Contribution of Nasally Emitted Sound to the Perception of Hypernasality of Vowels

JAMES R. ANDREWS, Ph.D.
DeKalb, Illinois

DAVID RUTHERFORD, Ph.D.
Evanston, Illinois

Acoustic correlates of hypernasality have been studied in synthetic vowel spectra generated by vocal tract analogs and in real vowel spectra produced by human subjects. Data from analog studies have suggested that because the nasal side branch is highly damped, little or no sound energy is emitted from this source. Consequently, investigators have emphasized the role of altered features of the orally emitted spectrum as a basis for the perception of hypernasality. Studies of real speech spectra produced by hypernasal human subjects have not altogether confirmed these findings, but have suggested that the perception of hypernasality is also associated with the addition of new resonances to the vowel spectrum.

Extra resonances between formants of vowels produced by hypernasal speakers have been reported by a number of investigators, including Coleman (*1*), Curry (*2*), Hattori, Yamamoto and Fujimura (*5*), Heffner (*6*), Joos (*7*), Peterson (*8*) and Potter, Kopp and Green (*9*). Dickson (*3*) analyzed the spectra of the vowels /u/ and /i/ of forty nasal subjects, and while he found no spectrum features consistently associated with the perception of nasality, his results suggest that added resonances between regular vowel formants may be associated with more severe degrees of hypernasality.

Since most investigators of hypernasality in human subjects have not made a distinction between the sound emitted from the nasal cavity and that emitted from the oral cavity, the contribution of the nasally emitted component to the severity of perceived hypernasality remains largely unknown. The objective of the present study was to distinguish between orally and nasally emitted sound and to determine the extent to which nasally emitted sound may contribute to a listener's perception of nasality.

Data for this investigation were gathered while the senior author was a trainee studying under NINDB Grant NB05420-05 at Northwestern University. The prosthesis was originally designed by Dr. Morton Rosen. It was modified for this investigation by Dr. Robert Wheeler. Both Drs. Rosen and Wheeler are on the staff of the Northwestern University Cleft Lip and Palate Institute. This report is based on a Ph.D. dissertation completed at Northwestern University under the direction of Professor David Rutherford. Portions of this paper were presented at the 1967 Convention of the American Speech and Hearing Association, Chicago, Illinois and at the 1968 Meeting of the American Cleft Palate Association, Miami Beach, Florida.

Method

Nasally and orally emitted components of vowel spectra were recorded separately, but simultaneously, as a normal speaker wearing a variable aperture prosthesis produced consonant-vowel syllables. The acoustic spectra of these vowels, produced at several different velopharyngeal aperture sizes, were analyzed and listeners judged the degree of perceived nasality.

SPEECH SAMPLES. Speech samples consisted of the vowels /i/, /u/, /a/ and /ɛ/ preceded by a glottal aspirate. Each syllable was repeated six times at each velopharyngeal aperture condition, three times without masking and three times with a masking noise presented binaurally to the subject.

ISOLATION OF THE NASALLY EMITTED SOUND. Figure 1 shows a sketch of the lead cylinder which served to isolate the nasally emitted sound from that coming from the subject's mouth. The tube was four feet long, three inches in diameter and had a one-quarter inch thick wall. The midportion of the subject's face, from the bridge of the nose to the middle of the upper lip, was placed against the end of the cylinder. A condenser microphone (Western Electric, Model 640 AA) placed in the tube eight inches from the subject's nose picked up the nasally emitted sound. Orally emitted sound was picked up by another condenser microphone placed eight inches from the lips. Signals from the two microphones were recorded separately on a dual channel tape recorder (Ampex 350-2).

In order to ensure isolation of the two energy sources, a gasket formed of modeling clay was molded to fit over the end of the tube, closely

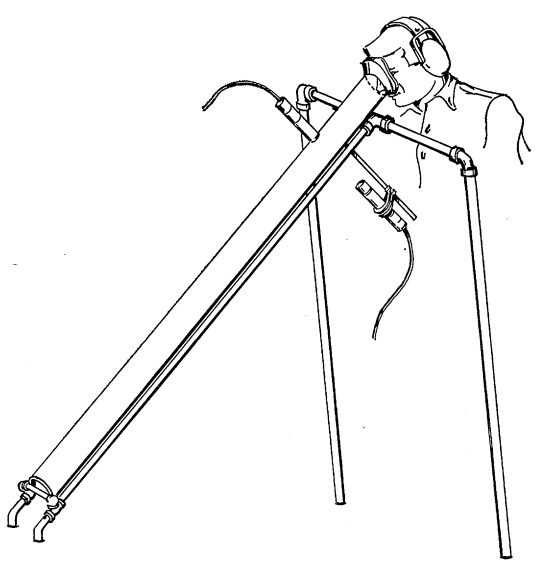

FIGURE 1. Lead cylinder with subject in position.

