The Relationship Between Nasalance Scores and Nasality Ratings Obtained With Equal Appearing Interval and Direct Magnitude Estimation Scaling Methods

Tami U. Brancamp, Ph.D., Kerry E. Lewis, Ph.D., Thomas Watterson, Ph.D.

Objectives: **To assess the nasalance/nasality relationship and Nasometer test sensitivity and specificity when nasality ratings are obtained with both equal appearing interval (EAI) and direct magnitude estimation (DME) scaling procedures. To test the linearity of the relationship between nasality ratings obtained from different perceptual scales.**

Stimuli: **Audio recordings of the Turtle Passage.**

Design: **Participants' nasalance scores and audio recordings were obtained simultaneously. A single judge rated the samples for nasality using both EAI and DME scaling procedures.**

Participants: **Thirty-nine participants 3 to 17 years of age. Across participants, resonance ranged from normal to severely hypernasal.**

Main Outcome Measures: **Nasalance scores and two nasality ratings.**

Results: **The magnitude of the correlation between nasalance scores and EAI ratings of nasality ($r = .63$) and between nasalance and DME ratings of nasality ($r = .59$) was not significantly different. Nasometer test sensitivity and specificity for EAI-rated nasality were .71 and .73, respectively. For DME-rated nasality, sensitivity and specificity were .62 and .70, respectively. Regression of EAI nasality ratings on DME nasality ratings did not depart significantly from linearity.**

Conclusions: **No difference was found in the relationship between nasalance and nasality when nasality was rated using EAI as opposed to DME procedures. Nasometer test sensitivity and specificity were similar for EAI- and DME-rated nasality. A linear model accounted for the greatest proportion of explained variance in EAI and DME ratings. Consequently, clinicians should be able to obtain valid and reliable estimates of nasality using EAI or DME.**

KEY WORDS: *DME, EAI, nasalance, nasality*

Psychophysical Scaling of Nasality

Nasality is a perceptual sensation that can be measured only by human judgment (Bzoch, 1979). Stevens (1975) has hypothesized that all sensations, such as nasality, are mentally processed as prothetic or metathetic sensations. Prothetic stimuli are described as having a degree of quantity, and metathetic stimuli are thought to vary in terms of quality. When prothetic stimuli are compared, a change in the stimulus would be perceived as an increase or decrease in the sensation under observation. Loudness would be an example of a prothetic sensation. When metathetic stimuli are compared, however, a new excitation is exchanged for an excitation that has been removed. Pitch would be an example of a metathetic sensation. Alternative theories dispute the prothetic/metathetic dichotomy (Warren and Sersen, 1958; Warren and Warren, 1963; Warren 1973). For example, the physical correlate theory proposes that all estimates of stimulus intensity are based upon "experience with the manner in which the levels of sensory excitation are correlated with some physical attribute of the stimulus" (Warren and Sersen, 1958, p. 700). For example, loudness judgments might be based on experience with distance, that is, one stimulus is judged half as loud as another because it sounds twice as far away.

Stevens (1975) has further proposed that different psychophysical scales are necessary for judging prothetic versus metathetic sensations (e.g., the ratio scale versus the category scale). Category scales require listeners to partition sensations into discrete categories. The equal appearing interval (EAI) scale is a type of category scale that is commonly used in studies of speech and resonance (e.g., Dalston et al., 1991; Hardin et al., 1992; Watterson et

---

Dr. Brancamp is Assistant Professor, and Drs. Lewis and Watterson are Professors, University of Nevada, School of Medicine, Reno, Nevada.

This paper was presented, in part, at the Annual Meeting of the American Cleft Palate Craniofacial Association, Scottsdale, Arizona, April 2009.

Submitted June 2009; Accepted December 2009.

Address correspondence to: Dr. Tami U. Brancamp, University of Nevada School of Medicine, Department of Speech Pathology and Audiology, Redfield Building/152, Reno, NV 89557. E-mail tbrancamp@medicine.nevada.edu.

