

Spring 2016

The effects of articulation errors on perceived nasality in speakers with repaired cleft lip and/or palate

Kristin Louise Dattilo
University of Iowa

Copyright 2016 Kristin Dattilo

This thesis is available at Iowa Research Online: <https://ir.uiowa.edu/etd/3067>

Recommended Citation

Dattilo, Kristin Louise. "The effects of articulation errors on perceived nasality in speakers with repaired cleft lip and/or palate." MA (Master of Arts) thesis, University of Iowa, 2016.
<https://doi.org/10.17077/etd.kztap03u>

Follow this and additional works at: <https://ir.uiowa.edu/etd>

Part of the [Speech Pathology and Audiology Commons](#)

THE EFFECTS OF ARTICULATION ERRORS ON PERCEIVED NASALITY IN SPEAKERS
WITH REPAIRED CLEFT LIP AND/OR PALATE

by

Kristin Louise Dattilo

A thesis submitted in partial fulfillment of the requirements for
the Master of Arts degree in Speech Pathology and Audiology
in the Graduate College of The University of Iowa

May 2016

Thesis Supervisor: Professor Jerald Moon

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

MASTER'S THESIS

This is to certify that the Master's thesis of

Kristin Louise Dattilo

has been approved by the Examining Committee for the thesis requirement for the Master of Arts degree in Speech Pathology and Audiology at the May 2016 graduation.

Thesis Committee:

Jerald Moon, Thesis Supervisor

Michael Karnell

Shawn Goodman

ACKNOWLEDGEMENTS

First, I would like to thank my friends and family for all of their support and encouragement throughout this process. I couldn't have done this without them. I am lucky to have people around me who are so understanding and have shown me the value of hard work and education. I would also like to thank Dr. Moon and my thesis committee for all of their guidance, opinions and ideas shared during this process. This has been such an excellent learning experience thanks to you.

ABSTRACT

The purpose of this study was to further investigate the effects of articulation errors on perceptual ratings of nasality in speakers with repaired cleft palates, specifically looking at the effect of varying magnitudes of articulation errors and education on perception of nasality. A group of expert listeners, speech-language pathologists with significant clinical experience in the area of cleft palate and resonance disorders, first rated the articulatory proficiency and nasality of a number of utterances produced by children with repaired cleft palates, on separate 6-point scales. Their ratings were then used to categorize stimuli into a three-by-three matrix (mild, moderate, severe) using articulation deficit and nasality as the two dimensions of interest. Untrained listeners (undergraduates and graduate students in a speech-language pathology training program) were then asked to rate the level of nasality on a 1 (normal) to 6 (severe) scale. Listener group ratings were compared to each other and to the expert listeners. Significant differences ($p=0.004$) were found between the undergraduate and graduate students' ratings when compared to the expert listeners. Graduates, had lower inter- and intra-rater reliability compared to the undergraduates. For both undergraduates and graduates, the difference between their ratings and those of the expert listeners was significantly lower for stimuli with mild articulation errors compared to those with moderate ($p<0.0001$) and severe ($p<0.0001$) articulation errors. No significant differences ($p=0.416$) were found between difference scores for stimuli with moderate versus severe articulation errors. The results were interpreted to suggest that the magnitude, and perhaps type, of articulation errors affects perceived nasality, and that there are group differences between perceptual nasality ratings. These findings support the importance of articulation therapy for children with repaired cleft palates to both decrease articulation errors and decrease perceived nasality. Further, this study highlights the need for

including awareness of this interaction in the training of speech-language pathologists in order for the “gold standard” of perceptual judgements to remain a valid and reliable measure.

PUBLIC ABSTRACT

Articulation errors and hypernasality, or nasal sounding speech, are common deficits seen in children with repaired cleft lip and/or palate. These two deficits affect intelligibility, but may also have an effect on each other. This study was designed to investigate how severity of articulation errors affects individual's perceived nasality and the influence of training on the ability to separate the two when listening to speech. Eight stimuli with varying degrees of articulation errors and nasality were chosen by expert listeners for the study. During the study, undergraduate and graduate students rated the stimuli for perceptual nasality. The degree of articulation errors was found to affect the amount of nasality listeners perceived. Listeners attended to articulation errors when nasality was mild, to both when nasality was moderate, and to nasality when the nasality was severe. Further, undergraduates' nasality ratings were more affected by articulation; however, graduate students' ratings were more variable. These findings support the importance of articulation therapy for children with repaired cleft palates to both decrease articulation errors and decrease perceived nasality. Further, this study highlights the need for including awareness of this interaction in the training of speech-language pathologists in order for the "gold standard" of perceptual judgements to remain a valid and reliable measure.

TABLE OF CONTENTS

LIST OF TABLES.....	viii
LIST OF FIGURES	ix
INTRODUCTION	1
Development of Speech Structures.....	1
Speech Production Process	3
Cleft Palate Speech	4
Intelligibility Measures in Speakers with Cleft Lip and/or Palate	6
Articulation Errors and Measurement.....	7
Hypernasality and Measures of Nasality	8
Interaction Between Articulation and Nasality.....	9
Factors that Affect Perceived Resonance.....	11
Summary.....	13
Research Study.....	13
Aims and Hypotheses	15
METHODOLOGY	16
Phase 1: Stimuli Generation.....	16
Raw Stimuli Selection.....	16
Study Stimuli Selection.....	17
Stimuli Selection Methods.....	18

Data Analysis	19
Phase 2: Testing	20
Stimuli.....	20
Subjects.....	20
Methods.....	21
Data Analysis	22
RESULTS	24
Within and Between Rater Agreement	24
Comparison of Graduates and Undergraduates.....	26
Effect of Articulation on Nasality	26
Response Time.....	30
DISCUSSION	31
Clinical Implications	36
Limitations	37
Future Studies	37
CONCLUSION.....	39
REFERENCES	41

LIST OF TABLES

Table 1: Nasality and Articulation Matrix	19
Table 2: Results of Kappa analysis of level of inter-rater agreement among undergraduate raters.....	25
Table 3: Results of Kappa analysis of level of inter-rater agreement among graduate raters.....	25
Table 4: Types and number of articulation errors present in stimuli separated by compensatory and developmental errors.....	33

LIST OF FIGURES

Figure 1: Instruction window	18
Figure 2: Rating window for expert raters	18
Figure 3: Rating window for graduate and undergraduate raters.....	22
Figure 4: Figure 4: Average (and standard deviation) rating difference scores for raters within each articulation error and nasality combination	27
Figure 5: Average (and standard deviation) raw scores for undergraduate and graduate raters within each articulation error and nasality combination	29
Figure 6: Average (and standard deviation) response time of graduates and undergraduates by stimuli ...	30

INTRODUCTION

The production of speech is an essential function of everyday life that often comes naturally. However, the reality is that speech is a complicated process, requiring planning, initiation, and coordination within and between many systems of the body. Abnormal structures or functioning of the face, nasal cavity, or throat may have negative effects on speech, which could lead to speech that is unintelligible. According to the Cleft Palate Foundation, approximately 1 in every 600 live births in the United States will have a cleft lip and/or palate (Cleft Palate Foundation, 2014). Individuals born with cleft lip and/or palate are likely to have both structural and functional deficits, often resulting in articulation and resonance disorders. Both articulation and resonance disorders may impact an individual's ability to communicate effectively with others. Not only may articulation and resonance balance affect oral communication separately, an interaction has been found between these two characteristics that may further impact the intelligibility of speech of individuals with cleft palate.

Development of Speech Structures

To appreciate the complexity of speech production and the effects of palatal clefting, it is necessary to first review both normal and abnormal development of the palate and surrounding structures. Beginning at about the fifth week of embryologic development, primitive regions of the middle and lateral aspects of the fetal face begin to fuse together to form the upper lip, upper jaw or alveolar arch, bony roof of the mouth, and the soft palate. The portion of this complex lying in the oral cavity, anterior to the incisive foramen is called the primary palate, which includes the alveolus and upper lip. The region posterior to the incisive foramen, including the hard and soft palate is called the secondary palate. During embryologic development, fusion of the left and right palatal segments proceed anteriorly from the incisive foramen to complete

creation of the primary palate. Fusion of the secondary palate proceeds posteriorly from the incisive foramen.

Clefting occurs when two or more of the embryonic facial structures that are supposed to fuse during this stage of embryologic development fail to do so. A cleft of the primary palate may affect only the lip, or may extend into the alveolus. Clefts of the primary palate may be bilateral, affecting both sides, or unilateral, affecting either the left or right side. A cleft lip can range from a small notch (incomplete) to a complete cleft through the base of the nose. Clefts affecting the alveolus may also be incomplete or complete depending on the size of the cleft.

The extent of clefting depends on when during this progressive process fusion is disrupted. If fusion extending posteriorly from the incisive foramen is disrupted at the beginning of the process, a complete cleft of the secondary palate (or hard and soft palate) will result, leaving the palate completely open from the incisive foramen posteriorly. If fusion is disrupted later in the process, the result will be an incomplete cleft. Normally, both complete and incomplete secondary palatal clefts affect the soft palate, as it is the last structure to fuse.

