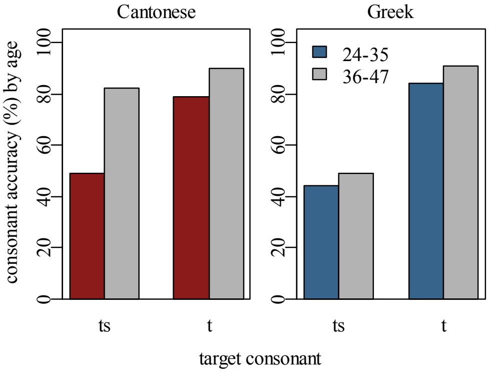
Starting with Nicolaidis, Edwards, Beckman, & Tserdanelis (2003), combine methods in a picture name repetition task …

Why use picture name repetition?

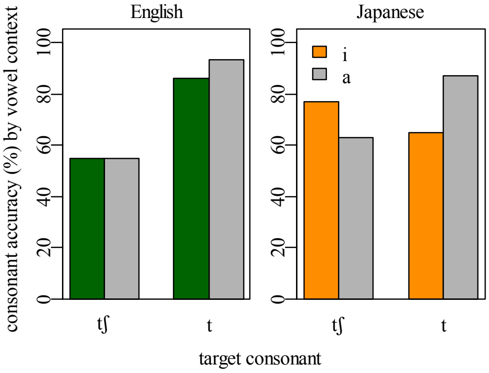
- We want to compare target consonants in a variety of vowel contexts across the languages, and some of these sequences are not attested in any real words in some of the languages — e.g., */su, tu, tʰu, tsu, tsʰu/ in Cantonese, */si/ in Japanese.
- Not all target consonant-vowel sequences that are attested in real words are attested in pictureable words, and even some of the pictureable words may not be familiar to the older children.
- We want to test children as young as 24 months.
- Many of the words that will be familiar to the 5-year-olds won’t be familiar to most of the 2-year-olds; young children are acquiring new words every day, so the unfamiliar words are effectively nonwords.
- One of the functions (the primary function?) of phonology for the young child is to be able to parse and remember new words.



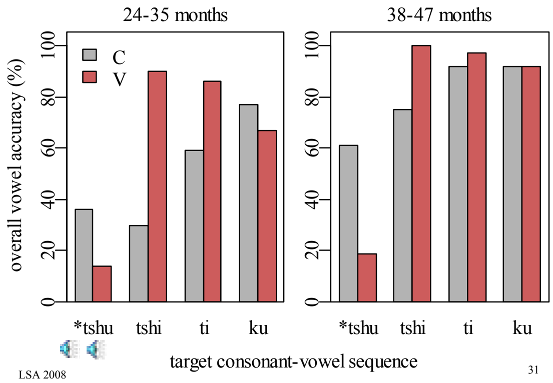
Edwards & Beckman (2008) — Cantonese vs Greek, /ts/ vs /t/



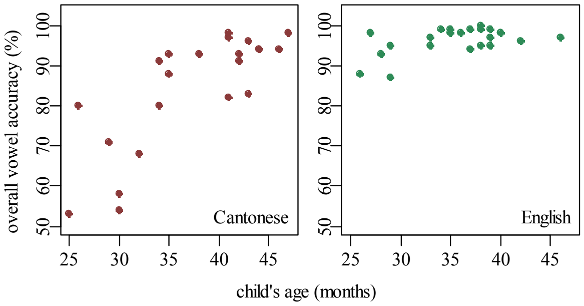
Edwards & Beckman (2008) — English vs Japanese, /tʃ/ vs /t/



Cantonese */tsʰu/ — the vowel is affected instead

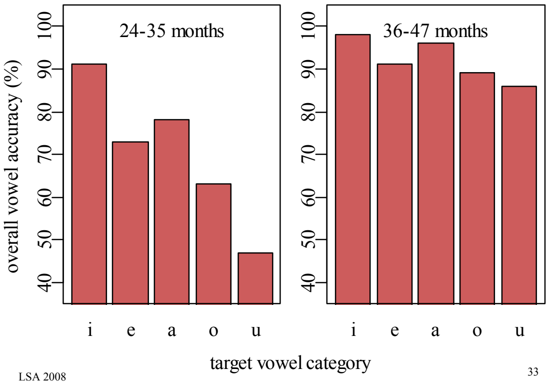


Cantonese vs English vowel accuracy, by age

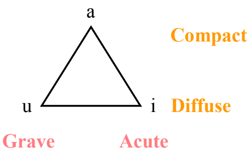


Cantonese: /iː, ɪ, ɛː, yː, uː, ʊ, œː, ɵ, ɔː, ɐ, aː/ and 11 diphthongs
English: /iː, ɪ, eː, ɛ, æ, a, ʌ, oː, ʊ, uː, ɹ/ and 4 to10 diphthongs

