Speech-Language Therapy students' auditory-perceptual judgements of simulated concurrent hypernasality and articulation disorders

Alice Lee[1], Sarah Potts[1] and Tim Bressmann[2]

[1]Department of Speech and Hearing Sciences, University College Cork, Cork, Ireland

[2]Department of Speech-Language Pathology, University of Toronto, Toronto, Canada

Address correspondence to: Alice Lee, Department of Speech and Hearing Sciences, University College Cork, Brookfield Health Sciences Complex, College Road, Cork, Republic of Ireland; e-mail: a.lee@ucc.ie; tel.: +353 (0)21 490 1540; fax: +353 (0)21 490 1542.

ABSTRACT

Auditory-perceptual judgements are regarded as the standard method for assessing speech disorders. However, the results of auditory-perceptual evaluations and rater reliability can be affected by various factors, such as concurrent problems in multiple speech subsystems. This study investigated the effect of a co-occurring articulation disorder on auditory-perceptual judgements of hypernasality and the effect of co-occurring hypernasality on judgements of an articulation disorder. The speech stimuli were sentences produced by a male speaker who simulated four levels of hypernasality (typical nasality, and mild, moderate, and severe hypernasality) at four levels of disordered articulation (typical articulation, and mild, moderate, and severe articulation disorder). Thirty speech and language therapy students used visual analogue scales to rate the severity of hypernasality and articulation disorder for each speech sample. Results showed that the hypernasality ratings were significantly higher when articulation disorder co-occurred compared to those without. However, there was no significant difference between mild, moderate and severe concurrent articulation disorder on hypernasality ratings. The speech samples with typical articulation and those with severe articulation disorder were rated as more severe in terms of articulation problem when combined with severe hypernasality. However, there was no significant hypernasality effect on articulation ratings for speech with mild or moderate articulation disorder. The present results generally agreed with previous findings regarding the effect of co-occurring speech problems on auditory-perceptual judgements. Clinicians are advised to be cautious of the potential impact. If possible, speech evaluation using instrumental techniques should be used to supplement auditory-perceptual judgements.

Keywords: Auditory-perceptual judgements, hypernasality, articulation disorders, speech disorders, rater reliability

INTRODUCTION

Auditory-perceptual judgements are the primary tool that speech and language therapists use to document speech disorders (Kent, 1996). Assessment of articulation, the balance of oral-nasal resonance and voice quality always begins with auditory-perceptual judgements. Clinicians rely largely on these judgements to diagnose the speech problem and to decide on the need for further instrumental measures. However, auditory-perceptual assessment is an inherently subjective method. Several authors have identified various factors that can affect the auditory-perceptual evaluation and one such factor is the type of speech materials used (e.g., Gerratt, Kreiman, & Garellek, 2016; Klintö, Salameh, Svensson, & Lohmander, 2011). It is also important to consider the nature of the perceptual dimension to be rated (whether it is a prothetic or metathetic continuum; see Stevens, 1975) and the appropriateness of the rating scale for the corresponding speech dimension (e.g., Baylis, Chapman, Whitehill, & TheAmericleftSpeechGroup, 2015; Whitehill, Lee, & Chun, 2002; Yiu & Ng, 2004). Auditory-perceptual ratings may also be influenced by specific characteristics of the listeners, such as the extent of their relevant clinical experience (e.g., Bunton, Kent, Duffy, Rosenbek, & Kent, 2007; Helou et al., 2010), the specific training they receive before the task (e.g., Eadie & Baylor, 2006; Ghio et al., 2015; Lee, Whitehill, & Ciocca, 2009), as well as listeners' linguistic background (e.g., Hartelius, Theodoros, Cahill, & Lillvik, 2003; Lee, Brown, & Gibbon, 2008; Yamashita, Borg, S., & Lohmander, 2018) and age (Goy, Pichora-Fuller, & van Lieshout, 2016). How much information about the speech samples or speakers is made available to the listeners can also make a difference on their auditory-perceptual ratings (Ramig, 1982). Finally, the listeners' auditory-perceptual assessment in one perceptual dimension, such as nasality, may be affected by concurrent disorders in other speech subsystems, such as voice quality or articulation (e.g., Dattilo, 2016; Imatomi, 2005; Starr, Moller, Dawson, Graham, & Skaar, 1984; Tardif, Berti, Marino, Pardo,

& Bressmann, 2018). The last factor is perhaps the most challenging to address and may be of high clinical relevance. Many speech disorders are characterised by symptoms in more than one speech subsystem. This is a common observation in speech disorders caused by congenital or acquired structural deficits or deviations (e.g., cleft palate and craniofacial anomalies, ablative surgery to structures in the oral cavity) as well as in developmental or acquired neurological impairments (e.g., motor speech disorders). The speech produced is the product of acoustic effects of affected and unaffected speech subsystems. Clinicians then have the difficult task of identifying individual atypical speech features in different subsystems and judging their severity independently from each other.

Previous studies have investigated the effect of concurrent disorders in different speech subsystems on auditory-perceptual judgements that also included ratings of hypernasality (Counihan & Cullinan, 1972; Dattilo, 2016; Hess, 1959; Imatomi, 2005; Imatomi & Arai, 2002; Imatomi, Arai, & Kato, 2000, 2003; Imatomi, Arai, Mimura, & Kato, 1999; Sherman & Goodwin, 1954). One study investigated the effects of hypernasality on ratings of articulation (Starr et al., 1984). Another study focused on the effects of hypernasality on ratings of different voice qualities, that is, breathiness, harshness, and hoarseness (Hess, 1959). Tardif et al. (2018) assessed the effect of simulated hypernasality on the perception of speech intonation. The findings of these studies are summarised in table 1.