DOI: 10.1597/09-106

al., 1998). Ratio scales require listeners to make judgments of magnitude. Direct magnitude estimation (DME) is a commonly used ratio scale (e.g., Toner et al., 1990; Schiavetti et al., 1994; Eadie and Doyle, 2002). Stevens believes that prothetic sensations are best judged with ratio scales and that metathetic sensations are best judged with category scales. To distinguish prothetic from metathetic, Stevens suggests that perceptual ratings from a category scale should be regressed onto ratings obtained from a ratio scale. If the best-fit line is linear, the perceptual sensation is metathetic. This linear relationship results from the listener's assigning EAI ratings that reflect equal interval perceptual discriminations. If the resulting regression line is curvilinear, the perceptual sensation is prothetic. In this case, the listener's EAI ratings do not reflect perceptual discriminations that are of equal intervals.

Two recent studies have presented data suggesting that nasality is a prothetic perceptual sensation (Zraick and Liss, 2000; Whitehill et al., 2002). Zraick and Liss had listeners rate nasality in one isolated, synthesized vowel (/i/). Progressive nasality was simulated by manipulating the vowel formant frequency and bandwidth and by adding nasal poles (extra formants). Twelve student listeners first judged the "nasal" vowels using an EAI scale on which 1 represented "least nasality" and 5 represented "most nasality." Then they rated the same stimuli using DME with modulus. A modulus is an example of the dimension being rated, and it is typically assigned a predetermined value in the middle of the scale. Results showed a significant curvilinear relationship when mean EAI ratings were regressed on the geometric mean of the DME ratings. According to Stevens (1975), this would indicate that the dimension being rated (nasality) is prothetic. The authors concluded, therefore, that nasality "cannot be validly rated using EAI scaling" (Zraick and Liss, 2000, p. 985). In another study, Whitehill et al. (2002) had listeners rate human speech using three rating scales. The speakers were hypernasal patients with a history of cleft lip/palate. All listeners rated the speech samples using a seven-point EAI scale. Then, one-half of the listeners rated the samples using a DME scale without modulus, and one-half rated the samples using a DME scale with modulus. Results showed a significant curvilinear relationship between mean EAI ratings and mean ratings for DME both with modulus and without modulus. In accordance with Stevens' theory, the authors concluded that "…EAI may not be a valid method for the evaluation of hypernasality" (Whitehill et al., 2002, p. 80).

NASALANCE AND NASALITY

The Nasometer is a popular computer-based software system that purports to measure a physical correlate of nasality called *nasalance*. A nasalance score is obtained by positioning head gear that contains two microphones separated by a metal plate. The acoustic signal at each microphone is filtered with a 300-Hz bandwidth filter that has a center frequency at 500 Hz. The filtered signals are compared by the ratio of the nasal energy divided by the oral plus nasal energy and multiplied by 100.

