Acoust. Sci. & Tech. **27**, 5 (2006)

ACOUSTICAL LETTER

Cue parsing between nasality and breathiness in speech perception

Takayuki Arai*

Department of Electrical and Electronics Engineering, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(*Received 23 March 2006, Accepted for publication 31 March 2006*)

Keywords: Speech perception, Cue parsing, Nasality, Breathiness
PACS number: 43.71.Es, 43.71.An [doi:10.1250/ast.27.298]

1. Introduction

Vowel nasalization is normally achieved by lowering the soft palate or velum, resulting in acoustic coupling between the main vocal tract and the nasal cavity. A simple model for a nasalized vowel could be a main vocal tract with a side branch, where different degrees of opening of the velopharyngeal port would produce different degrees of nasalization. According to acoustic theory [1–6], the basic difference between the transfer function of a vocal tract with a side branch and that of a non-nasal vowel is that additional poles and zeros are introduced to the vocal-tract transfer function as a consequence of acoustic coupling with the nasal tract.

The additional poles and zeros of nasal coupling cause several modifications in the spectrum. The amplitude of the first formant (F_1) is reduced, the bandwidth of F_1 is broadened, the frequency of F_1 shifts upward, and there is a relative strengthening of the spectrum at around 250 Hz [1–4,6–10]. Higher frequencies may also be affected by nasal coupling. The main effect of nasalization, however, is the perturbation of the low-frequency spectrum, which replaces the first formant with a shifted F_1 (F_1'), a nasal formant (F_n), and a nasal zero (F_z) [3,6,11]. As the cross-sectional area of the velopharyngeal opening is gradually increased, F_1 frequency shifts, F_1 bandwidth increases, and the spacing between the pole and zero introduced in the vicinity of the first formant increases.

A nasal peak between 250 and 450 Hz has been observed in various studies of vowel nasalization [6,7,12–14]. Most of these research studies show that the peak may be due to the pole-zero pair introduced by the paranasal sinuses [1,10].

On the other hand, a breathy vowel is characterized by a glottal configuration [15]. During a breathy vowel, the arytenoid cartilages are well separated at the back, but the vocal processes are sufficiently approximated so that the vocal folds vibrate when lung pressure is applied to the system. Because the glottis is never completely closed at the back over the vibratory period, there is considerable low-frequency airflow. As a result, the source signal has an increased open quotient (OQ: a proportion of a period during which the glottis is open) and a very strong fundamental component (H1). Additionally, the amplitudes of higher harmonics are substantially attenuated because of nonsimultaneous closure, and the increased airflow causes aspiration noise. Moreover, tracheal poles and zeros can be observed because of the acoustic coupling of the subglottal system [16]. Thus, the acoustic correlates of breathiness are increased OQ, an increase in the relative amplitude of the fundamental component, increased spectral tilt, and aspiration noise at frequencies above about 1.5 kHz [15].

Other studies have focused on acoustic cues found in the acoustic correlates of a breathy vowel. Hillenbrand *et al.* evaluated the effectiveness of several acoustic measures in predicting breathiness ratings [17]. The acoustic measures were signal periodicity, first harmonic amplitude, and spectral tilt. They concluded that signal periodicity measures provided the most accurate predictions and the relative amplitude of the harmonic correlated moderately with breathiness ratings. The measure H1-H2 (relative amplitude of the first two harmonics) has been used by several other researchers to reflect OQ and has been shown to be correlated with OQ [18,19]. Hanson *et al.* also tested H1-A1 (the relative amplitude of the first harmonic and the first-formant peak) as an indicator of B1 (the bandwidth of the first formant), and H1-A3 (the relative amplitude of the first harmonic and the third-formant peak) as a measure of spectral tilt [19].

Figure 1 shows the summary of the acoustic cues related to nasality and breathiness. This figure shows that nasality and breathiness have several acoustic cues in common, as well as some cues that are strongly correlated with one or the other. When a vowel sound is produced with a lowered velum and wide-spread glottis, the resulting acoustic signal will enable the listener to perceive nasality and/or breathiness depending on the interaction of the cues in Fig. 1.

