"Nasalance" vs. Listner Judgements of Nasality

SAMUEL G. FLETCHER, Ph.D.

Birmingham, Alabama 35294

Nasality is a major perceptual attribute of speech (16). It serves as a phonetic sign of the nasal consonants (13), as a vocal quality for speaker identification (12), and as a primary or secondary symptom of many disorders and disabilities affecting speech transmission (3, 5, 6, 7, 17). It is a criterion feature of "cleft palate speech" (14).

Scaling the magnitude of nasality has been found to be a difficult task which requires a sizeable group of listeners to achieve reliable results (2, 4). A persistent problem is that the scores derived in one laboratory are of comparatively little value to another since they don't have a common numerical relationship.

Many instrumental approaches have been used to measure facets of speech production that appear related to perceived nasality! The goals in this work were to reduce the complexities of the measurement procedures and at the same time derive scores that have more universal utility. A new instrument, TONAR II, developed during the past two decades (δ) is particularly promising in this regard. TONAR II is used to derive a numerical acoustic ratio score, expressed in percent "nasalance," which reflects the relative proportion of sound within a specified frequency band emitted from the mouth and nose during speech (ϑ). The present study was designed to evaluate the congruence between listener judgements of nasal voice quality, herein exclusively termed "nasality," and these instrumental measurements of nasalance.

Procedure

SUBJECTS. The subjects for this study were 14 boys and 9 girls from a group of 70 children who had undergone only primary palatal surgery for congenital clefts of the palate. Names for the larger group were drawn systematically from files of the Alabama Crippled Children's Service and private surgeons so they would be representative of the general speech competencies of such individuals between the ages of 5 and 15 years. The children were brought to the Biocommunication Laboratory at the University of Alabama in Birmingham

Dr. Fletcher is affiliated with the Department of Biocommunication, University of Alabama in Birmingham, Birmingham, Alabama.

This study was supported in part by USPHS research grant DE3580, National Institute of Dental Research and social and rehabilitation service grant No. 16 P5680 7/4-09, Rehabilitation Research and Training Center.

(UAB) for a series of speech and oral-pharyngeal evaluations. The subsample of subjects for the present study were drawn randomly from the larger pool.

The data base consisted of sound field and sound separated tape recordings from the 23 speakers who ranged in age from 5.2 to 15.0 years with a mean age of 9.7.

METHODS. The audiotape recordings were of two parallel sets of utterances: sound field recordings of the subjects speaking Fletcher's Zoo Passage (9) and TONAR II sound-separated recordings of the speakers saying the same passage. Forward and backward reproduced versions of the sound field recordings of the Zoo Passage utterances were duplicated from the master reel-to-reel tape to cassette tapes for scaling purposes.

An initial sample of 24 sound field recordings was selected randomly from the pool of 70 in the general study. These recorded materials were transferred to the cassette tapes, randomized, and replayed in individual listening sessions to 10 persons for scoring the perceived nasality. The scoring procedure is described below. Judgemental variation in the degree of nasality perceived in five of the recordings led to their rejection. In order to maintain an adequate speaker sample, four of these recordings were replaced through random order selections from the original pool. The 23 reproductions were then presented to seven male and thirteen female university students who served as paid volunteers to judge nasality in the forward reproduced recordings. Three male and five female members of the same group plus one additional male and one female served as paid volunteers to judge the backward reproduced recordings. Most of the listeners selected were currently enrolled in introductory speech classes. The hearing of all judges was screened and found to be within normal audiometric limits. None of them had had courses or training related to speech pathology. This stipulation was introduced to exclude systematic listener bias from prior training. The listeners were advised that the listening task would require approximately four hours to be completed.

In the nasality judgemental sessions, a different random order was used for each judge and each set of recordings.

NASALITY RATING. The following four steps were used in the nasality rating task progression:

Task I.

To gain insight into the "man-on-the-street" impression of nasality, the first task consisted of having the listeners sort the 23 recordings as "normal" or "abnormal" with respect to the amount of nasality perceived in the speakers' voices. No information was given concerning the age, sex, or medical background of the speakers being judged.

Task II.