The relationship between nasalance scores and nasality judgments has varied across studies from very good to poor. This variability may be related to differences in design such as different subject populations, different machines, different stimuli, and different methods of obtaining nasality judgments. Dalston, Warren, and Dalston (1991) compared judgments of nasality versus nasalance scores in speech samples from 96 subjects. Nasality was rated on a six-point EAI scale. They reported a correlation of $r = .82$ between nasalance scores and nasality. Test sensitivity was reported at .89, and test specificity was .95. Nellis, Nieman, and Lehman (1992) compared listener judgments of hyponasality and hypernasality versus nasalance scores. Judgments of nasality were obtained on two six-point EAI scales on which 1 indicated no hypernasality or hyponasality, and 6 indicated severe hypernasality or hyponasality. Scores were reported individually for seven different sentences, but the main finding of the study was a "lack of correlation" (p. 161) between nasalance scores and listener judgments of nasality. Correlation coefficients between nasalance and nasality for individual listeners ranged from a low of $r = .02$ to a high of $r = .43$. Hardin et al. (1992) compared nasalance scores and nasality judgments for 75 subjects. Nasality ratings were made using a seven-point EAI scale. Investigators did not report correlation coefficients, but test sensitivity was .76, and test specificity was .85. Watterson, McFarlane, and Wright (1993), using a five-point EAI scale, found a correlation of $r = .49$ between nasalance and nasality for a stimulus containing no nasal phonemes. Test sensitivity was .71, and test specificity was .55. Watterson, Hinton, and McFarlane (1996) compared nasalance scores on two passages (Zoo Passage and Turtle Passage) versus listener ratings of nasality. Both passages are devoid of nasal phonemes. Listeners rated nasality on a five-point EAI scale whereby 1 represented normal nasality, and 5 represented severe hypernasality. The correlation coefficient between nasality and nasalance for the Zoo Passage was $r = .70$, and for the Turtle Passage the correlation was $r = .51$. Test sensitivity was .72 for the Zoo Passage and .83 for the Turtle Passage. Watterson, Lewis, and Deutsch (1998) compared nasalance scores and nasality judgments for two different types of speech stimuli: sentences composed of only high-pressure consonants and vowels, and sentences composed of only low-pressure consonants and vowels. The sentences were rated for nasality using a seven-point EAI scale. The correlation coefficient between nasalance and nasality was $r = .78$ for the low-pressure stimulus and $r = .77$ for the high-pressure stimulus. Overall test sensitivity was .84, and test specificity was .88. Lewis, Watterson, and Houghton (2003) compared nasality ratings versus nasalance scores for three different groups of listeners clustered according to training and experience. Nasality ratings were made by using a five-point EAI scale.

Correlation coefficients between nasalance and nasality for individual listeners ranged from a low of $r = .29$ to a high of $r = .60$.

In all studies that have investigated the relationship between nasalance and nasality, nasality ratings were always obtained using an EAI scale. However, if nasality ratings obtained by EAI scaling are not valid estimates of nasality, as has been suggested by some recent studies, then it should be possible to show improved correspondence between nasalance and nasality by using a magnitude estimation procedure.

The purpose of this study, therefore, was to investigate the relationship between nasalance scores and nasality judgments when nasality ratings were obtained both with equal appearing interval (EAI) scaling procedures and with direct magnitude estimation (DME) procedures. Specific research questions included the following:

1. Is the relationship between nasalance and nasality stronger when nasality is rated using DME scaling than when nasality is rated using EAI scaling?
2. How are sensitivity and specificity measures affected when nasality is rated with DME versus EAI scaling?
3. Does the EAI/DME relationship depart significantly from linearity, suggesting that nasality is a prothetic sensation?

METHOD

Thirty-nine speakers produced the nonnasal Turtle Passage as nasalance scores and audio recordings were obtained simultaneously. Nasality was rated for each stimulus from the audio recordings using an EAI scale and using DME with modulus.

Participants

Speakers

The speakers were 39 children and adolescents ranging in age from 3.8 to 17.2 years. Twenty-five of the speakers had a history of hypernasality. The hypernasal speakers were patients followed by a cleft palate team. Fourteen speakers had normal speech, reported a negative history of hypernasality, and were not followed for speech or language deficits. Data were collected under a protocol approved by the Social Behavioral Institutional Review Board.

Judge

A single judge (T.W.) with more than 30 years' experience in assessing resonance disorders provided the perceptual ratings of nasality. The judge passed a hearing screening bilaterally at 1 kHz, 2 kHz, and 4 kHz at 25dB.

Stimulus

The Turtle Passage was used to obtain speech samples from the 39 speakers. The Turtle Passage consists of 29 syllables and contains no normally nasal phonemes. It has been shown to be syntactically and semantically within the competence of young children and comparable with the Zoo Passage in terms of nasalance scores obtained (Watterson et al., 1996).

