Perception of hypernasality and its physical correlates

Yukie KOZAKI. Noriko SUZUKI. Teruo AMAGASA. Masato AKAGI.

Maxillofacial Surgery, Graduate School, Tokyo Medical and Dental University
School of Information Science, Japan Advanced Institute of Science and Technology
First Department of Oral and Maxillofacial Surgery, School of Dentistry, Showa University
y-kozaki.mfs@tmd.ac.jp

Abstract

This paper discusses acoustic characteristics of hypernasality in Japanese vowel /i/ focusing on differences between hypernasality and nasal vowel by psychoacoustic experiments and simulations with vocal tract models. The psychoacoustic experiments were carried out with various modified vowels by STRAIGHT to investigate relationships between the degree of nasalization and the spectral envelope shape. Decreasing of the magnitude of F2 and appearing of a dip at the frequency region between F2 and F3 were related to perception of nasalization. It affected to perceive nasalization severely that these two spectral distortions were simultaneous. In the vicinity of F1, an additional peak above the frequency region of F1, broadening of the bandwidth of F1, and decreasing of the magnitude of F1 were related to perception of nasalization. The simulations with vocal tract models showed that the frequency region of dips spread from F1 region to high frequency region as the velopharyngeal open area was increased. When the velopharyngeal opening area was so small as to speak normal nasal vowel or mild hypernasality, a small dip appeared at F1 region only. When the velopharyngeal area was open as large as soft palate defect which may cause severe hypernasality, remarkable dips appeared at both F1 region and between F2 and F3 region. Therefore, nasal vowel and mild hypernasality were considered to be similar to each other acoustically. On the other hand, severe hypernasality was distinguished from nasal vowel and mild hypernasality by a remarkable dip between F2 and F3 region. The results suggest that severe degree of nasalization reflects the distortion of the spectral envelope between F2 and F3 region.

1. Introduction

Hypernasality is an acoustic distortion, which is represented as nasalization of vowels resulting from an acoustic coupling of oral and nasal cavity. Cleft palate patients and velum resection patients result in hypernasality due to velopharyngeal incompetence. Quantitative evaluation of hypernasality is important to observe development of speech, and to determine the goal of speech therapy. However, evaluation of hypernasality has depended on perceptual judgment by experimented listeners mainly. The goal of this research is to set up the standard characteristics to evaluate objectively the degree of hypernasal.

Previous researches with acoustic analysis of hypernasality and nasal vowel have shown several spectral characteristics related to nasalization. However, the difference between hypernasality and nasal vowel has not been clear. We focused on vowel /i/, and classified these spectral characteristics into 8 categories as shown in Table1.

In this work, we hypothesized that hypernasality and nasal vowel are different with the degree of nasalization. Moreover, we hypothesized that auditory impressions of nasalization are related to its physical characteristics. To prove this assumption, we conducted three things: (a) analyze the physical characteristics in the spectral envelope of hypernasality and imitative hypernasality (experiment 1); (b) clarify the relationship between spectral envelope features and auditory impressions of experimented listeners (experiment 2), and (c) clarify the relationship between spectral features and velopharyngeal open area (experiment 3).

2. Experiment1 : Acoustic analysis

2.1. Methods

2.1.1 Speech data

Speech data were obtained from 8 cleft palate patients and 2 oral tumor patients with diagnosed hypernasality by speech pathologists. Tumor patients had undergone surgical resection of the velum. The rating of hypernasal severity were 3 step score with “mild hypernasality (Ⅰ),” “moderate hypernasality (Ⅱ),” “severe hypernasality (Ⅲ).” As controls, 8 speech data were obtained from normal speakers, who were free from abnormalities in craniofacial anatomy and hearing. They were judged as normal phonation and resonance. The data of imitative hypernasality were obtained from 2 normal speakers pronounced intentionally with excess nasal resonance.

Speech data were sustained speech wave of Japanese vowel /i/ uttered by all subjects with natural pith and loudness. The data were recorded in DAT using a condenser microphone in a soundproof room.