The second task required the listeners to rank all 23 recordings by degree of perceived nasality. A paired comparison paradigm was followed. This consisted of selecting two specimens from a random ordered pool, playing them simultaneously or alternately, deciding which had *less* nasality, and placing it in a cassette storage tray with least nasality toward one end and most toward the other. Another recording was then selected and the

procedure repeated. The one selected this time as less nasal was to be placed in the tray in proper relation to the first. If in doubt those two were to be compared with each other. This procedure was followed until all 23 recordings were placed in rank order according to the perceived degree of nasality. All 23 recordings were then replayed in a step-wise paired comparison progression from the one ranked as least nasal to verify the nasality continuum and make final adjustments in the ranks assigned.

Although theoretically this task could have required n(n-1)/2 or 253 paired comparisons to arrive at an initial ranking, it was found that the listeners could shorten the task by placing the recordings within perceived ranges of responses during the initial sort. Errors were then rectified as the listeners replayed the total series and made adjustments within the final rank. The averaged times for a listener to rank the forward and backward played recordings were 62 and 60 minutes respectively. Task III.

In the third task, the listeners replayed the recordings beginning with the one with least nasality and classified each utterance as "normal," "mildly nasal," "moderately nasal," "severely nasal," or "very severely nasal." Specific definitions were given for each classification. For example, mildly nasal was defined as: "Nasality is apparent as the person speaks but is only mild in degree. It causes little distraction to you as a listener." Task IV.

Finally, the listeners replayed the recordings and assigned a numerical score of 5 to 69 to each of the utterances. A score of 5 represented the least possible nasality. A specific range of scores was given for each of the five classifications, i.e., normal nasality was to be assigned a score within the range of 5 to 14 according to the absolute amount of nasality judged to be present. No two recordings were to be given identical scores. The judges were allowed to relocate any recording now perceived to be misplaced in its rank and class before the final score was allotted.

Written instructions were provided and reviewed orally at the beginning of each task. Score sheets were then given to the listener to record his/her observations from each sorting or scoring task. The same four nasality rating tasks with similar instructions were used in forward and backward rating procedures. A minimum of two weeks time lapse was established between each of the sessions for the judges who rated both forward and backward reproduced recordings. The order of the recordings was rerandomized prior to each listening session in each mode of presentation.

The nasality rating sessions were conducted in a large therapy room with a carpeted floor, drapes on two walls, and acoustic tiled ceiling. The ambient noise level was approximately 40 dBA.

The participants were seated in individual sessions at a table with two Sanyo model RD4300 cassette tape recorders in front of them. The recorders were coupled through a stereo amplifier to separate speakers placed at each end of the table. The gain controls of the amplifier had been adjusted to produce equivalent outputs of 70 dBA from each recorder. This arrangement allowed paired

comparison judgements between two different recordings played either alternately or simultaneously. Different sets of cassette trays with appropriate labels were provided for each task to facilitate the sorting and classifying procedures.

In the original data collection, the sound separated recordings of the nasal and oral components of speech were obtained from each speaker immediately after the sound field recordings were made. The procedure for obtaining such recordings has been described previously (9, 11).

The sound-separated recordings were replayed through TONAR II, and two overlaid tracings of the outputs were charted on a tonagram for each of the speaker responses. One was produced with a .1 second time constant (T.C.) which reflected moment-by-moment fluctuations in nasalance. The other display was made using a 10-second T.C. filter. This produced a time averaged curve. In this mode, the ratio stays at the same level during pauses in vocalization, because both filtered signals maintain the same relationship to each other. The midpoint between the extremes of the curve toward the end of the 10-second averaged display was chosen to provide a single score estimate of nasalance after the effects of the filtering were attained. This nasalance score was used in all later comparisons with the ratings and scores from listener judgements of nasality.

Results

The mean nasalance measured from the sound separated recordings of the 23 speakers ranged from four to 60 per cent with a mean of 27.5 per cent. Three different estimates of the magnitude of nasality and one additional classification of nasality were obtained from the listener judgements. These results will be presented and contrasted by listener task.