Clefts of the secondary palate may also be submucous. Submucous clefts involve a failure of union of cleft palate musculature, even though fusion of the mucous membrane, or outer layer of soft palate tissue, occurs. Common signs of a submucous cleft include a bifid uvula, a notch in the hard palate and presence of a zona pellucida, or blue skin color due to lack of muscle bulk.

The principal muscle of the soft palate (levator veli palatini) typically forms a sling through the soft palate. The function of the levator veli palatine muscle, when contracted, is to elevate the palate in order to contact the posterior pharyngeal wall. In the case of a cleft of the

soft palate, the levator veli palatine inserts along the margins of the cleft. As a consequence, the ability to separate the nasal and oral cavity during attempts to produce speech will be compromised. Surgery to repair the palatal cleft attempts to reconstruct the levator sling, providing the musculature mechanism to elevate the soft palate contributing to closure of the velopharyngeal port. This ability to separate the oral and nasal cavities, or velopharyngeal closure, is essential to production of sounds that require oral pressure, such as /p/ and /b/ and normal oral resonance (Trost-Cardamone et al., 2006))

Speech Production Process

An appreciation of the normal process of speech production is essential to an understanding of the effects of clefting on speech. The process of speech production begins in the respiratory system. Air is taken into the lungs and then exhaled. When exhaled, the air is forced through the larynx, where the vocal folds are located. If the vocal folds are open, during speech production, voiceless sounds will be produced, such as /t/, /k/, or /p/. If the vocal folds are closed, or approximated, the exhaled air will cause the vocal fold to vibrate, creating voiced sounds. The velum, or the soft palate, is a component of the valve that controls whether the airstream is diverted from the pharynx into either the oral or nasal cavity. The velum is lowered during nasal breathing and when producing speech sounds such as /m/, /n/ and /ŋ/. If the velum is elevated, the passage to the nasal cavity is blocked, resulting in velopharyngeal closure. The airstream is diverted to the oral cavity and is necessary for producing oral sounds that require high intraoral pressure, such as stops, affricates and fricatives. Other oral sounds, such as liquids, glides and vowels, are produced with relatively little or no oral construction. All speech sounds are produced as a result of varying vocal tract shapes and manipulation of the air-flow moving through the tract.

When a cleft of the hard and/or soft palate occurs, there is a failure of separation between the oral and nasal cavity. As a result, there is an inability to route air through only the oral cavity, as well as the inability to create oral air pressure, both of which are necessary for normal vowel and oral consonant production. Even after surgical cleft palate repair, normal soft palate function may not be achieved.

Cleft Palate Speech

The structural abnormalities associated with a cleft lip and/or palate can result in associated speech problems both before and after repair. Speech difficulties may fall into one of three categories: resonance, aerodynamics, and articulation. Resonance is defined as “vibratory response of a body or air-filled cavity to a frequency imposed upon it” (Wood 1971). Hypernasality is a type of perceived resonance that occurs when the amount of nasal resonance is higher than expected for a given oral speech sound in context. This phenomenon is directly related to velopharyngeal inadequacy (VPI). VPI occurs when structural and/or physiologic deficits result in a failure of separation of the nasal and oral cavities. Hypernasality associated with VPI can be heard intermittently or consistently when the velopharyngeal port cannot close adequately for speech (Trost-Cardamone et al, 2006).

Audible nasal air emission is another common anomaly observed in speakers with cleft palate. Nasal emission is a nasal deviation of air flow which impacts the production of high pressure consonants, such as stops, fricative and affricates. Nasal emission is an inability to close the velopharyngeal port by lifting the soft palate into contact with the posterior pharyngeal wall during attempts to produce high pressure oral consonants. This may result in the shunting of air through the nasal passage. Nasal emission can be audible or inaudible and can adversely affect the speaker’s speech quality or intelligibility. Audible nasal emission is often associated

with a “nasal snorting” sound or an audible rush of air through the nose. Inaudible nasal emission can’t be heard but is often detected by holding a mirror below the nostrils and observing fog on the mirror. Nasal emission can also be a learned behavior in individuals with or without cleft palate. When it is produced on certain phonemes, or phoneme specific VPI, nasal emission occurs only during the production of certain phonemes, usually sibilants or fricative. (Trost-Cardamone et al, 2006).

The third issue associated with cleft palate involves articulation errors. Articulation is described as the ability to move the tongue, lips, and velum in order to create sounds for speech. According to Peterson-Falzone and colleagues (2001), it is estimated that about 75% of preschoolers who receive team cleft care had abnormal articulation, while some persisted into adolescence (Peterson-Falzone et. al., 2001). Articulation errors can be described as substitutions (e.g., “tat” for “cat”), omissions (e.g., “nana” for “banana”), distortions (e.g., nasalizing a sound), or additions (e.g., “doguh” for “dog”). All children, with or without a cleft palate, exhibit articulation errors as a part of the normal development of speech. For example, a three-year-old child is often expected to substitute “r” with “w,” such that “rabbit” is pronounced “wabbit.” Some children with cleft palate also make other articulation errors, referred to as compensatory articulation errors. These are learned productions believed to result from the child attempting to compensate for their structural abnormalities. Once a compensatory articulation is learned, the child tends to continue using it even after the structural abnormality has been corrected. Most compensatory strategies involve a change in the place of articulation. For example, stops like /b/ and /p/ may be produced as a glottal stop. The oral pressure needed to produce the bilabial oral stops is not possible to produce with a cleft given the inability to close

the velopharyngeal port. However, the child can build up pressure at the glottis and therefore replaces the oral stop with a glottal stop (Trost-Cardamone et al. 2006).

Intelligibility Measures in Speaker with Cleft Lip and/or Palate

All of these speech abnormalities are commonly seen in children with cleft lip and/or palate and may be addressed clinically by speech pathologists. The assessment of these disorders is heavily reliant on the clinicians' perceptual and formal assessment. One measure of speech used frequently by clinicians is an intelligibility judgment. Intelligibility is the amount of acoustic signal that a listener understands when a person is speaking (Duffy, 2013). Measures of intelligibility are used across all populations to quantify how much of the speaker's intended message is being correctly perceived by listeners. This can be used to gauge progress and make judgments about the functionality of the person's communication skills.

Factors of speech found to interact with intelligibility were explored by McWilliams (1954) in a study on the intelligibility of cleft palate speech. McWilliams (1954) found significant correlations between articulation and intelligibility ($r=.715$), and between articulation errors and nasality ($r=.821$). Moore and Sommers (1975) conducted an interval scaling study and found a correlation of .56 between intelligibility and nasality ratings. A study of patients with a repaired cleft and maxillary advancement was completed by Megawa et al. (1998). They found increased hypernasality in the group whose intelligibility decreased, suggesting that hypernasality interacts with intelligibility. As seen in the studies above, intelligibility is impacted by many factors. The two factors appearing to have the strongest impact on intelligibility are articulation errors and hypernasality.

Articulation Errors and Measurement

Due to a high incidence of developmental, persistent, and compensatory articulation errors, the articulation skills of cleft palate speakers are typically assessed to determine the severity of speech problems and to measure therapy progress. Articulation errors are typically assessed using standardized test instruments. Standardized measures, such as the Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986), involve the speaker naming a series of pictures. This word-naming task elicits a variety of consonants produced in a variety of positions within words. The test is standardized to normal speakers of similar age and gender. Although these tests have their benefits, such as easy administration and standardized scoring, there are many drawbacks to a standardized articulation test. Some of these negatives include the limited inventory of sounds and phonetic contexts standardized assessments test, as well as the context specific nature of the test. Some children may have errors in connected speech that may not be present in a one-word articulation test. Also, standardized tests may not show the functional effect that articulation errors may have on speech. For these reasons, informal assessment and perceptual severity ratings of articulation can be useful in a clinical setting. These measures give insight into how the articulation affects their speech overall, not just in specific contexts.

Many studies have attempted to compare the reliability and validity of perceptual judgments against more objective measures. Van Demark (1962) proposed that articulation judgments could be predicted by other variables. His study looked at 13 different types of articulation errors, as found by scoring standard sentences, and the judged ratings that accompanied those scores. It was found that the severity of judged articulation could be predicted by a combination of the errors investigated. Van Demark (1962) also concluded that

velopharyngeal closure was one of the primary factors that related to the severity of judged articulation errors.

Subtelny, Van Hattum and Myers (1972) looked at the differences between intelligibility judgments and articulation judgments in cleft palate speakers. Raters used an equal interval scaling method to rate both articulation and intelligibility. Thirty-four percent of the speech samples were rated as having severely abnormal articulation, but only four percent were rated as having unintelligible speech. The authors concluded that articulation and intelligibility judgments are not directly correlated. They also reported that judgments of intelligibility and articulation closely approximated the standardized test scores. From this, they concluded that perceptual judgments are equally as useful as objective measures.

Hypernasality and Measures of Nasality

There are two approaches to measuring nasality. The first approach is an objective measure using nasometry, a technique that uses the acoustic correlate of nasality to measure the oral and nasal acoustic energy present in the speech of an individual. Using microphones positioned in front of the nose and mouth, and separated by a metal plate, a ‘nasalance’ ratio is calculated by dividing the nasal acoustic energy by the sum of oral and nasal energy. Although an objective measure, previous studies have found that factors such as language, dialect, and the specific speech sample may affect the scores (Karnell, 1995).