Insert table 1 about here

The finding of these studies have been varied. Sherman and Goodwin (1954) found no significant pitch effect on hypernasality ratings for female speakers. However, the authors reported that passages produced at a lower-than-habitual pitch level by male speakers with history of cleft palate were perceived as significantly less nasal than the passages produced at habitual pitch and higher-than-habitual pitch level. In contrast, a later study reported significantly lower hypernasality ratings when the vowels were produced at a higher-than-habitual pitch level than the habitual pitch level of the speakers (Hess, 1959). Conflicting results were also reported regarding the effect of different intensity levels on perceptual ratings of hypernasality, with one study finding significantly lower nasality ratings for vowels produced at a higher intensity level by male speakers (Hess, 1959) but another study finding a significant intensity effect with a general trend of lower nasality ratings when vowels were produced at lower intensity levels by the female and male speakers (Counihan & Cullinan, 1972; see table 1 for detail).

Imatomi and colleagues described differential effects of co-existing breathiness or hoarseness on perceived level of nasality. When severe hoarseness or higher degree of breathiness co-occurred in the speech samples (synthesised vowels), the stimuli of moderate or severe hypernasality were perceived as less nasal but those of mild or no hypernasality were rated as more nasal (Imatomi, 2005; Imatomi & Arai, 2002; Imatomi et al., 2000, 2003; Imatomi et al., 1999). Dattilo (2016) investigated the effect of articulation errors on perceptual judgements of nasality by 20 undergraduate and 20 graduate students in speech-language pathology in her Master's thesis. The author reported that the nasality ratings for passages produced by children with different levels of hypernasality associated with cleft palate were significantly lower when mild articulation disorders co-occurred in the speech samples. However, there were no significant differences in the nasality ratings between the conditions of concurrent moderate and severe articulation disorders.

In turn, hypernasality can also affect the auditory-perceptual evaluation of other speech subsystems. One study showed that the ratings of articulation deviation, based on a passage reading task, were higher when mild or moderate (hyper)nasality co-existed (Starr et al., 1984). Another study reported that listeners rated sentences of different intonation levels (simulated by two voice actors) was monotonous when co-occurring hypernasality increased (Tardif et al., 2018). For auditory-perceptual ratings of breathiness, harshness, and hoarseness, Hess (1959) found that the speakers were perceived to be significantly less breathy when they produced the vowels at a higher vocal intensity level but there was no significant pitch effect on the breathiness ratings. In addition, the speakers were perceived to be significantly less harsh for the higher pitch and lower intensity conditions; and significantly less hoarse for the higher pitch and higher intensity conditions.

In individuals with repaired cleft lip and palate, hypernasality and articulation disorders often co-occur, and clinicians have to evaluate the nature and severity of the balance of oral-nasal resonance and articulation disorders. While the first study by Dattilo (2016) showed that nasality ratings may be influenced by a concurrent articulation disorder, the study only investigated possible cross-contamination effects in one direction, that is, it did not evaluate whether the presence of hypernasality would influence the assessment of severity of an articulation disorder. The study also relied on a convenience sample of clinical data that were rated by speech-language pathology students. While this ensured a degree of clinical realism of the study, it is difficult to curate clinical data to create exact and consistent gradations of the clinical speech features of interest. Tardif et al. (2018) used voice actors to create a controlled data set of speech samples with consistent levels of intonation and nasality. This allowed the authors to vary the severity of one auditory-perceptual dimension while holding the other aspect constant. The current study aimed to further investigate the relationship between ratings of articulation and hypernasality severity using methodology similar to the one used by Tardif et al. (2018), that is, using speech samples of co-occurring hypernasality and articulation disorder simulated by a typical speaker. The research had the following hypotheses: (1) increasing the severity of a simulated articulation disorder would result in more severe ratings of simulated hypernasality; and (2) increasing the severity of simulated hypernasality would results in more severe ratings of a simulated articulation disorder.

METHODS

Participants

Participants were recruited from a pool of 51 third and fourth year students of the BSc Speech and Language Therapy programme (i.e., the last two years of their study) at the University College Cork (UCC), Ireland. Students of Speech Language Therapy were recruited in order to accrue a sufficient number of participants within a reasonably short timeframe. Thirty of the students (26 females and four males) agreed to take part in the study. Fourteen of them were year 3 students and 16 of them were year 4 students. All participants were native speakers of English and had no hearing difficulties based on self-report. By the time of data collection, the students had already completed the curriculum on auditory-perceptual evaluations, oral-nasal balance disorders and articulation disorders. The general aim of the study was explained to each participant during participant recruitment. They had opportunities to ask questions related to the study, and written consent was obtained from each participant before data collection. Ethical approval for this research was granted by the Clinical Research Ethics Committee of the Cork Teaching Hospitals.

Materials

The speech samples were prepared by the third author. The speech stimulus based on the first three sentences of the Zoo passage (Fletcher, 1972): “Look at this book with us. It’s a story about a zoo. That is where the bears go.” The Zoo passage is an English lipogram devoid of nasal consonants. It is often used in nasometric evaluation of hypernasality. The third author (an adult male speaker with typical craniofacial structures) read the sentences using typical articulation and resonance and then simulated articulation disorders only, hypernasality only, and co-occurring articulation disorders and hypernasality when producing these sentences again. There were four levels for hypernasality – typical nasality, and mild, moderate, and severe hypernasality. There were also four levels for articulation – typical articulation, and mild, moderate, and severe articulation disorder. Thus, the four levels of hypernasality by four levels of articulation resulted in 16 different combinations of hypernasality and articulation severity.