NORMAL OR ABNORMAL NASALITY VS. NASALANCE SCORES. In the first task, the listeners classified the recordings as either normal or abnormal in nasal quality. As might be expected, the use of naive listeners produced considerable variation in the judgements of nasality. For example, in the perceived nasality judgements from the forward reproduced recordings (FRN), 20 of the 23 recordings were classified as "normal" in nasal quality by at least one of the listeners while 18 of the same responses were classified as "abnormal" by at least one other listener. In response to the backward reproduced recordings (BRN) only 12 speakers were classified as normal in nasality by any listener. Furthermore, each of the 23 responses were judged to be abnormally nasal by at least one listener.

Despite the variability, the responses of some of the speakers tended to be rated rather consistently with respect to perceived nasality. A 75 per cent listener agreement level was chosen to identify subjects with sufficient consistency in this initial task to make meaningful comparisons between perceived nasality and measured nasalance. 14 FRN and 11 BRN recordings met this criterion. Of the 14 FRN recordings, 7 were judged to be normal in nasal quality and 7 abnormal. The mean nasalance score of these FRN scored speakers was 12.9 per cent for those rated normal and 49.7 per cent for those rated abnormal in nasality. The mean nasalance for those not receiving consistent ratings was 21.8 per cent. Thus, listener agreement was apparently highest at each end of the nasal

35

resonance continuum and mixed in the midrange. All 11 of the BRN recordings with consistent listener judgements were classified as abnormally nasal. The mean nasalance from the sound separated recordings of these speakers was 40 per cent while that of the inconsistently judged responses was 16.1 per cent. A clear trend of agreement between the nasality judgement and nasalance measurement sets of data was evident in these data.

NASALITY VS. NASALANCE RANK. The second task required paired comparisons of perceived nasality from the backward reproduced and forward reproduced recordings.

Considerable variation was found in the judgemental rankings of nasality. This variability was documented by determining the median rank assigned by the different listeners and the semi-interquartile range of that rank for each of the 23 recordings in each mode of presentation. The among-listener semi-interquartile range for the FRN rankings of the individual recordings was 0.25 to 4.0 with a median of 1.5 ranks. The semi-interquartile range for BRN rankings was 0.75 to 6.25 with a median of 4.0 ranks. Thus, the judges were much more variable in ranking the BRN stimuli than in ranking the FRN stimuli.

Spearman Rho correlations (R) were used to reflect the degree of relatedness between the nasality and nasalance sets of data.

The R correlations between individual judge FRN rankings and the nasalance ranks ranged from .61 to .90 with a median of .81. Correlations between the individual BRN rankings and the nasalance ranks ranged from .37 to .82 with a median of .60.

To examine the general relationship between the nasality and nasalance data, the rankings from the individual judges were pooled and a median rank was derived for each speaker in each mode of stimulus presentation. The Rho correlation was then calculated between the sets of forward and backward nasality ranks and the corresponding nasalance rank. The R derived between the median nasality ranks from the forward reproduced recordings and the nasalance ranks was 88. That between the median ranks of the backward reproduced recordings and the nasalance ranks was .77. Thus, when the variability of individual judges was partially removed through use of grouped data, agreement between nasality and nasalance rankings increased sizeably. This increase was especially evident for the backward reproduced utterances.

NASALITY CLASS VS. NASALANCE SCORE. In the third listening task the judges were to relisten to the recordings arranged in a perceived nasality continuum in Task II and categorize the utterances within five classes: normal, mildly nasal, moderately nasal, severely nasal and very severely nasal.

For numerical treatment of the findings, the categories were later assigned values ranging from 1 to 5, with "normal" given a value of 1. Median scores for each of the 23 utterances were then derived from the pooled judgements of the 20 listeners who categorized the FRN materials and the 10 judges who categorized the BRN materials. A summary of the results is shown in Table 1.

It may be seen in Table 1 that the median FRN judgements were rather evenly distributed across all five categories of perceived nasality. Except for a minor reversal in the severely nasal and the very severely nasal categories the nasalance

Median Nasality Classification	No. Ss in FRN Class	Median Nasalance of Ss	No. Ss in BRN Class	Median Nasalance Ss
Normal	5	8	1	19
Mildly				
Nasal	5	18	6	9
Moderately				
Nasal	5	24	9	25
Severely				
Nasal	3	52	4	51
Very Severely				
Nasal	5	49	3	43

TABLE 1. Utterances of 23 speakers categorized by degree of nasality from forward reproduced (FRN) and backward reproduced (BRN) recordings. The corresponding median levels of nasalance are also given for the speakers in each category of nasality.

scores of the subjects in each category increased progressively with severity of nasality. Considerable overlap existed in the nasalance scores of subjects assigned to the upper two categories.

