


Development of a Weighted Accuracy Measure for implementation in Phon

Cynthia Core and Suresh Subramanian

The George Washington University

Introductions

- Cynthia Core
  - Assistant Professor, Department of Speech and Hearing Science, GWU
- Suresh Subramanian
  - Graduate student in Computer Engineering, GWU

Acknowledgements

Erika Hoff
Department of Psychology, Florida Atlantic University
http://psy.fau.edu/~hoff/LangDevLab/index.htm

Language Development Lab Managers
Melissa Senor, Rosario Rumiche

Graduate Students
Kelly Bridges, Marisol Parra, Silvia Place, FAU
Kristen Dorsey, GWU

Funding Sources:
Xx R03 grant number
NICHD HD054427-01S1
Florida Atlantic University

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON DC

Florida Atlantic University • Language Development Lab •
FAU

Collaborators

- Erika Hoff, Florida Atlantic University
- Jim Mahshie, George Washington University
- Tiffany Finnegan, Gallaudet University
- FAU Language Development Lab
- GWU students

My Projects

- First Language Bilingual Acquisition
  - Focus is word learning
- Speech Perception/Production interface in young children with cochlear implants
  - Focus is speech perception

Bilingual Projects

- Data collection in Florida
- Phonological Memory and Language Learning
  - Investigated the role of phonological memory in early vocabulary development in monolingual and bilingual children
  - 56 monolingual English speaking children
  - 47 bilingual Spanish-English speaking children
  - Data at 3 times, 22-, 25-, and 30-months
  - Real word and nonword repetition in English and Spanish
- Four-year follow-up
  - Real word and nonword repetition
  - Bilingual English Spanish Assessment (Goldstein, Peña)

Ongoing…

- Environmental Correlates of Bilingual Language Development
  - Looks at the role of linguistic experience in language development – family constellation, relative amount of input in each language
  - 170 bilingual children
  - Real word and nonword repetition
  - Bilingual English-Spanish Assessment

- Various spontaneous samples, 22-, 25-, 30-month, and 4-year-old bilinguals

Cochlear Implant Project

- Data collection ongoing in Washington, DC metro area
- Perception of phonetic features and production of those features
  - Consonant place, manner, voicing, vowel height, advancement, stress, syllable number, intonation
  - 40+ children with cochlear implants
    - 3 to 5 years
  - 2 ½ year longitudinal study (5 visits)
  - Speech production measures
    - Goldman-Fristoe Test of Articulation
    - Elicited multisyllabic words with varying stress patterns, questions, statements

Data to be analyzed

- 12 Real Words
  - 4 each, 1-, 2-, 3-syllable
  - From MacArthur-Bates CDI
- 12 Nonwords
  - Phonologically matched to real words for onsets/rhymes and stress patterns
- Elicited through toy naming game – imitation
  - Look, his name is ‘Kog” Can you say “Kog?”

Examples of Stimuli

- Dog, juice, cat, book
  - Kog, boos, dat, jook
- Pan, luz, tren, sol
  - Lan, trus, sen, pol

- Banana, telephone, lollipop, pajamas
  - Bajapop, tellina, lolemas, panaphone
- Muñeca, gallina, caballo, pelota
  - Gañeca, calota, peballo, mullina

Problems in analysis using PCC

- If a child doesn't repeat an elicited item
  - Score is 0
- If a child says [da:] for 'cat'
  - Score is 0
- If a child says [us] or [dus] for 'juice' those are the same

- This makes it impossible to differentiate non-repeaters from children who are actually building good phonological systems

Why quantify accuracy?