A second approach for assessing nasality in speech involves listener perception. In fact, Kuehn (1982, p. 518) highlighted the importance of perceptual assessment, stating that, “in a sense, a speech disorder does not exist until it is perceived by a listener.” Perceptual assessment is currently seen as the “gold standard” by many clinicians. Perceptual judgments can be performed utilizing a variety of stimuli and rating tools. As in articulation and intelligibility

measures, an equal appearing interval scale is the most common technique, although equal magnitude estimation and descriptive categories have also been used (Kuehn, 2000).

Although perceptual assessments are seen as an important tool for clinicians, the reliability and validity of these assessments have been questioned, especially in comparison with nasometry measures. A recent study by Sweeny and Sell (2008) studied the relationship between perceptual ratings of nasality and nasometry measures in children and adolescents with cleft palate or VPI. Correlation coefficients ranged from .69 to .74, suggesting that the perceptual ratings were strongly related to, or in agreement with, the nasalance measure. The specificity of the Nasometer ranged from .78 to .95 with a sensitivity of .83 to .88. In other words, the nasometer was 78-95% accurate in detecting hypernasality when it was present and 83-88% accurate in detecting a lack of hypernasality when there was none. Overall, Sweeny and Sell (2008) concluded that there was a strong relationship between the perceptual assessment and the acoustic assessment. They also proposed that perceptual assessment is a valid assessment tool, but that the nasometer can be used as a supplemental assessment approach.

Interaction Between Articulation and Nasality

Articulation and nasality are two important factors that affect intelligibility. Not only can these characteristics of speech impact intelligibility, but they also have an effect on each other. As described above, McWilliams (1954) reported a correlation between articulation and nasality, suggesting that not only do they both individually have an effect on intelligibility judgments, but also that nasality may affect articulation judgments and vice versa.

Van Hattum (1958) examined the relationship between judgments of articulation and judgments of nasality. Connected speech samples were gathered using 10 sentences. Isolated consonant-vowel combinations were produced using the combinations seen in the sentences. In

total, the participants produced 51 vowels and 73 consonants. The 20 participants (aged 13 – 25 years) all had repaired cleft palates, not otherwise specified, and 13 also had lip involvement. The listeners were three graduate students who were chosen based on their accuracy in a training task, although the definition of accuracy was not described in the study. A seven-point scale was used for perceptual ratings and articulation was judged as either correct or incorrect. In total, four measures were taken for the 20 subjects: sentence nasality, nasality on vowels, sentence articulation, and syllable articulation. Correlations between the nasality and articulation measures were calculated. The severity of nasality in sentences and the articulation in sentences were found to be significantly negatively correlated ($r=-.64$). The correlation between articulation in syllables and nasality in sentences was also found to be significant with $r=-.67$. In other words, as severity of nasality increased, the perceived articulation accuracy tended to decrease.

Several conclusions were offered, the first being that articulation and nasality appear to be independent when the interaction of articulation on nasality is minimized. That is, when very few articulation errors were present, perceived nasality didn't appear to be influenced. Second, good articulation tends to be associated with judgments of less nasality.

Sherman (1970) specifically investigated the interaction of articulation and nasality. The purpose of this study was to see if there was a tendency for the degree of perceived hypernasal resonance to vary depending on the ratings of articulation errors. A 7-point scale was used to rate articulation and nasality. However, articulation and nasality were scaled in separate listening sessions. The 37 listeners were only described as “sophisticated” listeners who were all communication disorders majors. The speech recordings used were from 154 children with cleft palate. The original recordings contained a randomized order of 13 sentences, and therefore the first five seconds played varied in content. The first five seconds of the tape was played

backward during the nasality rating session. Sherman theorized that by playing it backwards, the interaction of articulation and resonance would be lessened. Today, however, it is now thought that the naturalness of speech, which is completely disrupted when playing backwards, is more important and aligned with typical listening conditions.

A correlation of .34 was found between articulation and nasality. The authors interpreted the results as strong evidence in support of a moderate tendency for defectiveness of articulation and nasality to be functionally related in the speech of children with cleft palate.

Factors that Affect Perceived Resonance

As stated previously, perceptual assessments are seen as the “gold standard” in clinical assessment, but they are, at their core, not objective. Research has identified many factors that affect perceived resonance ratings, some of which have to do with listeners and their environment.

One such factor is the effect that education and experience has on perceptual assessments. Lewis, Watterson and Houghton (2003) examined this using listeners with varying levels of experience and training with speakers of cleft palate rate the amount of perceived nasality on a 1 to 5 scale. The groups consisted of elementary and middle school teachers, graduate students in the field, speech language pathologists with training and experience, and craniofacial surgeons who had experience, but no academic training in rating nasality. The highest agreement in ratings was found amongst the speech language pathologist and surgeon listeners. The authors concluded that experience, not academic training, had the greatest effect on the ratings. Results also showed that speech language pathologists and surgeons produced overall lower nasality ratings than the teachers and graduate students. This suggests that listeners with little experience perceive nasality to be more severe than those with greater experience.

Another study (Brunnegard and colleagues, 2009) compared untrained listeners' ratings with trained listeners' ratings. This study used descriptive statements to rate the severity of nasality. For example, "The child has a blocked nose" or "The child had difficulties with pronunciation." The results indicated good agreement (mean $r=0.62$) between the untrained and trained listeners. The authors also found that untrained listeners were able to hear hypernasality when it was in the moderate to severe range and that the trained listeners were more sensitive to a lower level of nasal emission.

The number of listeners has also been shown to have an effect on judgments of nasality. Multiple listener judgments are generally thought to be more accurate than single listener judgments. Early studies researching judgements of nasality used ratings of a single clinician (Counihan, 1960; McWilliams, 1954). Later studies, by authors such as Starr and colleagues (1984), Moller and Starr (1984), as well as the majority of more recent studies by other authors all used multiple listener ratings. This seems logical because judgments of nasality are subjective and when using a larger listening population, an average or consensus can be gathered.

The speech stimuli itself can affect the perceived nasality ratings. Moore and colleagues (1973) explored the interaction of phonetic context and perceived nasality. Listeners judged both the intelligibility and nasality of speakers with repaired palatal clefts on an equal interval scale for 16 consonants and 6 vowels in various isolated consonant-vowel combinations. The results indicated that perceived nasality increased in the following order: glides, glottal fricative /h/, plosives and fricatives/affricatives. They also found a correlation of .56 between the ratings of nasality and intelligibility; however, the correlation seemed to be very context specific. This result illustrates that, in general, there is a moderate relationship between nasality and

intelligibility; however, certain sounds, such as glides, have a greater impact on perceived than others, such as fricatives.

Summary

Articulation errors and hypernasality are common speech deficits in children with cleft palate. Perceptual judgments of intelligibility, articulation and nasality have been widely studied in speakers with cleft palate and are currently seen as the “gold standard” for assessing speech within this population. Several authors have indicated that perceptual judgments are affected by listener experience, education, and number of listeners. Intelligibility ratings show a high correlation with ratings of articulation and nasality individually (McWilliams, 1954). Ratings of these two variables have been proven valid when judged separately by listeners; however, the speech of individuals with cleft palate often contains both articulation errors and hypernasality. It is difficult, if not impossible, to separate the two factors during testing, and some interaction is likely to occur. Van Demark (1962) found that the judgments of articulation severity could be estimated using a combination of variables, one of which was nasality. Sherman (1970) found a moderate correlation between articulation and nasality. She suggested that they were “functionally related.” Van Hattum (1958) also found a significant correlation in his study. The discrepancy in the strength of correlations between the studies of Sherman and Van Hattum, with correlations of 0.34 and -0.67, respectively, may have been due in part to differences in methodology, stimuli, and scales used during testing. Also, in general, it can be hypothesized that perceptual judgments will produce variable results due to their subjective nature

Research Study

As previously stated, the correlation of articulation and nasality in speakers with cleft palate has been studied in the past (Van Hattum, 1958; Sherman, 1970). Others have mentioned a

possible interaction as well (McWilliams, 1954; Van Demark 1962). However, these studies did not explore the magnitude to which these factors interact. Van Hattum (1958) and Sherman (1970) only looked for whether or not there was any correlation at all; they did not explore the amount to which they are related, and further, whether one had a greater effect on the other. The present study not only explored the existence of an interaction, but how the magnitude of articulation deficits affected the perceived degree of nasality. In addition, the previous two studies on this topic used a very controlled listener population. The listeners who participated in the study by Van Hattum (1958) were selected because of their perceptual accuracy. Perceptual ratings are subjective in nature and the study did not state how the accuracy was determined. Sherman (1970) used “sophisticated” listeners, but didn’t describe further the training or experience the listeners had.

The present study used a variety of listeners; undergraduate and graduate listeners. The listener population in the present study had various levels of education and experience with speakers with cleft palate. Graduate students received 1-2 classes in the area of cleft palate and some (minimal) clinical experience with the population. Undergraduate students had approximately 2 lectures regarding the topic and likely no clinical experience. The trained professionals had an extensive amount of clinical experience as well as education regarding speakers with cleft palate and are actively involved in the area. The variety of listener groups more closely mimics the various listeners who interact with cleft palate speakers, such as parents, teachers and beginning or expert speech-language pathologists, who also have various degrees of education and experience.