The BRN judgements showed the recordings to be clustered toward the middle categories of nasality. Only one subject was classified as having normal nasality in the BRN judgements and only three were classified as very severely nasal. The median nasalance scores of the subjects in each class showed a general increase with severity of nasality although the single subject classified as normal and the median of the three classified as very severely nasal were exceptions to this trend.

NASALITY MAGNITUDE ESTIMATE VS. NASALANCE SCORE. The final task was to estimate the magnitude of the nasality perceived in the 23 utterances of the Zoo Passage. The instructions were designed to utilize both the listener's categorization of nasality and his/her estimate of the relative magnitude of nasality within the stipulated category. The judges were permitted to relocate a recording if the degree of nasality now perceived indicated an earlier classification was in error.

The average scores from the 20 FRN judgements of the 23 speakers ranged from 9.6 to 65.2 with a mean of 34.3. Average scores from the 10 BRN judges ranged from 14.1 to 62.2 with a mean of 35.4.

Despite the care taken to structure the listening tasks, considerable variation still existed in the ratings given the same utterance. The scores assigned by different judges typically spanned two categories of nasality. In no instance was the range of scores assigned any given recorded utterance of the Zoo Passage by the different judges less than 12 points for the FRN recordings nor less than 15 points for the BRN recordings. Thus, although fewer judges were used to evaluate the BRN recordings, the range within the ratings assigned was greater for the BRN than for the FRN utterances.

To examine the mathematical relationship between the nasality ratings and the nasalance measurements, correlation coefficients (r) were computed. The resultant correlations between FRN ratings by the individual judges and nasalance scores, shown in Table 2, ranged from .74 to .92 with a mean of .85. Correlations between individual BRN ratings and nasalance scores ranged from .45 to .93 with a mean of .66.

The scores from the 20 listeners who rated forward reproduced recordings and from the 10 listeners who rated the backward reproduced recordings were next pooled within the two groups. Pooled FRN and BRN mean scores were then derived for each of the 23 speakers. Correlation coefficients were computed between these mean nasality scores and the nasalance scores. The resultant r's were .91 between the mean FRN group ratings and nasalance scores and .81 between the mean BRN group ratings and nasalance scores. Since different recordings of the speakers were used to elicit listener judgements and to measure nasalance, the level of agreement is felt to be excellent. Furthermore, in each instance the correlations were considerably higher between grouped nasality ratings and nasalance scores than for the ratings of the individual judges. Thus, partial removal of the individuality of the ratings enhanced agreement with the instrumental measurement.

Figure 1 is a scatter diagram displaying the relationship between FRN mean nasality scores and corresponding nasalance scores. To facilitate the comparison of the two sets of data the nasality scores were transformed to the same range as that of the nasalance scores. The linear relationship between the two sets of data is portrayed by the slope of the calculated regression line extending diagonally through the graph. Rather close agreement between the data arrays is apparent although the number of points in the midrange is rather sparse. The rightward shift in origin of the line at the low end of the scale likely reflects the lower limit of "4" imposed on the transformed listener scores. Since the listeners varied in the actual values assigned any given response, their mean score could be anticipated to be somewhat higher than the minimum possible.

Finally, it appeared evident that agreement between the listener judgemental scores and instrumental measurements increased as the listeners moved through the task sequence. To examine this apparent trend more closely Fisher's z

	CORRELATION COEFFICIENTS			
	Individual Ratings $(N = 23)$		Group Ratings	
	Range	Mean	(N = 23)	
Backward Reproduced Nasality Judgements (J = 10) vs. Nasalance Scores	45 93	66	91	
Forward Reproduced Nasality Judgements $(J = 20)$ vs.	.4575	.00	.01	
Nasalance Scores	.74–.92	.85	.91	

TABLE 2. Correlations between nasality ratings and nasalance measurements of utterances by 23 speakers. Nasality was judged from both forward and backward reproduced recordings.