The simulation of hypernasal speech was guided by nasalance scores, as measured using the Nasometer II Model 6450 (Kay Pentax, Lincoln Park, NJ), which quantifies the ratio of nasal to nasal plus oral acoustic energy during speech production. Higher nasalance scores correspond to more severe hypernasality. The nasalance scores ranged from 11 to 18 for the four speech samples of typical nasality, 28-35 for mild hypernasality, 44-50 for moderate hypernasality, and 58-67 for severe hypernasality. The nasalance scores were generally comparable to the nasalance data and the corresponding perceived level of hypernasality reported in the literature (e.g. Vallino-Napoli & Montgomery, 1997).

For articulation disorder, the mild level was represented by an interdental lisping. The moderate and severe levels were simulated using two tongue contortions manoeuvres described in detail in Bressmann (2012). Moderate articulation disorder was achieved by anchoring the tip of the tongue behind the lower incisors. Severe articulation disorder was simulated by holding the left lateral free margin against the lateral floor of mouth.

The recordings were made in a sound booth using a Zoom Q3 digital audio-recorder (Zoom North America, Hauppauge, NY) with a sample frequency of 44.1 kHz and a signal resolution of 16 bit, saved to the *.wav file format. The audio recorder was mounted on a camera tripod to ensure a constant mouth-to-microphone distance. After the recording session, the third author and a doctoral student listened to the 16 speech samples through Philips SNL3000RD headphones (Philips Canada, ON) in a quiet office space, to make sure each sample was perceived as adequate for representing the four levels of articulation and nasality conditions. This was followed by independent listening and judgements by the first author. No disagreement was noted. Furthermore, two speech samples of voice disorders (a hoarse voice and a high-pitched voice), which were not the speech dimensions investigated in this study, were simulated by the second author. These two speech samples were used in the practice trials to familiarise the participants with the format of the auditory-perceptual judgement task before the start of the experiment.

Procedures

All participants rated the 16 speech samples two times in order to establish intra-rater reliability. A randomised sequence was generated for presenting the 32 speech samples. The sequence was checked to make sure that no speech sample was presented two times consecutively. The task of the auditory-perceptual judgement was presented individually to each participant using Microsoft PowerPoint Presentation and a headphone (Sony MDR-V150) in the Speech and Hearing Lab in Department of Speech and Hearing Sciences, UCC. The participants were instructed to (1) listen to the entire speech sample, (2) rate the severity of hypernasality, (3) listen to the speech sample a second time, and (4) rate the severity of articulation disorders. The participants were informed that they were free to take a break whenever they needed.

The participants' ratings were recorded on a response sheet which included two 10 cm visual analogue scales (VAS) – one for hypernasality and one for articulation disorder – for each speech sample. VAS was used in the present study because a recent research by Baylis et al. (2015) showed that VAS is a valid rating scale for rating hypernasality and that their listeners, who were experienced speech and language therapists, showed better reliability within and between the listeners compared to that of equal-appearing interval (EAI) scale. The left end of the VAS represented "Normal" and the right end "Very severe". These descriptions were printed on the response sheet. The description "Moderate" was indicated at the mid-point of the scale at the top of each response sheet. The participants were instructed to record their ratings by marking "X" on the 10 cm lines at a point that corresponded to their perceived level of severity for hypernasality and articulation disorders.

Data analysis

The second author manually measured the marking on the VAS to obtain numerical values of the ratings. All ratings were measured to one decimal point. When ratings were located between millimetres, the ratings were rounded up. Mean ratings and standard deviation for the perceptual judgements of hypernasality and articulation disorders were calculated across the 30 listeners (based on their first rating) for each of the 16 speech samples. Statistical analysis was conducted using IBM SPSS Statistics Version 21. A linear mixed effect (LME) model (with maximum likelihood estimation) was used to analyse the resonance ratings and articulation ratings, with resonance condition (four levels), articulation condition (four levels), and two-way interactions, resonance × articulation, as fixed effects and listener (n = 30) as random effect. The intra-class correlation coefficient (ICC) was calculated to assess the intra-rater and inter-rater reliability. A two-way mixed effects ICC for single measurement and absolute agreement was used for the intra-rater reliability (Koo & Li, 2016). For the inter-rater reliability, a two-way random-effects ICC for average-measures and absolute agreement was used (Hallgren, 2012).

RESULTS

Effect of articulation on hypernasality judgements

The mean resonance ratings for the 16 speech samples by the 30 listeners (based on their first rating) are displayed in figure 1, with the numerical results for mean and standard deviation detailed in table 2. The results of the statistical analysis showed that there were significant main effects of hypernasality [$F(3,480) = 74.24$, $p < 0.001$] and articulation [$F(3,480) = 47.11$, $p < 0.001$] and a significant interaction effect for hypernasality × articulation [$F(9,480) = 2.13$, $p = 0.03$]. Analyses of the main effect of articulation within hypernasality revealed significant effect of articulation on hypernasality ratings in each level of resonance conditions (see table 2). Pairwise comparisons using Bonferroni adjustment showed that, for the four speech samples with typical nasality but different level of articulation conditions, the resonance ratings for the ones with concurrent moderate and severe articulation disorder were significantly higher than that of the one with concurrent typical articulation (significance level adjusted to 0.008 because of six multiple comparisons; see also table 2). For the four speech samples with mild hypernasality, the ones with concurrent mild, moderate and severe articulation disorder had significantly higher hypernasality ratings than the speech sample with concurrent typical articulation ($p < 0.008$). Similar results were obtained for the category of moderate hypernasality, where the nasality ratings for the speech samples with concurrent mild, moderate and severe articulation disorder were significantly higher than that of the one with typical articulation ($p < 0.008$). For severe hypernasality, the resonance ratings of the speech sample with concurrent moderate articulation disorder were significantly higher than that of the one with typical articulation ($p < 0.008$).

Insert figure 1 and table 2 about here

Regarding the listeners' reliability in perceptual judgements of hypernasality, the ICC was 0.70 with 95% CI (0.65 to 0.74) for intra-rater reliability and 0.96 with 95% CI (0.93, 0.98) for inter-rater reliability.