conforming to the contour of the subject's face. Tests utilizing both live voice and pure tones indicated that attenuation through the tube walls amounted to at least 30 dB over the frequency range of interest. After loosely packing the cylinder with cotton batting, the frequency response characteristic of the nasal microphone was determined, and an equalizing network was constructed to correct for the acoustic reaction of the tube. Analysis of the glottal spectra recorded intra-orally by means of a probe tube microphone with and without the lead cylinder in place indicated that acoustic reactance of the tube upon the laryngeal harmonics was negligible.

Variable Aperture Prosthesis. A variable aperture prosthesis was constructed and fitted according to standard prosthetic procedures. The prosthesis, shown in Figure 2, differed from a standard speech appliance in two ways. The center portion of the speech bulb was drilled out, leaving an irregularly shaped oval aperture with a cross sectional area of 240 mm^2. In addition, a wire basket projected upward from the outer perimeter of the bulb to prevent the subject's palate from occluding the aperture.

Aperture size was controlled for the various conditions by selectively occluding the opening with beeswax. The experimental conditions included five different aperture sizes. These were 0 mm^2, 60 mm^2, 120 mm^2, 180 mm^2, and 240 mm^2.

Masking Noise. During three of the six productions of each syllable at each aperture size a masking noise, consisting of white noise mixed with a 125 Hz pure tone, was presented binaurally to the subject at a level sufficiently intense to completely mask his awareness of the sound of his own voice. This procedure was used because Coleman (*1*) has suggested that under conditions similar to those of the present study, a subject might compensate for nasality on the basis of auditory feedback. In order to investigate this possibility, the sound pressure of the nasally and orally emitted sound was determined for each syllable. Measurements of the

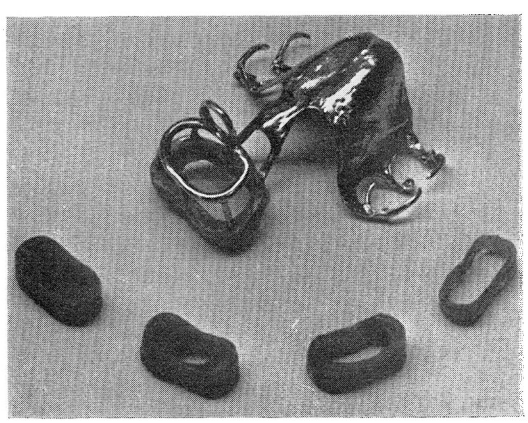

FIGURE 2. Variable aperture prosthesis and inserts.

sound pressure of the nasally emitted sound relative to the orally emitted sound were examined and comparisons were made between syllables recorded using different aperture sizes. Since relative sound pressure of the nasally emitted sound did not vary systematically with increase in aperture size under the "no masking" condition, these samples were not considered for further analysis. Only the 120 syllables produced under the "masking" condition were used in the remainder of the investigation.

Recording Procedures. All recordings were made with the subject seated in a sound treated room (Industrial Acoustics Company, Model 1201). The sequence of aperture conditions was randomly selected with the size of aperture in use at any given time unknown to the subject. The subject was asked to produce the syllables "in a natural manner" rather than sustaining the vowel. No attempt was made to control the loudness or the pitch of the subject's voice.

Results

Listener Judgments

Reliability of Judgments and Interjudge Variability. A panel of three judges, each accustomed to hearing and judging hypernasal voices, was selected to rate, on a five point scale, the degree of nasality of both the combined oral and nasal signals and the oral component alone. The listening tape consisted of 120 different speech samples which had been arranged randomly and equalized for intensity ± 1 dB. To make the judges' scale ratings equivalent, the mean scale rating assigned to the samples by each judge was computed. Following this the standard deviation of each individual judge's distribution of ratings was derived. The mean was subtracted from each scale rating and the difference, divided by the standard deviation, yielded a score which was multiplied by ten and added to fifty in order to avoid minus numbers. For each speech sample the three standardized units were averaged and the mean unit was used to represent the degree of severity of nasality.

