Journal of Communication Disorders 76 (2018) 11–20

Contents lists available at ScienceDirect

Journal of Communication Disorders

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

Perceptual evaluation of hypernasality, audible nasal airflow and speech understandability using ordinal and visual analogue scaling and their relation with nasalance scores

Kim Bettens[a,*], Laura Bruneel[a], Youri Maryn[a,b,c,d], Marc De Bodt[a,e], Anke Luyten[a], Kristiane M. Van Lierde[a,f]

[a] Department of Speech, Language and Hearing Sciences, Ghent University, Ghent, Belgium
[b] Department of Otorhinolaryngology, Speech-Language Pathology, Sint-Augustinus General Hospital, Wilrijk, Antwerp, Belgium
[c] Department of Speech-Language Therapy and Audiology, Faculty of Education, Health and Social Work, University College Ghent, Ghent, Belgium
[d] Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium
[e] Rehabilitation Centre for Communication Disorders, Antwerp University Hospital, Wilrijk, Belgium
[f] Department of Speech-Language Therapy and Audiology, University of Pretoria, Pretoria, South Africa

ARTICLE INFO

Keywords:
Perceptual evaluation
Rating scales
Resonance
Hypernasality
Understandability
Speech

ABSTRACT

Purpose: Perceptual assessments remain the most commonly utilized procedure to diagnose and evaluate resonance disorders. However, the discussion continues about which rating scale has to be applied. Therefore, this study aimed to compare the reliability and validity of ordinal and visual analogue scales to rate hypernasality, audible nasal airflow and speech understandability.
Methods: Four experienced speech-language pathologists rated 35 speech samples of children with a range of hypernasality, audible nasal airflow and speech understandability, using an ordinal scale and a visual analogue scale. Intraclass correlations coefficients determined intra- and inter-rater reliability. The model of best fit was determined by plotting both rating scales against each other. A Pearson correlation coefficient verified the relationship between both rating scales and nasalance scores determined by a Nasometer.
Results: Good intra- and inter-rater reliability was found for both rating scales. A multiple regression analysis revealed a curvilinear relationship between both rating scales, indicating a slight preference to rate all parameters by a visual analogue scale. Comparable correlations with nasalance scores were found.
Conclusions: This study confirms that visual analogue scale ratings form a reliable and valid alternative for ordinal ratings in the perceptual judgments of hypernasality, audible nasal airflow and speech understandability. A combination of both rating scales may even combine the advantages and eliminate their limitations. However, further research is necessary to verify how this new approach can be implemented in available protocols for clinical practice, audits and research.

* Corresponding author at: Department of Speech, Language and Hearing Sciences, Ghent University, Corneel Heymanslaan 10, 2P1, BE-9000 Gent, Belgium.
E-mail address: Kim.Bettens@Ugent.be (K. Bettens).

https://doi.org/10.1016/j.jcomdis.2018.07.002
Received 28 April 2017; Received in revised form 4 July 2018; Accepted 20 July 2018
Available online 22 July 2018

1. Introduction

To diagnose resonance disorders, perceptual judgments are still considered the gold standard since no instrumental measurement can transcend the capabilities of a trained ear (Henningsson, Kuehn, Sell, Sweeney, Trost-Cardamone, & Whitehill, 2008). Hence, perceptual judgments need to be valid and reliable. Recently, initiatives have been taken to standardize perceptual assessments and to explore both validity and reliability of perceptual assessment protocols for resonance disorders in patients with cleft palate. In order to meet the demand for a uniform speech sample, Henningsson et al. (2008) proposed universal parameters that can be applied in several languages to compose a consistent speech sample. However, the reliability and validity of those samples are not yet confirmed. Another speech analysis protocol to judge hypernasality, nasal airflow and articulation, more specifically the Cleft Audit Protocol for Speech-Augmented (CAPS-A), was developed by John, Sell, Sweeney, Harding-Bell, and Williams (2006). Validity, reliability and acceptability of this protocol were proven by several studies (Britton, Albery, Bowden, Harding-Bell, Phippen, & Sell, 2014; Chapman et al., 2016; John et al., 2006). All these perceptual assessment protocols include categorical scales with clear description of the different grades to judge hypernasality, audible nasal airflow (ANA) and intelligibility. Moreover, categorical scaling was applied in 74% of the studies that included a perceptual assessment of cleft palate speech, as reported in the critical review of Lohmander and Olsson (2004).