The stimuli used were also more controlled in this study. Trained listeners initially rated the articulation deficits and degree of nasality of various recordings using a six-point scale.

Using this method, recordings with agreement among the professionals that represented various combinations of articulation errors and degree of nasality were identified. Sherman (1970) and Van Hattum (1958) took recordings of the speakers and didn't control for the amount of nasality or articulation errors. As such, the stimuli may have consisted of all mild or moderate nasality speakers with mild articulation. The present study utilized stimuli that include a variety of combinations of articulation errors and nasality.

Aims and Hypotheses

The first aim was to investigate the degree to which the severity of nasality and articulation errors are associated. A second aim was to confirm previous findings that education and experience affect perceived nasality by investigating differences between undergraduate and graduate listeners.

The main hypothesis that drove this study was that there is a significant interaction between the severity of articulation deficit and hypernasality as reflected in listener ratings of nasality. When comparing stimuli with similar nasality rating, but differing ratings of articulation error severity, it was expected that utterances with more articulation errors would have higher perceived nasality. Further, it was hypothesized that untrained listeners, the undergraduate and graduate groups, would perceive higher hypernasal resonance on the stimuli that also have articulation errors when compared to the trained (expert) listeners. It was also expected that the undergraduates would perceive a greater degree of hypernasal resonance overall than the graduate students due to their lack of education and experience.

METHODOLOGY

This study (IRB number 201503719) was reviewed and approved by the Institutional Review Board and the Human Subjects Committee at the University of Iowa on April 6, 2015.

The current study was comprised of two phases: Stimuli generation and testing. The University of Iowa Hospitals and Clinics Cleft Clinic database contains a wide variety of speech stimuli available for inclusion in this study. The initial phase, stimuli generation, was designed to identify stimuli that represented various degrees of severity of both articulation error and nasality. This allowed for the inclusion of stimuli that spanned various combinations (mild to severe) of articulation errors and nasality. Trained expert raters rated the recordings based on severity of articulation errors and nasality, with a goal of gathering one representative recording for each of nine possible combinations (mild, moderate, severe) of articulation and nasality impairment. The second study phase then used these stimuli to test the study hypotheses. Untrained raters rated perceived nasality, with the articulation ratings gathered in the generation phase used to control for articulation errors. The untrained listener's ratings were then compared with the ratings of the trained raters.

Phase 1: Stimuli Generation

Raw Stimuli Selection

The Cleft Clinic database of the University of Iowa Hospitals and Clinics (UIHC), Department of Otolaryngology-Head and Neck Surgery, was searched to gather speech stimuli for the study. The following search criteria were used:

1. Recordings made after January 1st, 2012
2. Children ages 3-12,
3. Repaired cleft lip and/or palate,

4. Zoo passage recording.

Using the ratings in the clinic database, as well as the principle investigator's judgment, 35 recordings were selected that included various degrees of articulation errors and nasality. The stimuli were not controlled for by type of cleft or other speech characteristics, such as nasal emission.

Each of the recordings selected included the following phrases: "Look at this book with us. It's a story about a zoo. That's were bears go. Today it's very cold out of doors. But we see a cloud overhead. That's a pretty white fluffy shape." All of the recordings included the clinician's speech prompts in between phrases, as this was performed as a repetition task. Each raw recording was imported into Adobe Premiere (Adobe Systems, Inc., San Jose, CA). All clinician prompts were removed and the child's utterances spliced together.

Study Stimuli Selection

The 35 selected recordings were then rated by four trained raters. Three trained raters were recruited from the Speech and Swallowing Clinic of the University of Iowa Hospitals and Clinics, while the fourth was recruited from the University of Iowa Department of Communication Sciences and Disorders. All four trained raters had more than 7 years of clinical experience working with children with cleft lip and/or palate, rating nasality, and/or advanced education and training in the area.

Stimuli Selection Methods

A custom MATLAB (Mathworks, Inc., Matick, MA) computer program was created and used to present stimuli to raters and record their responses. Testing was conducted in a quiet computer lab, and stimuli were presented via headphones (Yamaha model CM500). The 35 recordings were presented in a random order. Instructions were given both visually and auditorily (Figure 1). The raters were asked to rate the severity of articulation and nasality of each recording on a 1 to 6 scale (1=normal, 2=mild, 3=mild-moderate, 4=moderate, 5=moderate-severe, 6=severe) (Figure 2). Unlimited time was given to the rater to rate each recording. Responses were logged by the MATLAB program and exported to a spreadsheet file.

Figure 1: Instruction window

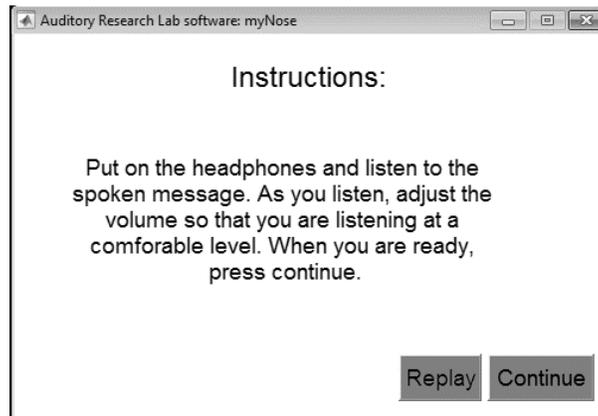
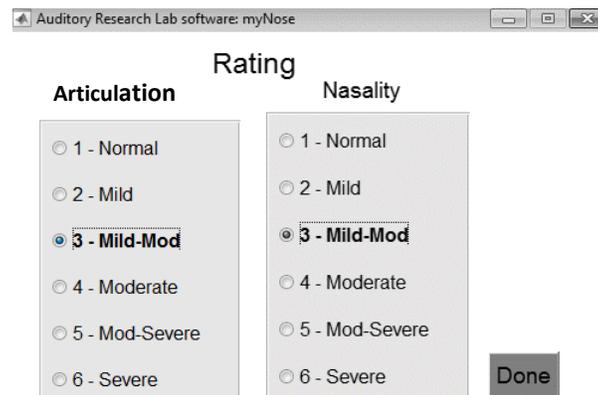


Figure 2: Rating window for expert raters



Data Analysis

The recorded articulation and nasality ratings were then analyzed for agreement between the four expert raters, with the goal of obtaining stimuli for all possible combinations of articulation error and nasality severity.

It is to be noted that at this point of the study the 6-point rating scale was reduced to a 3-point scale, with ratings of 1 and 2, becoming 1, 3 and 4 becoming 2, and 5 and 6 becoming 3. The 6-point scale used was consistent with how the trained clinicians traditionally rated patients seen in Cleft Clinic. However, a high degree of agreement among the trained raters was not found for such a precise scale. Given that requiring a high level of agreement for sentences to be placed with a 6 x 6 articulation errors and nasality matrix was highly unlikely, ratings were successfully collapsed into a 3 by 3 matrix (mild, moderate, severe) (Table 1). Using the 3-point scale still allowed for variability across the stimuli and allowed for a higher threshold of agreement to be used for inclusion of the stimuli.

Table 1: Nasality and Articulation Matrix

	<u>Articulation</u>		
<u>Nasality</u>	a1	a2	a3
n1	1 recording	1 recording	1 recording
n2	1 recording	1 recording	1 recording
n3	1 recording	1 recording	1 recording

A mean and mode ratings across the four raters was calculated for each stimulus. Ratings were deemed to have high agreement when there was a mode and the mean and mode were within one point of each other. For example, a particular recording was rated as follows by the 4 raters: Articulation- 4, 4, 3, 2; Nasality-1, 2, 1, 1. For the articulation ratings, the mode is 4 and

the mean is 3.25. The nasality rating mode is 1 and the mean is 1.25. Because the mean and mode were within 1 point of each other for nasality and articulation, the recording was deemed to have high agreement. Those with no mode or highly discrepant mean and mode were deemed to have moderate agreement.

After the first round of ratings, seven of the thirty-five stimuli ratings were found to be variable. Those stimuli, as well as four others for control purposes, were presented to three of the four trained raters to rate again. One of the four trained raters was unavailable to participate in this step. This increased the agreement on several stimuli, especially those in the severe categories. Stimuli with moderate or high agreement were then placed into the three-by-three matrix (Table 1) based on their nasality and articulation ratings. The recording with the highest level of agreement within each matrix cell were selected as the stimuli for the second phase of the study. Unfortunately, no recordings in the mild articulation- severe nasality (a1n3) category reached the minimum criteria for selection as a stimulus for phase 2.

Phase 2: Testing

Stimuli

The stimuli used in this phase included the eight stimuli from Phase 1 representing eight of the nine possible combinations of articulation errors and nasality severity depicted in Table 1.

Subjects

Subjects participating in this phase were recruited from the University of Iowa Speech-Language Pathology graduate program and University of Iowa Communication Sciences and Disorders undergraduate program. These subjects, deemed the “untrained raters”, were recruited using the following criteria:

1. Graduate or undergraduate students in speech-language pathology
2. 18-50 years old
3. No known speech, language or hearing problems.