In order to determine the reliability of judgments, the first forty samples presented on the listening tape were presented again at the end of the session in a different random order. The number of speech samples on the listening tape, then, totaled 160. A Pearson Product Moment correlation was computed to determine the relationship between the ratings for the first and last forty samples. A reliability coefficient of .91 was obtained.

Difference in Degree of Nasality Between the Combined Signal and the Orally Emitted Component. Tabulations showing the number of times the rated nasality of the combined signal exceeded that of the oral signal are shown in Table 1. For the vowel /u/, the combined signal was consistently the more nasal. On 11 of the 15 productions, the degree of nasality of the combined signal exceeded that of the oral signal alone. In approximately one-half of the productions of the vowels /i/, /a/, and /ɛ/, the combined

TABLE 1. Relative nasality of oral and combined signals

	vowel			
	/u/	*/i/*	*/a/*	*/ɛ/*
combined signal judged more nasal	11 (73%)	7 (47%)	9 (60%)	6 (40%)
oral signal judged as nasal or more nasal	4 (27%)	8 (53%)	6 (40%)	9 (60%)

TABLE 2. Relative nasality of oral and combined signals at each aperture condition

	aperture size in mm²				
	0	*60*	*120*	*180*	*240*
combined signal judged more nasal	6 (50%)	6 (50%)	6 (50%)	7 (58%)	9 (75%)
oral signal judged as nasal or more nasal	6 (50%)	6 (50%)	6 (50%)	5 (42%)	3 (25%)

signal was rated as more nasal. No significance should be attributed to the findings of any of these three vowels since they probably indicate the inability of the judges to hear systematic differences in nasality between the two signals.

Data showing the relative severity of nasality of the oral and the combined signals for the four vowels produced three times at each aperture condition are displayed in Table 2. It should be noted that while degree of nasality increased as expected when a velopharyngeal aperture was introduced, the present investigation was primarily concerned with perceptual *differences* between the oral and combined oral and nasal signals. Therefore, only perceptual data describing those differences are presented. For all samples combined, the effect of adding the nasally emitted component to the orally emitted component increased markedly at the 240 mm^2 aperture condition. The rated nasality of the combined spectrum exceeded the rated nasality of the oral component in approximately one-half of the cases at aperture sizes from 0 mm^2 to 180 mm^2. This percentage increased to 75% at the 240 mm^2 aperture condition.

Although the samples were not drawn randomly, the exact binomial probability can be used as a guide in evaluating these percentages. If only chance were operating and there was no actual difference in the relative hypernasality of the oral and combined components, one would expect the number of "more nasal" judgments to occur about equally between the two signals. With a sample size of twelve, the exact binomial probability of obtaining as many as 75% or more of the judgments favoring one category is less than 7%. Consequently, it might be concluded that when

TABLE 3. Relative nasality of high and low vowels at each aperture condition

	aperture size in mm²				
	0	*60*	*120*	*180*	*240*
combined signal judged more nasal for /i/ and /u/	3 (50%)	3 (50%)	4 (67%)	4 (67%)	5 (83%)
combined signal judged more nasal for /a/ and /ɛ/	3 (50%)	3 (50%)	2 (33%)	3 (50%)	4 (67%)

the combined signal was judged to be more nasal at least 75% of the time, as it was at the 240 mm² aperture condition, the nasal component was, in fact, adding to perceived nasality.

The trend for the number of "more nasal" judgments to increase for the combined signal with increase in aperture size was especially apparent for the two high vowels. Table 3 shows the relative nasality of the two high and the two low vowels at each aperture condition.

Spectrographic Measures

As discussed earlier, vowels produced by human nasal speakers have tended to contain extra regions of resonance in both relatively narrow frequency bands and more widely dispersed regions throughout the spectrum. These resonances may be emitted from the nose. Their effect on severity of nasality is largely unknown since most investigators have not differentiated between orally and nasally emitted sound. The presence of extra resonances may account for the perceptual differences already described. In order to verify this assumption, to specify more precisely the possible effect of frequency and relative intensity of nasal resonances, and to compare these with distortions of the orally emitted signal spectrographic measurements were made. The vowel spectra of both the combined oral and nasal signal and the oral component alone were analyzed for each production in order to identify spectrum changes that were associated with changes in degree of perceived nasality. Comparisons were made of the first, second and third formant frequencies, intensity of formants two and three relative to the first formant, and relative intensity of the third and seventh harmonics. In addition, the spectra were analyzed for obvious differences such as the addition of new resonances between regular vowel formants.

