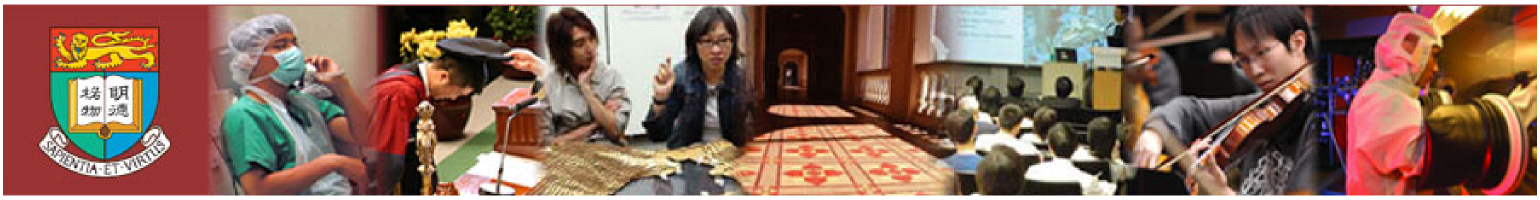

The effects of stimulus and modulus on perceptual rating of hypernasality

Cheung, Sum Chi Joanna

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, May 7, 2004.

ABSTRACT

This study investigated the effects of using different speech stimuli and direct magnitude estimation (DME) modulus types in the perceptual rating of hypernasality. The speakers were fourteen children with repaired cleft palate whose ages ranged from 5;01 years to 15;04 years (mean age 8;11 years). The listeners were twenty-four undergraduate students in the Division of Speech and Hearing Sciences. The speech stimuli included two isolated vowels (/a/ and /i/), ten monosyllabic words and six nonnasal sentences. DME with modulus (DME M) and DME with free modulus (DME FM) were used for rating. The result showed that using nonnasal sentences as stimuli would lead to significantly higher DME scores than isolated vowels. Using nonnasal sentences would also lead to significantly higher intra-listener reliability than isolated vowels and monosyllabic words. The inter-listener reliability was the highest in rating nonnasal sentences. Moreover, using DME M would lead to significantly higher DME scores than DME FM. The intra- and inter-listener reliability of DME M were higher than those for DME FM in most conditions. Possible reasons for the findings are discussed.

INTRODUCTION

Hypernasality refers to an excessively undesirable amount of perceived nasal cavity resonance due to the coupling of the nasal and oral cavities in speech production (Boone & McFarlane, 1994). Velopharyngeal insufficiency (VPI) is the primary cause of hypernasality. There are three contributing factors to VPI, they are: structural anomalies (e.g. submucous cleft palate), neurological disease (e.g. neuromuscular disorder which leads to dysarthria) and behavioral issues (e.g. hearing loss which results in lack of internal monitoring ability during speech production) (Peterson-Falzone, Hardin-Jones & Karnell, 2000; Willging, 1999). As a number of factors can lead to hypernasality, many speakers would be affected by this problem. It is thus crucial to develop methods to measure hypernasality so as to classify the severity of the problem, to design treatment accordingly and to measure the treatment effectiveness. It is also important to find a reliable and valid method for measuring the resonance property so that research can investigate the problem accurately.

Currently, both instrumental and perceptual measurements are used in assessing hypernasality or aspects related to hypernasality. For the instrumental method, the nasalance score, which refers the ratio of nasal to nasal-plus-oral acoustic energy in speech (Peterson-Falzone et al., 2000), can be used as it can reflect the severity of hypernasality. Examples of instruments which can measure nasalance are TONAR, Nasometer and Nasal View. However, differing results have been found in terms of the correlation between nasalance and perceived nasality. Dalston, Neiman & Gonzalez-Landa (1993) reported a correlation of $r = 0.78$ between nasalance and perceptual judgment of hypernasaliy in reading a nonnasal passage. Also, R.M. Dalston, Warren & E.T. Dalston (1991) reported a correlation of $r = 0.82$ in reading or repeating a nonnasal passage. However, some other studies suggested a less ideal association between nasalance and perceptual judgment of hypernasaliy, for example, Chun & Whitehill (2001) reported a correlation of $r = 0.55$ ($p < 0.01$) in repeating nonnasal sentences. Moreover, Watterson, Mcfarlane & Wright (1993) reported a correlation of $r = 0.49$ ($p = 0.06$) in reading a nonnasal passage. The discrepancies in the correlations could be as a result of differences in clinical experiences among the raters (Lewis, Watterson & Houghton, 2003) or due to the number of raters used in the study. For example, Dalston et al. (1991), who achieved the highest correlation ($r = 0.82$), involved only one rater. Watterson et al. (1993) used ten raters which resulted in greater variability among the judges and presumably lower correlation with nasalance ($r = 0.49$, $p = 0.06$).

In addition to using nasalance, it is possible to use other instrumental measurements to investigate hypernasaility. Examples are videofluoroscopy, nasopharyngoscopy, endoscopy and spectrography. Yet, the relationship between instrumental measurement, especially the acoustic measurement, and perceived nasality was unclear (Kuehn & Moller, 2000). Due to the limitation of instrumental measurement and the fact that resonance quality is fundamentally perceptual in nature (Kreiman, Gerratt, Kempster, Erman & Berke, 1993), perceptual judgment of hypernasality remains the golden standard for evaluating hypernasality.

Among the studies that have focused on perceptual judgment of hypernasality, different speech stimuli were used. For instance, Lewis et al. (2003) used connected speech, i.e. sentences which contained a variety of vowels, Zraick and Liss (2000) used isolated vowel /i/, Sherman and Hall (1978) used connected speech, i.e. reading passage, and Lintz and Sherman (1961) used CVC (where C was a nonnasal consonant and V was a vowel) syllables. However, the reason for choosing a particular type of stimuli was rarely specified.