Effect of hypernasality on articulation judgements

The mean articulation ratings for the 16 speech samples by the 30 listeners (based on their first ratings) are shown in figure 2. The numerical results for mean and standard deviation are summarised in table 3. There were significant main effects for hypernasality [$F(3,480) = 11.88$, $p < 0.001$] and articulation [$F(3,480) = 224.21$, $p < 0.001$] and a significant interaction effect for hypernasality × articulation [$F(9,480) = 3.22$, $p = 0.001$]. The results of analyses of the main effect of hypernasality within articulation showed that there was significant effect of hypernasality on articulation ratings for typical articulation and severe articulation disorder ($p < 0.001$). However, the effect of hypernasality on articulation ratings was not significant when the articulation disorder was mild or moderate ($p > 0.0125$; significance level was adjusted for four comparisons; see table 3). Pairwise comparisons showed that for the four speech samples with typical articulation, the articulation rating of the one with concurrent severe hypernasality was significantly higher than those of the ones with typical resonance, mild and moderately hypernasality ($p < 0.008$; see table 3). For the category of severe articulation disorder, the articulation ratings of the speech sample with concurrent severe hypernasality were significantly higher than those of the ones with typical nasality and mild hypernasality. The articulation ratings of the speech sample with concurrent moderate hypernasality were significantly higher than that of the speech sample with concurrent mild hypernasality ($p < 0.008$).

Insert figure 2 and table 3 about here

For intra-rater reliability and inter-rater reliability, the ICCs were 0.82 with 95% CI (0.79 to 0.85) and 0.99 with 95% CI (0.96 to 0.99), respectively.

DISCUSSION

This study investigated the effect of different levels of concurrent articulation disorders (typical articulation, and mild, moderate, and severe articulation disorders) on speech and language therapy students' perceptual judgements of hypernasality of different levels (typical nasality, and mild, moderate, and severe hypernasality) and vice versa. The results showed that the speech samples were judged as more nasal when articulation disorders were also present compared to those without any articulation problems. However, the effect of concurrent articulation disorders on nasality ratings did not differ between mild, moderate and severe levels. This means that increasing the severity of the co-occurring articulation disorder did not increase the hypernasality ratings significantly. The present results generally supported our first hypothesis. The results were also in congruence with the findings reported by Dattilo (2016) that hypernasality ratings were significantly higher when moderate or severe articulation disorder co-occurred compared to concurrent mild articulation disorder. The author also reported that there was no further significant increase in hypernasality ratings when the severity of the concurrent articulation disorder was increased from moderate to severe. Despite some differences in the methodology between Dattilo's study and the current study (such as speech samples from children with repaired cleft palate versus speech samples simulated by a typical adult; speech rated using EAI scale versus VAS), both studies reported a significant effect of a concurrent articulation disorder on perceptual ratings of hypernasality.

The second finding of this study was that the speech samples with typical articulation and those with severe articulation disorder were rated as more severe in terms of articulation disorder when severe hypernasality co-occurred. However, there was no converse significant effect of concurrent hypernasality on the articulation ratings for speech with mild or moderate articulation disorder. The effect of concurrent hypernasality found in this study was inconsistent and only partially supported the second hypothesis. The finding was generally in agreement with the results of the study by Starr et al.'s (1984), who found a significant nasality effect and a general trend of lower ratings of articulation disorder for typical nasality, followed by mild and moderate nasality.

The results of this study seemed to show that there was a stronger effect of concurrent articulation disorder on auditory-perceptual judgements of hypernasality than the converse effect of concurrent hypernasality on auditory-perceptual judgements of articulation disorder. There are two possible explanations for this finding. First, the results may have been influenced by the sequence of speech dimensions that the listeners rated. They were instructed to listen to a speech sample, rate hypernasality, then listen to the speech sample again and rate articulation disorder. Perhaps, the added exposure to the speech sample before rating articulation helped the listeners rate the severity of this speech dimension. However, as the listeners were instructed to focus the listening on a single speech dimension each time, the advantage of added exposure for rating articulation was probably minimal. Nonetheless, counterbalancing the order of rating the two speech dimensions is recommended for future studies. A second, more plausible reason for this result may have been that the listeners, who were speech and language therapy students, were relatively more proficient in rating articulation disorder than hypernasality. This may have been related to the listener's clinical experience gained so far through their undergraduate training – there were probably more opportunities to evaluate articulation disorder than oral-nasal balance problems in their clinical practice education. Hence, completing auditory-perceptual judgements of the articulation disorders with co-occurring hypernasality was perhaps relatively easier than the reverse task. An inspection of the data in table 2 and 3 shows that the standard deviations – were numerically lower for the articulation ratings than for the hypernasality ratings. The higher intra- and inter-rater reliability for articulation ratings was also higher than for the hypernasality. Both of these observations seem to support the speculation that the listeners may have been more proficient at rating articulation than oral-nasal balance.

The current study included only speech and language therapy students as listeners. While they had the requisite academic and clinical experience to participate in the study, it is unclear in how far the current findings can be generalised to clinicians who have specific clinical experience with the speech characteristics of individuals with cleft lip and palate. Some of the previous studies reviewed in the Introduction involved experienced speech and language therapists as the listeners (Imatomi, 2005; Imatomi & Arai, 2002; Imatomi et al., 2000, 2003; Imatomi et al., 1999; Starr et al., 1984). Similar to the present study, these previous studies reported significant effects of a concurrent atypical speech feature on the auditory-perceptual judgements of another speech subsystem. Hence, one might predict that similar results would have been obtained had the present study been conducted with experienced listeners. None of the previous studies reviewed above (and in table 1) compared the ratings between less experienced listeners and experienced listeners, with the exception of Starr et al. (1984). They included six groups of listeners, including four groups of adult listeners: speech and language therapists at a cleft palate clinic, speech and language therapists at public schools, parents of children with repaired cleft palate, and parents of typically developing children. The authors found that the school clinicians gave significantly higher (worse) articulation ratings on average than the parents of children with repaired cleft palate. There was no significant different in ratings between the other groups of adult listeners. It is unclear what factors might have led to this result, and the authors offered no explanation.