In this study, we investigate how a listener parses these acoustic cues for nasality and breathiness. We systematically synthesize vowels by changing the acoustic parameters using the Klatt's synthesizer [15,20] and then we conduct a perceptual experiment.

2. Perceptual Experiment

2.1. Stimuli

All stimuli were vowels synthesized using XKL, a revision of the software package developed by Klatt [15,20]. The length of each stimulus was 200 ms. The fundamental frequency (F_0) started at 120 Hz, reached 125 Hz in 150 ms, and ended at 90 Hz. The amplitude of voicing (AV) started at 54 dB, reached 60 dB in 50 ms, stayed at the same level for 100 ms, and ended at 48 dB. Because the vowel was assumed to be /a/, the first and second formant frequencies (F_1 and F_2) were set to 800 and 1,200 Hz, respectively. The rest of the

*e-mail: arai@sophia.ac.jp

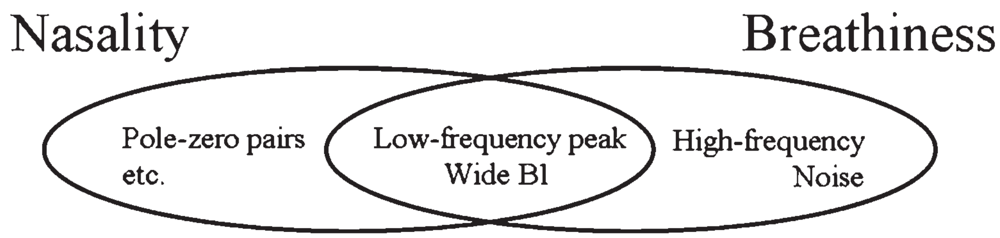

Fig. 1 Cues related to nasality and breathiness.

Table 1 Values of parameters for 27 synthesized stimuli and their average ratings on nasality and breathiness.

Stimulus No.	FNZ [Hz]	OQ [%]	AH [dB]	Average rating	
				Nasality	Breathiness
1	500	50	0	0.5	1.0
2	500	50	50	0.0	0.5
3	500	50	60	0.5	3.0
4	500	70	0	0.5	0.5
5	500	70	50	1.0	1.0
6	500	70	60	0.0	4.0
7	500	90	0	1.5	0.0
8	500	90	50	1.0	2.5
9	500	90	60	0.0	3.0
10	600	50	0	3. 0	1.0
11	600	50	50	2.0	2.5
12	600	50	60	2.0	4.0
13	600	70	0	2.5	0.0
14	600	70	50	1.5	1.5
15	600	70	60	2.0	3.0
16	600	90	0	2.0	0.5
17	600	90	50	2.5	2.0
18	600	90	60	2.5	3.0
19	700	50	0	3.5	0.5
20	700	50	50	3.5	1.0
21	700	50	60	4.0	3.0
22	700	70	0	3.5	0.5
23	700	70	50	4.0	0.5
24	700	70	60	3.5	3.5
25	700	90	0	3.5	0.5
26	700	90	50	3.5	1.5
27	700	90	60	3.5	3.5

parameters were kept at default values, specifically, the higher formant frequencies remained constant at $F_3 = 2{,}500$, $F_4 = 3{,}250$, $F_5 = 3{,}700$, and $F_6 = 4{,}990$ Hz. The sampling rate (SR) was 10 kHz. The parameters above were kept constant for all stimuli.

To affect the degree of nasality and breathiness, we changed the parameters as shown in Table 1. While the frequency of the nasal pole (FNP) was fixed at 500 Hz, the frequency of the nasal zero (FNZ) was variable from 500 to 700 Hz based on the technique used by Hawkins *et al.* [8]. When FNZ = 500, FNP and FNZ cancel each other out, and no nasal coupling is simulated. When FNZ = 600 or 700, it simulates different degrees of nasalization. We changed the open quotient (OQ) from 50 to 90% and the amplitude of aspiration (AH) from 0 to 60 dB. When the glottal configuration is modal, OQ = 50 and AH = 0 (these are the default values of XKL software). The higher values of OQ and AH simulate the conditions when the glottis is widely spread.

