Journal of Fluency Disorders xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Fluency Disorders

journal homepage: www.elsevier.com/locate/jfludis

Effect of control samples and listener attributes on speech naturalness ratings of people who stutter

Brenda Carey, Shane Erickson*, Susan Block

Discipline of Speech Pathology, La Trobe University, Bundoora, Victoria, 3088, Australia

ARTICLE INFO

Keywords:
Stuttering
Speech naturalness ratings
Speech restructuring treatment
Listener attributes
Controls

ABSTRACT

Purpose: Speech restructuring treatment can effectively reduce stuttering but the resultant speech may sound unnatural. Martin et al. (1984) speech naturalness scale is widely used by clinicians and researchers, yet little is known about whether including normally fluent speech samples alters the judgement of the naturalness of speech samples of people who stutter, and whether attributes of listeners – specifically training and sex – influence ratings.

Methods: In this study 20 untrained listeners (male and female) and 19 speech language pathology students (female only) rated either the naturalness of 21 speech samples from adults who stutter obtained post-treatment, or the same 21 post-treatment samples randomly mixed with samples of 21 samples from normally fluent speakers matched for age and sex. The independent variables were sample composition (addition of fluent controls) and listener training. The dependent variable was listener naturalness rating.

Results: A two-factor ANOVA with listener training and sample composition as independent variables and naturalness ratings as the dependent variable was performed. Untrained listeners rated samples as significantly less natural than trained listeners. The addition of control samples did not significantly impact scores assigned to post-treatment samples. A comparison of male and female listeners was completed using the Mann Whitney *U* Test. A significant group difference was observed with female listeners rating the samples more leniently (more natural) than male listeners.

Conclusion: Based on this preliminary research, the addition of controls does not appear necessary in evaluating speech naturalness, however the composition of the listener group may affect results.

1. Introduction

Speech restructuring treatment methods have been shown to be efficacious in reducing stuttering frequency (Bothe, Davidow, Bramlett, & Ingham, 2006). However, the success of these methods has been tempered by claims that their use results in unnatural sounding speech (O'Brian, Onslow, Cream, & Packman, 2003). Such speech may continue to draw attention to, and ultimately be rejected by, its user. Hence, there is broad recognition of the need to incorporate speech naturalness as a therapeutic goal and outcome measure in the treatment of stuttering in adults (Bloodstein, 1987; Ingham & Cordes, 1997).

Martin, Haroldson, & Triden's (1984) 9-point naturalness scale is widely used both as a clinical tool (e.g. O'Brian et al., 2017) and to evaluate treatment outcome (e.g. Carey, O'Brian, Lowe, & Onslow, 2014; Teshima, Langevin, Hagler, & Kully, 2010). In spite of the differences in the way the naturalness scale has been used, some recurring findings have emerged. First, it appears that with this

* Corresponding author at: La Trobe University School of Allied Health, Discipline of Speech Pathology, Bundoora, Melbourne, Victoria, 3088, Australia.
E-mail address: s.erickson@latrobe.edu.au (S. Erickson).

https://doi.org/10.1016/j.jfludis.2017.11.004
Received 6 March 2017; Received in revised form 25 October 2017; Accepted 28 November 2017

scale, adults who stutter have speech naturalness which is usually distinguishable from that of matched controls, irrespective of whether the samples evaluated are pre-treatment, during treatment or following treatment. This finding has resulted regardless of listener sophistication (Ingham, Martin, Haroldson, Onslow, & Leney, 1985; Onslow, Adams, & Ingham, 1992), speech sample length (Onslow et al., 1992), speech task (Martin et al., 1984; Metz, Schiavetti, & Sacco, 1990; Van Borsel & Eekhout, 2008), stage of treatment (Ingham et al., 1985; Metz et al., 1990), or even type of treatment undertaken (Runyan, Bell, & Prosek, 1990; Van Borsel & Eeckhout, 2008). Although this finding has been replicated for groups of people who stutter, there has been within-group variation, and recognition that for some subgroups or individuals, post-treatment speech is indistinguishable from that of matched controls (Carey et al., 2014; Carey, O'Brian, Onslow, Packman, & Menzies, 2012; Carey, O'Brian, Onslow, Block, & Jones, 2010; O'Brian et al., 2003; O'Brian, Packman, & Onslow, 2008; Runyan et al., 1990; Teshima et al., 2010). Mean naturalness scores for post-treatment speech samples have typically fallen between 3.5 (Tasko, McClean, & Runyan, 2007) and 6.0 naturalness scale points (Finn & Ingham, 1994), where a naturalness rating of 1 = highly natural sounding speech, and 9 = highly unnatural sounding speech. Mean naturalness scores for matched fluent control speech samples have typically fallen between 2.1 (Martin et al., 1984) and 3.6 naturalness scale points (O'Brian et al., 2003; Metz et al., 1990).