2.1.2 Acoustic analysis

The data were analyzed using the unbiased estimation
Table 1. Acoustical characteristics of nasal vowel and hypernasality in previous research. V: nasal vowel. H: hypernasality

<table>
<thead>
<tr>
<th></th>
<th>F1 width</th>
<th>between F1andF2</th>
<th>F2 between F2andF3</th>
<th>F3 anti resonance</th>
<th>nasal peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>House, et al. [1]</td>
<td>level □</td>
<td>□</td>
<td>level □</td>
<td>□</td>
<td>700-1800Hz</td>
</tr>
<tr>
<td>Hattori, et al. [2]</td>
<td>level □</td>
<td>diffuse</td>
<td>level □</td>
<td>□</td>
<td>550Hz</td>
</tr>
<tr>
<td>Dickson [3]</td>
<td>shift</td>
<td>level □</td>
<td>level □</td>
<td>□</td>
<td>300-400Hz</td>
</tr>
<tr>
<td>Fujimura, et al. [4]</td>
<td>level □</td>
<td>flat</td>
<td>□</td>
<td>level □</td>
<td>F1near</td>
</tr>
<tr>
<td>Takeuchi, et al. [5]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>450Hz</td>
</tr>
<tr>
<td>Maeda [6,7]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Dickson [3]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Chen [10,11]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Feng [12]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Imai, et al. [13]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Kataoka [14]</td>
<td>level □</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Fig1a. hypernasality (+++)  
Fig1b. hypernasality (+)  
Fig1c. hypernasality (˴)  
Fig1d. imitative hypernasality

2.2. Results and Discussion

Fig.1 shows typical spectral envelopes of hypernasality and normal speech. By the comparison of hypernasality with normal speech, the characteristics of hypernasality were classified into 7 categories. Concerning as the formant peaks, decreasing of the magnitude of F1, F2, and F3 were observed. Concerning as dips, two remarkable dips were observed between formant peaks, which were a dip between F1 and F2 [D1], and a dip between F2 and F3 [D2]. Broadening of the bandwidth of F1, and an additional peak above the frequency region of F1 [P1] were observed. Table 2 illustrated which categories were also observed in each speech data.

The characteristics of F1 region, which is decreasing

of F1 magnitude or P1 or broadening of the bandwidth, were observed in all speech data.

Decreasing of F2 magnitude and D2 were observed in speech data with (+++) and (+), while they were not observed in (˴). The results indicated that these two distortions of spectral envelope were related to the severe degree of hypernasality. Decreasing of the magnitude of F2 and F3 were always observed with D2. The results indicated that D2 was related to decreasing of the magnitude of F2 and F3.

On imitative hypernasality, P1 was observed clearly, while D2 and decreasing of F3 magnitude were observed slightly. Decreasing of F2 magnitude was not observed. That is, spectral distortions of high frequency region above F2 were not remarkable on imitative hypernasality.

3. Hypothesis

We hypothesized the spectral characteristics of hypernasality from acoustic analysis and previous research, which were (1) broadening of F1 bandwidth,(2)
<table>
<thead>
<tr>
<th>Condition</th>
<th>Frequency [Hz]</th>
<th>Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) F1 bandwidth ₗ</td>
<td>400~700</td>
<td>400Hz'</td>
</tr>
<tr>
<td>(2) F1 magnitude ₗ</td>
<td>200~500</td>
<td>10dB down</td>
</tr>
<tr>
<td>(3) P0</td>
<td>200~300</td>
<td>10dB up</td>
</tr>
<tr>
<td>(4) P1</td>
<td>800~1000</td>
<td>20dB up</td>
</tr>
<tr>
<td>(5) F2 magnitude ₗ</td>
<td>1800~2300</td>
<td>20dB down</td>
</tr>
<tr>
<td>(6) D2</td>
<td>2300~3500</td>
<td>20dB down</td>
</tr>
<tr>
<td>(7) F1 - F2 ₗ</td>
<td>1000~1800</td>
<td>0</td>
</tr>
<tr>
<td>(8) F2 - F3 ₗ</td>
<td>1000~1800</td>
<td>10dB up</td>
</tr>
</tbody>
</table>

Table3. Condition of spectral modification

decreasing of F1 magnitude, (3) additional peak below F1 [P0], (4) additional peak above F1 [P1], (5) decreasing of F2 magnitude, (6) dip between F2 and F3 [D2], (7) decreasing of between F1 and F2 magnitude, (8) increasing of between F1 and F2 magnitude.

