Nasality Perception of Vowels in Different Language Background

Shahina HAQUE and Tomio TAKARA

Department of Information Engineering
University of the Ryukyus, Okinawa, Japan
takara@ie.u-ryukyu.ac.jp

ABSTRACT

Nasality is a distinctive feature of Bangla vowels. In this paper, we describe our study on nasality perception of Bangla vowels by Bangla and Japanese listeners. We discuss an interesting result that Japanese listeners perceive most Bangla nasal /˜i/ as Japanese non-nasal /u/. As the amount of nasalization of /˜i/ was increased synthetically, perception of this vowel change was found to be more categorical for Japanese listeners than that of Bangla listeners' vowel perception.

Index Terms: nasality perception, vowels, language background, Bangla, Japanese

1. INTRODUCTION

Bangla is a language of more than 120 million people of Bangladesh. Nasality is a distinctive feature of Bangla vowels. All seven vowels of Bangla have nasal counterpart [1]. Nasality of vowels changes the meaning of words in Bangla. Therefore, Bangla is useful for the study of the nasality feature of vowels.

Nasality introduces poles and zeros in the spectrum of nasal vowel [2]. Nasal zero is one of the principal spectral characteristic of nasal vowels. Cepstral method [3] and log magnitude approximation (LMA) filter [4] effectively approximates spectral peaks formed by poles and spectral dips formed by zeros of the vocal tract transfer function. Therefore, we use cepstral method and LMA filter for the analysis by synthesis of Bangla vowels to analyze and to synthesize, respectively.

As a study on nasality perception, we used natural and synthetic vowel data. Using natural and synthetic stimuli of Bangla vowels with varying nasality, we performed language dependent listening test with Japanese and Bangla listeners. From our study, we observed that Japanese listeners' perceive natural Bangla /˜i/ as /u/ by confusing nasal pole of /˜i/ with F2 of /u/. As nasality of /˜i/ is increased synthetically, perception of this vowel change was found to be more categorical for Japanese listeners than that of Bangla listeners' vowel perception.

2. SPEECH ANALYSIS-SYNTHESIS METHOD

The experimental part consists of recording each of the isolated 14 Bangla vowels uttered three times at a normal speaking rate by four native Bangla male speakers. The recording was done in a sound proof room on a DAT tape at a sampling rate of 48 kHz and 16 bit resolution. These digitized speech sounds are then down-sampled to 10 kHz for the purpose of analysis. For each speaker, the best one of these three sounds is chosen for our work.

The short-term cepstral analysis method is used to extract the speech parameters. The speech wave is segmented to 25.6ms frame length. A time-domain Blackman window is used. Frame shifting time is 10ms. We use cepstrum for spectral parameters, which is the inverse Fourier transform of the short time logarithmic amplitude spectrum of the speech waveform [2]. The resulting parameters of the speech unit include the number of frames and, for each frame, voiced/unvoiced decision, pitch period and cepstral coefficients. The first cepstral coefficient is the power content of the frame.

The speech synthesis section shown in Fig. 1 is designed on the source-system model [3]. The vocal tract features can be suitably represented for all speech sounds by the pole-zero LMA filter proposed by Imai. An LMA filter is a part of a homomorphic vocoder. In our speech synthesizer, an LMA filter is made from a cascade of 30 elemental second order filters. An LMA filter representing the vocal tract is driven by an adequate excitation source. In case of producing voiced speech, the excitation source is a series of unit impulses separated by fundamental period intervals. In case of producing unvoiced speech, the excitation source is noise having unit amplitude and random polarity.

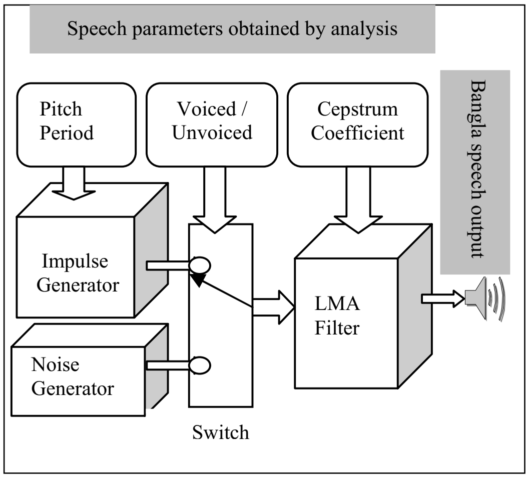


Fig. 1: Speech synthesis sub-system.