Several rating scales can be used to rate hypernasality perceptually, e.g. direct magnitude estimation, equal-appearing intervals scale. Most previous studies have compared speech stimuli of different lengths using equal-appearing intervals (EAI) scale. Spriestersbach and Powers (1959) compared ratings of isolated vowels and a thirty-second conversational speech sample played backward from speakers with cleft palate and hypernasality. Conversational speech yielded a higher intra-listener reliability than isolated vowels ($r = 0.97$ and $r = 0.81$ respectively, p level not specified). Counihan and Cullinan (1970) studied four types of stimuli, i.e. isolated vowels, CVC syllables, nonnasal sentences played backward and nonnasal sentences played forward, produced by speakers with cleft palate and hypernasality. Nonnasal sentences played forward yielded the highest inter-listener reliability as the coefficient r of the ratings for possible pairs of judges were found to be 0.93 for isolated vowels, 0.95 for CVC syllables, 0.94 for sentences played backward and 0.96 for sentences played forward accordingly.

Daniel (1971) compared the perceived nasality of single (monosyllabic) words, three-word phrases and ten seconds of unstructured running speech. There was a statistically significant difference in severity rating for different stimuli (F = 4.18, p = 0.05). The ratings for running speech were significantly higher than those for single words (t = 2.72, p = 0.05), but no significant difference was found between single words and three-word phrases (t = 1.30, p = 0.05) or between three-word phrases and running speech (t = 1.61, p = 0.05). The study claimed that three-word phrases were a better predictor of perceptual judgment of hypernasality in unstructured running speech than single words due to the high correlation with running speech samples in terms of EAI scores (r = 0.83 for the correlation between three-word phrases and running speech and r = 0.71 for the correlation between single words and running speech, p level not specified). These results suggested that longer speech stimuli would lead to higher scores in perceptual judgment of hypernasality. Longer speech stimuli, i.e. three-word phrases, also appeared to have greater validity as they are a better predictor of hypernasality in connected speech.

Previous studies also used different types of scaling methods in the perceptual judgment of hypernasality. Most studies have used EAI scale, for example, Lewis et al. (2003) used a five-pointed EAI scale. In using EAI scale, listeners have to assign a number to each stimulus presented along a linear partition of the continuum, usually from an odd numbered scale (Schiavetti, 1992). Zraick, Liss, Dorman, Case, LaPointe & Beals (2000) used paired comparison. In this method, listeners are given two stimuli each time and they are asked to judge which the stimuli are similar or different in terms of nasal voice quality. Jones, Folkins and Morris (1990) used direct magnitude estimation (DME). DME refers to the listener assigning a number directly to the perceived magnitude of each stimulus (Engen, 1972). There are traditionally two types of DME. In DME with modulus (DME M), the experimenter presents a standard modulus to the listeners with a subjective value, usually 10 or 100 (Schiavetti, 1992). In DME without modulus (DME WM), the listeners have to select a number that they find appropriate for the first stimulus presented (Schiavetti, 1992). In both DME methods, the listeners then have to rate all subsequent stimuli in accordance with the first stimuli. The modulus may be repeated at a specified interval. For example, McHenry (1999) repeated the modulus after every stimuli and Prather (1960) presented the modulus once in the beginning of the task. Recently, a new type of DME, i.e. DME with free modulus (DME FM), has been proposed (Lee, Whitehill & Ciocca, 2003). In this procedure, the experimenter still presents a modulus to the listener, but without giving it a number. Listeners are asked to assign the standard a number, and the modulus is played at regular intervals for making subsequent judgments.

There were also studies aiming at comparing different scaling methods on judgment of hypernasality. Whitehill, Lee and Chun (2002) carried out a study on comparing DME and EAI by using connected speech, i.e. nonnasal sentences, produced by speakers with repaired cleft palate. Also, Zraick and Liss (2000) carried out a study on comparing DME and EAI by using isolated synthetic vowel /i/. Both studies revealed that DME was a more valid measure to measure hypernasality as they demonstrated that hypernasality is a prothetic continuum. Prothetic continuum is considered to have a degree of intensity or quantity. EAI is not considered as a valid rating procedure for a prothetic continuum because listeners would tend to further divide the lower end of the continuum into smaller intervals than the upper end, which results in unequal intervals (Schiavetti, 1992). The research done by Whitehill et al. (2002) further suggested that DME WM might be a better modulus type than DME M as it yielded a higher intra-listener reliability of $r = 0.95$ ($p < 0.05$) as compared with $r = 0.67$ for DME M. Previous studies which compared different speech stimuli using EAI might not have given a valid result as the scaling method used might not have been valid.

To sum up, there have been several previous studies comparing speech stimuli of different lengths in perceptual judgment of hypernasailty. However, all these studies have used EAI, which is now under question as a valid scaling method for hypernasality. Moreover, although previous studies have compared DME M and DME WM, there was no study which focused on comparing DME M and DME FM. This dissertation aims at comparing three different types of speech stimuli in perceptual judgment of hypernasality. This can help to find out if there is difference in severity when different stimuli from the same speaker are used in judging hypernasality. This could also help to find out if a particular type of stimuli leads to a more reliable judgment in terms of intra- and inter-listener reliability. Moreover, a comparison between DME M and DME FM could also help to find out if there is significant difference in severity and reliability when different modulus types in DME are applied.

Two research questions were addressed in this study:

1. Is there significant difference in listeners' perceptual judgment of hypernasality, in terms of the scores provided and, intra- and inter listener reliability, when different stimuli are used?
2. Is there significant difference in listeners' perceptual judgment of hypernasality, in terms of the scores provided and, intra- and inter listener reliability, when different direct magnitude estimation modulus types, i.e. DME WM and DME FM, are used?

It was hypothesized that longer speech stimuli would yield significantly higher DME scores, consistent with the findings of Daniel (1971). Intra- and inter-listener reliability were also predicted to be significantly higher for the sentences than the isolated vowels or monosyllabic words. This is presumably because more acoustic cues associated with nonnasal consonants are provided in longer speech stimuli and listeners were more competent to contrast the nasal quality by listening to nonnasal consonants and the relatively more hypernasal vowels adjacently (Westlake & Rutherford, 1966, cited in Counihan & Cullinan, 1970).