FIGURE 1. Scatter diagram displaying the relationship between the nasalance scores and the final nasality judgements from the forward reproduced recordings. Prior to plotting the data, the nasality scores were transformed to the same range as the nasalance scores. The linear relationship between the two sets of data is portrayed by the calculated, dashed regression line extending diagonally through the graph.

transformation was carried out on the Task II and Task IV correlations. This normalization of the distributions of the correlations permitted a direct comparison of the relative agreement between listener judgements and instrumental measurements as a function of the task conditions.

The mean z score from the transformed correlations between FRN Task II judgements and TONAR II measurements was 1.13 with a s.d. of .21. The mean z score from the transformed correlations between Task IV FPN judgements and the instrumental scores was 1.29 with a s.d. of .17. A paired "t" test was carried out to determine the significance of this change. The resultant "t" of -5.57 with 19 d.f. was highly significant (p < .01). This indicates that as the listeners gained experience in judging nasality and as they were given additional leeway in the precision with which they could score the degree of the nasality perceived, their scores both individually and as a group moved toward increasingly close agreement with the instrumental measurements of nasalance.

A similar transformation and statistical comparison was carried out using the BRN correlations. In this instance the mean z score from the Task II correlations was .74 with a s.d. of .26. From the Task IV correlations it was .86 with a s.d. of .34. The paired "t" score was -1.96 with 9 d.f. which was also significant (p < .05). Thus, although the group of listeners was smaller and their judgements were more variable, the agreement between the listener judgements and instrumental measurements again increased significantly with experience and scoring leeway.

Discussion

A principal assumption examined in this paper is that the vocal resonance characteristics measured physically by TONAR II and termed "nasalance" correspond rather closely with those identified as "nasality" by a human listener. If so, the nasalance scores which can be precisely and reliably determined through instrumental procedures may be used as a valid alternative means of assessing the presence and degree of nasal resonance.

Since nasality is the criterion measure, proof of the foregoing assumption rests largely on the accuracy of the nasality observations. One might argue that the most accurate observations of nasality could be obtained by using speech pathologists who are by training prepared to detect speech disturbances. Such individuals could then be specifically trained toward agreement in judging nasality. The fallacy in this logic would seem to be that our major question is not whether or not a person can be trained to agree or disagree with any given instrumental measurement or even whether or not the instrument is measuring that which the speech pathologist has been trained to hear as nasality. Rather, the crucial question is how well do the instrumental measurements of the vocal properties correspond to the general phenomenon of perceived nasality? The answer to this question would presumably be most completely answered if the judges were biased only by their natural experiences with the language.

Another important assumption of this study is that if an instrument measures the physical factors which underlie perceived nasality correctly, agreement between the instrumental measurements and listener perceptual experiences should increase as the accuracy of the listener's judgements of the nasality improve. This, of course, embodies an underlying assumption that a given set of untrained listeners can, in fact, detect nasality and judge its severity.

The initial task assigned to the listeners in this study consisted of having them label recorded speech samples as "normal" or "abnormal" with respect to the degree of nasality they perceived in the speakers' voices. Considerable variation was found in the results. This variability is illustrated by the fact that a majority of the recorded utterances were labelled as both normal and abnormal by different listeners. Such scatter does not necessarily mean that the listeners were unable to detect nasality during this early task. It might be interpreted as evidence of variability in the listeners' tolerance of nasality. The perception of speech as acceptably or excessively nasal in its resonance characteristics likely depends to a large extent upon the experiential framework and values of the individual perceiver.

Interlistener agreement at the two extremes of the nasality range suggested that the normal-abnormal dichotomy actually represented a continuum of perceived nasality. Low agreement among different listeners was confined largely to the midrange. This characteristic enabled us to identify a subgroup of speakers with conspicuously normal nasality and a subgroup with conspicuously abnormal nasality. The recordings of these speakers were judged to be either normally or abnormally nasal by at least 75 per cent of the listeners. Identification of these subgroups presented our first opportunity to compare listener judgements of nasality with instrumental measurements of nasalance.