3. POLE-ZERO (PZ) MODEL

Studies involving direct manipulation of spectrum indicate that greatest number of nasal judgments tended to occur when a pole-zero pair is introduced in the vicinity of the first formant region of an oral vowel spectrum [6]. In another study, the frequencies and spacing of the inserted pole and zero in an all pole transfer function, around the vicinity of the first formant was varied to change the degree of nasalization [7].

In our study, we introduce nasalization by inserting a pole-zero pair in a pole-zero transfer function, at various frequencies in the vicinity of the first formant. Then we find a suitable value of frequency, spacing and amplitude of the pole-zero pair at which the oral vowel is judged to be nasal. Using this suitable value of frequency and spacing, amplitude of the pole-zero pair is varied systematically in steps to change the degree of nasalization.

From nasal vowel spectra, we observed that the nasal pole (P) and zero (Z) pair have a pattern like sine curve i.e. a combination of a peak pattern and a dip pattern. Therefore, in our previous study, we model nasal pole-zero pair by a sine curve in pole zero (PZ) model. Result obtained by listening tests for this model gave a correct score of nasal vowel recognition of 93% [8]. Pole-zero and zero-pole (ZP) rule of PZ model are shown in Fig. 2. Parameters of PZ model are amplitude and position of nasal pole or zero from the origin of the spectrum. Amplitude of PZ or ZP = amplitude of sine curve = $A1$[dB] or $A2$[dB]. Position of pole of PZ or ZP = $P1$[kHz] or $P2$[kHz].

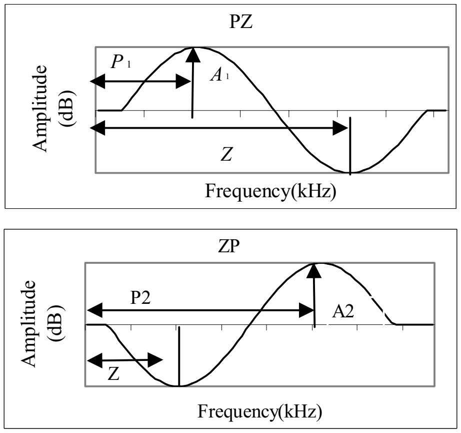


Fig. 2: PZ and ZP rule for vowel nasality generation by PZ model.

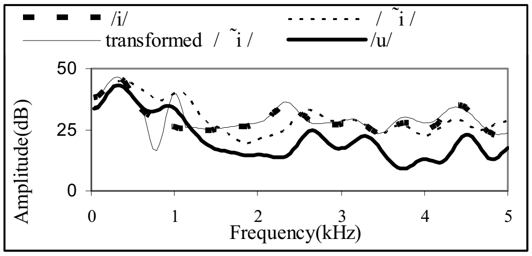


Fig. 3: For speaker 1, spectra of /i/, /˜i/, transformed /˜i/ by PZ model, and /u/.

For introducing nasalization using PZ model, an analyzed frame is chosen from the stable portion of the oral vowel spectrum. PZ model is applied to each frame with varying $A1$ or $A2$ to produce synthetic nasal vowel with varying nasality. An example of transformed spectrum /˜i/ by PZ model is shown in Fig. 3.

4. LANGUAGE DEPENDENT PERCEPTION OF /˜i/

Japanese has no oral-nasal phonemic contrast in vowel. There are few exceptions, in Kanto or North Kanto dialect, vowels preceded by /g/ consonant in continuous speech are occasionally nasalized, but it is not phonemic, whereas Bangla has oral-nasal phonemic contrast for all its seven vowels. Therefore, in sub-section 4.1, Japanese listeners are tested using natural stimuli of all Bangla oral-nasal vowels to observe how they perceive nasal vowels in their own language background. In sub-section 4.2, we further confirm the result obtained in sub-section 4.1 by listening tests of synthetic stimuli with varying nasality.

LMA filter can directly take into account the effect of spectral poles and zeros. We used a PZ model which has the peak pattern as well as the dip pattern to represent nasal pole and zero. Therefore, use of LMA filter with the PZ model can produce nasal vowel, with the full effect of spectral peak and dip pattern.

4.1. Perceptual test with natural stimuli

4.1.1. Speech materials

In this study, we use speech stimuli as described in Section 2. For each vowel and four speakers, there are four data. Total number of data presented to each listeners for 14 vowels are 56(= 14*4). The lengths of the stimuli are around one second as recorded.

