Journal of Communication Disorders 64 (2016) 91–102

Contents lists available at ScienceDirect

Journal of Communication Disorders

# Deriving gradient measures of child speech from crowdsourced ratings

Tara McAllister Byun*, Daphna Harel, Peter F. Halpin, Daniel Szeredi

*New York University, New York, NY, USA*

ARTICLE INFO

*Article history:*
Received 19 December 2015
Received in revised form 17 June 2016
Accepted 1 July 2016
Available online 6 July 2016

*Keywords:*
Crowdsourcing
Speech rating
Research methods
Speech sound disorders
Speech perception
Visual analog scaling
Covert contrast

ABSTRACT

Recent research has demonstrated that perceptual ratings aggregated across multiple non-expert listeners can reveal gradient degrees of contrast between sounds that listeners might transcribe identically. Aggregated ratings have been found to correlate strongly with acoustic gold standard measures both when individual raters use a continuous rating scale such as visual analog scaling (Munson et al., 2012) and when individual raters provide binary ratings (McAllister Byun, Halpin, & Szeredi, 2015). In light of evidence that inexperienced listeners use continuous scales less consistently than experienced listeners, this study investigated the relative merits of binary versus continuous rating scales when aggregating responses over large numbers of naive listeners recruited through online crowdsourcing. Stimuli were words produced by children in treatment for misarticulation of North American English /r/. Each listener rated the same 40 tokens two times: once using Visual Analog Scaling (VAS) and once using a binary rating scale. The gradient rhoticity of each item was then estimated using (a) VAS click location, averaged across raters; (b) the proportion of raters who assigned the "correct /r/" label to each item in the binary rating task ($\hat{p}$). First, we validate these two measures of rhoticity against each other and against an acoustic gold standard. Second, we explore the range of variability in individual response patterns that underlie these group-level data. Third, we integrate statistical, theoretical, and practical considerations to offer guidelines for determining which measure to use in a given situation.



## 1. Introduction

### 1.1. Gradience in the acquisition of phonemic contrasts

Both typically-developing children and children with phonological delay or disorder produce speech patterns that differ systematically from adult inputs. These patterns are often described in terms of substitution of one phoneme for another. For instance, in a child with the phonological pattern termed *stopping,* the words 'sea' and 'tea' may be transcribed identically ([ti]). However, adults may be biased by their existing phonological knowledge in transcription tasks, favoring categories familiar from their own fully developed phonology (Amorosa, Benda, Wagner, & Keck, 1985; Buckingham & Yule, 1987). Given the anatomical, motoric, and cognitive-linguistic differences between children and adults, these transcriptions may

* Corresponding author at: Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Room 914, New York, NY 10012, USA.
*E-mail address:* tara.byun@nyu.edu (T. McAllister Byun).

http://dx.doi.org/10.1016/j.jcomdis.2016.07.001
0021-9924/

actually stand at some distance from phonetic reality. The term *covert contrast* is used to describe cases where a typical listener would transcribe a categorical error, such as substitution of one phoneme for another, but instrumental measurements reveal a reliable phonetic difference. In one well-known illustration, children who are perceived to neutralize the voiced-voiceless contrast in initial position can be seen to maintain a statistically reliable phonetic distinction in voice onset time (VOT) between voiced and voiceless targets (Hitchcock & Koenig, 2013; Macken & Barton, 1980; Maxwell & Weismer, 1982). The phenomenon of covert contrast in child speech has now been amply documented over more than three decades of research (e.g. Edwards, Gibbon, & Fourakis, 1997; Gibbon, 1999; Munson, Johnson, & Edwards, 2012; Munson, Schellinger, & Urberg Carlson, 2012; Richtsmeier, 2010; Scobbie, 1998; Tyler, Edwards, & Saxman, 1990; Tyler, Figurski, & Langsdale, 1993; Weismer, 1984; Young & Gilbert, 1988). Several accounts have suggested that covert contrast could represent a typical or even universal stage in the normal course of acquiring a phonological contrast (Munson, Johnson et al., 2012; Munson, Schellinger et al., 2012; Richtsmeier, 2010; Scobbie, 1998).