- Can answer questions of how children change in production abilities over time
- Can answer questions of how groups of children might be different
  - 'Are bilingual children as accurate in productions of English words as monolingual English speaking children?'
  - 'Does a child have a speech sound disorder?'
    - Describe severity of disorder
- Can help with item analysis
  - Might be able to inform us about variability of production within-child
- Need accuracy measures for statistical analysis for group analyses
  - Do children with small vocabularies have poor production skills relative to children with large vocabularies

Need for accuracy measures

- Manual calculations of accuracy are insufficient – time-consuming and prone to inaccuracies in calculation for large databases
- Need ways to efficiently calculate accuracy
  - Lots of computational power is required
- By child
  - Individuals or groups
- By target
  - Subsets of targets (real v. nonwords, 1 v. 3-syllables, iambs v. trochees, etc)
- By word-position – onset/rhyme, coda consonants, even by features

Wishlist

- Reflect complexity of items produced
- Incorporate principles of markedness and normal development
- Variability in assigned values
- Account for all kinds of errors – deletions, substitutions, additions
  - For consonants and vowels
- Useful for elicited samples and connected speech
- Validated, psychometrically sound
- Grounded in theory

What's available?

- Standardized tests of articulation
- Percent Consonants Correct (PCC)
  - and its variants (PCC-R, PCC-A, PPC)
    - (e.g., Shriberg et al., 1997)
- Phonological Mean Length of Utterance; Proportion of Whole Word Proximity
    - (Ingram & Ingram, 2001; Ingram, 2002)
- PPC – Percent Phonemes Correct
    - (Dollaghan, Biber, & Campbell, 1993)
- DIY - Excel

Percent Consonants Correct

- Scoring of a consonant segment is binary
  - Whether incorrect is omitted or differences in phonetic features
  - All consonants are treated equally
  - Omissions, distortions, deletions all rated equally
  - Common/uncommon substitutions/deletions counted the same
- Additions are not counted as incorrect
- Vowels aren't considered
- Word shape is not considered
- No way to differentiate noncompliance (refusal to name or repeat) from inaccurate production

PMLU and PWP

- PMLU = Phonological Mean Length of Utterance
  - Each segment produced by child gets a point
    - Up to number of segments in target word
  - Each correct consonant gets another point
  - ‘cat’ produced correctly would get 5 points
  - If child said [kati] it would still be 5 points

- PWP = Proportion of Whole Word Proximity
  - Child’s PMLU/Target word PMLU

- PWC = Percent Whole Word Correctness
  - How many words in sample a child produces accurately

Limitations of PMLU/PWP

- Doesn't consider complexity of segments – adjacencies or clusters, just total number of consonants and accuracy
- Does not address vowel accuracy
- PMLU doesn't track growth over time very well
  - (cf Taelman, 2005; Saaristo-Helin, 2009)
- PWP is better than PCC at information about word shape, but it doesn't differential common from unusual errors
- Not sure how these measures deal with distortions which are clinically relevant

The vision:

- A tool that allows us to conduct a more fine-grained analysis would allow
  - Differentiation of repeaters and non-repeaters
  - Following growth over time
  - A way to capture common v. unusual patterns
- Combine the power of Phon's powerful search abilities (e.g. to identify all aligned obstruent coda productions) with a calculator that will tell you quickly whether stops are produced more accurately than fricatives

Weighted Speech Sound Accuracy Measure - WSSA

- Adaptation of a measuredeveloped by Oller & Ramsdell (2006)
  - Designed for LIPP (Logical International Phonetics Program, Kim Oller)
- Kim Oller, Heather Ramsdell, Jonathan Preston, Mary Louise Edwards, Stephen Tobin
- Feature-based approach (grounded in phonetics/phonology)
- “Common” errors involve small penalties “Atypical” errors involve larger penalties (markedness, developmental patterns)
- Considers both consonants and vowels

Basics

- An algorithm based on two tiers
  - Word shape match and segment accuracy
- Assigns a value to a child's production based on its match to a target
- Two tiers of accuracy production
  - Global Structural Agreement - Word shape match
  - Featural Agreement – feature match at the segmental level
- WSSA = Global Structural Agreement x Featural Agreement
- Some principles are established for alignment
  - Matching nuclei, then consonants with minimal discrepanices and no reordering