Changes in the Spectrum of the Orally Emitted Component Associated with Increases in Degree of Nasality. Considering first the spectra of the oral component alone, for the vowel /u/ the relative intensity of both the third and seventh harmonics decreased as severity of nasality increased. When Spearman rank correlations were computed between severity of nasality and relative intensity of the third and seventh harmonics, coefficients of −.86 and −.87 respectively, were obtained. Both of these coeffi-

cients were significant at the .01 level, suggesting that these apparent antiresonances were systematically associated with degree of nasality. These harmonics fell within the frequency region of the first and second formants. Other spectrum features of the orally emitted component of the vowel /u/ which tended to be associated with an increase in nasality included a reduction of first formant frequency and a rise in second formant frequency. These features correlated −.76 and .67 respectively, with degree of nasality, and both were significant at the .01 level.

For the vowel /i/, still considering only the orally emitted spectrum, a Spearman rank correlation of −.81, statistically significant at the .01 level, was obtained between relative intensity of the third harmonic and degree of nasality, with the third harmonic decreasing in intensity with increases in nasality. As for /u/, the third harmonic for the vowel /i/ fell within the frequency region of the first formant. The intensity of the seventh harmonic, which fell between the first and second formants, was unrelated to degree of nasality for the vowel /i/. Another spectrum characteristic of /i/, negatively correlated with increasing severity of nasality, was a decrease of first formant frequency. The Spearman Rank Correlation Test yielded a coefficient of −.48, significant at the .05 level, between degree of nasality and frequency of the first formant. All productions of /i/, even those at the "no aperture" condition, had some interformant resonances. These interformant resonances appeared to be what Hattori *et al.* (*5*) referred to as diffuse resonances between regular vowel formants on nasalized vowels. Such diffuse resonances seemed to have little to do with degree of nasality in the vowels of the subject used in this study.

Few systematic changes were seen in the oral spectra of the vowels /a/ and /ɛ/ with increases in degree of nasality. The lack of systematic changes in spectrum features may be in part due to the fact that degree of nasality for these two vowels did not vary greatly over the range of the velopharyngeal aperture sizes employed.

Changes in the Spectrum of the Combined Orally and Nasally Emitted Components Associated with Increase in Degree of Nasality. Since the conventionally heard signal of a hypernasal speaker is the combination of the orally and nasally emitted components, the spectrum of the combined signal is essentially the same as that of the oral signal except when the intensity of a nasally emitted harmonic equals or exceeds that of the same harmonic in the orally emitted signal. In cases where the degree of nasality of the combined sound might substantially exceed that of the oral component alone, the expected spectrum difference would be the presence of relatively intense nasally emitted harmonics.

For the vowel /u/ the intensity of the nasally emitted signal exceeded that of the orally emitted only in the frequency range from approximately 1800 to 2400 Hz, that is, in the region of the third formant of the vowel as produced by this subject. The intensity of the nasally emitted component was greater than the intensity of the orally emitted signal in this region

on eight of the twelve productions made with some degree of velopharyngeal aperture. In each instance the combined spectrum was rated as more nasal than the oral spectrum alone. The feature tending to be associated with relatively large increases in degree of nasality of the combined spectrum was the combination of increased intensity and reduced frequency of the third formant. In several instances the third formant resembled Fant's (*4*) description of a split formant.

For the vowel /i/ nasally emitted harmonics tended to be relatively intense in a region from 900-1500 Hz in the productions at the 120 mm^2 and 240 mm^2 aperture conditions, producing resonances between the first and second formants. In addition, the intensity of the nasally emitted harmonics slightly exceeded those of the orally emitted spectrum in the frequency region just above the first formant on two other productions of the vowel. The combined spectrum was rated as more nasal than the orally emitted spectrum in only four of these eight productions.

For both the vowels /a/ and /ɛ/, the intensity of nasally emitted harmonics rarely exceeded those of the oral spectrum. In instances when this did occur the intensity of the entire signal was relatively low. These spectrum features were not related to degree of nasality. Several low intensity resonances were apparent, however, and it seems likely that had the structure of the speech appliance allowed for a larger velopharyngeal aperture, these components might have reached an intensity great enough to affect degree of nasality.

Discussion

Consistent with the findings of both analog and human studies, the two high vowels, /i/ and /u/, were found to be more nasal than the two low vowels, /a/ and /ɛ/. Differences between the high vowels and the low vowels were evident both in terms of spectrum features and perceptual judgments. Even the most severely nasal low vowel samples did not greatly exceed the degree of nasality of the high vowels produced at the smallest aperture condition. Seemingly even the largest velopharyngeal aperture possible in the obturator used in this study was not great enough to cause the low vowels to be extremely nasal.