Cantonese vowel accuracy, by vowel



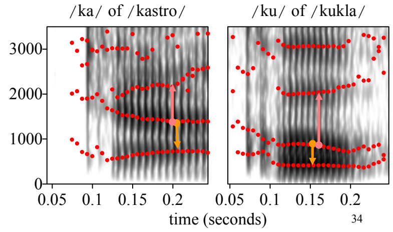
The specific property, according to Jakobson, Fant, & Halle



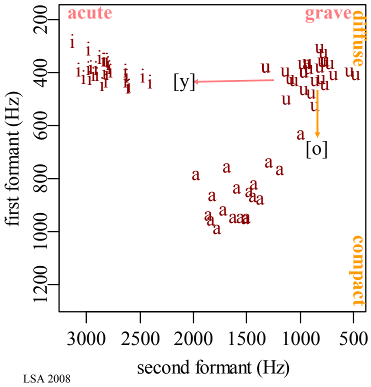
Compact phonemes are characterized by the relative predominance of one centrally located formant region. ...

In the case of vowels this feature manifests itself primarily by the position of the **first formant** …

The position of the **second formant** … is the most characteristic index of this feature.



Cantonese adult female point vowels after /k/ or /kʰ/

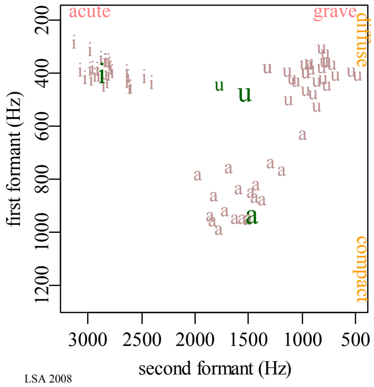


(from Chung, 2008)

The Cantonese children's errors on the sequence */tsʰu/ are substitutions of one of the legal sequences /sy/ or /so/.

By Jakobson's principle of maximal contrast, the /u/→[y] is very strange.

Comparing English adult female point vowels & effect of /s/



(from Chung, 2008)

u /u/ in *cougar*
u /u/ in *soup*
i /i/ in *key*
a /a/ in *soccer*

Ohio English /u/ is "sharp" relative to Cantonese /u/, particularly in the context of coronals.

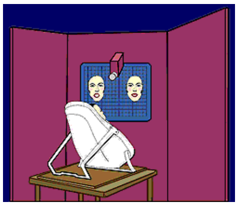
Why are vowels generally early?

- Somatosensory feedback from child’s own production is simultaneous with the auditory feedback from child’s own production.
- Therefore, there is a fairly immediate and “transparent” feedback loop.
- For sighted child, visual image from watching another’s production is simultaneous with the auditory image of hearing another’s production.
- This supports a triangulation between acoustics and “what I feel in my mouth” and “what I see of your mouth” in order to be able to better map from “what I hear from your mouth” to “what I hear from my mouth” (even though my vocal tract is so very different from yours).

Kent (1984) on the relationship of production to perception

4) 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.

- Kuhl & Meltzoff (1982): 4-mo-old looks longer at face that matches heard vowel.



- Kuhl & Meltzoff (1996): Listeners judge infant’s coos to be more like the vowel that infant hears/watches.

Evidence for early ambient influence on the vowel space

- Kuhl, Williams, Lacerda, Stevens, & Lindblom (1992) and Werker & Polka (1994) measure 6- or 7-mo-old infants’ perceptual sensitivity to vowel differences near prototypical members versus peripheral members of categories such as English /i/ versus Swedish /y/ and German /y/ versus /u/, and show ambient language effects on perception.
- de Boysson-Bardies, Hallé, Sagart, & Durand (1989) measure formants in vowel-like vocalizations of 10-mo-old babies acquiring Arabic, Cantonese, English, or French, and show different distributions that mimic different distributions of vowels in adult speech corpora for the languages.