Instrumentation

The Nasometer II, Model 6400 (Kay-Pentax Elemetrics, Inc., Lincoln Park, NJ) was used to obtain a mean nasalance score for each speech sample. Speech samples were audio recorded with a Shure Model SM48 microphone connected to a CD recording system (Marantz Professional Model CDR300, D & M Professional, Itasca, IL). The speech samples were reproduced for rating of nasality on a compact disc player (Pioneer Model CDJ-100MK3, Pioneer Electronics, Tokyo, Japan) with dual speakers (Yamaha Model MSP3, Yamaha Corporation, Buena Park, CA).

Procedures

Nasalance and Audio Recordings

Before each recording session, the Nasometer was calibrated according to the manufacturer's specifications. Each participant was seated in a comfortable chair in a sound-isolated room. The examiner placed the Nasometer headgear according to the manufacturer's specifications. The microphone was positioned along the same plane as the Nasometer headgear separation plate with an apparatus that held it at a fixed distance of 13 cm. Speech samples were recorded simultaneously with the nasalance score. Participants who could read fluently read the Turtle Passage. Those who could not read fluently recited after the examiner. Nasalance scores were entered onto a data sheet, and a numeric code was assigned to corresponding nasalance scores and audio recordings for data analysis.

Audio Disks for Nasality Ratings. Two audio disks were created for obtaining the nasality rating data. One disk was used for EAI scaling, and the other for DME scaling. The 39 speech samples were dubbed, in different random order, onto the two audio disks. In addition, on each disk, eight of the 39 speech samples were selected at random and were duplicated at the end of each disk for later assessment of the judges' intrarater agreement. The audio disk used to make DME ratings of nasality also contained a reference speech sample selected to represent moderate hypernasality. Two of the investigators (T.B. and K.L.) chose the reference sample by consensus from an available pool of Turtle Passage recordings not otherwise used in the present study. This pool of speech samples was obtained from children with hypernasal speech. The reference sample, which had a nasalance score of 42%, appeared as the first sample on the DME audio disk and again after every five experimental speech samples.

Interrater and Intrarater Agreement on Nasality Ratings. Interrater agreement was established for the five-point EAI scale. Scale values ranged from 1, representing normal resonance, to 5, representing severe hypernasality. The study judge (T.W.) and a clinician with more than 15 years' experience in rating nasality (K.L.) each independently rated the degree of nasality in 10 samples selected at random from the 39 experimental samples. Interrater agreement was computed using percent agreement and the chance corrected agreement statistic Cohen's kappa (Cohen, 1960). The judge and the clinician achieved exact scale point agreement on 8 of the 10 samples (80%), and the remaining two paired ratings differed by one scale point. This level of agreement produced a kappa of .75, which represents substantial strength of the chance corrected agreement (Landis and Koch, 1977).

Levels of intrarater agreement on the EAI scale were calculated from eight repeated ratings obtained at the time the judge rated all 39 speech samples for data collection. Intrarater agreement for the judges' first and second ratings of the eight samples was 88%. Ratings were identical for seven of the eight samples and differed by one scale point for the eighth sample. This level of intrarater agreement produced a chance corrected kappa of .83.

The judge and the clinician also established interrater agreement for the DME scale by independently rating the 10 samples using DME procedures. Ratings were made in relation to the moderately hypernasal reference sample, which was assigned a modulus value of 100. For example, if a sample was judged to be twice as nasal as the reference sample, a rating of 200 would be assigned. The range of normal resonance was defined as ratings between 1 and 10. Ratings above 10 represented degrees of hypernasality. A strong association was found between the paired ratings of the judge and of the clinician ($r = .89, p < .01$). As with the EAI agreement procedure, the level of intrarater agreement for DME was calculated from repeated ratings obtained at data collection. A strong association was found between the judges' first and second DME ratings of eight samples ($r = .99, p < .01$).

Nasality Ratings of Experimental Samples. The judge first listened to the audio disk prepared for EAI scaling and rated the 39 speech samples for degree of nasality. Two days later, the judge listened to the audio disk prepared for DME ratings and again rated the nasality in the 39 speech samples.