The results of this investigation should help to clarify the relationship between nasally emitted sound and the perception of nasality as it has been described in studies utilizing synthesized speech and live voice. Changes in oral spectra which correlated with increasing severity of nasality did in fact demonstrate the importance of distortions of regular vowel features upon degree of hypernasality, most particularly those produced by the partial coincidence of an antiresonance and a regular vowel formant. While the addition of the nasally emitted component significantly affected severity of nasality only on the vowel /u/, the presence of low intensity resonances in the nasally emitted spectra of low vowels and

the tendency for the combined signal to be judged as increasingly more nasal than the oral component alone with increase in aperture size, indicates that with greater nasal tract coupling the nasally emitted component could be expected to contribute proportionately more to judgments of hypernasality. Thus, at small degrees of coupling (i.e. when the nasal cavity input impedance is high relative to that of the oral cavity) perception of hypernasality probably is based on the distorted oral output, but under conditions of greater coupling, the nasally emitted component may contribute significantly to the overall perception of severity of nasality. When the intensity of nasally emitted harmonics exceeds that of corresponding harmonics of the orally emitted component in the frequency range near a formant, degree of nasality of the combined signal is greater than nasality of the oral signal alone. On the other hand, when the intensity of harmonics of the nasal component exceeds those of the oral component in regions of intensity minima (i.e. between vowel formants), degree of nasality of the combined spectrum is not systematically changed from that of the oral spectrum alone.

Hypernasality, whether associated with an antiresonance or a new resonance, is apparently heard when a listener recognizes a distortion in the expected frequency, intensity or bandwidth characteristics of the vowel formants. The presence of new resonances did not systematically affect degree of nasality when the resonances did not alter the characteristics of the vowel formants. The identification of hypernasality apparently is not based on the presence of specific spectral peaks, as seems to be the case with "nasality" as a distinctive feature of the nasal consonants, /m/, /n/, and /ŋ/. Rather, hypernasality on vowels seemingly is perceived when some feature of the learned vowel spectrum is distorted. Listeners apparently base their judgments of nasality upon their prior experience with the sounds of the specific language involved, and listen for distortions in vowel spectra they have learned. The implications of this hypothesis for the identification of hypernasality in a language or dialect with which the listener is not extremely familiar need to be explored.

Summary

The extent to which nasally emitted sound contributes to the perception of hypernasality was studied in a single normal speaker fitted with a speech appliance having an aperture which could be varied from 0 mm^2 to 240 mm^2 in 60 mm^2 steps. The speech samples consisted of consonant-vowel syllables in which the consonant /h/ was paired with the vowels /u/, /i/, /a/, and /ɛ/. The nasally emitted component was isolated from the oral component by a lead cylinder, allowing the two signals to be recorded separately but simultaneously. Data indicated that when the intensity of harmonics of the nasally emitted spectrum exceeded the intensity of corresponding harmonics of the orally emitted spectrum in the

frequency region near a regular vowel formant, the degree of nasality of the combined signal was perceived to be greater than that of the oral component alone.

reprints: *James R. Andrews, Ph.D.*
Speech and Hearing Clinic
Northern Illinois University
DeKalb, Illinois 60115

References

1. Coleman, R., The effect of changes in width of velopharnygeal aperture on acoustic and perceptual properties of nasalized vowels. Unpublished Ph.D. dissertation, Northwestern University, 1963.
2. Curry, R., The Mechanism of the Human Voice. New York: Longmans Green, 1940.
3. Dickson, D., An acoustic and radiographic study of nasality. Ph.D. dissertation, Northwestern University, 1961.
4. Fant, G., Acoustic Theory of Speech Production. 's-Gravenhage: Mouton & Co., 1960.
5. Hattori, S., Yamamoto, K., and Fujimura, O., Nasalization of vowels in relation to nasals, *J. of Acoustical Society of America, 30*, 267–274, 1958.
6. Heffner, R. M. S., *General Phonetics*. Madison: University of Wisconsin-Press, 1960.
7. Joos, M., Acoustic phonetics, *Language, 24*, 1–136, 1948.
8. Peterson, G. E., Parameters of vowel quality, *J. of Speech and Hearing Research, 4*, 10–29, 1961.
9. Potter, R. K., Kopp, G. A., and Green, H. C., Visible Speech, New York: D. van Nostrand and Co., 1947.