Global Structural Agreement

- Number of aligned segments produced/total number of segments
- [da] / dog GSA Value .66
  - 2/3 segments produced
  - Reflects omission of final consonant
- [di] for ‘kitty’ GSA Value .5
  - 2/4 segments produced
  - Reflects omission of CV
- Additions are scored as 0
  - ‘kog’ -> [kagi] then total segments are 4 and target contained 3 segments, so GSA value is .75

Mean Featural Agreement

- Each segment receives a value for featural agreement
  - Consonants and vowels start with a value of 1 each
  - Deductions are made according to type of error
- Segment values are averaged
- This gives Mean Featural Agreement

'juice' -> [du]

MFA = .84 +1 = 1.84/2 = .92

WSSA Weights: Consonants

CONSONANT FEATURE		Penalties
Manner (0.333)	**Huge Manner** -uncommon errors, damaging to intelligibility	-.3333
	Big Manner - Less common in phonological development	-.25
	Small Manner -Common errors in phonological development	-.1666
	Teeny Manner -minor phonetic errors	-.0833
Place (0.333)	**Huge Place:** -Uncommon, very damaging to intelligibility	-.333
	Big Place - Less common in phonological development	-.25
	Small Place - Typical errors in phonological development	-.1666
	Teeny Place -Phonetic errors in English, based on small changes in tongue placement.	-.0833
Voicing (0.333)	**Huge Voicing** -Uncommon	-.3333
	Small Voicing -Common	-.2222
	Teeny Voicing -Phonetic changes	-.1111

WSSA Weights: Consonants

For example: Place of articulation

Huge Place	-.333	Dorsal Glottal	↔ ↔	Labial Non-Glottal
Big Place	-.25	Coronal Coronal Alveolar Palatal Retroflex	↔ → → → ↔	Labial Dorsal Palatal Dental Not Retroflex
Small Place	-.1666	Linguadental Dental Palatal Dorsal	↔ ↔ → →	Labiodental Alveolar Alveolar Coronal
Teeny Place	-.0833	Bilabial Labialization Blading Tongue Advance/Retract	↔	Labiodental

WSSA Weights: Vowels

Vowel Feature	Weight	Penalties			Example
Height	0.40	Huge Height	-.40	4 step height change	/i/ ↔ [a]
		Big Height	-.30	3 step height change	/ɪ/↔ [a]
		Small Height	-.20	2 step height change	/i/↔[e]
		Teeny Height	-.1	1 step height change	/a/↔[ɛ]
Advancement	0.40	Big Front	-.40	Front ↔Back	/o/↔ [e]
		Small Front	-.20	Front↔Central or Back↔Central	/i/↔ [ə]
Nasalization	0.1	Small Nasal	-.10	Not Nasal → Nasal	/a/→ [ã]
Rounding	0.1	Small Rounding	-.10	Round ↔Not Round	/ʌ/↔ [ɔ]

Computational Example (WSSA)

Gloss	“	l	o	l	l	i	p	o	p	“
Target	/	l	a	l		i	p	a	p	/
Child Production	[	j	a	j		i	p	a	p	]

- Weighted Speech Sound Accuracy = Global structural agreement x Featural agreement
  - GSA = 1
  - MFA = .95
  - WSSA = .95
  - PCC = .40 and PWP = .63

Computational Example (WSSA)

Gloss	“	t	e	l	e	p	h	o	n	“
Target	/	t	ɛ	l		f		o	n	/
Child Production	[	k	ɛ			p		o	ŋ	]

- GSA = .71
- MFA = .86
- WSSA = .6
- PCC = .25 and PWP = .55

Computational Example (WSSA)

Gloss	" d	o	g	"
Target	/ d	a	g	/
Child Production	[g	a		]