2.2. Procedure

Two experienced speech pathologists participated in the perceptual experiment. Twenty-seven icons of a loudspeaker were displayed on a PC screen. Initially, the icons were ordered randomly; each icon corresponded to one of 27 stimuli. For as many times as a subject double-clicked on the target icon, the sound was played from the PC via the digital-to-analog (D/A) converter of an Onkyo MA-500U digital audio amplifier through Sennheiser HD600 headphones. The subjects were asked to put the icons in order according to their perceived nasality and breathiness, and rate each stimulus for nasality and breathiness on a five-point scale.

2.3. Experimental results

The right half of Table 1 shows the results of this experiment. From this table, we can observe the general tendencies as described below: 1) when the glottal configuration is modal (OQ = 50 and AH = 0) and there is no nasal coupling (FNZ = 500), listeners do not perceive nasality or breathiness; 2) perceived nasality increases as the values of FNZ depart from FNP (= 500); and 3) breathiness increases as AH increases.

3. Discussions

Klatt and Klatt [15] reported that the following acoustic parameters of natural speech are correlated with subjective judgments of breathiness: degree of aspiration noise intruding in frequencies above 1.5 kHz in vowels, and the relative strength of the fundamental component [15]. Using a formant synthesizer, Klatt and Klatt also confirmed the importance of aspiration noise [15]. Although they mainly focused on manipulating the parameters related to breathiness, in the present study, we further introduced a parameter directly related to nasality, that is, a pole-zero pair, and investigated how a listener parses the cues related to both nasality and breathiness depending on the accompanying cues.

From Table 1, we observe a general tendency wherein breathiness judgments increase as AH increases. This is consistent with Klatt and Klatt's findings that the strongest single cue to breathiness was the amplitude of the aspiration noise added to the spectrum [15].

When AH is low (AH = 0), there was no demonstrable tendency of breathiness judgments for varying FNZ and/or OQ. When AH is high (AH = 50, 60), however, we found that breathiness judgments tend to be higher with higher OQ values. These observations are also consistent with those of Klatt and Klatt [15] when they concluded that the preferred stimulus in terms of breathiness (and naturalness) is one in which all of the following cues are present: add aspiration noise, increased spectral tilt, increased open quotient to increased H1, and widened first-formant bandwidth. It is as if the listener is aware of all the systematic changes that go into breathy phonation, and the listener uses these expectations during perception in such a way that no single cue is as effective as all of them in conjunction [15].

When the pole-zero pair became slightly separated (i.e., FNZ = 600), there was a tendency for breathiness judgments to increase. However, in a heavily nasalized situation (i.e., FNZ = 700), breathiness judgments decreased again. This indicates the existence of an interaction between nasality and breathiness when listeners parse the cues.

From Table 1, we observe a tendency for nasality judgments to increase as FNZ increases, due to the separation of the pole-zero pair (FNP is fixed at 500 Hz.) This is consistent with Hawkins *et al.*'s findings [8] that a wider spacing of the pole-zero pair that introduces a prominent extra peak was found to be necessary for the perception of a nasal vowel, particularly for [e], [ɑ], and [u].

In general, the higher the OQ, the more judgments for nasality were observed. This is mostly true for FNZ = 500 and 600, particularly when AH = 50. When the phonation is modal and there is no nasal coupling (FNZ = 500, AH = 0, and OQ = 50), there is no perception of nasality. However, nasality was perceived when OQ increased with a high AH. When there is already nasal coupling (FNZ = 600, 700), the increased OQ seems to be perceived as more nasal.

A general flattening of the spectrum at low frequencies occurs when there is vocal fold vibration with a spread glottis [19]. The glottal opening contributes acoustic losses to the lowest vocal tract resonance and it enhances the amplitude of the first harmonic in relation to higher harmonics [21]. Thus, some of the acoustic consequences of spreading the glottis are similar to the acoustic correlate of nasalization. Moreover, Keyser and Stevens [21] supposed that a spread glottis contributes to the relative "enhancement" of the harmonic for nasal vowels. The tendency observed in the present study that the increased OQ contributes to the perception of nasality supports the finding that the "spread glottis" can be the enhancing gesture for nasality, as pointed out by Keyser and Stevens [21].