Despite a broad recognition of the importance of measuring speech naturalness, the concern is raised here about the validity of measuring this construct. It has become apparent that individuals use different criteria to make their judgments of speech naturalness (Finn & Ingham, 1994; Ingham, Ingham, Onslow, & Finn, 1989) and that speech naturalness encompasses a number of dimensions (Franken, Boves, Peters, & Webster, 1995). For example, anecdotal evidence has suggested that adults who stutter may focus more on their feeling of anxiety while rating their own speech samples, while other listeners may focus on pitch, intensity, and rate (Teshima et al., 2010), prosody (Franken, Boves, Peters, & Webster, 1992) or voice onset time and vowel duration (Metz et al., 1990). This has led Finn and colleagues to state that researchers should not assume that the listeners in their study share their concept of 'speech naturalness' (Finn & Ingham, 1994). Still other researchers have suggested that other methods of measuring speech naturalness may yield more useful information on post treatment speech, when the instrument identifies which perceptual characteristics of speech are deviant both before and after speech restructuring treatment (Franken et al., 1992).

One variable that may impact the validity of the naturalness scale is the range or breadth of naturalness samples presented. Previous research using other perceptual scales has shown that the exact rating a speech sample is assigned is determined by its quality relative to other samples in that set (Boves, 1984; Kreiman, Gerrat, Kempster, Erman, & Berke, 1993). It has been hypothesised that listeners use an entire perceptual scale when assigning scores, so the anchor points on a scale become defined by the most (or least) extreme sample in the set. It has also been argued that including samples from speakers who do not stutter is essential, as these define the anchors of the scale (Franken et al., 1995). However, this assumption has not been tested using the speech naturalness scale of Martin et al. (1984). It is not known whether providing a more varied range of speech samples for evaluation of the speech naturalness of people who stutter alters the scores assigned. Yet this information is crucial to establish validity of this instrument because an accurate measure should result in the same rating regardless of the range of samples being presented for evaluation. There are significant implications if the scale does not yield the same result when a wider range of speech samples are present; describing a naturalness score as an absolute value would be meaningless. It would only be meaningful to discuss the naturalness of a speech sample as *better* as or *worse* than the pool from which it was drawn. Yet this would defeat the point of a quantitative scale, reducing it to a bimodal measure. Further, knowing whether manipulating the range of speech naturalness samples provided for evaluation affects scores for those samples also has implications for the ease and speed with which post stuttering treatment speech naturalness can be assessed in treatment outcome studies; if the range of samples for evaluation does not impact naturalness scores, the time consuming and laborious task of recruiting matched control speakers to post treatment participants is not necessary.

This study aimed to observe: (1) whether manipulating the range of speech naturalness samples presented for evaluation, by including control samples, influenced the rating scores assigned by untrained listeners and final year SLP students using the speech naturalness scale of Martin et al. (1984); (2) listener group differences in post-treatment speech naturalness evaluation; and (3) listener sex differences in speech naturalness evaluation.

2. Materials and methods

Ethics approval for this research was obtained from the La Trobe University Faculty of Health Sciences Human Ethics Committee (FHEC 14/5). Informed consent was obtained from all participants.

2.1. Participants

Thirty-nine listeners were recruited to judge speech samples for naturalness. Twenty of these were untrained listeners from the general community and nineteen were final year speech language pathology (SLP) students. Prospective participants were asked to self disclose the presence of hearing abnormality. None were excluded based on this. No participant had prior knowledge of the purpose of the study.

The untrained listener group consisted of 10 men and 10 women with a mean age of 26.7 years (SD = 9.2). These participants were administration and marketing staff, and university students undertaking degrees not related to the health sciences, recruited from La Trobe University in Melbourne, Australia. The group of SLP students were final year Masters students consisting of 19 women with a mean age of 23.8 years (SD = 2.7). Having only women is representative of SLP student cohorts at this university where 1–5% of student SLPs are men. These participants were recruited from the La Trobe University Communication Clinic following completion