A few ideas for future research emerged from the findings of the present study. First, further study on the interaction between the effect of concurrent speech disorders on auditory-perceptual judgements and the amount of clinical experience of the listeners would be useful. Second, a number of previous studies have shown that training can improve listeners' auditory-perceptual judgements and/or intra- and inter-rater reliability (e.g., Lee et al., 2009). However, whether training can reduce the impact of concurrent speech disorders on auditory-perceptual ratings of the speech feature that was the object of the training is unknown. Third, the present study included only one male speaker. It is unclear how vocal characteristics of male and female speaker might interact with the effect of co-occurring speech disorders on auditory-perceptual judgements. Further research on this topic is warranted.

In conclusion, the presence of disordered features in one speech subsystem can have an effect on the auditory-perceptual ratings of a different speech subsystem. Clinicians need to be aware of the potential impact of concurrent atypical speech features on their auditory-perceptual judgements. Further speech evaluation using instrumental techniques, such as acoustic analysis, should be used to supplement and corroborate auditory-perceptual judgements.

ACKNOWLEDGEMENTS

The authors would like to thank the participants who took part as the listeners in this study and to Gillian de Boer, PhD, who assisted in the preparation of the speech stimuli. Part of the work reported in this study constituted a final-year honours project of Sarah Potts, University College Cork, Ireland, 2017. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

DECLARATION OF INTEREST

The authors report no conflicts of interest.

References

Baylis, A., Chapman, K., Whitehill, T. L., & The Americleft Speech Group. (2015). Validity and reliability of visual analog scaling for assessment of hypernasality and audible nasal emission in children with repaired cleft palate. *Cleft Palate-Craniofacial Journal, 52*(6), 660-670.

Bressmann, T. (2012). An ultrasonographic study of lingual contortion speech. *Journal of Phonetics, 40*(6), 830-836.

Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F. (2007). Listener agreement for auditory-perceptual ratings of dysarthrias. *Journal of Speech, Language & Hearing Research, 50*(6), 1481-1495.

Counihan, D. T., & Cullinan, W. L. (1972). Some relationships between vocal intensity and rated nasality. *Cleft Palate Journal, 9*(2), 101-108.

Dattilo, K. L. (2016). *The effects of articulation errors on perceived nasality in speakers with repaired cleft lip and/or palate (Master's thesis).* Retrieved from https://ir.uiowa.edu/cgi/viewcontent.cgi?article=6411&context=etd

Eadie, T. L., & Baylor, C. R. (2006). The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *Journal of Voice, 20*(4), 527-544.

Fletcher, S. G. (1972). Contingencies for bioelectronic modification of nasality. *Journal of Speech & Hearing Disorders, 37*, 329-346.

Gerratt, B. R., Kreiman, J., & Garellek, M. (2016). Comparing measures of voice quality from sustained phonation and continous speech. *Journal of Speech, Language & Hearing Research, 59*(5), 994-1001.

Ghio, A., Dufour, S., Wengler, A., Pouchoulin, G., Revis, J., & Giovanni, A. (2015). Perceptual evaluation of dysphonic voices: Can a training protocol lead to the development of perceptual categories? *Journal of Voice, 29*(3), 304-311.

Goy, H., Pichora-Fuller, K., & van Lieshout, P. (2016). Effects of age on speech and voice quality ratings. *Journal of the Acoustical Society of America, 139*(4), 1648-1659.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34.

Hartelius, L., Theodoros, D., Cahill, L., & Lillvik, M. (2003). Comparability of perceptual analysis of speech characteristics in Australian and Swedish speakers with multiple sclerosis. *Folia Phoniatrica Et Logopaedica, 55*, 177-188.

Helou, L. B., Solomon, N. P., Henry, L. R., Coppit, G. L., Howard, R. S., & Stojadinovic, A. (2010). The role of listener experience on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ratings of postthyroidectomy voice. *American Journal of Speech-Language Pathology, 19*(3), 248-258.

Hess, D. A. (1959). Pitch, intensity, and cleft palate voice quality. *Journal of Speech & Hearing Research, 2*(2), 113-125.

Imatomi, S. (2005). Effects of breathy voice source on ratings of hypernasality. *Cleft Palate-Craniofacial Journal, 42,* 641-648.

Imatomi, S., & Arai, T. (2002, 2002, September). *The relation between perceived hypernasality of cleft palate speech and its hoarseness.* Paper presented at the Forum Acusticum Sevilla, Sevilla, Spain.

Imatomi, S., Arai, T., & Kato, M. (2000, 2000, August). *How hoarseness affects on rating of hypernasality.* Paper presented at the 8th conference of International Clinical Phonetics and Linguistics Association, Edinburgh.

Imatomi, S., Arai, T., & Kato, M. (2003). Effects of hoarseness on ratings of hypernasality. *音声言語医学 [The Japan Journal of Logopedics and Phoniatrics], 44*(4), 304-314.

Imatomi, S., Arai, T., Mimura, Y., & Kato, M. (1999, 1999, September). *Effects of hoarseness on hypernasality ratings.* Paper presented at the European Conference on Speech Communication and Technology, Budapest.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*(3), 7-23.

