Child-Level Factors & Acquisition of the /t/-/k/ Contrast: Perception

Sara (Bernstein) Cline[1]

Allison Johnson[2] Mary Beckman[3] Jan Edwards[2]

and Benjamin Munson[1]

[1]Department of Speech-Language-Hearing Sciences, University of Minnesota, Twin Cities, [2]Department of Communication Sciences and Disorders, University of Wisconsin, Madison, [3]Department of Linguistics, Ohio State University

learning to TALK

UNIVERSITY OF MINNESOTA Driven to Discover℠

Introduction

Background and motivation

- Children acquire the ability to produce speech sound contrasts gradually [1,2,3]
- Phonetic transcription is the traditional perceptual rating system
- Continuous rating scales such as those utilizing Visual Analog Scaling (VAS) have been applied to the acquisition of fricative contrasts[4,5]
- VAS ratings correlate well with acoustic measures and are more gradient and hence potentially more informative than phonetic transcriptions[4,5]

Aims of this study

1. Apply a continuous rating scale to characterize adults' perception of children's production for the /t/-/k/ contrast and derive measures of how robustly children's productions differed based on listeners' ratings
2. Examine predictors of child-by-child differences in the VAS-derived measures of robustness of /t/-/k/ contrast

Methods

Child Talkers

Talker Participants

- 63 children, aged 28-39 months (part of larger study: www.learningtotalk.org)
- Monolingual (Mainstream American English & African American English)
- Range of maternal education levels
- Passed hearing screening, some late talkers, no other diagnoses

Hypothesized Predictor Variables

- Executive Function: Fruit Stroop, Behavioral Rating Inventory of Executive Functions (BRIEF) - Preschool
- Vocabulary: Expressive Vocabulary Test (EVT-2), Peabody Picture Vocabulary Test (PPVT-4), MacArthur Bates Communication Development Inventory (CDI)
- Speech Perception: Minimal pair discrimination task
- Home Language Input: Language Environment Analysis (LENA) measures
- Maternal Education Level: Caretaker questionnaire

Stimulus Preparation

Speech Recording

- Picture prompted auditory word repetition task
- 8 /t/-initial words, 9 /k/-initial words as part of a longer word repetition task
- Recorded in a sound treated booth

Acoustic Event Tagging

- Initial consonant transcribed as:
  [t], [k], [t:k]=intermediate more "t-like", [k:t]=intermediate more "k-like"
- Release of stop burst and onset of vocal fold vibration tagged in Praat
- Consonant-vowel sequence extracted for 1564 total tokens

Adult Listeners

Listener Participants

- 47 native English speakers, aged 19-39 years
- No history of speech, language or hearing disorders

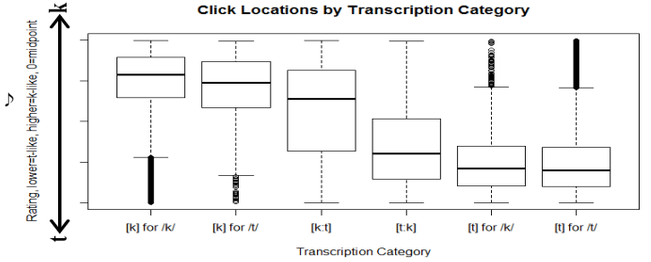
Perception Testing

- Experiment split into 3 versions (about 20 talkers, 500 tokens each)
- 5 training items, 20 repeated items in each version
- One talker presented across all versions
- Click along VAS to rate consonant-vowel sequence; click location logged automatically

Results: Listener Ratings

Category Differentiation

- Linear mixed-effects regressions showed that listeners rated all transcription categories differently (a<0.05) except [t] for target /t/ vs. [t] for target /k/
- Ratings for intermediates were most widely dispersed along the VAS



Intra-rater Reliability

- Reliability was measured by the distance between clicks for repeated tokens
- Reliability varied by experiment version, listener age, and listener sex
- Overall reliability was poor to good depending on listener: need for training?