Three comparisons were made between nasalance and initial FRN judgements of nasality. The subgroup consistently judged as normal in nasality had the lowest nasalance scores, and those consistently judged as abnormally nasal had the highest scores. The ones in the range between these extremes were those the listeners were unable to classify consistently as either normally or abnormally nasal. The nasalance scores of this group were in the midrange. Such nasalance scores would thus be especially advantageous in clinical or experimental situations where an examiner must assess rather precisely the degree of nasal resonance in a broad sampling of speakers.

Although the BRN judgements were in agreement only at the abnormal end of the nasality continuum, the nasalance scores of those so rated were considerably higher than those in the non-determined range. Thus, gross congruence was found between both sets of nasality data and the nasalance scores. In each instance, the apparent continuum in judgements of nasality paralleled the direction of the nasalance scores.

The second task was designed to force the listeners to make a more precise evaluation of the degree of nasality in the different recordings. Through an arduous series of step-by-step comparisons and recomparisons, the listeners were required to arrange the recordings into an explicit continuum with respect to the degree of perceived nasality. That there was considerable variability from listener to listener in the results was not surprising. Even with "trained" listeners, earlier studies had shown great variability among different judgements of nasal voice quality (2, 4).

Although considerable variability was found among individual judgements made concerning the different spoken responses, a strong central core of agreement was also present. This was evident even in the subgroups at the extremes of the first task. The trend became clear across all 23 speakers in the succeeding assignments. It should be emphasized that at no time were the listeners given feedback concerning the accuracy of their judgements. Nevertheless, judgemental agreement increased as the listeners completed the various tasks assigned. Apparently simply requiring them repeatedly to focus close attention upon the spoken responses through the series of tasks and also to arrange the recordings into certain classes and progressions of "nasality" was sufficient to expose an underlying competency to detect and quantify nasality. It was not a matter of teaching them to hear the nasality. That ability was clearly present prior to the study despite the fact that they were chosen as "naive" listeners. The fourth task was designed to produce as precise an assessment as possible of the degree of nasality. The listeners were required to assign a discrete score within a limited range of scores. Furthermore, they were not permitted to assign the same score to any two individuals. They were forced to make an exclusive judgement of the degree of nasality present in the vocal output of each speaker. The results from this task followed the trend described in the second task. The more closely the listeners agreed with each other and the more listener variability was removed before comparisons were made, the closer was the agreement found between the scores derived from the perceptual judgements of the listeners and the instrumental measurements of nasalance.

Attention will now be focused more specifically upon the effects of the variability upon the scores assigned. The results from each of the listener tasks demonstrated a consistent pattern. As the listeners gained experience in the listening tasks and as the tasks themselves demanded greater precision. variability decreased. Also as variability among the nasality judgements lessened, thus suggesting increased accuracy in the scores derived, the perceptual ratings moved more and more closely toward the nasalance scores obtained instrumentally. This was true for both modes of stimulus presentation despite the fact that the nasality and nasalance assessments were derived from two different sets of utterances by each speaker. For example, the individual judgements were most variable when they were made in response to the backward reproduced recordings, and the correlations between the nasalance scores and the BRN ratings by individual judges were lowest of all. Conversely, when the recordings were played in the normal forward direction, which allowed the listeners to make full use of their prior experience with the language, variability was reduced. Least variability was found when the recordings were presented in the normal forward direction, data for all listeners were pooled, and mean scores were calculated compared with those from instrumental measurements of nasalance. The highest correlations were also reached between these two sets of data. Hence, the correlations were highest when judgemental variability was least.

In light of the many factors that can influence listener judgements, the final correlation of .91 between the nasalance scores and the grouped data from FRN ratings appears to be excellent. This level of correlation would indicate that the instrumental measurement could account for over 80% of the variance in the listener judgements of nasality. From this, it seems reasonable to conclude that, for most purposes, the instrumental scores derived from TONAR II may be used as a valid alternative means of assessing clinically important nasal resonance. This conclusion is especially meaningful in situations where the cumbersome procedure of assembling a large group of listeners, eliciting their ratings of nasality, and compiling average judgemental scores is not specifically demanded.

Some additional background on TONAR measurements may be helpful to avoid confusion between current scores and procedures for measuring nasalance and results from earlier versions of TONAR.