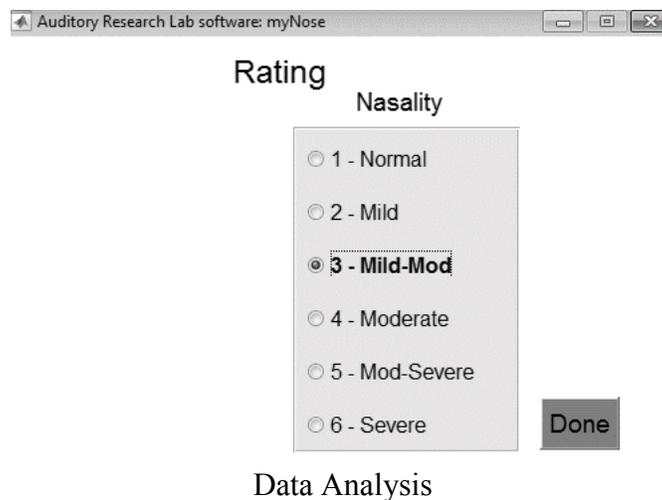
The undergraduate students were college aged students who had minimal education (approximately two lectures) about clefting and nasality, and had minimal to no experience in rating aspects of speech production such as nasality. Graduate students in the program had received one or two classes regarding palatal clefts and nasality and had some clinical experience with perceptual judgements of speech. There were 20 raters recruited from both groups, with a total of 40 “untrained raters.”

Methods

The custom MATLAB computer program developed for Phase 1 of this study was used for the testing phase, with a few modifications. Again, testing took place in a quiet computer lab with the raters wearing headphones (Yamaha model CM500). The first modification was the addition of a four stimuli training module. The training module provided the raters practice using the 1-6 nasality rating scale and oriented them to how the program functioned (Figure 3). It was chosen to continue to use the 6-point scale in order to be consistent with what the expert raters used. During the training session, no feedback was given; however the stimuli used in the training contained minimal articulation errors and contained various degrees of nasality. This was designed to expose the raters to a variety of nasality in speech without the influence of articulation errors. None of the recordings used in the training session were used in the rating task.

The training session was followed immediately by the rating session. The eight stimuli were presented in random order and repeated four times. The stimuli were randomized within each set of eight and were re-randomized for each of the four repetitions. After each stimulus was played, a rating menu appeared and the listener selected the severity of perceived nasality using the same 1-6 rating scale used on Phase 1. Unlimited time was given to the listener to rate each recording. A three second countdown was displayed before the next stimulus was played. In total, the raters listened to 32 recordings. Each rating responses was exported to a spreadsheet document.

Figure 3: Rating window for graduate and undergraduate raters



Raw nasality rating data from each rater was exported to a spread sheet document. Ratings on the 6-point rating scale were converted to a 3-point rating scale (1=mild, 2=moderate, 3=severe). Using Matlab, the data was organized by undergraduate versus graduate raters and organized into 4 (ratings) by 20 (listeners) matrices for each of the 8 stimuli. Difference scores (listeners ratings-expert ratings) were also calculated using the MATLAB program.

Inter- and intra- rater reliability statistics using Cohen's Kappa were calculated the using mean raw scores for a given stimulus (Brennan, 1981). Kappa takes into account the agreement and also the percent chance of the agreement.

Comparisons between the undergraduates and graduates were calculated using the average difference score of the four ratings. The statistical significance of this difference was calculated using a Wilcoxon rank sum test exact p-value (Wilcoxon & Wilcox, 1970), a non-parametric test used for repeated measures.

Evaluation of the relationship between articulation score and perceived nasality was conducted using Friedman's test (Higgins, 2003). Using difference scores (listener nasality score – expert nasality score), this test evaluated whether the magnitude of articulation error affected the perception of nasality both within and across raters.

The mean and standard deviation of difference scores were calculated, as were the mean and standard deviation of response times, defined as the amount of time it took raters to record their rating following presentation of the stimulus.

RESULTS

Within and Between Rater Agreement

During the study, 20 undergraduates and 20 graduates each listened to eight stimuli, four times. Agreement within raters and between raters was determined using the Kappa statistic for each rater group using the average of the raw ratings.

Undergraduates showed a mean Kappa of 0.63 (95% C.I.: 0.59, 0.67) for intra-rater agreement or when comparing each single participants' ratings for a given stimuli. Graduates' intra-rater Kappa was 0.48 (95% C.I.: 0.45, 0.52) when comparing the four ratings within any single rater. As such, undergraduates were judged to exhibit substantial intra-rater agreement, while graduates showed moderate agreement.

Analysis of the inter-rater agreement for the undergraduates yielded a mean Kappa of 0.51 (95% C.I.: 0.44, 0.57). Table 2 shows the p-values for each specific stimulus. When calculating the average ratings of a stimulus, there was the possibility for an average that was not an integer, or was between two ratings. For the purpose of this inter-rater agreement analysis, these intermediate ratings (i.e., 1.5, 2.5) were used. As seen, the p-values were variable, with the lowest level of agreement occurring for the moderate articulation (2), mild nasality (1) stimulus. The overall probability of agreement between any 2 raters was 0.69.

Table 2: Results of Kappa analysis of level of inter-rater agreement among undergraduate raters.

Artic	Nasal	1	1.5	2	2.5	3	p_value
1	1	16	3	1	0	0	0.65
1	2	18	1	1	0	0	0.81
2	1	1	2	10	4	3	0.29
2	2	0	0	4	4	12	0.41
2	3	0	0	0	0	20	1.00
3	1	0	0	4	3	13	0.46
3	2	0	0	0	0	20	1.00
3	3	0	0	0	1	19	0.90

For graduates, the inter-rater mean Kappa result was 0.23 (95% C.I.: 0.19, 0.27), with a probability of two raters agreeing equal to 0.45. Table 3 shows the p-values for each stimuli rated by graduates. Compared to the undergraduates, who had moderate agreement, the graduates exhibited fair inter-rater agreement.

Table 3: Results of Kappa analysis of level of inter-rater agreement among graduate raters.

Artic	Nasal	1	1.5	2	2.5	3	p value
1	1	15	2	2	0	1	0.56
1	2	14	2	4	0	0	0.52
2	1	1	2	13	1	3	0.43
2	2	0	1	4	2	13	0.45
2	3	1	0	5	1	13	0.46
3	1	5	0	5	2	8	0.26
3	2	0	2	5	2	11	0.35
3	3	1	0	2	2	15	0.56

Comparison of Graduates and Undergraduates

After calculating the average of each listener's four ratings of each stimulus, a comparison of graduate and undergraduate ratings was calculated using median difference scores (grad/undergrad rating - expert rating). A Wilcoxon rank sum test was used to assess the level of agreement between graduates and undergraduates for each stimulus. Controlling for articulation and nasality, undergraduate and graduate listeners differed significantly ($p=0.004$) in their accuracy of rating nasality compared to the expert raters. As seen in Figure 6, these differences were apparent for stimuli with moderate articulation errors and severe nasality, and for all three stimuli with severe articulation errors. Undergraduates and graduates did not differ significantly for those stimuli with mild articulation errors and for those with moderate articulation errors and mild to moderate nasality.

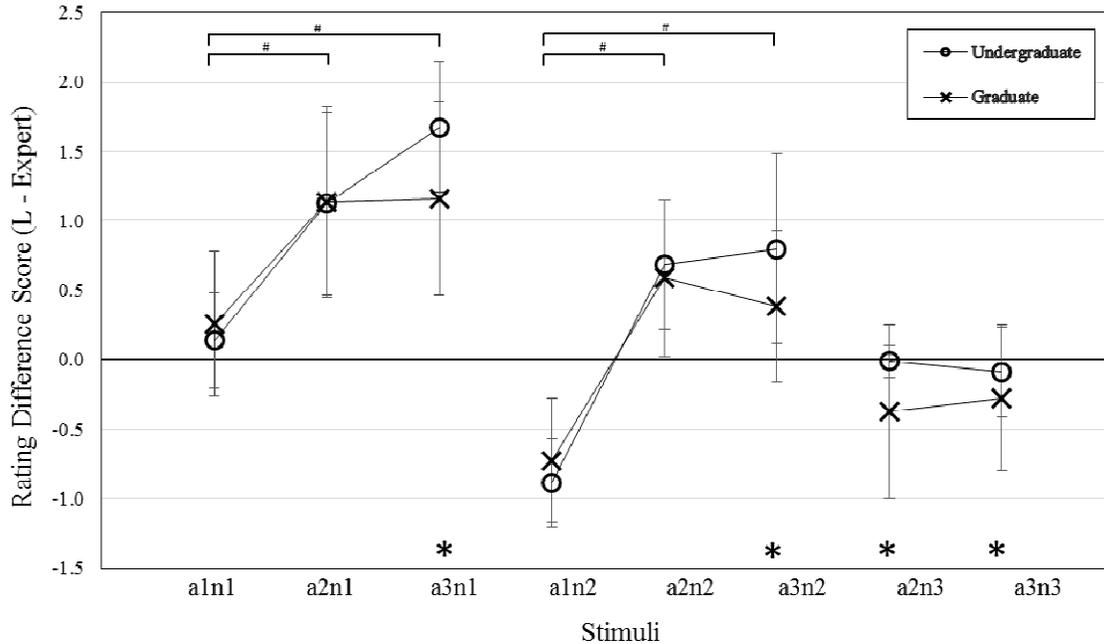
Effect of Articulation on Nasality

Group mean and standard deviation difference scores for each articulation and nasality combination studied are shown in Figure 6, organized by increasing degree of articulation deficit within each nasality rating category (mild, moderate, severe).