To obtain EAI nasality ratings, the judge rated each speech sample using the five-point EAI scale. Scale values ranged from 1, representing normal resonance, to 5, representing severe hypernasality. For DME ratings, the judge rated each speech sample in relation to the reference speech sample, which was assigned a modulus value of 100. The reference was repeated after every five speech samples, and the judge could, at any time, request that the reference be played.

Data Analyses

Separate bivariate correlations (Pearson's r) were computed to assess the strength of the relationship between nasalance scores and nasality ratings made using EAI scaling, as well as the strength of the relationship between nasalance scores and nasality ratings made using DME scaling. In keeping with the procedures of Barker and Rose (1979), Nasometer test sensitivity and specificity were calculated for nasalance scores and EAI nasality ratings, and for nasalance scores and DME nasality ratings. Polynomic fitting procedures with linear, quadratic, and cubic functions were applied to determine whether the association between EAI nasality ratings and DME nasality ratings deviated significantly from linearity (Engen, 1971).

RESULTS

Descriptive Statistics

The mean of the EAI nasality ratings across all 39 samples was 2.15 (standard deviation [SD] = 1.22). The judges' ratings ranged from 1 to 5, with scale values 2 through 5 representing some degree of hypernasality. The mean of the DME nasality ratings was 105.13 (SD = 76.73). With the use of DME scaling, assigned values ranged from 10 to 225, with ratings above 10 representing degrees of hypernasality. Nasalance scores ranged from 2% to 64%, with a mean of 28.21% (SD = 19.64).

Nasalance and Nasality Association

A moderate correlation ($r = .63, p < .01$) was found between nasalance scores and EAI nasality ratings. This represents a substantial relationship between the two variables (Guilford, 1956). A moderate correlation ($r = .59, p < .01$), also representing a substantial relationship, was found between nasalance scores and DME nasality ratings. A difference test (McNemar, 1955) showed no significant difference ($t\ (36) = -.50, p > .05$) between the coefficients .63 (nasalance and EAI ratings) and .59 (nasalance and DME ratings).

Sensitivity and Specificity

Nasometer test sensitivity and test specificity were determined for each rating scale (Barker and Rose, 1979). To calculate sensitivity and specificity, it was necessary to determine what nasalance score should be used to provide the division between normal nasalance and abnormal nasalance. From Watterson et al. (1996), a nasalance score of 22% was initially chosen as the division between normal and excessive nasalance. Therefore, nasalance scores above

TABLE 1 Nasometer Test Sensitivity and Specificity Measures: EAI Nasality Ratings

	Judge's Rating of Nasality With EAI		
Nasalance	*Hypernasal EAI ≥ 2*	*Normal EAI = 1*	*Total*
>22% (hypernasal)	17	4	21
≤22% (normal)	7	11	18
Total	24	15	39
	Sensitivity = .71	Specificity = .73	Overall = .72

TABLE 2 Nasometer Test Sensitivity and Specificity Measures: DME Nasality Ratings

	Judge's Rating of Nasality With DME		
Nasalance	*Hypernasal DME ≥ 25*	*Normal DME = 10*	*Total*
>22% (hypernasal)	18	3	21
≤22% (normal)	11	7	18
Total	29	10	39
	Sensitivity = .62	Specificity = .70	Overall = .64

22% were classified as abnormal, and scores equal to or below 22% were classified as normal.

EAI Nasality Ratings

Speech samples with nasality ratings of 1 were categorized as normal, and samples with a rating of 2 or greater were categorized as hypernasal. As is shown in Table 1, the judge rated 24 of the 39 speech samples as hypernasal. The Nasometer agreed with the judge on 17 occasions, resulting in a sensitivity measure of .71 (17 divided by 24). For the 15 samples judged as normal, the Nasometer agreed with the judge on 11 occasions, for a specificity rating of .73 (11 divided by 15). Overall efficiency is the sum of the number of times the Nasometer agreed with the judge divided by the total number of opportunities. Overall efficiency was .72 (28 divided by 39). Sensitivity and specificity were then recalculated to determine whether different nasalance cutoff scores would yield better results. However, the best sensitivity and specificity measures occurred with a nasalance cutoff score of 22%.