It was also hypothesized that no significant difference would be found in either the rating scores or inter-judge reliability between DME M and DME FM, since the listeners are referring to the same modulus. However, the intra-listener reliability for DME FM was hypothesized to be higher as the listeners are allowed to assign their own values to the modulus and use the scale that they find the most appropriate to rate the samples.

METHOD

Subjects

The speakers were fourteen Cantonese-speaking children with repaired cleft palate. This group of speakers has been previously described (Chun, 1999; Chun & Whitehill, 2001; Whitehill et al, 2002; Whitehill & Chun, 2002). See Appendix A for details. The ages of the six males and eight females ranged from 5;01 years to 15;04 years (mean age 8;11 years). All the speakers had primary repair of the palate between twelve and eighteen months old. They were recruited from the Cleft Lip and Palate Centre, Prince Philip Hospital, The University of Hong Kong. All speakers had been identified as hypernasal during a speech-language screening and were referred for videonasopharyngoscopy evaluation of velopharyngeal status as well as nasometer evaluation of nasalance. They had normal hearing abilities (as determined by previous audiometric examination). Moreover, no other neurological disease, no syndrome associated with cleft palate, no voice disorders and no hyponasality problem was reported.

The listeners were twenty-four native Cantonese speakers, twenty-three females and one male, who were fourth year undergraduate students in the Division of Speech and Hearing Science, The University of Hong Kong. They were recruited on a voluntary basis. All listeners passed a pure-tone audiometric screening from octave frequencies 250 Hz to 8000 Hz at 25dB HL according to the guidelines of The American Speech-Language-and-Hearing Association and the standards of American National Standard Institute. All of them had previous exposure to hypernasal speech during coursework and some might have had exposure during clinical placements. Eleven of the twenty-four had participated in a previous study of perceptual judgment of hypernasality using EAI, DME M and DME WM, with a different group of speakers, one and a half years ago (Whitehill et al., 2002).

Speech stimuli and data collection

All speech stimuli had been previously collected by Chun (1999). The stimuli and procedures for collecting the speech samples are summarized here. Three types of speech stimuli were used due to their common use in the previous studies. For details, please refer to Appendix B. The first was two isolated vowels, prolonged /a/ and /i/. The second out of three trials produced by the speakers and the middle 1.5 seconds of the production were selected to ensure maximum steadiness.

The second type of stimuli was ten monosyllabic words which were in C_1V or C_1VC_2

structure (where C_1 or C_2 was a nonnasal consonant and V was a vowel or diphthong). They were part of the Cantonese Single-Word Intelligibility Test (Whitehill, 1998). They were chosen since 98% of Cantonese words are in those syllabic structures (Wang, 1941, cited in Lau & So, 1988).

The third stimuli type was a set of six nonnasal sentences, varying from seven to ten syllables. They were part of the standard assessment protocol used by the Division of Speech and Hearing Sciences, The University of Hong Kong, for measuring nasalance using the Nasometer.

All speech samples were collected in a quiet room by Chun (1999), using a Sony 241 minidisk player, a Bruel & Kjaer Type 2812 pre-amplifier and a Bruel & Kjaer low noise unidirectional microphone (Model 4003), which was maintained at a mouth-to-microphone distance of 10 cm. The speakers were asked to repeat the stimuli after the experimenter since some of the younger subjects could not read fluently.

The speech samples were then filtered using a low pass filter and digitized using the computer program of Cool Edit with a sampling rate of 441K Hz and resolution of 16-bit to a Pentium III 866 desktop computer (Model no: GENIE-IV-533) by the present author. The loudness of each sound file was adjusted so as to avoid differences in judgment of hypernasality due to varying intensity levels (Counihan & Cullinan, 1972 & Zraick et.al, 2000).

Listening Procedure

The experiment was carried out with individual listeners in a sound-proof booth. The speech samples were presented to the listeners through an AKG K135 headphone which was connected to an Apple PowerMacintosh G3 computer. A HyperCard Program was used for running the experiment.

The listeners were randomly divided into two groups. The first group took part in DME M first then DME FM while the other group experienced the modulus types in the reverse order. The tasks were separated by two sessions in one week's time. Moreover, the sequence of presentation of the three types of stimuli was arranged in six different orders. In other words, the twenty-four listeners were randomly assigned to twelve different conditions. Within each type of stimuli, the speech samples were randomized by the HyperCard Program before each listening task.

In DME M task, a modulus, which was produced by a 12;11 years old Cantonese-speaking female with a repaired cleft palate, was presented in the beginning of the task. The girl was not one of the fourteen speakers described previously. She was judged to have moderately hypernasality by four experienced speech therapists. Moderately hypernasal speech was chosen for the modulus to avoid the listeners' tendencies of rating towards either the upper or lower end of the continuum (Weismer & Laures, 2002). A rating of 100 was given to the modulus. Listeners were instructed to give ratings to the subsequent stimuli proportional to the modulus. In other words, giving a rating which is higher than 100 indicates that the stimulus is more hypernasal than the modulus, vice versa. Listeners were asked to ignore the articulation errors, if any. The modulus was presented again before every trial. Listeners could replay the modulus and stimulus once in each trial.

In DME FM, the same modulus which was used in DME M was presented at the beginning of the task. However, no numerical rating was given. Listeners were asked to give the modulus a value themselves, which was not zero or a negative number. Then, they needed to rate the subsequent stimuli proportional to the value they had assigned to the modulus. Similar to DME M task, giving a rating which is higher than the modulus value indicates that the stimulus is more hypernasal than the modulus, vice versa. Listeners were asked to ignore the articulation errors, if any. The modulus again was presented before every trial and listeners could replay the free modulus and stimulus one more time in each trial.