4.1.2. Listening test method

The listening test is done in a sound proof room using headphones. Five native speakers of Bangla and ten native speakers of Japanese, with normal hearing ability participated in the listening test. At first, they are familiarized with the listening test system and are allowed to hear a few examples of data. Then the test is initiated. Each 56 data of the constructed speech corpus is played once randomly in three seconds interval. Japanese listeners are asked to select them among the 5 Japanese vowels. Our purpose was to check how the Japanese listeners perceive Bangla vowels in their own language background. So, for the Japanese listeners, there was no training for Bangla vowel perception and no option for selection of Bangla vowels. Bangla listeners are asked to select them as one of 14 Bangla vowels.

Table 1: Listening test result of Bangla vowels by ten Japanese listeners.

Response Input	/ i/	/ e/	/a/	/ o/	/u/
/i/	40				
/e/		40			
/æ/		38	2		
/a/			40		
/ɔ/			6	31	3
/o/			1	33	6
/u/					40
/ĩ/	**17**				**23**
/ẽ/		37		2	1
/æ̃/		38	2		
/ã/			40		
/ɔ̃/			6	31	3
/õ/				31	9
/ũ/				3	37

4.1.3 Result and Discussions

For Bangla listeners, the result of oral-nasal vowel detection gives 100% correct recognition score because Bangla listeners are familiar with vowel nasality. From the listening test result of Table 1, we observe that Japanese listeners perceive most of the nasal vowel as their corresponding oral vowel counterpart. As there is no /æ/ and /ɔ/ vowel in Japanese, so they perceive them as /e/ and /o/ respectively, which are nearest vowel quality for them.

But interesting case arises among the two language groups in case of perceiving /˜i/ as indicated by circle in Table 1. All Bangla listeners perceive /˜i/ as / ĩ/, whereas Japanese listeners perceive most (more than 50%) of the natural /˜i/ as /u/. As we observe this perceptual difference between the two language groups, so we analyze the spectra of /˜i/ and /u/ shown in Fig.3. We observe that: First and second speakers' /˜i/ are perceived as /u/ by the Japanese listeners. Third speakers' /˜i/ data are perceived as /i/ and also as /u/. Fourth speakers' /˜i/ data are perceived all as /i/.

For speaker 1, 2 and 3, the second formant of /u/ is near about the same frequency region (around 1 kHz) of / ĩ/'s prominent nasal formant. As Japanese has no nasal vowel, so it is easily predictable that, Japanese listeners may confuse / ĩ/'s nasal formant with /u/'s second formant, thereby confusing /˜i/ with /u/. So, for speaker 1 and 2, all /˜i/ data are perceived as /u/. For speaker 3, because the nasal formant is not so prominent, /˜i/s are perceived both as /i/ and /u/. But, in case of speaker 4, no nasal formant of /˜i/ is observed which is near about /u/'s second formant. So, /˜i/ is perceived as /i/ by the Japanese listeners. /˜i/'s F1 and F2 are in about the same frequency of /u/'s F1 and F3. Nasal pole (Fnp) of /˜i/ is in about the same frequency region (around 1 kHz) of /u/'s F2. As Japanese listeners have no perceptual axis of nasality, so, we may say that Fnp is superimposed on F2 axis in F1-F2 space. Therefore, it is easily predictable that Japanese listeners may confuse / ĩ/'s nasal formant Fnp with /u/'s second formant F2 in F1-F2 space, thereby perceiving /˜i/ as /u/.

It may be known that nasal pole of /˜i/ is located around F2 region of /u/. But as far as it is known, there is no experimental study which describes that similar frequency location of F2 of /u/ and nasal formant of /˜i/ may be related to the perceptual effect of confusion of the two vowels for the non-natives.

From the analyzed spectra of /e/, / a/, / o/, / u/ and their corresponding nasal counterpart, it is observed that no nasal formant of nasal vowels is near about the same frequency location of other oral vowels' second formant. So for other nasal vowels, confusion between oral and nasal vowel does not occur. Only in case of / ĩ/, the nasal formant is confused with second formant of /u/, thereby making phoneme confusion with /u/.