Friedman's test was used to test if, overall, nasality ratings (expressed as difference scores relative to the experts) were associated with the level of articulation errors across all stimuli. For both undergraduates and graduates, difference scores were significantly lower for stimuli with mild articulation errors compared to moderate ($p<0.0001$) and severe ($p<0.0001$) articulation errors. No significant differences were found between difference scores for stimuli with moderate versus severe articulation errors ($p=0.416$). In other words, the magnitude of articulation disorder had a significant effect on nasality ratings when comparing moderate and

severe articulation levels with mild articulation, but there was no difference in this effect between moderate and severe articulation. This result can be clearly visualized in Figure 6.

Figure 4: Average (and standard deviation) rating difference scores for raters within each articulation error and nasality combination (a=articulation, n=nasality, 1=mild, 2=moderate, 3=severe). *statistically significant ($p < 0.05$) difference between raters # statistically significant ($p > 0.05$) difference

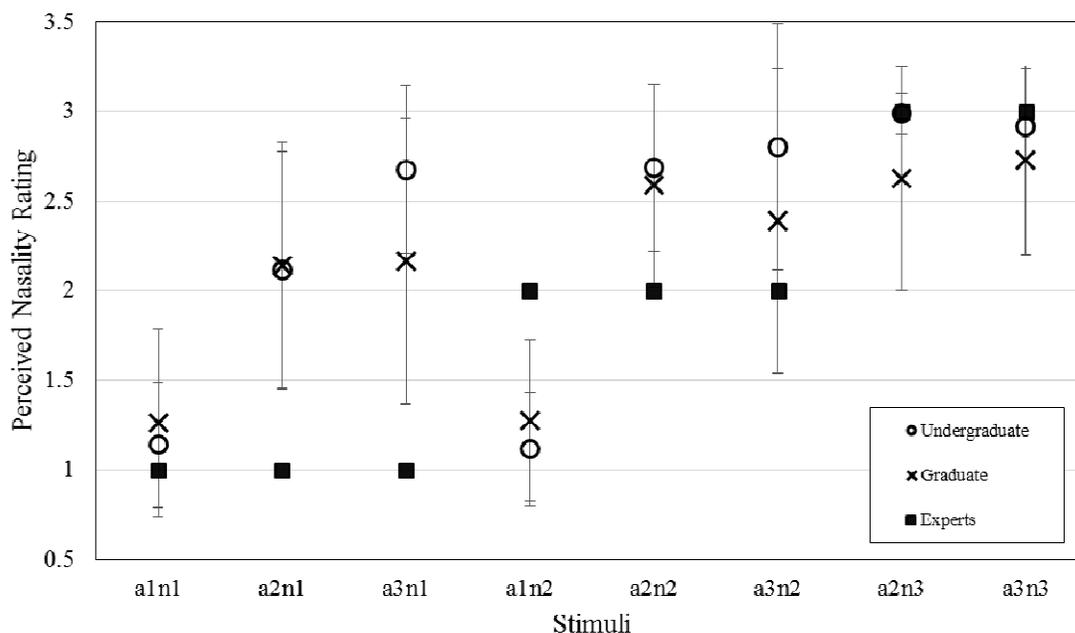


For mild nasality stimuli (i.e., n1), both groups rated the degree of nasality much higher than the experts when the articulation errors were considered by the experts to be moderate (2) or severe (3). However, on average, both groups' ratings were similar to the experts for stimuli with a mild degree of articulation errors. This indicated some effect of articulation errors on perceived nasality. For the stimuli with mild nasality and severe articulation errors (a3n1), the undergraduates rated the nasality as significantly ($p=0.014$) higher than the graduates. When considering group variability (expressed as standard deviation), graduates were much more variable on the stimuli with mild articulation and severe articulation errors, but were similar in rating variability with the undergraduates on the moderate articulation error stimulus, where both groups were highly variable.

For stimuli with a moderate degree of nasality (i.e., n2), both groups were affected similarly by stimuli with mild and moderate articulation errors. Both groups underestimated the degree of nasality for the mild articulation deficit stimuli, and overestimated it for the moderate and severe condition. A significant difference ($p = 0.011$) between graduates and undergraduates was observed for the severe articulation error stimuli, consistent with the pattern observed for mildly nasal stimuli containing severe articulation errors. Overall, the magnitude of variation from the expert ratings was less than that observed for the mildly nasal stimuli. Graduates again exhibited more variability on the mild articulation stimulus, but were similar to the undergraduates for the moderate stimulus. Both groups exhibited high variability for the severe articulation error stimulus.

For stimuli with a severe degree of nasality (i.e., n3), both undergraduates and graduates rated stimuli with moderate and severe articulation errors to possess levels of nasality close (< 0.5 difference on average) to the expert's ratings. The graduates rated these stimuli as being significantly ($p < .01$ and $p < .05$, respectively) less nasal than the undergraduates, although the absolute average difference was less than 0.5 rating scale points. Further, the undergraduate raters had much less variability in their ratings for both stimuli when compared to the graduates.

Figure 5: Average (and standard deviation) raw scores for undergraduate and graduate raters within each articulation error and nasality combination (a=articulation, n=nasality, 1=mild, 2=moderate, 3=severe).



To better appreciate the undergraduate and graduate ratings relative to the experts, Figure 7 shows raw nasality ratings organized by increasing level of articulation errors within each nasality severity category. For stimuli with mild nasality (n1), the perceived level of nasality increased as the severity of articulation errors increased. For stimuli with moderate (a2) or severe (a3) articulation errors, but mild nasality (n1), the raw scores more closely resembled the level of articulation error than nasality. For example, stimulus a2n1 had an average perceived nasality of 2.1 when the expert nasality rating was 1.

When looking at the moderate nasality stimuli (n2) the stimulus with mild articulation errors (a1n2) was rated by the graduates and undergraduates lower than the expert ratings. Again, this appeared to resemble the mild articulation error severity and not the expert rater nasality rating. The stimuli with moderate and severe articulation errors (a2 and a3) were both rated higher than the expert ratings of nasality. Compared to the articulation severity level, the

perceived nasality for the stimulus with moderate articulation errors and nasality (a2n2) was higher than 2, and for the stimulus with severe articulation errors and moderate nasality was lower than 3.

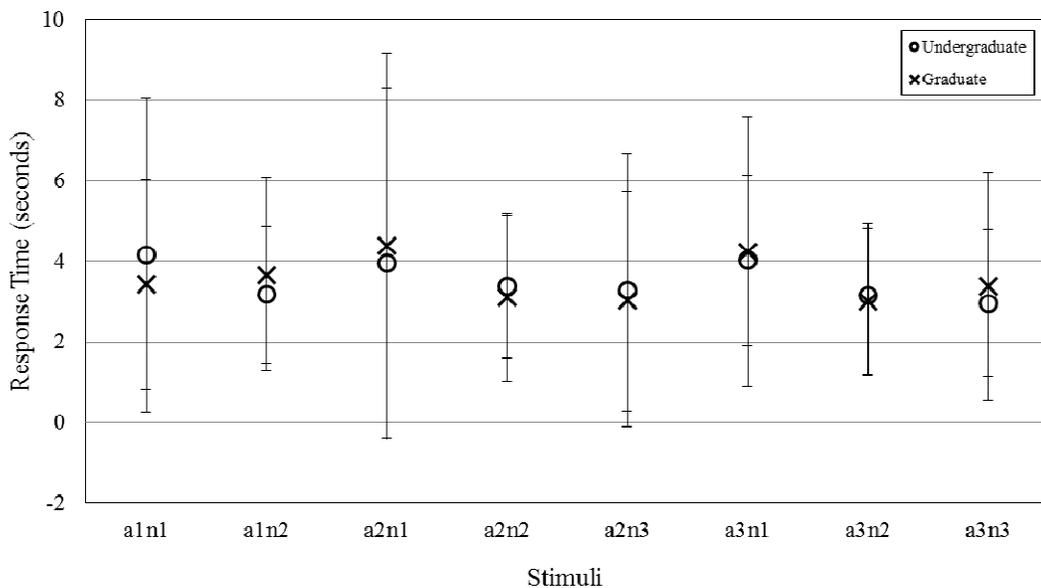
The severe nasality (n3) stimuli were rated as relatively close to the expert scores, regardless of articulation deficit severity.

Response Time

In order to assess whether the two rater groups differed in the time taken to make a judgement either in an overall sense, or as a function of articulation or nasality severity, average response times were calculated for each of the stimuli and rater groups, as seen in Figure 8.

There were no obvious differences between rater groups or across the stimuli. When comparing the response times to differences between groups seen in the ratings, there was no observable trend.