In both tasks, all of the stimuli were repeated once in order to evaluate intra-listener reliability. Therefore, in each session, listeners listened to the eighteen stimuli (two isolated vowels, ten monosyllabic words and six nonnasal sentences) from each of the fourteen speakers two times, resulting in a total of 504 trials for each listener. Each session took about one and a half hours to finish.

Analysis

As some of the listeners had participated in a previous study concerning perceptual judgment of hypernasality, it was necessary to determine whether this would be a confounding factor. Therefore, Mann-Whitney *U* test was performed on the intra-listener reliabilities of the two groups in the six conditions to see if the listeners were from a homogeneous group. A non-parametric test was chosen since the number of listeners was unequal in the two groups, i.e. eleven listeners in the group which had participated in the previous study and thirteen listeners in the group which had not participated in that study. A significance level of 0.01 was taken in this test to ensure that the listeners were not from two different populations. If no significant difference was found between the two groups, they will be combined as one group to performing the following statistical analyses.

The arithmetic mean of each listener's judgment of each speaker for each type of stimuli was calculated. The scores of all listeners for the same speaker in the six conditions, i.e. isolated vowels with modulus (Vowel M), isolated vowels with free modulus (Vowel FM), monosyllabic words with modulus (Monosyllabic word M), monosyllabic words with free modulus (Monosyllabic word FM), nonnasal sentences with modulus (Sentence M) and nonnasal sentences with free modulus (Sentence FM), were then equalized according to the procedure of Egen (1972). This was done in order to take account of the differences in the modulus value chosen by each listener in the DME FM task, as well as possible differences caused by different listeners working with different number ranges (Egen, 1972). Means and standard deviations of the equalized scores were calculated. Then, a within-subject two-way

ANOVA was performed on those equalized scores to see if there was significant difference in the ratings when different stimulus and DME modulus types were used.

For calculating intra-listener reliability, the raw DME scores were used instead of the equalized scores as the analysis aimed at finding out the consistency within each individual listener. Pearson-Moment Correlation Coefficient was calculated between listeners' first and second ratings of the same speech stimulus in all six conditions. Means and standard deviations of the correlation coefficient across listeners were calculated. Then, a within-subject two-way ANOVA was performed on the correlation coefficients of all listeners in the six different conditions to see if there was a significant difference in intra-listener reliability for the different conditions.

For inter-listener reliability, intraclass correlation (ICC) type 3, k was calculated. Mean ratings rather than individual ratings were used for calculation.

RESULTS

Table 1 shows the result of the Mann-Whitney *U* test which was performed on the intra-listener reliability coefficients between the group of listeners who had participated in a previous study on perceptual judgment of hypernasality and the group which had not participated in it. All the *U* values in the six conditions were larger than the critical value of 27, with a significance level of 0.01. This indicated that there was no statistically significant difference between the two groups in terms of intra-listener reliability. The two groups were confirmed to be homogeneous and were considered together for all subsequent analyses.

Table 1.
Results of the Mann-Whitney *U* test on the effect of participation in a previous study on perceptual judgment of hypernasality.

	Vowel M	Vowel FM	Mono M	Mono FM	Sent M	Sent FM
U value	59	71	33	62	54	41
Significance level	0.47	0.98	0.03	0.58	0.31	0.08

Figure 1 shows the means and standard deviations of the equalized scores of the six conditions (Vowel M, Vowel FM, Monosyllabic word M, Monosyllabic word FM, Sentence M and Sentence FM). The means varied from 36.75 for Vowel FM to 135.01 for Sentence M. The standard deviations varied from 6.18 for Monosyllabic word FM to 56.47 for Sentence M.

The within-subject two-way ANOVA revealed there was a significant main effect for stimuli type, $F(2,26) = 3.66$, $p < 0.05$. The Tukey HSD post-hoc test found that the scores for nonnasal sentences were significantly higher than that of isolated vowels ($p < 0.05$). The differences between isolated vowels and monosyllabic words, and between monosyllabic words and nonnasal sentences were not statistically significant ($p > 0.05$ in the two comparisons).

There was also a significant main effect for DME modulus type, $F(1,13) = 216.13$, $p < 0.01$. The equalized scores in DME M were significantly higher than those for DME FM.

No significant interaction effect was found between the variables of stimuli type and modulus type ($F[2,26] = 2.58$, $p > 0.05$).