Klintö, K., Salameh, E.-K., Svensson, H., & Lohmander, A. (2011). The impact of speech material on speech judgement in children with and without cleft palate. *International Journal of Language & Communication Disorders, 46*(3), 348-360.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Lee, A., Brown, S., & Gibbon, F. E. (2008). Effect of listeners' linguistic background on perceptual judgements of hypernasality. *International Journal of Language & Communication Disorders, 43*(5), 487-498.

Lee, A., Whitehill, T. L., & Ciocca, V. (2009). Effect of listener training on perceptual judgment of hypernasality. *Clinical Linguistics & Phonetics, 23*(5), 319-334.

Ramig, L. A. (1982). Effects of examiner expectancy on speech ratings of individuals with cleft lip and/or palate. *Cleft Palate Journal, 19*(4), 270-274.

Sherman, D., & Goodwin, F. (1954). Pitch level and nasality. *Journal of Speech & Hearing Disorders, 19*(4), 423-428.

Starr, C. D., Moller, K. T., Dawson, W., Graham, J., & Skaar, S. (1984). Speech ratings by speech clinicians, parents and children. *Cleft Palate Journal, 21*(4), 286-292.

Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.

Tardif, M., Berti, L. C., Marino, V. C. C., Pardo, J., & Bressmann, T. (2018). Hypernasal speech is perceived as more monotonous than typical speech. *Folia Phoniatrica Et Logopaedica, 70*(3-4), 183-190.

Vallino-Napoli, L. D., & Montgomery, A. A. (1997). Examination of the standard deviation of mean nasalance scores in subjects with cleft palate: Implications for clinical use. *Cleft Palate-Craniofacial Journal, 34*, 512-519.

Whitehill, T. L., Lee, S. Y. A., & Chun, C. J. (2002). Direct magnitude estimation and interval scaling of hypernasality. *Journal of Speech, Language, & Hearing Research, 45*(1), 80-88.

Yamashita, R. P., Borg, E., S., G., & Lohmander, A. (2018). Reliability of hypernasality rating: Comparison of 3 different methods for perceptual assessment. *Cleft Palate-Craniofacial Journal, 55*(8), 1060-1071.

Yiu, E. M.-L., & Ng, C.-Y. (2004). Equal appearing interval and visual analogue scaling of perceptual roughness and breathiness. *Clinical Linguistics & Phonetics, 18*(3), 211-229.

Figure 1. Mean hypernasality ratings of the 16 speech samples (four resonance conditions by four articulation conditions) made by the 30 listeners using visual analogue scale where “0” represents “Normal” and “10” represents “Very severe”.

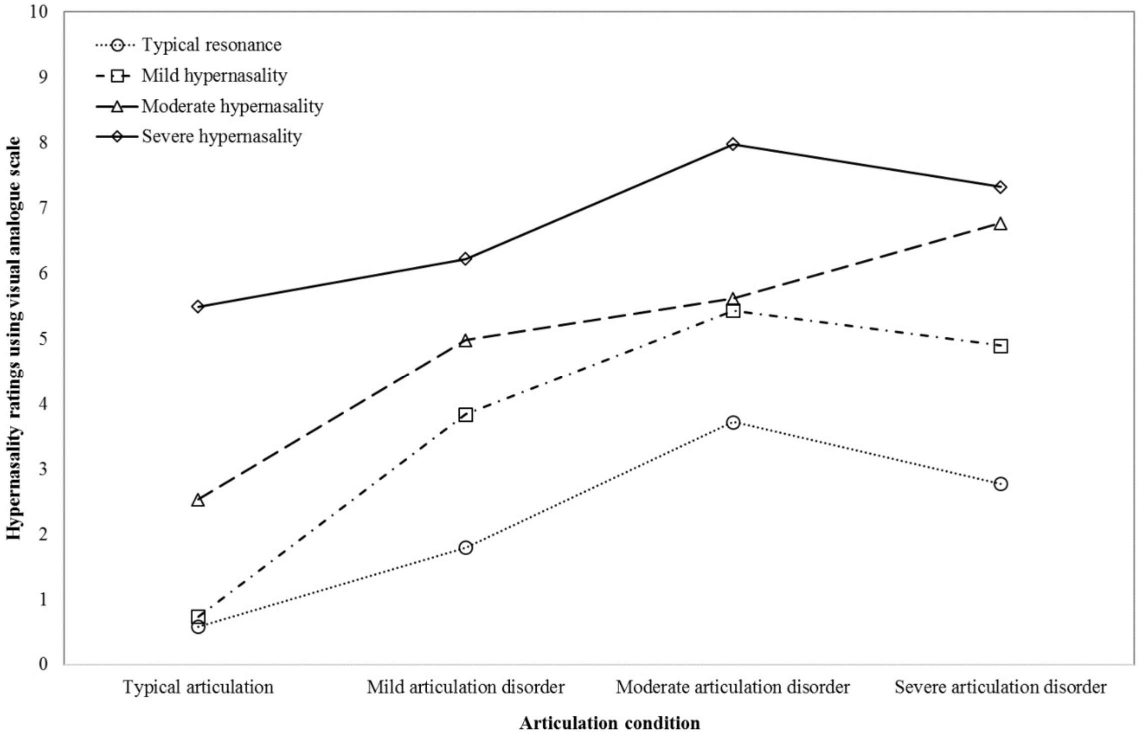

Figure 2. Mean articulation ratings of the 16 speech samples (four resonance conditions by four articulation conditions) made by the 30 listeners using visual analogue scale where “0” represents “Normal” and “10” represents “Very severe”.