How then should we model the fact that there are vowel space universals and the fine-grained differences across languages?

Modeling vowel space universals (and differences)

- Liljencrants & Lindblom (1972) show that maximizing sum of pair-wise distances fairly accurately predicts the most common distribution of vowels in spaces with 3, 4, 5, 6, or 7 vowels.
- de Boer (2000) shows how imitation among social agents predicts more varied systems, but with clusters in the same regions as the points predicted by MAXIMAL CONTRAST.
- Oudeyer (2006) shows how models of neural maps built by perception-action feedback loops also predicts these clusters.
- Callan, Kent, Guenther, & Vorperian (2000) show that this kind of auditory feedback can accommodate to the changing shape of vocal tract over the first four years of life.
- Rvachew, Mattock, Polka, & Ménard (2006) apply a model of vocal tract development to vowels produced by 1-yr-olds sorts out residue to be explained by motor control development.

Interim summary — what we know about vowel spaces

After decades of modeling vowel production and perception going back to Chiba & Kajiyama (1945), Fant (1960), etc., we understand the specific properties of static vowel postures well enough that we can …

- Build models to compare how well principles such as maximal contrast versus quantal stability in production do in predicting vowel space universals.
- Build models of how neural maps that are set up by the auditory-articulatory feedback loop in babbling predict similar vowel spaces well before child has any words.
- Begin to model how ambient language patterns can shape production and perception through imitation.
- Begin to build predictive models of what kinds of vowel errors toddlers might make in acquiring different languages.

A more challenging task …

What none of the models mentioned above do is talk about formant dynamics, although ….

- Lindblom (1992) suggests how principles of maximal contrast and minimization of movement trajectory can predict a common inventory of stop+vowel syllables:

pi	pu	ti	tu	k^{j}i	ku
pe	po	te	to	k^{j}e	ko
pa		ta		ka	

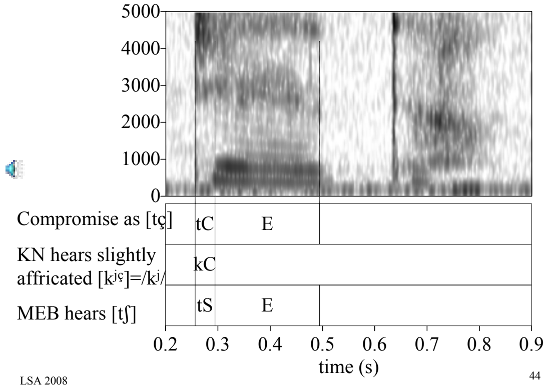
- Oudeyer (2001) shows how the “imitation game” can lead to this kind of inventory in social agents who have no lexicons.

However, the phonetic space in these models is vowel formants.

The challenge of consonants

- Lingual postures for stops are “hidden” acoustically during the stop closure, and cues are distributed onto adjacent segments.
- Perceiving these postures in the CV or VC formant transitions means parsing out the contribution of the stop from the contribution of the co-produced vowel posture.
- Fricatives (particularly sibilant fricatives) give more clues to the lingual posture during the fricative constriction, but …
- Even so, the posture behind the constriction is “hidden” acoustically, making for the same parsing problem.
- Speakers of different languages weight cues (such as spectrum during stop burst or fricative turbulence versus CV formant transitions) differently, as shown by McGuire talk yesterday, Munson, Li, et al. talk tomorrow, etc.
- This affects adult transcriptions of children’s productions.

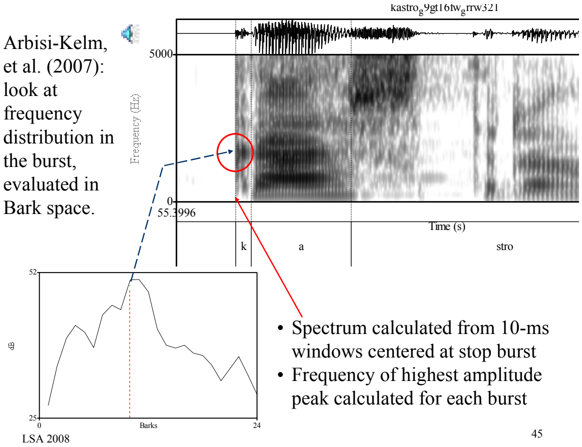
Transcriptions for g2nf07’s production of *κέντρο* ‘center’