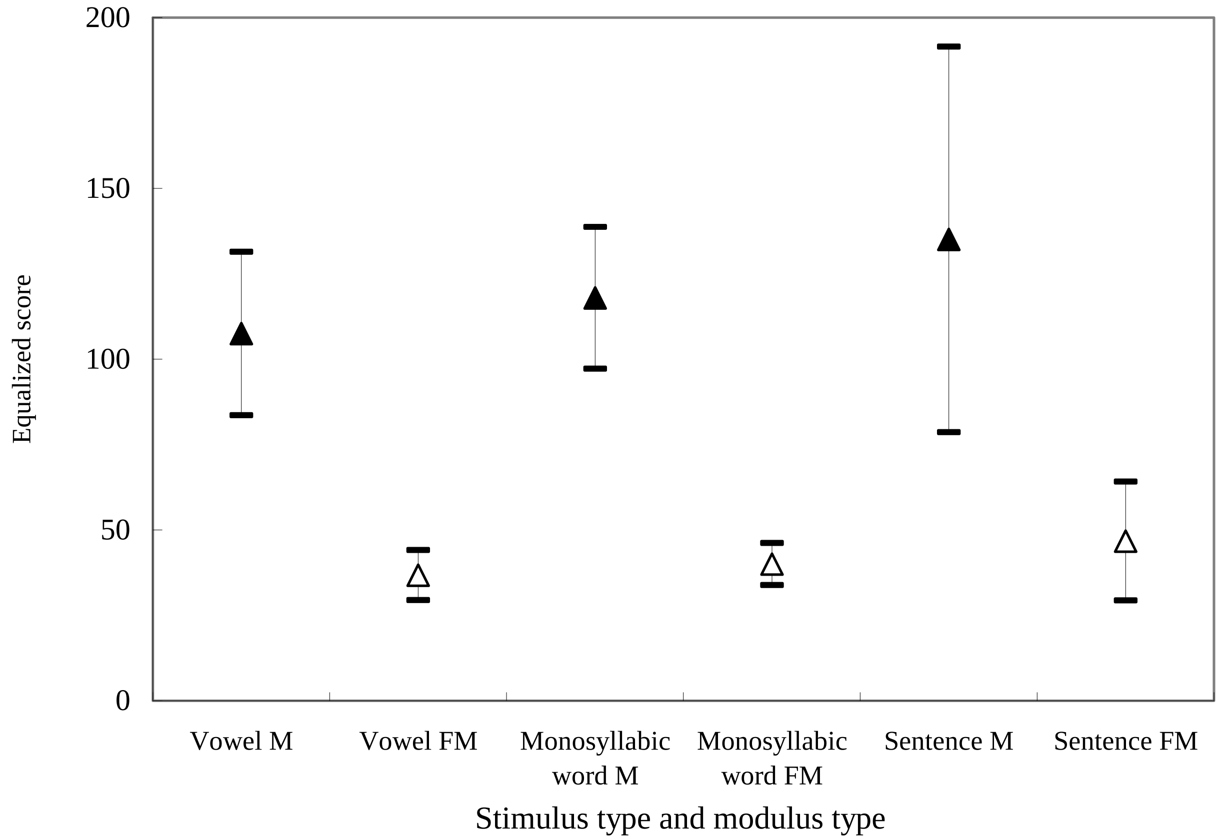

Figure 1. Means and standard deviations of the equalized scores.

Figure 2 shows the means and standard deviations of the correlation coefficients in the six conditions. The means varied from 0.59 for Vowel M to 0.79 for Sentence M. The standard deviations varied from 0.11 for Sentence FM to 0.23 for Vowel M. A within-subject two-way ANOVA was performed on the correlation coefficients of the six different conditions. There was a significant main effect for stimuli type, $F(2,46) = 14.90$, $p < 0.01$. Tukey HSD post-hoc tests showed that the intra-listener reliability of nonnasal sentences was significantly higher than that of isolated vowels ($p < 0.01$) and monosyllabic words ($p < 0.01$). There was no significant difference between isolated vowels and monosyllabic words ($p > 0.05$).

There was no significant main effect for modulus type, $F(1,23) = 0.02$, $p > 0.05$. In other words, using DME M versus DME FM did not lead to a significant difference in intra-listener reliability.

There was also no significant interaction effect between stimulus type and DME modulus type in intra-listener reliability ($F[2,46] = 1.37$, $p > 0.05$).

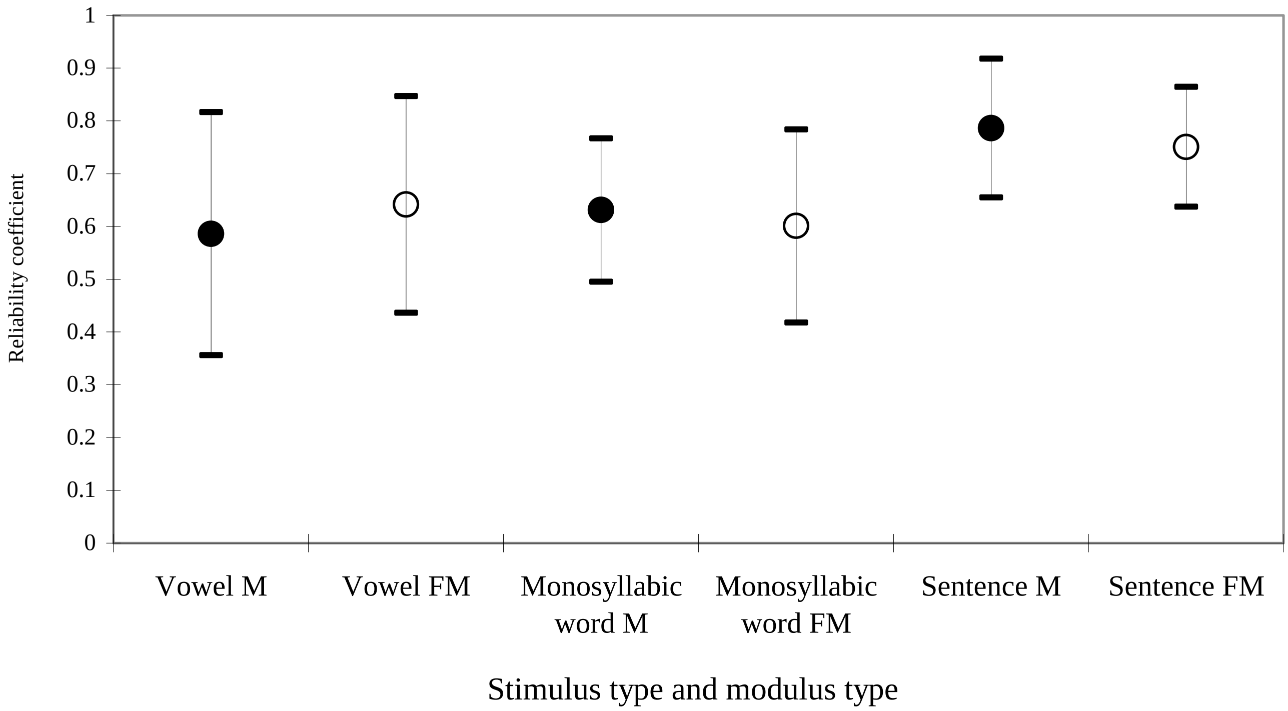

Figure 2. Means and standard deviations of the intra-listener reliability coefficients.