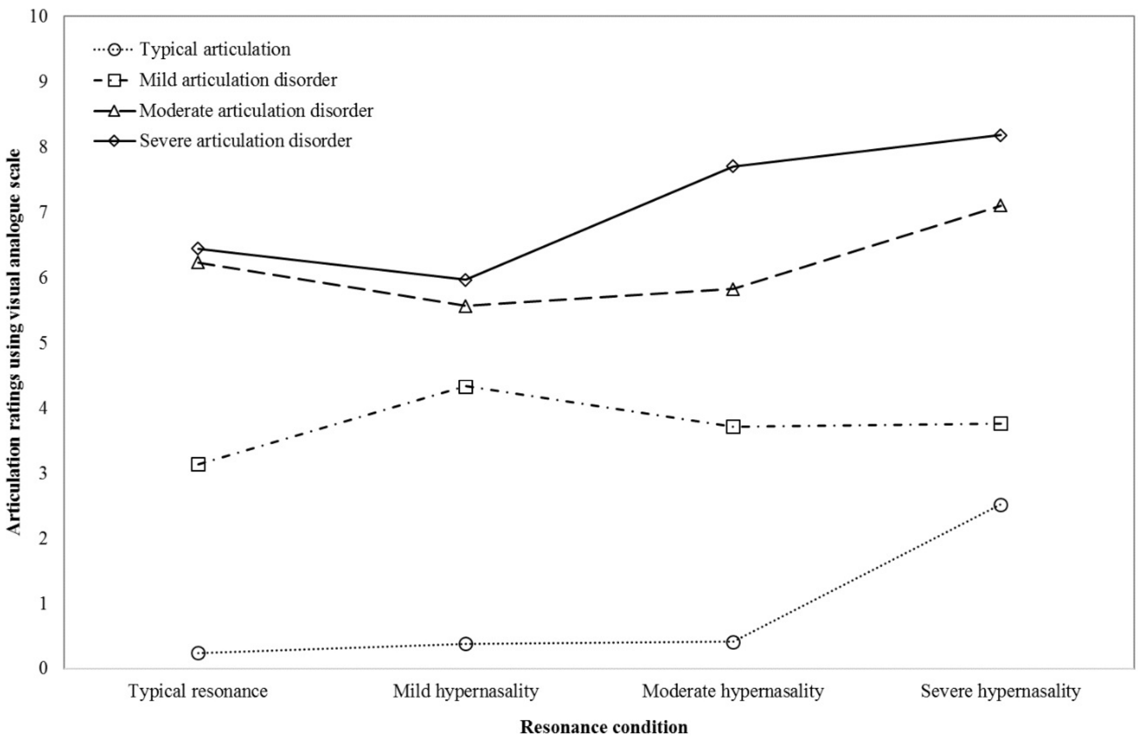

Table 1. Summary of previous findings regarding the effect of concurrent articulation, resonance or voice disorders or attributes on listeners' auditory perception of a speech dimension.

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
Nasality	Pitch (3 levels: lower-than-habitual, habitual, higher-than-habitual)	First paragraph of the Rainbow Passage read by 10 men and 10 women with 'nasal voices' of unknown causes (p. 424)	7-point equal appearing interval (EAI) scale, rated by 30 'seniors' and graduate students in speech pathology	Men: significantly lower nasality ratings for lower-than-habitual level than habitual or higher-than-habitual level. Women: no significant pitch effect.	Sherman & Goodwin (1954)
Nasality	Pitch (2 levels: habitual and a pitch level 1.4 times higher); intensity	6 vowels (/i/, /u/, /e/, /o/, /æ/, /ɑ/) sustained for 2.5 seconds by 15 male mid-teens and	7-point EAI scale, rated by four graduate students experienced in	Significantly lower nasality ratings for higher pitch and higher intensity levels.	Hess (1959)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
	(2 levels: 75 and 85 dB SPL)	adults with history of cleft palate	perceptual judgements of voice quality		
Nasality	Intensity (4 levels: 70, 75, 80, 85 dB SPL)	Vowels /u/ and /ɑ/ sustained for 4 seconds by 10 male and 10 female mid-teens and adults with history of cleft palate, and 20 age- and gender-matched typical speakers	7-point EAI scale, rated by 11 graduate students	Significant intensity effect for /u/ produced by the two groups of females and typical males, and /ɑ/ by males with history of cleft palate; a general trend of lower nasality ratings for lower intensity levels for all except /ɑ/ produced by typical males and females.	Counihan & Cullinan (1972)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
Nasality	Breathiness (6 degrees of breathiness)	21 synthesised stimuli each of vowels /a/ and /i/: 6 breathy source by 3 levels of hypernasality (none, mild-to-moderate, severe), plus 3 original speech samples	5-point EAI scale, rated by 3 experienced speech and language therapists	Concurrent high degree of breathiness lowered the nasality ratings for mild-to-moderate and severe hypernasality; nasality was perceived for stimuli with typical resonance when breathiness co-occurred; stronger effect of concurrent breathiness observed for vowel /i/.	Imatomi & Arai (2002)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
Nasality	Breathiness (6 degrees of breathiness)	21 synthesised stimuli each of vowels /a/ and /i/: 6 breathy source by 3 levels of hypernasality (mild, moderate, severe), plus 3 original hypernasal vowels	5-point EAI scale, rated by 13 speech and language therapists of various amount of clinical experience	Differential effect of breathiness on nasality ratings: significant effect observed for higher degree of breathiness; concurrent high degrees of breathiness raised the nasality ratings for the category of mild hypernasality but lowered those for moderate and severe hypernasality.	Imatomi (2005)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
Nasality	Hoarseness (2 levels: with and without hoarseness)	18 synthesised stimuli each of vowels /a/ and /i/: 3 with and 3 without hoarseness by 3 levels of hypernasality (none, moderate, severe)	5-point EAI scale, rated by 4 experienced speech and language therapists	The nasality ratings decreased for severe hypernasality, varied among listeners for moderate hypernasality, and increased or did not change for typical resonance when hoarseness co-occurred.	Imatomi, Arai, Mimura, & Kato (1999)
Nasality	Hoarseness (2 levels: with and without hoarseness)	4 synthesised stimuli: 2 levels of hoarseness by 2 levels of hypernasality (none, severe)	Rated by ‘several’ speech and language therapists using an unspecified rating procedure	Hypernasality was perceived as less severe when hoarseness co-occurred.	Imatomi, Arai, & Kato (2000)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
Nasality	Hoarseness/roughness (3 levels: none, moderate, severe)	6 synthesised stimuli each of vowels /a/ and /i/: 3 levels of hoarseness/roughness by 2 levels of hypernasality (none, severe)	5-point EAI scale, rated by 4 experienced speech and language therapists	Hypernasality was perceived as less severe when severe hoarseness co-occurred.	Imatomi, Arai, & Kato (2003)
Nasality	Articulation (3 levels: mild, moderate, severe)	The Zoo passage read by 8 children with history of cleft lip and/or palate; the stimuli represented 3 levels of articulation disorders by 3 levels of	6-point EAI scale, rated by 20 undergraduate and 20 graduate students in speech-language pathology	Significantly lower nasality ratings when mild articulation disorders co-occurred; no significant difference in nasality ratings between	Dattilo (2016)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
		hypernasality, except for mild articulation disorders-severe hypernasality		moderate and severe articulation disorders.	
Articulation	Nasality (3 levels: typical resonance, mild nasality, moderate nasality)	A paragraph read by 15 children and young adults with history of cleft palate; the stimuli represented 3 levels of nasality by 3 levels of articulation disorders (none, mild, moderate)	8-point EAI scale, rated by 3-6 experienced ‘clinic clinicians’, 20 school-based ‘speech clinicians’, 12 parents of children with cleft palate, 12 parents of typically developing children, 12 children with nasal speech	Significant nasality effect, with higher ratings of articulation deviation when mild or moderate nasality co-occurred.	Starr, Moller, Dawson, Graham, & Skaar (1984)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
			associated with cleft palate, and 12 typically developing children (pp. 287-288)		
Breathiness		(See item 2 above)		Significantly lower breathiness ratings for higher intensity level; no significant pitch effect.	Hess (1959)
Harshness		(See item 2 above)		Significantly lower harshness ratings for higher pitch and lower intensity levels.	Hess (1959)