Another tendency we observed was that higher AH decreases perceived nasality. This can be observed when FNZ = 500 and OQ = 70 and 90. This observation is consistent with Imatomi and Arai's findings [22]. In their study, the perceived nasality ratings were investigated with normal and breathy voice sources. They synthesized stimuli with different degrees of nasal coupling on the basis of the source-filter model and concluded that the ratings of nasality of the synthesized stimuli with breathy voice sources were lower, in most cases, than those with normal voice counterparts. A similar tendency was also observed in the present experiment.

Thus, there is an interaction between breathiness and nasality judgments as pointed out by Klatt and Klatt [15]. They showed that increases in the strength of the fundamental component were not always a sign of perceptual breathiness, but rather may have induced a sensation of increased nasality unless accompanied by aspiration noise. They call this a paradox. In other words, one cue, an increase in the amplitude of the first harmonic, is interpreted either as signaling nasality or breathiness depending on the values of the other cues present in the signal [15]. The observation in the present study is consistent with their findings, and we consider it to be a type of the cue parsing phenomenon.

Ohala and Amador [23] tested a hypothesis that vowels produced with a slightly open glottis might have acoustic characteristics that would mimic the effects of nasalization. Ohala [24] studied a phenomenon in Indo-Aryan, in which nasal vowels appear in certain words where the following consonant is voiceless (so-called "spontaneous nasalization"). This process is of interest because these words never had a nasal consonant in their prehistory. The nasalization of a vowel can be viewed as a process of enhancing this acoustic property by further decreasing the prominence of F_1 [21].

The relationship between the lowering of the velum and the adjustment or articulation at the larynx is more prevalent than is generally realized. Matisoff [25] named the affinity between the feature of nasality and the articulatory involvement of the glottis "rhinoglottophilia." When someone is exhausted, he/she might naturally produce a voice with nasalization and aspiration noise. This reflects that the glottis and velopharyngeal port are open at the same time. Moreover, we unconsciously control these organs to breathe, that is, we open the glottis and the velopharyngeal port at the same time [26]. As a result, spreading the glottis and nasality mutually enhance each other.

4. Summary

Nasality and breathiness have several acoustic cues in common, as well as some cues that are strongly correlated with one or the other. When we produce a vowel sound with a lowered velum and/or a wide-spread glottis, the resulting acoustic signal might contain these cues. In this study, we investigated how a listener parses these cues for nasality and breathiness. As a result of the perceptual experiment, we observed that a listener perceives nasality and breathiness depending on the interaction of such cues. The general tendencies were as follows: 1) perceived nasality increases as the spacing of the nasal pole and zero becomes wider; 2) perceived breathiness increases as aspiration increases; 3) with strong aspiration, breathiness is higher as open quotient increases; 4) the higher the open quotient, the higher the judgment for nasality; and furthermore, when there is already nasal coupling, the increased open quotient with aspiration yields more perceived nasality; and 5) strong aspiration decreases perceived nasality.

Acknowledgments

This study was carried out while I was a Visiting Scientist of the Speech Communication Group in the Research Laboratory of Electronics, Massachusetts Institute of Technology (Cambridge, MA, USA) from 2000 through 2004. I would like to thank all the people who helped me in various ways, especially Kenneth N. Stevens, Joseph S. Perkell, Stefanie Shattuck-Hufnagel, Sharon Manuel, Janet Slifka, Helen Hanson, other members of the Speech Communication

Group at MIT, Ben Gold of MIT Lincoln Laboratory, John J. Ohala of the University of California, Berkeley, Setsuko Imatomi of Sophia University, and Kyoko Takeuchi of Kokugakuin University. This research was supported in part by Grant-in-Aid for Scientific Research (A-2, 16203041) from the Japan Society for the Promotion of Science.

References

[1] K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, Mass., 1998).

[2] A. S. House and K. N. Stevens, "Analog studies of the nasalization of vowels," *J. Speech Hear. Disord.*, **21**, 218–232 (1956).

[3] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).