Table 2 shows the inter-listener reliability. The reliability varied from 0.40 for Vowel FM to 0.96 for Sentence M. When the type of stimuli is considered, the highest inter-listener reliability was found when nonnasal sentences were used (0.96 and 0.92). Inter-listener reliability was also higher for monosyllabic words (0.91 and 0.83). The inter-listener reliability was poorer for isolated vowels (0.76 and 0.40). When the type of DME modulus was concerned, DME M was found to yield a higher inter-listener reliability than DME FM in all three stimulus types.

Table 2.
Inter-listener reliability (ICC) in the six conditions.

	Vowel M	**Vowel FM**	**Monosyllabic word M**	**Monosyllabic word FM**	**Sentence M**	**Sentence FM**
ICC	0.76	0.40	0.91	0.83	0.96	0.92

DISCUSSION

When the effect of using speech stimuli of different lengths is considered, the equalized scores of sentences were significantly higher than those in isolated vowels. The finding was consistent with the hypothesis. Daniel (1971) reported that EAI scores were significantly higher in unstructured running speech than single words. The discrepancies in the means of the equalized scores between nonnasal sentences and monosyllabic words agreed with the findings of Daniel (1971) that longer stimuli tended to yield higher ratings, although the scores in nonnasal sentences were not significantly higher than those of monosyllabic words. This indicated that longer speech stimuli, such as sentences, would be perceived as more hypernasal than shorter stimuli, such as monosyllabic words and isolated vowels, when produced by the same speakers.

This finding may be attributable to the acoustic cues associated with consonants found in nonnasal sentences (Westlake & Rutherford, 1966, cited in Counihan & Cullinan, 1970). In judging nonnasal sentences, the listeners tended to be influenced by the large contrast between the nonnasal consonants and the relatively more hypernasal vowel adjacently. The occurrence of consonant-vowel contrast in nonnasal sentences might cause listeners to perceive the sentences as more hypernasal compared with isolated vowels thus resulted in significantly higher scores on sentences compared with isolated vowels. As similar consonant-vowel contrast appears on both nonnasal sentences and monosyllabic words, the difference in the ratings between these two types of stimulus were found to be not statistically significant.

Another possible reason for the present finding is due to the physiology of velar movement for speech. In producing nonnasal sentences, the physiological demand for maintaining appropriate velopharyngeal closure is higher compared with the demand during producing isolated vowels or monosyllabic words. Since there are relatively more articulatory movements in the production of sentences than monosyllabic words and isolated vowels, the increase in movement of other articulators may lead to a trade-off on maintaining an appropriate closure of the velopharyngeal port. So, the DME ratings on nonnasal sentences were generally higher than the other types of stimuli.

Yet another possible explanation is speakers in this study had articulation errors in their production of nonnasal sentences. For example, 35.71% (5/14) of the speakers had stopping in their production of sentences. For details, please refer to Appendix C. As suggested by Sherman & Hall (1978), a decrease in articulatory precision would lead to an increase in perceived nasality. It is suspected that the judgment of the listeners could have been affected by the speakers' articulation even though the listeners were instructed to ignore the articulation errors. Therefore, the higher hypernasality ratings on nonnasal sentences could partially be due to the interference of the articulation errors. However, no significant difference was found between the equalized scores of monosyllabic words, which were free of articulation error, and nonnasal sentences. The present finding suggested that articulation error might not be a main factor which interferes perceptual judgment of hypernasality.

The significantly higher intra-listener reliability found for the judgment of nonnasal sentences compared with isolated vowels and monosyllabic words supported initial hypothesis. This also supported the findings of Spriesterbach & Powers (1959) that judging longer speech stimuli, i.e. 30 seconds of conversational speech played backward, yielded higher intra-listener reliability than shorter samples, i.e. isolated vowels. A possible reason for the significantly higher intra-listener reliability in nonnasal sentences is that acoustic cues associated with nonnasal consonants are important elements for perceptual judgment of hypernasality (Westlake & Rutherford, 1966, cited in Counihan & Cullinan, 1970). Therefore, when relatively more nonnasal consonants appear in sentences, listeners tended to contrast the nonnasal consonants and the relatively more hypernasal vowels consistently in judging hypernasality. This results in higher reliability within listeners. In future research on the perceptual judgment of hypernasality, investigators may wish to consider the significantly higher intra-listener reliability on nonnasal sentences found in this study.

Concerning the inter-listener reliability, nonnasal sentences were found to have the highest reliability, followed by monosyllabic words and then isolated vowels. The finding agreed with the hypothesis and was generally consistent with the study by Counihan & Cullinan (1970) who found that the inter-listener reliability of nonnasal sentences played forward was higher than that for CVC syllables, followed by nonnasal sentences played backward, and vowels. Again, listeners might be helped by the acoustic cues associated with the nonnasal consonant-vowel contrast in the judgment of monosyllabic words and nonnasal sentences. Therefore, there tends to be a higher reliability among listeners in rating monosyllabic words and nonnasal sentences. The use of nonnasal sentence in future research is suggested to increase the inter-listener reliability.

When the effect of using different modulus types was considered, hypernasality ratings for DME M were significantly higher than those for DME FM. This did not support the initial hypothesis. When the modulus values which were chosen by the listeners in DME FM task was considered, it was found that three listeners gave the modulus a single-digit value in rating isolated vowels and monosyllabic words. Four listeners used a single-digit modulus value in judging nonnasal sentences. Although an equalizing procedure was done to balance the differences due to different modulus values chosen by the listeners, the extreme numbers selected by individual listeners might still have affected the values of equalized scores.