However, the discussion continues about the type of rating scale that has to be applied to judge resonance reliably (Baylis, Chapman, Whitehill, & Group, 2015; Brancamp, Lewis, & Watterson, 2010; Whitehill, Lee, & Chun, 2002; Zraick & Liss, 2000). This discussion is not only restricted to the perception of resonance, but also occurs in the domain of other speech dimensions like voice (Karnell, Melton, Childes, Coleman, Dailey, & Hoffman, 2007; Wuyts, De Bodt, & Van de Heyning, 1999; Yiu & Ng, 2004) and stuttering (Schiavetti, Sacco, Metz, & Sitler, 1983). The discussion is related to the mental processing behind the perception of resonance (Stevens, 1975). Stevens (1975) differentiates two kinds of perceptual continua: metathetic and prothetic. Metathetic dimensions involve qualitative perceptual changes, while prothetic dimensions involve quantitative perceptual changes. As a result, metathetic stimuli are more concerned with 'what kind' (e.g., pitch), whereas prothetic stimuli are concerned with 'how much' (e.g., loudness) (Roeckelein, 1998). Due to this difference in mental processing, Stevens (1975) postulated that different perceptual phenomena should be rated by different scales. More specifically, he proposed to rate metathetic stimuli by partition measures, whereas prothetic stimuli should be rated by using magnitude measures. Examples of partition measures are equal appearing interval (EAI) and ordinal scales, in which the listener chooses between a finite set of numbers (0 to N) or categories to rate a specific stimulus, representing the quality of the stimulus. Magnitude measures, on the other hand, represent quantitative measures by assigning numbers to stimuli in proportion to their magnitude, for example direct magnitude estimation (DME) and visual analogue scales (VAS) (Yiu & Ng, 2004).

For judgments based on EAI scales, whole numbers are used to divide the scale in equal intervals. The endpoints are fixed in which odd-numbered scales with 5–7 points are usually applied to rate resonance disorders in clinical practice. When adjectives or definitions are added to the different numbers, intervals are not equally appearing anymore, but are defined by the descriptions, resulting in an ordinal scale. Although both scales are fundamentally categorical (Wuyts et al., 1999), Castick, Knight, and Sell (2017) cautioned that researchers should not use the terms 'equal appearing interval' and 'ordinal' interchangeably.

Categorical (or partition) scales are widely used to rate resonance disorders as they are accessible, easy to use and interpret and it is easy to compare the results of different listeners or patients (Baylis et al., 2015). Nevertheless, the validity of this type of rating scales is questioned as several psychophysical experiments denoted that listeners divide the lower end of the scale into smaller intervals (Stevens, 1975). This results in a systematic bias towards the lower part of the continuum (Brancamp et al., 2010). Additionally, only categorical data are provided which limits the subsequent statistical analysis (Brancamp et al., 2010). Hence, the use of magnitude measures scales, such as DME and VAS, was explored to rate resonance. DME can be applied with a modulus, or standard for comparison (DME-M), or without a modulus (DME-WM). During DME-M, a specific value is assigned to a standard speech sample (i.e. the modulus) after which the listener rates all speech samples relative to the magnitude of the modulus. For example, if the value of the modulus was set at 100 and a subsequent speech sample is judged to be twice as nasal as the modulus, a value of 200 will be given to that sample. When a DME-WM is used, the listener has to assign a value to the first sample, after which all other samples are compared with this first. As this is a ratio scale, no systematic bias associated with categorical scales occurs. Moreover, there are more options regarding statistical analyses (Brancamp et al., 2010). Nevertheless, the procedure is complex because it requires more explanation and training of the listeners, a more complicated presentation of the stimulus and complex statistical analysis (Baylis et al., 2015). Therefore, implementation of this rating procedure in a clinical setting seems difficult. Hence, VAS was explored by Baylis et al. (2015) as an alternative magnitude measures scale to judge resonance disorders since VAS and DME ratings of hypernasality seem to be strongly correlated (Cheng, 2006). A 100 mm bar is presented to the listeners who have to place a mark on the bar going from 0 (normal) to 100 (most severely disturbed). This results in a continuous level of measurement suggesting an easier implementation in daily clinical practice because of its ease of use and the more convenient analysis (Baylis et al., 2015).

To determine whether a perceptual phenomenon is prothetic or metathetic, Stevens (1975) postulated to explore the relationship between partition and magnitude scale ratings based on judgments of the same speech sample. If the ratings are related to each other in a linear way, the rated phenomenon can be considered to be metathetic. If a quadratic or curvilinear relationship is found, the rated phenomenon would be prothetic. In the literature, most of the studies reported a curvilinear relationship between partition and magnitude measure scale ratings of hypernasality, suggesting that hypernasality is rather a prothetic phenomenon that can be rated more validly by using magnitude measures scales such as DME or VAS (Baylis, Munson, & Moller, 2011, 2015; Whitehill et al., 2002; Zraick & Liss, 2000, see Supplementary material). Brancamp et al. (2010), on the other hand, reported no statistically significant differences between ratings of hypernasality based on EAI and DME scales based on the judgments of one rater and found a linear