DME Nasality Ratings

By design, speech samples with ratings between 1 and 10 were categorized as normal. The next DME rating in the distribution was 25. The judge did not assign any rating between 10 and 25, so ratings ≥25 were considered hypernasal. Sensitivity and specificity were calculated by using a nasalance cutoff score of 22%. As is shown in Table 2, the judge rated 29 of the 39 speech samples as hypernasal. When a cutoff score of 22% was used, the Nasometer agreed with the judge on 18 occasions, resulting in a sensitivity measure of .62. For the 10 samples judged to be normal, the Nasometer agreed with the judge on seven occasions, for a specificity rating of .70. Overall efficiency was 0.64. As with EAI-scaled nasality ratings, the best sensitivity and specificity measures occurred with a nasalance cutoff score of 22%.

Test of Linearity

Polynomial fitting procedures with linear, quadratic, and cubic functions were applied. The regression of EAI nasality ratings on DME nasality ratings resulted in significant trends for both linear and curvilinear (quadratic and cubic) models. However, the linear model produced the best-fit line (Fig. 1), accounting for the greatest proportion of explained variance: linear model $F(1,37) = 74.23$, $p < .001$, $R^2 = .66$; quadratic model $F(2,36) = 43.75$, $p < .001$, $R^2 = .71$; and cubic model $F(3,35) = 28.41$, $p < .001$, $R^2 = .70$. Differences in linear, quadratic, and cubic functions were nonsignificant based on t test analyses.

DISCUSSION

It has been suggested that EAI scaling is not a valid method for obtaining ratings of nasality because nasality is a prothetic sensation that is best judged with DME scaling (Zraick and Liss, 2000; Whitehill et al., 2002). In the present study, however, EAI ratings were comparable with DME ratings with regard to the relationship between nasalance and nasality. Further, DME ratings did not improve Nasometer test sensitivity and test specificity. Thus, for the purposes of comparing nasalance scores versus judgments of nasality, it appears that ratings from the two methods did not demonstrate statistically significant differences for Nasometer test sensitivity and specificity.

Stevens (1975) has suggested that all stimuli may be categorized as prothetic or metathetic. The present data were regressed as suggested by Stevens, and significant trends were found for two curvilinear models; however, those models did not represent a significant improvement in variance explained over the linear model. If this finding is interpreted relative to Stevens' theory, it would mean that nasality is a metathetic sensation as opposed to prothetic. Said differently, nasality would be a qualitative sensation as opposed to quantitative. This interpretation would contrast with that of two previous studies that have identified nasality as a prothetic sensation (Zraick and Liss, 2000; Whitehill et al., 2002). However, differences in design between our study and previous studies could account for such differences in results and interpretation. In the study by Zraick and Liss, stimuli were derived from a single, isolated, synthesized vowel. Vowel formant frequency and bandwidth were manipulated to simulate nasality. But it is unclear how these synthetic manipulations of a single vowel relate to hypernasality in connected human speech; therefore, it is not possible to make a meaningful comparison between our data and their data. The study by Whitehill et al. differed from the present study in that

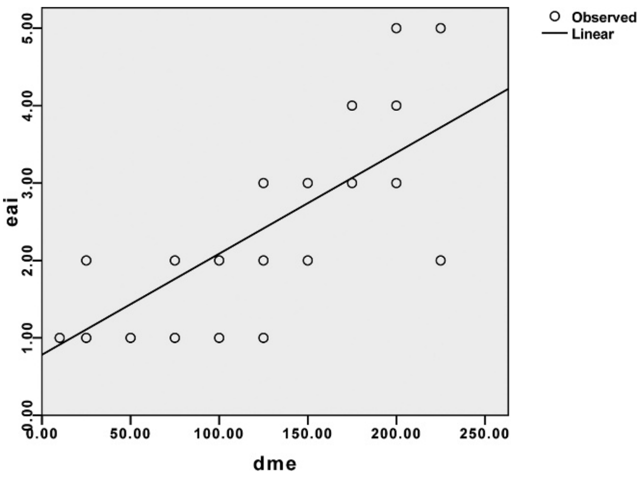

FIGURE 1 EAI ratings regressed on DME ratings with linear best-fit line.