There was no significant difference in the intra-listener reliability when DME M and

DME FM were compared. The result did not support the initial hypothesis that DME FM could yield higher intra-listener reliability. However, the reliability was numerically higher in DME M than in DME FM in two out of the three stimuli conditions, i.e. in monosyllabic words and nonnasal sentences. The finding that DME M tended to yielded higher intra-listener reliability than DME FM may be possibly due to the listeners in the present study have relatively less experience in perceptual judgment of hypernasality using DME. Therefore, they might be likely to make consistent judgment when a modulus with a given value was provided by the experimenter. These findings suggest that DME M tends to yield higher intra-listener reliability and thus should be considered in the future.

Moreover, DME M was found to yield higher inter-listener reliability than DME FM in all three stimulus types. These findings again suggested that DME M might be a preferable rating method due to its higher inter-listener reliability in different stimuli conditions. The discrepancy in the inter-listener reliability between DME M and DME FM was the most obvious for vowels. This result suggested that using DME FM and isolated vowels to rate hypernasality is the least desirable method, as this results in the lowest inter-listener reliability.

The present study focused only on using DME as the rating method. The comparison on the effect of using three different speech stimuli, i.e. isolated vowels, monosyllabic words and nonnasal sentences, and two different DME modulus types, i.e. DME M and DME FM, for the perceptual judgment of hypernasality revealed that using nonnasal sentences and DME M appears to be the most suitable choice in terms of the intra- and inter-listener reliability. Further study can be carried out comparing the reliability of using other rating methods for perceptual judgment of hypernasality, such as paired comparisons or visual analog scale.

One point should be considered in interpreting the present results. Only one modulus from a female speaker was used in the study, it was suspected that the listeners might find the present modulus to be easier to refer to female speakers due to similar pitch level. Choosing one modulus from a female speaker and another modulus from a male speaker, and further dividing the tasks in blocks based on gender, may lead to a difference to the result.

CONCLUSION

To summarize, the aim of this study was to investigate the effects of stimulus type and modulus type on listeners' ratings of hypernasality in children with cleft palate. When the equalized ratings of isolated vowels, monosyllabic words and nonnasal sentences were compared, the hypernasality ratings for nonnasal sentences were significant higher than those for isolated vowels. Intra-listener reliability for nonnasal sentences was found to be significantly higher than those for monosyllabic words and isolated vowels. The highest value for inter-listener reliability was also found for nonnasal sentences in all two modulus types.

When DME M and DME FM were compared, the equalized ratings using DME M were significant higher than those for DME FM. No significant difference in intra-listener reliability was found between the two moduli. However, the intra-listener reliability appeared numerically higher for DME M than DME FM in two out of three conditions, i.e. monosyllabic words and nonnasal sentences. The inter-judge reliability in DME M was higher than that for DME FM in all three stimulus conditions.

The findings of this study suggested that using nonnasal sentences and DME M in perceptual judgment of hypernasality should be encouraged as this combination could result in a higher consistency of judgment within individual listeners as well as more consistency across different listeners.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Tara Whitehill for her guidance and support throughout the process of this study. I would also thank Alice Lee, Joan Ma and Joyce Chun for their advice in the development of this dissertation. Thanks are also extended to the technicians, Donald Chan and Raymond Wu, of the Division of Speech and Hearing Sciences for their technical support. Finally, I thank all the subjects who participated in this study for their time devoted.

REFERENCES

Boone, D.R. & Mcfarlane, S.C. (1994). *The voice and voice therapy (5th Ed.)*. New Jersey: Prentice Hall.

Chun, J.C. (1999). *The relationship between nasalance, nasality and intelligibility of Cantonese children with cleft palate*. Unpublished undergraduate thesis, University of Hong Kong.

Chun, J.C. & Whitehill, T. L. (2001). The relationship between nasalance and nasality in Cantonese children with cleft palate. *Asia Pacific Journal of Speech, Language and Hearing, 6,* 135-147.

Counihan, D.T., & Cullinan, W.L. (1970). Reliability and dispersion of nasality ratings. *Cleft Palate Journal, 7,* 261-270.

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

Dalston, R.M., Neiman, G.S., & Gonzalez-Landa, G. (1993). Nasometric sensitivity and specificity: A cross-dialect and cross culture study. *Cleft Palate-Craniofacial Journal, 30,* 285-291.

Dalston, R.M., Warren, D.W., & Dalston, E.T. (1991). Use of nasometry as a diagnostic tool for identifying patients with velopharyngeal impairment. *Cleft Palate-Craniofacial Journal, 28,* 184-189.

Daniel, H.J. (1971). Nasality ratings of single words, phrases, and running speech samples obtained from cleft palate children. *Folia Phoniatrica et Logopaedica, 23,* 41-49.

Engen, T. (1972). Psychophysics II. Scaling methods. In J. W. Kling & L. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology.* (pp. 47-86). New York: Holt, Rinehart, & Winston.

Jones, D.L., Folkins, J.W., & Morris, H.L. (1990). Speech production time and judgments of disordered nasalization in speakers with cleft palate. *Journal of Speech and Hearing Research, 33,* 458-466.

Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., & Berke, G.S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for further research. *Journal of Speech and Hearing Research, 36,* 21-40.

Kuehn, D.P. & Moller, K.T. (2000). Speech and language issues in the cleft palate population: The state of the art. *Cleft Palate-Craniofacial Journal, 37,* 348-1-348-35.

Lau, C.C. & So, K.W. (1988). Material for Cantonese speech audiometry constructed by appropriate phonetic principles. *British Journal of Audiology, 22,* 297-304.

Lee, S. Y. A., Whitehill, T. L., & Ciocca, V. (2003 November). *The effect of listener training on perceptual ratings of hypernasality*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Chicago.

Lewis, K.E., Watterson, T.L. & Honghton, S.M. (2003). The influence of listener experience and academic training on ratings of nasality. *Journal of Communication Disorders, 36*, 49-58.

Lintz, L.B. & Sherman, D. (1961). Phonetic elements and perception of nasality. *Journal of Speech and Hearing Research, 4,* 381-398.

McHenry, M.A. (1999). Aerodynamic, acoustic, and perceptual measures of nasality following traumatic brain injury. *Brain Injury, 13,* 281-290.