TONAR II is the outgrowth of work extending over approximately the last decade. In an early prototype (8), a simple computational formula consisting of

nasal signal divided by oral signal was used. The frequency bandwidth of the two signals included in this computation was not specifically limited. In 1968, an experiment was conducted to examine the level of agreement between such instrumental measurements and listener ratings of nasality (11). Ten listeners judged tape-recorded speech samples of 20 children who had originally had clefts of the palate. The Rho correlation computed between the resultant sets of scores was .74. The recordings were also filtered in different ways in an effort to identify a frequency bandwidth which might increase agreement between listener judgements and ratio measurements of the output. Closest agreement seemed to be achieved when the filter band was centered in the region of 500 Hz. Further studies indicated that the agreement was optimized for most speakers with a 300 Hz bandwidth around the central frequency of 500 Hz. This characteristic was incorporated in TONAR II.

The computation formula of TONAR II was also changed to compare the nasal signal with the combined nasal plus oral signals. This permitted inclusion of the entire possible range of nasal vs. oral output in a single numerical system. That is, absence of nasal signal would be expressed by a nasalance ratio of "0." Absence of an oral signal would be expressed as a ratio of "1." Any variation between these limits would be expressed in an appropriate decimal fraction. The scores were then instrumentally multiplied by 100 to convert them into percentage values, thus eliminating decimals. The term "nasalance" was coined to remove possible confusion between the instrumental measurements and perceptual judgements of nasal resonance. The present study was designed to establish the current level of agreement between listener judgements of nasality and the nasalance scores. The results indicate that nasalance scores now agree rather closely with scores from listener ratings of nasality.

Finally, the discrepancies found in this study between listener judgements of nasality from forward reproduced and backward reproduced speech are worthy of special comment.

Sherman (15) first proposed the use of backward play reproduced recordings with the rationale that this procedure "should eliminate most of the irrelevant factors which might influence observers." She assumed specifically that playing the speech backward would remove distractive influences such as phonetic information and thereby enable a listener to concentrate more exclusively on vocal quality such as nasality. If this assumption were valid, we should expect to find greater consistency among judgemental responses of different listeners to this type of material. Normal speech should, of course, still be perceived to be normal. Our findings did not support those assumptions.

The listener judgements of backward played materials were consistently more variable than those for the same utterance heard in a forward direction. This trend was shown in each task. In Task I, interjudge agreement was very high in the classification of the utterances of seven of the speakers judged from forward reproduced recordings to be normal in nasality. When these same recordings were played backwards, thereby eliminating much of the listeners' past experience in dealing with spoken English, one of the seven utterances was consistently classified as abnormally nasal. The other six were still judged to be normal by over half of the listeners, but the variability among the judgements of the backward played stimuli was so great that none of them reached the minimum criterion of 75 per cent interjudge agreement.

In Task 2, the semi-interquartile range of FRN scores spanned only one and a half ranks. When the utterances were presented backwards, the semi-interquartile range spanned four ranks. In Task 3, the semi-interquartile range of the BRN scores was half again that of the FRN. In Task 4, the S.D.'s of the scores from the BRN ratings were almost twice those from the FRN ratings. Thus, although in all four tasks considerable variation was found among listener judgements, in each instance the variability was greater for material reproduced backwards.

Black (1) has noted that backward played speech retains some of the phonetic attributes of normal speech while others are grossly distorted. The speech is not only gibberish but can also be described as sounding bizarre. Seemingly, the listeners in the present study reacted to this bizarreness by judging the nasality more harshly. The increase in variability also suggests that the task of adapting to this new type of listening experience was not an easy one. It seems evident that, for judgements of nasality, these additional problems outweigh any possible theoretical advantages which might encourage the use of backward played stimuli.

Summary

In this article, judgements of nasality from sound field recorded utterances of 23 children with repaired palatal clefts are compared with "nasalance" values derived from TONAR II analyses of parallel sound separated recordings. For the nasality judgements, the sound field recordings were presented in forward reproduced mode to 20 "naive" listeners and in backward reproduced mode to 10 naive listeners. A four-level task progression was followed in each mode of presentation: sorting the responses by "normal" or "abnormal" nasality, ranking them by severity of nasality, classifying them within five degrees of nasality, and, finally, assigning a discrete score to the magnitude of nasality perceived in each recording.