Whitehill used 20 naïve listeners to obtain nasality ratings, but the present study used only one experienced listener. This difference in method could have affected the data obtained in two ways. First, experienced listeners may rate nasality differently than inexperienced listeners (Fletcher, 1976; Dalston et al., 1991; Lewis et al., 2003). For example, Lewis et al. found that experienced listeners were more reliable than inexperienced ones, and they rate nasality significantly lower. Second, data points in the Whitehill et al. study were means averaged over 20 listeners. For EAI ratings, arithmetic means were used to represent each speech sample, and for the DME scale, investigators used geometric means. Mean ratings are used to protect against random error by judges, especially inexperienced judges, but means may change the fundamental relationship between the two sets of ratings in two ways. First, averages cause gravitation of ratings toward the center of the scale, and this effect would be more profound for the EAI ratings because the EAI scale was limited to five scale points. Second, mean ratings cause the high end of the EAI scale to become underrepresented relative to the low end because listeners are more likely to agree that a sample is normal than to agree that it is severely hypernasal. This gives the statistical impression in the regression that the EAI scale is not linear relative to the DME scale. Consider, for example, a speech sample that is severely hypernasal. For that speech sample to receive an average rating of "5," all listeners must rate it a "5." If half of the listeners rate it "4" and half rate it "5," then the stimulus becomes represented in the data analysis by the rating of "4.5." Thus, with 20 listeners, it is very unlikely that any stimulus would ever receive an average rating of "5." With a single listener, this is not an issue. Further, an average rating frequently produces a representative rating that is not on the scale and was never used by any listener. In reality, a rating of "4.5" does not and could not exist because listeners did not have the option of assigning that rating. Thus, the overall effect of group judges and mean ratings is to cause the relationship between the two scales to become more prothetic than it really is. Even Stevens (1975) was uncertain about the use of arithmetic means to represent stimulus intensity on an EAI scale. He suggested that they should "practically never" be used (p. 270).

In previous studies involving judgments of other speech qualities, it has not been the case that every speech dimension studied turned out to be clearly prothetic or metathetic, as would be predicted by Stevens (1975). For some speech dimensions, the data obtained showed that listener perceptions fell into a gray area that was somewhere between the two extremes (Toner et al., 1990; Schiavetti et al., 1994; Eadie and Doyle, 2002). This would be consistent with Eisler's theory (1963) that sensations are not binary but rather lie on a prothetic continuum, where metathetic is the limiting case. With this continuum, Eisler predicts that when sensations are not strongly prothetic, they will show a curvilinear relationship with category scaling, but the curvilinear model will not be a significant improvement over the linear model. This was exactly the case with the present data. Given that theoretical possibility, perhaps nasality may be a psychological dimension that has only a small degree of prothetic-ness. If this is the situation, nasality could be more prothetic on one occasion than on another, depending on study design and data interpretation.

From a practical perspective, the most important issue in this quandary is what procedure should be used in a typical clinical setting to obtain nasality ratings? Historically, clinicians have used EAI scales because they are better suited to a clinical situation. EAI scales are intuitive, and the ratings obtained are relatively easy to compare across different scales and judges. DME scales, on the other hand, are cumbersome in a clinical setting because the speech sample in question must be compared with a standard, the task is foreign to most clinicians, and mathematical equalization procedures may be necessary to compare ratings across different judges. The present study simulated a clinical situation in that a single, experienced judge rated the speech of patients. With this procedure, the ratings were reliable and valid and bore a statistical relationship to nasalance scores. It would be our expectation, therefore, that experienced listeners may continue to confidently use EAI scales to rate nasality for clinical purposes. Future studies should continue to explore the validity and reliability of clinical procedures used to assess resonance by perceptual and acoustic measures.