Peterson-Falzone, S. J., Hardin-Jones, M.A., & Karnell, M.P. (2000). *Cleft palate speech (3th Ed.).* St. Louis: Mosby.

Prather, E.M. (1960). Scaling defectiveness of articulation by direct magnitude-estimation. *Journal of Speech and Hearing Research, 3,* 380-392.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R.D. Kent (Ed.), *Intelligibility in Speech disorders: Theory, measurement and management* (pp.11-34). Amsterdam: John Benjamins Publishing Co.

Sherman, D. & Hall, P.K. (1978). Nasality and precision of articulation. *Perceptual and Motor Skills, 46,* 115-118.

Spriestersbach, D.C. & Powers, G.R. (1959). Nasality in isolated vowels and connected speech of cleft palate speakers. *Journal of Speech and Hearing Research, 2,* 40-45.

Watterson, T., McFarlane, S.C., & Wright, D.S. (1993). The relationship between nasalance and nasality in children with cleft palate. *Journal of Communication Disorders, 26,* 13-28.

Weismer, G. & Laures, J.S. (2002). Direct magnitude estimates of speech intelligibility in dysarthria: Effects of a chosen standard. *Journal of Speech, Language, and Hearing Research, 45,* 421-433.

Whitehill, T.L. (1998). *Speech intelligibility in Cantonese speakers with congenital dysarthria*. Unpublished doctoral dissertation. Hong Kong: University of Hong Kong.

Whitehill, T.L., & Chun, J.C. (2002). Intelligibility and acceptability in speakers with cleft palate. In F. Windsor, M. L. Kelly, & N. Hewlett (Eds.), *Investigations in clinical phonetics and linguistics* (pp.405-416). New Jersey: Lawrence Erlbaum Association.

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

Willging, J.P. (1999). Velopharyngeal insufficiency. *International Journal of Pediatric Otorhinolaryngology, 49,* 307-309.

Zriack, R.I. & Liss, J.M. (2000). A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *Journal of Speech, Language and Hearing Research, 43,* 979-988.

Zraick, R.I., Liss, J.M., Dorman, M.F., Case, J.L., LaPointe, L.L., & Beals, S.P. (2000). Multidimensional scaling of nasal voice quality. *Journal of Speech, Language and Hearing Research, 43,* 989-996.

APPENDIX A

Speaker Information

Name of Speakers	Age	Gender
CLK	15;04	Female
WPS	12;00	Male
LWC	10;02	Male
LPK	10;01	Female
MHM	9;11	Male
THM	9;08	Male
LBY	8;11	Female
LNF	8;03	Female
FYT	8;02	Female
HHC	8;01	Male
HHK	8;00	Female
LCL	5;11	Male
LWY	5;08	Female
LTM	5;01	Female
CKW (modulus)	12;11	Female

APPENDIX B

Stimuli used for perceptual rating of hypernasality

Isolated vowels:

1. /a/
2. /i/

Monosyllabic words:

1. 代 tɔi22
2. 該 kɔi55
3. 教 kau33
4. 靠 k^{h}au33
5. 開 hɔi 55
6. 蝦 ha55
7. 呵 hɔ55
8. 叔 suk5
9. 飛 fei55
10. 稅 sœy33

Nonnasal sentences:

1. 布袋有四十四塊大石頭
   /pou33 tɔi35 jɐu23 sei33 sɐp2 sei33 fai33 tai22 sek2 t^{h}ɐu21/
2. 這是一束白菊花
   /tsɛ35 si22 jɐt5 ts^{h}uk5 pak2 kuk5 fa55/
3. 伯伯有一個大鼻哥
   /pak3 pak3 jɐu23 jɐt5 kɔ33 tai22 pei22 kɔ55/
4. 就快落大雨帶把遮出街
   /tsɐu22 fai33 lɔk2 tai22 jy23 tai33 pa35 tsɛ55 ts^{h}œt5 kai55/
5. 婆婆叫哥哥餵雞仔
   /p^{h}ɔ21 p^{h}ɔ35 kiu33 kɔ21 kɔ55 wɐi33 kɐi55 tsɐi35/
6. 爸爸最怕排隊搭車
   /pa21 pa55 tsœy33 p^{h}a33 p^{h}ai21 tœy35 tap3 ts^{h}ɛ55/

APPENDIX C

Articulation proficiency of speakers in producing nonnasal sentences

Name of Speakers	Percentage of Segment Correct (%)	Error Pattern (Number of Incidence)
CLK	100	Nil
WPS	99.39	Stopping, i.e. /ts/ → [t] (1)
LWC	98.79	Stopping, i.e. /ts/ → [t] (2)
LPK	100	Nil
MHM	99.39	Stopping, i.e. ts/ → [t] (1)
THM	100	Nil
LBY	100	Nil
LNF	100	Nil
FYT	98.18	Diphthong Reduction, ie. /ai/ → [a] (2), Vowel Substitution, i.e. /y/ → [i] (1)
HHC	100	Nil
HHK	96.36	Stopping, i.e. /ts/ → [t] (5), Vowel Substitution, i.e. /y/ → [i] (1)
LCL	93.29	Backing, i.e. /t/ → [k] (5), Stopping, i.e. /s/ → [t] (5), /s/ → [t] (1)
LWY	93.94	Aspiration, i.e. /ts/ → [ts^h] (4), Fronting, i.e. /t/ → [p] (2), Backing, i.e. /t/ → [k] (1), /t^h/ → [h] (1), Frication, i.e. /t^h/ → [h] (1), Diphthong Reduction, i.e. /ai/ → [a] (1), Nasalization, i.e. /ts/ → [m] (1)
LTM	100	Nil
CKW (modulus)	99.39	Vowel Substitution, i.e. /ɔ/ → [u] (1)

Note. A segment refers to an initial consonant, a final consonant, a vowel, a diphthong, or a lexical tone.