Figure 6: Average (and standard deviation) response time of graduates and undergraduates by stimuli



DISCUSSION

This study attempted to determine the magnitude of any interaction between articulation and nasality during perceptual judgements of speakers with repaired cleft lip and/or palate, as well as the ability of undergraduate versus graduate students to accurately rate perceived nasality in the presence of varying degrees of articulation disorders.

The first hypothesis of this study stated that there is a significant interaction between the severity of articulation deficit and perceived hypernasality as reflected in listener ratings of nasality. The results of the present study showed that such an interaction exists. This result is consistent with previous studies by Sherman (1970) and Van Hattum (1958) who both found correlations between articulation and nasality. In this study, there were significant differences in difference scores of ratings between recordings in the mild and moderate articulation error subgroups and mild and severe articulation error subgroups. The effect of articulation errors on nasality did not differ significantly between those recordings with moderate and severe nasality ratings when compared to each other. In other words, moderate and severe articulation levels were found to affect perceptual nasality to the same degree. Because the study did not include a population with normal articulation, it can't be determined if this effect would be seen between normal articulation and mild articulation errors. It could be speculated that the effect might be present, but not as strong as in those with moderate or severe errors.

When looking at the raw nasality ratings (Figure 6), it appears that when raters are listening to stimuli with mild nasality (n1), but varied degrees of articulation errors, they may be attending more to the articulation. This causes them to perceive levels of nasality that are influenced by articulation levels. That is, the articulation errors, not the nasality, are what drive the perception.

When nasality is moderate (n2), but articulation errors are mild (1), raters again may be attending to the articulation, not nasality. However, when the stimuli contains moderate (a2) or severe (a3) articulation errors, a combined influence of both the articulation errors and the nasality drive listener perception. These two stimuli were rated similarly, slightly higher than the expert scores, but not as dramatically different than those with mild nasality (n1). This suggests that the nasality plus the moderate to severe articulation errors have a combined influence, therefore making the perceived nasality seem worse to the rater.

The severe nasality stimuli (n3) seemed to not be affected by articulation, but just by the nasality. The raw data correlated with the nasality ratings and didn't change when the level of articulation errors was different.

Differences seen between stimuli might also be related to the types of articulation errors produced by the children. An informal analysis was conducted to determine the types of errors present in the recordings, with a particular focus on developmental versus compensatory errors (Table 4). The following developmental errors were heard in one or more stimuli: final consonant deletion, cluster reduction, initial consonant deletion, gliding, fronting, devoicing and weak syllable deletion. The following compensatory errors were heard in one or more stimuli: glottal stop, nasal substitution, audible nasal emission and backing.

A general trend was observed when comparing these types of errors and increased perceived nasality ratings. Those stimuli with more and a higher percentage of compensatory errors, such as nasal substitutions and audible nasal emission, tended to have higher perceived nasality ratings. Those with only, or predominantly, developmental errors tended to have lower perceived nasality. Specifically, stimuli with few articulation errors, such as stimuli a1n1 and a1n2, were rated to have the lowest perceptual nasality, regardless of the expert nasality rating.

The stimulus with mild nasality, but with only many developmental speech errors (a2n1), was judged to have moderate perceived nasality. Mild and moderate nasality stimuli that had both compensatory and developmental errors present were rated with even higher degrees of perceptual nasality (a3n1, a2n2 and a3n2). Stimuli with severe nasality (a2n3 and a3n3) all had an increased number of compensatory errors and total errors. Both of these stimuli were rated with high perceived nasality. The increased number of compensatory errors may have an impact on the perceived nasality, especially for moderately nasal utterances. This supports the notion that the type of error, not just articulation errors in general, affects perceptual nasality.

Table 4: Types and number of articulation errors present in stimuli separated by compensatory and developmental errors.

Stimuli	a1n1	a2n1	a3n1	a1n2	a2n2	a3n2	a2n3	a3n3
Compensatory Errors								
GS			3			5	3	13
Nasal Substitution			1		1	5	9	11
ANE					5		5	5
Backing					3	1	2	
Total Compensatory Errors	0	0	4	0	9	11	19	29
% of Total Errors	0	0%	27%	0%	60%	48%	56%	58%
Developmental Errors								
FCD		1			2	9	12	12
CR		1	5		2			5
ICD			2			3		
Gliding		3	4		2		3	
Fronting								2
Devocing				1				
Syllable Deletion								2
Total Developmental Errors	0	5	11	1	6	12	15	21
% of Total Errors	0	100%	73%	100%	40%	52%	44%	42%
Total Number of Errors	0	5	15	1	15	23	34	50

When performing an informal analysis of the articulation errors, it was also observed that the stimulus with ratings of moderate articulation, mild nasality presented with a hoarse and breathy vocal quality. While data regarding incidence and prevalence of voice disorders in

individuals with cleft lip and/or palate is widely varied among studies, it is generally agreed that voice disorders occur more frequently in these individuals than the rest of the population. A retrospective case study by Haamming and colleagues of 185 patients with various combinations of cleft lip and/or palate found that 20% of children ages 3-4 have vocal hoarseness and 18.4% of children ages 6-7 have hoarseness (Hamming, Finkelstein & Sidman, 2009).

Looking at the data from this study, this stimulus (a2n1) had the closest nasality rating mean between groups, but also one of the highest standard deviations compared to the other stimuli. This could indicate that both undergraduates and graduates were generally affected the same by the articulation, vocal quality and the nasality for this stimulus, but at the same time, were highly variability. It also had a difference score of over one, meaning the majority of the listeners perceived moderate nasality when there was mild nasality.

Imatoni (2005) investigated the effect of a breathy voice quality on ratings of hypernasality. The author found that when synthesizing breathy vocal quality from speech samples of children with clefts, breathiness tended to moderate the hypernasality. Those recordings with slight hypernasality were judged slightly higher with breathiness and those with severe hypernasality were judged slightly lower. Based on these results, it may be possible that the vocal quality present in the stimulus with articulation 2 and nasality 1, could have contributed to increased perceived nasality.

The second hypothesis of this study was that untrained raters, the undergraduate and graduate groups, would perceive higher nasality on the stimuli that also had articulation errors when compared to the trained (expert) raters. This hypothesis has been proven to be correct in previous studies, where high experience, not necessarily education increased reliability of perceptual judgements (Brunnegard & Van Doorn, 2009; Lewis, Watterson & Houghton, 2003).

This same effect was not fully observed in the results of this study. Graduate students' perceptual nasality ratings were not as affected by articulation errors when compared to the undergraduate's ratings. However, the graduate student raters showed lower inter- and intra-rater reliability compared to the undergraduates.

Lewis, Watterson and Houghton (2003) found that listeners with greater experience, not necessarily education, showed the highest agreements. Graduate students and undergraduate students in this study had differing education levels and experience in regards to cleft palate which would account for this difference seen between the groups. Compared to the undergraduates, the graduates were more familiar with the nasality aspect of these stimuli, due to their education and clinical experience. Because of this, they were less affected by the articulation aspect of the stimuli. The undergraduates, on the other hand, had little experience with nasality and could have been much more influenced by, or distracted by, the articulation errors.

Graduate students, however, were more variable in their ratings compared to the undergraduates. There is some amount of experience built into the curriculum, but one graduate student may have had a clinical assignment to a child with hypernasality and another never had seen an individual with hypernasality. It was not possible to quantify the graduate students' experience with this population, and the results may reflect that some of the graduates within the group differed in their experience level. This, according to Lewis and colleagues (2003), would make the results within the graduate student group variable, which was seen in this study.

Another possible explanation for the differences observed between the graduate and undergraduate groups is that the graduate raters were "over-thinking" the task, therefore leading to variability within and between the groups. To objectively test this hypothesis, the response

time was analyzed comparing the groups and the stimuli. If a noticeable difference in response time was seen between the groups or for a given stimuli, it could be that this hypothesis was true. However, no patterns or significant differences were noted when looking at the comparison of response times. Using response time, however, can't fully rule out the possibility of the group differences and high variability being affected by an "over-thinking" phenomenon. Other measures, such as a confidence rating, may better explain differences seen between the groups.

Clinical Implications

An overarching clinical implication of this study involves the issue of how to address articulation and nasality in therapy. The results of this study would suggest that addressing articulation errors may impact the perceived nasality of the speaker, especially those with mild and moderate articulation errors. In turn, since both nasality and articulation affect intelligibility, the speaker's overall intelligibility could be improved, which is the main goal of speech therapy. While physical management is necessary for an individual with structural problems and persistent VPI, speech therapy for articulation errors may reduce the amount of nasality others hear, especially for mild and moderate articulation errors.

Quality of life is significantly impacted by the ability to communicate, which is a relevant concern for children with cleft lip and/or palate. If a child is hard to understand or has speech that is different from others, it is very noticeable and may be embarrassing. Many children with cleft lip and/or palate have associated medical problems, are away from school more often, and are more likely to experience social obstacles. Adding a speech disorder that sets them apart from their peers can be difficult for that child. Their lives could be improved if articulation therapy better focused on the sounds that impact intelligibility and may help them feel less different compared to their peers.