Scores from the individual judges were highly variable, especially in the first listening task. They were consistently more variable when nasality was judged from backward reproduced recordings. As the listeners gained experience and as the listening tasks demanded more precision in nasality judgements, variability was reduced. Correlations computed between the nasality ratings and nasalance measurements increased and variability decreased. Highest agreement between physical and perceptual measurements was found when variability among the judges was reduced by pooling the scores and comparing the mean scores of nasality with the nasalance scores. Using this procedure, a correlation of .91 was obtained between listener judgemental scores of forward reproduced speech and nasalance scores. Under these conditions, the instrumental score could apparently account for over 80 per cent of the variability in the listener judgements of nasality.

From the findings of this study, it was concluded that listeners without prior

training are capable of judging nasality reliably and that nasalance scores provide a valid correlate of perceived nasality. The use of backward reproduced recordings was questioned as a means of assessing nasality.

Acknowledgements: Thanks are given to Mrs. Sharon Carrigan, UAB Department of Biocommunication, who served as research assistant during data collection, and to Dr. Pat Taylor and Mrs. Margot Cawthon, UAB School of Humanities, for their assistance in obtaining listeners.

References

- 1. BLACK, J. M., The "phonemic" content of backward-reproduced speech. J. Speech Hear. Res., 16, 165–174, 1973.
- 2. BRADFORD, J. L., A. R. BROOKS, and R. L. SHELTON, JR., Clinical judgments of hypernasality in cleft palate children, *Cleft Palate J.*, 1, 329-335, 1964.
- 3. COLTON, R. H. and H. S. COOKER, Perceived nasality in the speech of the deaf, J. Speech Hear. Res., 11, 553-559, 1968.
- 4. COUNIHAN, D. T., and W. L. CULLINAN, Reliability and dispersion of nasality ratings, *Cleft* Palate J., 7, 261-270, 1970.
- 5. DARLEY, F. L., A. E. AARONSON, and J. R. BROWN, Clusters of deviant speech dimensions in the dysarthrias, J. Speech Hear. Res., 12, 462-495, 1969.
- 6. FARMAKIDES, M. N., and D. R. BOONE, Speech problems of patients with multiple sclerosis, J. Speech Hear. Dis., 25, 385-390, 1960.
- 7. FLETCHER, S. G., Hypernasal voice as an indication of regional growth and developmental disturbances, Logos, 3, 3-12, 1960.
- 8. FLETCHER, S. G., Theory and instrumentation for quantitative measurement of nasality, *Cleft Palate J.*, 7, 601-609, 1970.
- 9. FLETCHER, S. G., Contigencies for bioelectronic modification of nasality, J. Speech Hear. Dis., 37, 329-346, 1972.
- 10. FLETCHER, S. G., and M. E. BISHOP, Measurement of nasality with TONAR, Cleft Palate J., 7, 610-621, 1970.
- 11. FLETCHER, S. G., I. SOOUDI, and S. D. FROST, Quantitative and graphic analysis of prosthetic treatment for "Nasalance" in speech, J. Prosth. Dent., 32, 284-291, 1974.
- 12. GLENN, J. W., and N. KLEINER, Speaker identification based on nasal phonation, J. Acoust. Soc. Am., 43, 368-372, 1968.
- 13. MALECOT, A., Acoustic cues for nasal consonants, an experimental study involving a tapesplicing technique, *Language*, 32, 274-284, 1956.
- 14. MOLL, K. L., Speech characteristics of individuals with cleft lip and palate, In Cleft Palate and Communication, D. C. Spriestersbach and D. Sherman, Ed. Academic Press, New York, 1968.
- 15. SHERMAN, D., The merits of backward playing of connected speech in the scaling of voice quality disorders, J. Speech Hear. Dis. 19, 312-321, 1954.
- 16. SINGH, S., A step towards a theory of speech perception, Paper presented at Speech Communication Seminar, Stockholm, Aug. 1–3, 1974.
- 17. WEINBERG, B., J. F. BOSMA, J. C. SHANKS, and W. DEMEYER, Myotonic dystrophy initially manifested by speech disability, J. Speech Hear. Dis., 33, 51-58, 1968.