Set Effects

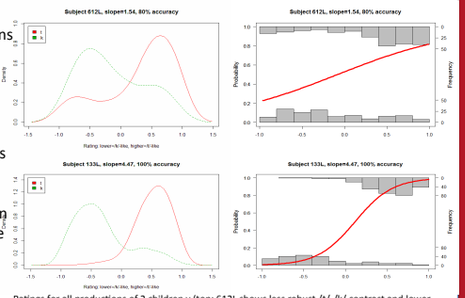
- One talker's productions were presented in all three experiment versions, to determine whether ratings were stable across different sets of stimuli
- Some transcription categories were rated differently by experiment version

Robustness of Contrast

LEFT: Density plots of click locations for /t/ attempts (red) and /k/

RIGHT: Histograms of ratings of attempts at target /k/ (bottom) and /t/ (top).

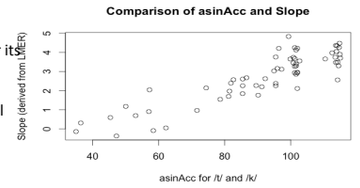
A logit mixed-effect model predicted target (dummy-coded as {1,0}) from VAS rating, with listeners as random effects. The slope of each talker's logit function was estimated. The slope of these functions indicates robustness of contrast. Higher absolute slope = more robust /t/-/k/ contrast



Ratings for all productions of 2 children y (top: 612L shows less robust /t/-/k/ contrast and lower accuracy , bottom: 133L shows more robust contrast and higher accuracy)

Accuracy vs. Slope

Accuracy = Percent of child's attempts at a sound that were transcribed as that sound or its intermediate counterpart (i.e., [t] or [t:k] for target /t/). Rationalized Arcsine transform applied to Accuracy. Slope provides additional separation in data for talkers with high Accuracy



Results: Talkers

Child-level Factors

Linear regression models were analyzed to determine predicting factors of speech production (arcsine transformed Accuracy and Slope).

All vocabulary measures (EVT, PPVT, CDI) were significant (p<0.05) in determining both Accuracy and Slope. Growth Scale Value (GSV) scores were used for the EVT and the PPVT. No other factors were significant in determining speech production

Predictor	Slope p-value	Slope partial r^2 (controlling for age)	Accuracy p-value	Accuracy Partial r^2 (controlling for age)
EVT-2 GSV	0.02*	8%	0.02*	9%
PPVT-4 GSV	0.02*	9%	<0.01**	12%
CDI	<0.01**	15%	<0.01**	13%
Maternal Ed.	>0.05		>0.05	
LENA	>0.05		>0.05	
Executive Fn	>0.05		>0.05	
Minimal Pair	>0.05		>0.05	

Conclusions

- Listeners were able to use a VAS to differentiate all transcription categories except [t] for /t/ vs. [t] for /k/
- Reliability varied by experiment version and listener characteristics. Can reliable listeners be selected based on characteristics, or trained to be more reliable?
- Set effects were present in the experiment versions: context of surrounding tokens influences listeners' ratings
- The robustness of contrast measure "slope" characterizes the difference in VAS ratings for a talker's /t/ and /k/ attempts. Slope provides more information for talkers who have high production accuracy
- Vocabulary size is significant in models predicting speech accuracy. Slope and Accuracy behave similarly in these models

Acknowledgements

References

[1] Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. Journal of Child Language, 7(01), 41-74.

[2] Forrest, K., Weismer, G., Elbert, M., & Dinnsen, D. A. (1994). Spectral analysis of target-appropriate /t/and/k/produced by phonologically disordered and normally articulating children. Clinical linguistics & phonetics, 8(4), 267-281.

[3] Gibbon, F. (1990). Lingual activity in two speech-disordered children's attempts to produce velar and alveolar stop consonants: evidence from electropalatographic (EPG) data. International Journal of Language & Communication Disorders, 25(3), 329-340.

[4] Julien, H. M., & Munson, B. (2012). Modifying speech to children based on their perceived phonetic accuracy. Journal of Speech, Language, and Hearing Research, 55(6), 1836-1849.

[5] Munson, B., Johnson, J. M., & Edwards, J. (2012). The Role of Experience in the Perception of Phonetic Detail in Children's Speech: A Comparison Between Speech-Language Pathologists and Clinically Untrained Listeners. *American Journal of Speech-Language Pathology*, *21*(2), 124-139.

Acknowledgements

This work would not have been possible without the efforts of the entire Learning to Talk team, the children who provided the speech stimuli and their families, and the adult listeners. This work was supported by the Learning to Talk Grant from the National Institutes of Deafness and other Communication Disorders (NIH DC02932) to Jan Edwards, Mary E. Beckman, and Benjamin Munson