Compromise as [tç]

KN hears slightly affricated [k^{jç}]=/k^j/

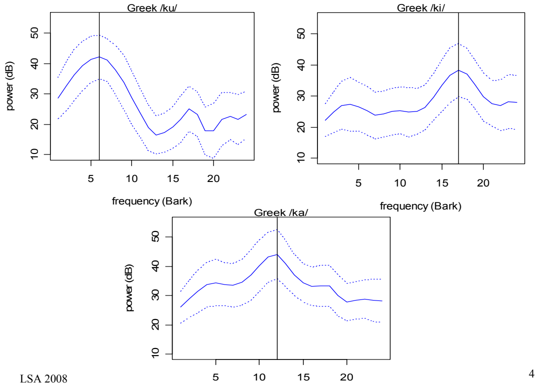
MEB hears [tʃ]



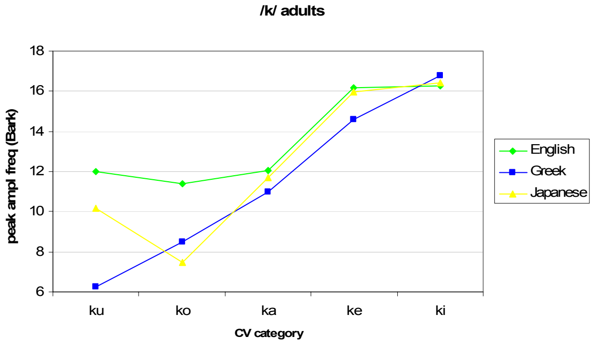
Arbisi-Kelm, et al. (2007): look at frequency distribution in the burst, evaluated in Bark space.

- Spectrum calculated from 10-ms windows centered at stop burst
- Frequency of highest amplitude peak calculated for each burst

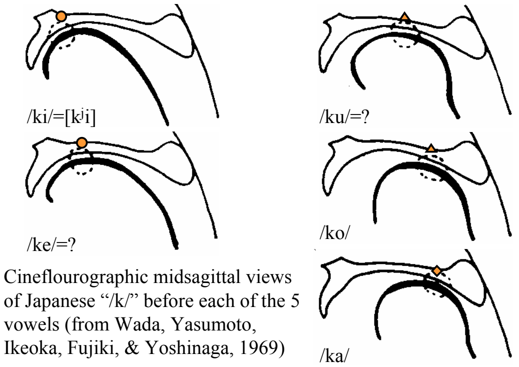
Greek dorsals before point vowels



Context effects across languages

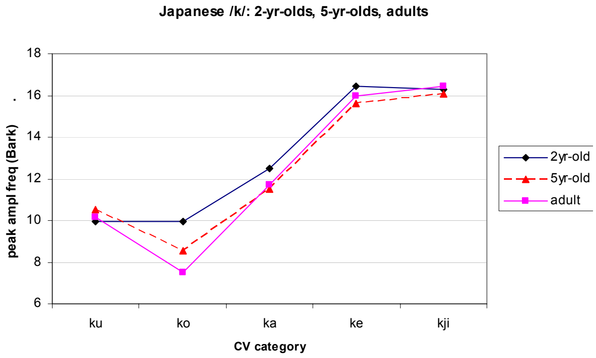


Arbisi-Kelm et al. (2007) — a continuum from /ka/ to /kʲi/?



Cineflourographic midsagittal views of Japanese “/k/” before each of the 5 vowels (from Wada, Yasumoto, Ikeoka, Fujiki, & Yoshinaga, 1969)

Japanese-acquiring children



LSA 2008

Summary and conclusion

- Acquisition of vowels has been a challenging, and at the same time a very fruitful testing ground for our models of the relationship between lexically-contrastive categories and the specific properties in the speech signal that adult speakers and listeners command.
- Acquisition of consonants is an even more challenging but potentially even more fruitful testing ground for our understanding of what “segments” are and how sequential constraints come about in the parsing of coarticulated gestures.
- To realize this potential, we need to move beyond transcription, and try to develop representations of the specific properties for consonant contrasts that are as workable as our current representation of the vowel space.

LSA 2008