4. Experiment 2: Psychoacoustic experiment

4.1. Methods

4.1.1 Stimuli

The speech data for preparing stimuli was Japanese vowel /i/ without nasalization uttered by adult male with natural pith and loudness. Modifications of the spectral envelope were carried out by STRAIGHT [16]. At first, we modified the spectral envelopes with one condition as shown in Table3. Secondly, focusing on D2 in combination with decreased F2, we modified with this combination and other condition. The number of stimuli was 16 (1 no modified, 8 modified with one condition, 1 modified with decreased F2 and D2, and 6 modified with decreased F2 and D2 and other condition). All stimuli was recorded randomly on DAT and presented to the subjects through headphones at 3 times.

4.1.2 Procedure

The subjects were 5 experimented listeners who often diagnose and treat hypernasality patients. The judgment procedure was to give a score with the range from 0 to 5 based on the standard as follows: 0 is no nasal resonance; 1 is mild; 2 is mild-moderate; 3 is moderate; 4 is moderate-severe; 5 is severe-hypernasal resonance.

4.2 Results and Discussion

As the score of no modified stimulus was not 0 in all subjects, new score was calculated by subtraction of no modified stimulus score from modified stimulus score on each subject. Table4 illustrates the averaged new score for each stimulus type. The stimulus for which the score is larger was identified as severe hypernasal.

Analyses of variance with 5 percent were carried out on the score of 7 stimuli modified with focusing on decreased F2 and D2. The stimulus, which is D2 in combination with decreased F2, did not have the affection of subjects. The average score was 2.3. The results indicated that D2 in combination with decreased F2 was related to perceive hypernasality severely. We consider it to be related to masking. When the dip appeared between F2 and F3 region without decreasing of F2, the dip was difficult to perceive due to masking. However, when the dip appeared with decreased F2, the dip was free from masking and perceived as hypernasal.

Mixing this combination into other condition, the score showed large compared with the score without the combination. In analysis of variance, 2 stimuli, which are decreased F1 magnitude and decreased between F1 and F2, did not have the affection of subjects.

The hypothesis was confirmed on (2) broadening of the bandwidth of F1, (4) P1, (5) decreasing of F2 magnitude, and (6) D2. To clarify the findings, we introduced the simulation with vocal tract model.

5. Experiment 3: Simulation

5.1.Methods

The spectral envelope of normal speech and hypernasality are very different. We consider it results from the different degree of nasal resonance. To clarify this, we simulated using vocal tract model, which had been described by Sondhi and Schroeter [17].

We divided nasal cavity into 16 sections on the model, and set up each section as shown in table5 based on Chen’s model [11]. The velopharyngeal area when normal speaker open largely was set up at 0.8 mm. To clarify the affection of increasing of velopharyngeal

<table>
<thead>
<tr>
<th>subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>average</th>
</tr>
</thead>
</table>

Table4. Score of hypernasality for each subject (1~5).
area, we simulated by opening velopharyngeal area gradually.

5.2. Results and Discussion

The peaks of vocal tract transfer function were considered as F1, F2, F3 from low frequency region by turns. The vocal tract transfer function without nasal cavity was showed as Fig.2.

When the velopharyngeal area opened slightly, a dip appeared at F1 region only and F1 magnitude decreased as shown in Fig.3A. When the velopharyngeal opening area became large, and a new dip appeared between F2 and F3 region, and F2 magnitude decreased as shown in Fig.3B. When the velopharyngeal opening area became remarkably large, the dips became remarkably deep and wide, and F2 magnitude decreased more as shown in Fig.3C. That is, the simulations showed that the frequency region of dips spread from F1 region to high frequency region, and the dips became wide and deep as the velopharyngeal opening area was increased.

6. General Discussion

We consider the difference between hypernasality and nasal vowel. In the case of nasal vowel which is pronounced by normal speaker and mild hypernasal, velopharyngeal open area is small, and a dip appears at only F1 region or a small dip appear at between F2 and F3 region. In the case of severe hypernasality with severe velopharyngeal incompetence, velopharyngeal open area is large, and dips appeared clearly at F1 region and between F2 and F3 region. That is, acoustical characteristics of mild hypernasality are almost the same as nasal vowel. And severe hypernasality shows clear difference from mild hypernasality as the distortion of spectral envelope at between F2 and F3 region.

7. Acknowledgements

We would like to thank Yukari Yamashita and Masako Kato of Showa University, and Michiko Mibu of Tokyo Medical and Dental University and Setsuko Imatomi of Sophia University, for serving as the subjects who judged hypernasality.

8. Reference