[4] O. Fujimura, "Spectra of nasalized vowels," *Res. Lab. Electron. Q. Prog. Rep.*, No. 58, pp. 214–218 (1960).

[5] O. Fujimura, "Analysis of nasalized vowels," *Res. Lab. Electron. Q. Prog. Rep.*, No. 62, pp. 191–192 (1961).

[6] O. Fujimura and J. Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," *J. Acoust. Soc. Am.*, **49**, 541–558 (1971).

[7] S. Hattori, K. Yamamoto and O. Fujimura, "Nasalization of vowels in relation to nasals," *J. Acoust. Soc. Am.*, **30**, 267–274 (1958).

[8] S. W. Hawkins and K. N. Stevens, "Acoustic and perceptual correlates of the non-nasal- nasal distinction for vowels," *J. Acoust. Soc. Am.*, **77**, 1560–1575 (1985).

[9] S. Maeda, "Acoustic of vowel nasalization and articulatory shifts in French nasal vowels," in *Phonetics and Phonology*, M. K. Huffman and R. A. Krakow, Eds., 5, Nasals, Nasalization, and the Velum (Academic Press, San Diego, Calif., 1993), pp. 147–167.

[10] M. Y. Chen, "Acoustic correlates of English and French nasalized vowels," *J. Acoust. Soc. Am.*, **102**, 2360–2370 (1997).

[11] K. N. Stevens, G. Fant and S. Hawkins, "Some acoustical and perceptual correlates of nasal vowels," in *In Honor of Ilse Lehiste*, R. Channon and L. Shokey, Eds. (Foris, Dordrecht, Holland, 1986), pp. 241–254.

[12] J. Lindqvist-Gauffin and J. Sundberg, "Acoustic properties of the nasal tract," *Phonetica*, **33**, 161–168 (1976).

[13] S. Maeda, "The role of the sinus cavities in the production of vowels," *Proc. Int. Conf. Acoust. Speech Signal Process.*, Paris, pp. 911–914 (1982).

[14] J. Dang and K. Honda, "An investigation of the acoustic characteristics of the paranasal cavities," *Proc. XIIIth Int. Congr. Phonetic Sciences*, Vol. 1, pp. 342–345 (1995).

[15] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.*, **87**, 820–857 (1990).

[16] D. H. Klatt, "Acoustic correlates of breathiness," *J. Acoust. Soc. Am.*, **82**, Suppl. 1, S91 (A) (1987).

[17] J. Hillenbrand, R. A. Cleveland and R. L. Erickson, "Acoustic correlates of breathy voice quality," *J. Speech Hear. Res.*, **37**, 769–778 (1994).

[18] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. Guiod and S. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *J. Speech Hear. Res.*, **38**, 1212–1223 (1995).

[19] H. M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *J. Acoust. Soc. Am.*, **101**, 466–481 (1997).

[20] D. H. Klatt, "The new MIT speech VAX computer facility," *Speech Communication Group Working Papers IV*, Research Laboratory of Electronics, MIT, Cambridge, Mass., pp. 73–82 (1984).

[21] S. J. Keyser and K. N. Stevens, "Enhancement revisted," in *Ken Hale: A Life in Language*, M. Kenstowicz, Ed. (MIT Press, Cambridge, Mass., 2001), pp. 271–291.

[22] S. Imatomi and T. Arai, "The relation between perceived hypernasality of speech and its hoarseness," *Proc. Forum Acusticum*, Sevilla (2002).

[23] J. J. Ohala and M. Amador, "Spontaneous nasalization," *J. Acoust. Soc. Am.*, **69**, S54–S55 (1981).

[24] J. J. Ohala, "The phonological end justifies any means," *Proc. Int. Congr. Linguistics*, Tokyo, pp. 232–243 (1982).

[25] J. A. Matisoff, "Rhinoglottophilia: The mysterious connection between nasality and glottality," in *Nasalfest: Papers from a Symposium on Nasals and Nasalization*, C. Ferguson, L. M. Hyman, and J. Ohala, Eds. (Stanford University Language Universals Project, Stanford, Calif., 1975), pp. 265–287.

[26] K. N. Stevens, *Personal communication* (2004).