- GSA = .66
- MFA = .75
- WSSA = .57
- PCC = 0 and PWP = .4

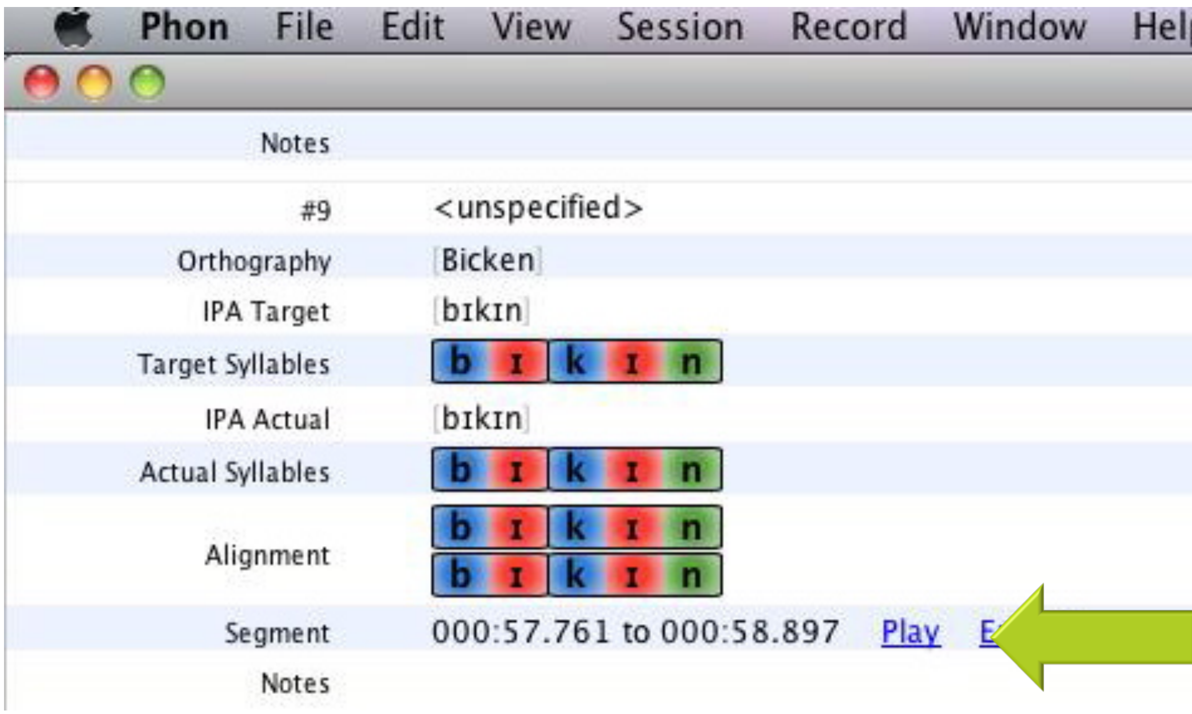
Psychometric Information

- Validity
  - Correlates with existing measures of phonetic accuracy in toddlers and adolescents
  - Distinguishes productions by children with speech sound disorders and typically developing children
  - Sensitive to growth in phonetic accuracy as a child gets older
- Reliability
  - Between transcribers
  - Between different word lists

Phon implementation of WSSA

- Would work as a plug-in to Phon
- Will allow users to select participants, targets, and hopefully word position
- Implementing the measure in Phon would allow more users to validate the measure
  - Develop similar measures for other languages
  - Validate on larger groups of children

It could look like this…



New Tier for Accuracy Calculation

Challenges

- Programming
  - Initial interface ☺
  - Teaching a programmer about linguistics
- Measurement
  - How to differentiate weighting by word position
    - Consider prevocalic voicing and final devoicing
  - How to deal with stress placement errors
  - What about harmony or errors? Cluster reductions? Epenthesis? Fusion?
  - How to assess intermediate productions?
  - Allophonic variation?
  - Should consonants and vowels be weighted equally?
- Other languages - Spanish

Where to go next?

- Implementing WSSA first
- Should weightings be adjustable/customizable?
- Feedback?
  - Utility?
  - Adjustments?