Studies of covert contrast in child speech have resulted in a major shift in how speech development is conceptualized. The traditional view, in which acquisition of a new phonemic contrast was framed as an abrupt, categorical change, has given way to a new consensus that speech sound development is phonetically gradient (Hewlett & Waters, 2004; Li, Edwards, & Beckman, 2009). Research on covert contrast also has significance for clinical applications. First, the presence or absence of covert contrast is often interpreted as evidence about the level of processing at which a child's speech errors apply: categorical substitutions imply a grammatical process, whereas covert contrast is often interpreted as evidence that errors are occurring at a more peripheral or articulatory level (Gibbon, 1999; but see McAllister Byun, Richtsmeier, & Maas, 2013). Second, the presence of covert contrast can be interpreted as a positive prognostic indicator for subsequent stages of a child's speech development. Children who maintain a covert contrast between two sound categories are more likely to progress to an overt contrast without receiving formal intervention (Forrest, Weismer, Hodge, Dinnsen, & Elbert, 1990). Once enrolled in treatment, children who show covert contrast are reported to reach the criterion for dismissal in a smaller number of treatment sessions (Tyler et al., 1990).

### *1.2. Measuring continuous speech development with listener ratings*

Despite the well-documented importance of covert contrast in understanding a child's phonological system, it remains rare in clinical practice and some areas of clinical research to collect the instrumental measures that could reveal gradient differences between sound categories. This largely reflects the nontrivial difficulty associated with obtaining and verifying acoustic measurements of child speech. A further deterrent is posed by the inconclusive nature of negative results in a study seeking acoustic evidence of covert contrast in child speech. A phonemic contrast can be realized in any of a large number of acoustic dimensions, and young children may mark contrast in phonetically unusual ways (Scobbie, Gibbon, Hardcastle, & Fletcher, 2000). If a given study reports no evidence of contrast in the limited set of acoustic parameters that the investigator chose to quantify, it remains possible that a difference was present in another domain that was not measured. Unlike acoustic measures, listeners' perceptual ratings can be collected in a relatively efficient fashion, and the human ear is simultaneously sensitive to numerous acoustic dimensions. These factors have led to a literature investigating whether gradient measures of child speech can be obtained from listener ratings.

Traditional descriptions of human speech perception have emphasized its discontinuous, categorical nature (Liberman, Harris, Hoffman, & Griffith, 1957). From this point of view, listeners tend to overlook covert contrast in child speech because they are perceptually predisposed to ignore phonetic differences within phonemic categories. However, subsequent research has amply demonstrated that within-category discrimination is possible (Gerrits & Schouten, 2004; Massaro & Cohen, 1983; Pisoni & Tash, 1974; Werker & Logan, 1985). These studies emphasize that listeners' responses can be manipulated to be more continuous or more categorical depending on the nature of the task used to elicit perceptual judgments (Pisoni & Tash, 1974; Werker & Logan, 1985). A task that has been successfully used to elicit continuous perceptual ratings is visual analog scaling (VAS). In VAS, the endpoints of a line are defined to represent two different speech categories, and listeners mark any location to indicate the gradient degree to which they perceive a speech token as belonging to one category or the other (Massaro & Cohen, 1983; Munson, Schellinger et al., 2012). A recent body of work by Munson and co-workers (Julien & Munson, 2012; Munson, Johnson et al., 2012; Munson, Schellinger et al., 2012) has documented the relationship between VAS ratings and continuous acoustic measures of child speech. In a task of rating children's sibilants on a continuum from /s/ to /ʃ/, Julien and Munson (2012) found that mean VAS click location was significantly correlated with a relevant acoustic measure, centroid frequency, for 22 out of 22 listeners. In a multiple regression examining mean VAS click location as a function of both centroid frequency and fricative duration, the acoustic measures accounted for 53% of variance in mean VAS click location. Munson, Johnson et al. (2012) examined mean VAS click location as a function of acoustic measures for three contrasts in child speech: /s/-/θ/,/d/-/g/, and /t/-/k/. The correlations were significant for the first two contrasts, but not for the /t/-/k/ contrast.

The use of a continuous rating scale, as in VAS, is not the only option to derive continuous measures of child speech from perceptual ratings. Binary responses aggregated over multiple listeners can also be used to derive continuous or near-continuous measures. Perhaps the simplest metric is $\hat{p}$, which estimates the proportion of raters who classify a token as belonging to a particular category (Ipeirotis, Provost, Sheng, & Wang, 2014). McAllister Byun, Halpin, and Szeredi (2015) reported that the proportion of listeners classifying children's productions as a "correct /r/ sound" was well-correlated with