REFERENCES

Barker DJP, Rose G. *Epidemiology in Medical Practice*. London: Churchill Livingstone; 1979:34–35.

Bzoch K. Measurement and assessment of categorical aspects of cleft palate speech. In: Bzoch K, ed. *Communication Disorders Related to Cleft Lip and Palate*. Boston: Little, Brown; 1979.

Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960:37–46.

Dalston RM, Warren DW, Dalston E. The use of nasometry as a diagnostic tool for identifying patients with velopharyngeal impairment. *Cleft Palate Craniofac J*. 1991;28:184–189.

Eadie TL, Doyle PC. Direct magnitude estimation and interval scaling of naturalness and severity in tracheoesophageal (TE) speakers. *J Speech Lang Hear Res*. 2002;45:1088–1096.

Eadie TL, Doyle PC. Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers. *J Acous Soc Am*. 2002;112:3014–3021.

Eisler H. How prothetic is the continuum of smell? *Scand J Psych*. 1963;4:29–32.

Engen T. Psychophysics II. Scaling methods. In: Kling JW, Riggs L, eds. *Woodworth and Schlossberg's Experimental Psychology*. New York: Holt, Rinehart, & Winston; 1971.

Fletcher SG. Nasalance versus listener judgments of nasality. *Cleft Palate J*. 1976;13:31–44.

Guilford JP. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill; 1956.

Hardin MA, Van Demark DR, Morris HL, Payne MM. Correspondence between nasalance scores and listener judgments of hypernasality and hyponasality. *Cleft Palate Craniofac J*. 1992;29:346–351.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.

Lewis KE, Watterson TL, Houghton SM. The influence of listener experience and academic training on ratings of nasality. *J Commun Disord*. 2003;36:49–58.

McNemar Q. *Psychological Statistics*. New York: John Wiley & Sons; 1955.

Nellis JL, Nieman GS, Lehman JA. Comparison of Nasometer and listener judgments of nasality in the assessment of velopharyngeal function after pharyngeal flap surgery. *Cleft Palate Craniofac J*. 1992;29:157–163.

Schiavetti N, Martin RR, Haroldson SK, Metz DE. Psychophysical analysis of audiovisual judgments of speech naturalness of nonstutterers and stutterers. *J Speech Hear Res*. 1994;37:46–52.

Stevens SS. *Psychophysics*. New York: John Wiley & Sons; 1975.

Toner MA, Emanuel FW, Parker D. Relationship of spectral noise levels to psychophysical scaling of vowel roughness. *J Speech Hear Res*. 1990;33:238–244.

Warren RM. Anomalous loudness function for speech. *J Acoust Soc Am*. 1973;54:390–396.

Warren RM, Sersen EA. A basis for loudness-judgments. *Am J Psych*. 1958;71:700–709.

Warren RM, Warren RP. A critique of S.S. Stevens' "New Psychophysics." *Percept Motor Skills*. 1963;16:797–810.

Watterson T, Hinton J, McFarlane SC. Novel stimuli for obtaining nasalance measures from young children. *Cleft Palate Craniofac J*. 1996;33:67–73.

Watterson T, Lewis KE, Deutsch C. Nasalance and nasality in low pressure and high pressure speech. *Cleft Palate Craniofac J*. 1998;35:293–298.

Watterson T, McFarlane SC, Wright DS. The relationship between nasalance and nasality in children with cleft palate. *J Commun Disord*. 1993;26:13–28.

Whitehill TL, Lee ASY, Chun JC. Direct magnitude estimation and interval scaling of hypernasality. *J Speech Hear Res*. 2002;45:80–88.

Zraick RI, Liss JM. A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality. *J Speech Hear Res*. 2000;43:979–988.