Another important implication of this study is the awareness of this effect. Clinicians, teachers and parents must be aware that articulation may impact perceptual nasality. This study supported the fact that raters with the greatest experience and education (expert raters) are better able to make judgements with less influence of articulation. Clinical professionals and graduate programs should consider how to train individuals to develop perceptual skills that take into account both nasality and articulation errors. Education regarding this interaction is the first step, but ensuring that all professionals and graduate students in the area of cleft palate have experience is needed. This will ensure that using perceptual judgements of nasality will continue to be the "gold standard" of practice that is valid and reliable.

Limitations

There are two limitations that should be highlighted when interpreting the results of this study. One obvious limitation is that a stimulus with mild articulation errors and severe nasality was not able to be obtained for use in the study due to lack of consensus among the expert listeners. Having a stimulus missing from the matrix prevented a full analysis of the possible effects of articulation on nasality within the mild articulation error subgroup.

The second limitation involved the number of expert listeners. Four listeners qualified as both highly experienced and trained listeners who were readily available and willing to participate in this study. While they all had considerable expertise and fairly consistent ratings, the need for re-rating in order to obtain a consensus might not have been needed if more expert raters were used initially.

Future Studies

Analysis of the data for this study brought to light ideas and suggestions for future study. First, it might be interesting to look more comprehensively at the effect of articulation error type

on perceptual nasality. A general trend was seen between compensatory errors and development errors, but a more controlled investigation of articulation error types might provide valuable information regarding the interaction between articulation and nasality. A future study looking at the different error types and their effect on nasality could provide some clinical evidence as to which errors should be targeted first during therapy, since they would have the most significant impact on intelligibility.

When looking at study design, it might be useful to add normal articulation and normal nasality to the scale. This might require the stimuli population to go beyond individuals with cleft lip and/or palate, but could reveal how normal articulation compared to mild articulation affects perceptual nasality, which was missing from this study.

The addition of a confidence rating after individual's rate the nasality of each stimulus could have provided valuable information relative to the variability observed within and between the listener groups. It was concluded that response time was not a factor and that articulation error type may have had an impact, but a confidence rating might give more insight into which stimuli individuals were unsure about. Also, differences between graduate and undergraduate raters might be reflected in terms of how confident they are in their ratings. It might also be interesting to see how the differences in education and experience translate to confidence during rating.

One of the stimuli used in this study had a deviant vocal quality. While Imatoni (2005) used recording with synthesized voice qualities to assess the relationship between voice quality and nasality, it would be interesting to see which voice characteristics (e.g. roughness, breathiness or hoarseness) in natural speech samples affect perceptual nasality and by how much.

CONCLUSION

The first aim of this study was to investigate the degree to which the severity of nasality and articulation errors are associated and the strength of that association. For both undergraduate and graduate listeners, the error level in perceived nasality (compared to experts) differed significantly as the magnitude of articulation errors increased from mild to moderate and severe, but were similar when comparing moderate to severely affected articulation. Of particular interest was the observation that raters appeared to attend to articulation errors when nasality was mild. A combined influence of nasality and articulation errors was evident when nasality was moderate. Raters appeared to attend to nasality when it was severe, regardless of the articulation errors. These results may have been due to the differences in perception of compensatory and developmental errors in the stimuli.

The second aim was to confirm previous findings that education and experience affect perceived nasality by investigating differences between undergraduate and graduate listeners. Controlling for articulation and nasality, undergraduate and graduate listeners differed significantly in their accuracy of rating nasality compared to the expert raters. Overall, the graduates' ratings of nasality appeared less affected by articulation errors. The increased education of the graduates may be why they reported lower perceived nasality. Graduates had less inter-rater agreement compared to the undergraduates, perhaps as a consequence of variability in experience within the graduate.

This study provides evidence in support of articulation therapy for children with articulation errors due to a cleft lip and/or palate. Articulation therapy will not only improve their articulation skills, but it may, in turn, reduce the amount of nasality perceived by listeners and result in improved intelligibility. Future studies focusing on how specific types of errors

may impact nasality more than others may provide valuable guidance as to which articulation errors to target first.

The interaction between articulation errors and nasality has been well documented. This study provides evidence that, not only is there an interaction between the two, but that the magnitude of the articulation errors and nasality affect perceptual nasality differently. The effect of articulation errors and nasality must be considered if perceptual judgements are to remain the “gold standard” of practice, valid and reliable.

REFERENCES

- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement*, 41(3), 687-699.
- Brunnegard, K., Lohmander, A., & Van Doorn, J. (2009). Untrained listeners' ratings of speech disorders in a group with cleft palate: A comparison with speech and language pathologists' ratings. *International Journal of Language and Communication Disorders*, 44(5), 656-674.
- Cleft Palate Foundation. (2014). Prenatal Diagnosis of Cleft Lip and Cleft Palate. <http://www.cleftline.org/wp-content/uploads/2012/03/PRE-01.pdf>
- Counihan, D. T. (1960). Articulation skills of adolescents and adults with cleft palates. *The Journal of speech and hearing disorders*, 25, 181.
- Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. Elsevier Health Sciences. Pg 84.
- Goldman, R., & Fristoe, M. (1986). *Goldman Fristoe test of articulation*. American Guidance Service.
- Hamming, K. K., Finkelstein, M., & Sidman, J. D. (2009). Hoarseness in children with cleft palate. *Otolaryngology--Head and Neck Surgery*, 140(6), 902-906.

- Higgins, J. J. (2003). *Introduction to modern nonparametric statistics*.
- Imatomi, S. (2005). Effects of breathy voice source on ratings of hypernasality. *The Cleft palate-craniofacial journal*, 42(6), 641-648.
- Kantner, C. E. (1948). Diagnosis and prognosis in cleft palate speech. *Journal of Speech and Hearing Disorders*, 13(3), 211-222.
- Karnell, M. P. (1995). Discrimination of hypernasality and turbulent nasal airflow. *Cleft Palate Craniofacial Journal*. 32, 145-148.
- Kuehn, D. P. (1982). Assessment of resonance disorders. In: Lass, N., McReynolds, L., Northern, J., Yoder, D., eds. *Speech, Language and Hearing* Vol. II, Pg. 518. Philadelphia: W.B. Saunders.
- Kuehn, D. P., & Moller, K. T. (2000). Speech and language issues in the cleft palate population: The state of the art. *The Cleft palate-craniofacial journal*, 37(4), 348-348.
- Lewis, K. E., Watterson, T. L., & Houghton, S. M. (2003). The influence of listener experience and academic training on ratings of nasality. *Journal of Communication Disorders*, 36(1), 49-58.

- McWilliams, B. J. (1954). Some factors in the intelligibility of cleft-palate speech. *Journal of Speech and Hearing Disorders*, 19(4), 524-527.
- Maegawa, J., Sells, R. K., & David, D. J. (1998). Speech changes after maxillary advancement in 40 cleft lip and palate patients. *Journal of Craniofacial Surgery*, 9(2), 177-182.
- Moller, K. T., & Starr, C. D. (1984). The effects of listening conditions on speech ratings obtained in a clinical setting. *Cleft Palate Journal*, 21(2), 65-69.
- Moore, W. H., & Sommers, R. K. (1973). Phonetic contexts: their effects on perceived nasality in cleft palate speakers. *Cleft Palate J*, 10(1), 72-83.
- Peterson-Falzone, S. J., Hardin-Jones, M. A., Karnell, M. P., & McWilliams, B. J. (2001). *Cleft palate speech* (pp. 266-291). St. Louis: Mosby.
- Trost-Cardamone, J. E., Karnell, M. P., & Hardin-Jones, M. A. (2006). The clinician's guide to treating cleft palate speech (pp. 69-104). Philadelphia: Mosby/Elsevier.
- Sherman, D. (1970). Correlation Between Defective Articulation and Nasality in Cleft Palate Speech. *Cleft Palate Journal*. Vol. 7, 626-629.
- Starr, C. D., Moller, K. T., Dawson, W., Graham, J. and Skaar, S., (1984), Speech ratings by speech clinicians, parents and children. *Cleft Palate Journal*, 21, 286-292.

Subtelny, J. D., Van Hattum, R. J., & Myers, B. B. (1972). Ratings and measures of cleft palate speech. *The Cleft palate journal*, 9(1), 18-27.

Sweeney, T., & Sell, D. (2008). Relationship between perceptual ratings of nasality and nasometry in children/adolescents with cleft palate and/or velopharyngeal dysfunction. *International Journal of Language & Communication Disorders*, 43(3), 265-282.

Van Demark, D. R. (1962). *Misarticulations and listener judgments of the speech of individuals with cleft palates* (Doctoral dissertation, State University of Iowa.)

Van Hattum, R. J. (1958). Articulation and nasality in cleft palate speakers. *Journal of Speech, Language, and Hearing Research*, 1(4), 383-387.

Whitehill, T. L. (2002). Assessing intelligibility in speakers with cleft palate: a critical review of the literature. *The Cleft Palate-Craniofacial Journal*, 39(1), 50-58.

Wilcoxon, F., Katti, S. K., & Wilcox, R. A. (1970). *Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. Selected tables in mathematical statistics*, 1, 171-259

Wood, K.S. (1971). Terminology and nomenclature. In Handbook of speech pathology and audiology. (L.E. Travis). New York, Appleton-Century-Crofts.