Speech dimension rated	Concurrent attribute	Speech samples	Perceptual rating	Relevant results	Reference
Hoarsenesss		(See item 2 above)		Significantly lower hoarseness ratings for higher pitch and higher intensity levels.	Hess (1959)
Intonation	Hypernasality (4 levels: none, mild, moderate, severe)	48 sentences: 2 sentences by 2 voice actors (1 male, 1 female), each simulated 3 levels of intonation (monotone, normal, exaggerated) by 4 levels of hypernasality	Visual analogue scale, rated by 15 female speech-language pathology students and 15 male naïve listeners	Significant hypernasality effect; generally, sentences were perceived as more monotonous as hypernasality level increased.	Tardif, Berti, Marino, Pardo, & Bressmann (2018)

Table 2. Mean hypernasality ratings and standard deviations (in brackets) of the 16 speech samples (four resonance conditions by four articulation conditions) made by the 30 listeners using visual analogue scale where “0” represents “Normal” and “10” represents “Very severe”, with the results of analyses of simple main effect of articulation and pairwise comparisons.

	Articulation				Analyses of simple main effects	Pairwise
Resonance	Typical (0)	Mild (1)	Moderate (2)	Severe (3)	of articulation	comparisons
Typical	0.6 (1.1)	1.8 (1.9)	3.7 (3.2)	2.8 (2.8)	$F(3,480) = 9.06$, $p < 0.001$	0 < 2, 3
Mild	0.7 (0.7)	3.8 (2.7)	5.4 (2.5)	4.9 (3.0)	$F(3,480) = 22.17$, $p < 0.001$	0 < 1, 2, 3
Moderate	2.5 (1.8)	5.0 (2.6)	5.6 (2.7)	6.8 (2.9)	$F(3,480) = 16.08$, $p < 0.001$	0 < 1, 2, 3
Severe	5.5 (2.5)	6.2 (2.2)	8.0 (2.6)	7.3 (3.0)	$F(3,480) = 6.19$, $p < 0.001$	0 < 2

Note. For the pairwise comparisons, the mean difference was significant at the 0.008 level (0.05 by six multiple comparisons).

Table 3. Mean articulation ratings and standard deviations (in brackets) of the 16 speech samples (four resonance conditions by four articulation conditions) made by the 30 listeners using visual analogue scale where “0” represents “Normal” and “10” represents “Very severe”, with the results of analyses of simple main effect of resonance and pairwise comparisons.

Articulation	Resonance				Analyses of simple main effects of articulation	Pairwise comparisons
	Typical (0)	Mild (1)	Moderate (2)	Severe (3)		
Typical	0.2 (0.2)	0.4 (0.4)	0.4 (0.4)	2.5 (2.2)	$F(3,480) = 8.62$, $p < 0.001$	0, 1, 2 < 3
Mild	3.1 (1.7)	4.3 (2.3)	3.7 (1.9)	3.8 (1.9)	$F(3,480) = 1.74$, $p = 0.157$	0 = 1 = 2 = 3
Moderate	6.2 (2.0)	5.6 (2.5)	5.8 (2.9)	7.1 (2.6)	$F(3,480) = 3.29$, $p = 0.021$	0 = 1 = 2 = 3
Severe	6.4 (2.3)	6.0 (2.3)	7.7 (2.5)	8.2 (2.4)	$F(3,480) = 7.87$, $p < 0.001$	1 < 2; 0, 1 < 3

Note. For the pairwise comparisons, the mean difference was the 0.008 level (0.05 by six multiple comparisons).