4.2 Perceptual test with synthetic stimuli

4.2.1 Speech materials

In this study, we use speech stimuli of /i/ and /˜i/ uttered by the first three native Bangla speakers as described in Section 2. In order to create a synthetic speech corpus with varying nasality, steps as discussed in Section 3 are applied. The above procedure is repeated for each oral vowel for the three speakers. For each speaker, speech corpus was made of 21 data (an oral analysis-synthesis (AS), a nasal AS and 19 synthetic speech generated with varying vowel nasality using PZ model). Therefore, total number of speech data for three speakers are 63(= 21*3). An example of parameters of PZ model for all vowels of speaker 1 is shown in Table 2.

4.2.2 Listening test method

The listening test is done in the same way with same five listeners from two language groups as described in Section 4.1.2. Bangla listeners are asked to select the speech data among /i/, / ĩ/, and /u/. The Japanese listeners are asked to select them between /i/ and /u/. As our purpose was to check how the Japanese listeners perceive varying nasality of / ĩ/, in their own language background. So, for the Japanese listeners, there was no training for Bangla vowel perception and no option for selection of / ĩ/.

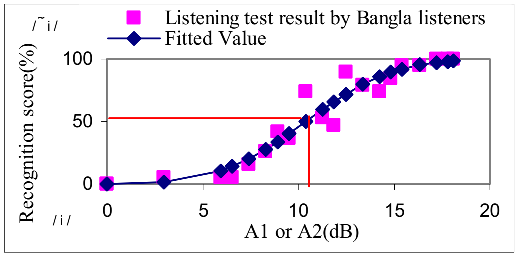
4.2.3 Result and Discussion

For speaker 1, listening test result for Bangla and Japanese listeners is shown in Fig. 4. As the nasality of /i/ is increased, Japanese listeners tend to hear /i/ as /u/ (i.e. across vowel category), whereas Bangla listeners perceives /i/ gradually as /˜i/. This result of perceiving /˜i/ as /u/ by Japanese listeners has similarity with the result obtained with natural stimuli in sub-section 4.1. From this experiment we observe:

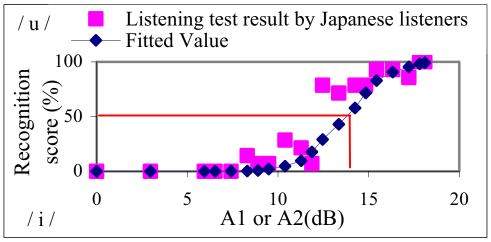
For Japanese listeners, the change of perception of vowels is more categorical: The graph of the listening tests of Fig. 4 is fitted with a normal cumulative density function. We measure the threshold of phoneme change (50% crossover point of nasal vowel recognition) and slope of the curve at threshold.

Table 2 : Parameters of PZ model for speaker 1

	*A*1(dB)	*P*1(kHz)	*Z*(kHz)	*P*2(kHz)	*A*2(dB)
/i/			0.790	1.090	11.5
/e/			0.796	1.058	8.7
/æ/			0.990	1.252	5.8
/a/	11.29	0.272	0.480		
/ɔ/			0.561	0.981	3.95
/o/			0.756	1.176	5.64
/u/			0.561	0.981	11.29



(a) Native Bangla listeners has slope of 11.2%/dB



(b) Japanese listeners has slope of 17.3%/dB

Fig. 4: Fitting the listening test result with a normal cumulative density function curve (data: Speaker 1).

We can take these two quantities as a measure of comparing the difference in perception of the listeners of the two language groups. Average value of 3 speakers' threshold of phoneme change is observed to be higher (10.6dB) for Japanese than Bangla listeners (9.3dB). We observe that at threshold, Japanese listener's perception from /i/ to /u/ has steeper slope (19%/dB, more categorical) than change of slope of /i/ to /˜i/ (11.8%/dB, continuous) of Bangla listeners. Therefore, as vowel nasality increases, Japanese listeners perceive from /˜i/ to /u/ which is more categorical than the change of perception from /i/ to /˜i/ of Bangla listeners. Because Japanese are used to hear consonant nasality which is categorical, we believe that this phoneme change is similar to the nasal consonant perception (categorical change).

We may also explain the above fact as follows: It is known that differences can be perceived as abrupt having sharp boundaries, i.e. categorical (consonants) or as continuous (vowels) [9]. But, categorical perception still occurs for vowels [10] though not as abrupt change in categories as consonants along a stimulus continuum that spans from one category to another. So we may say that, since Japanese language does not have /˜i/ category, the Japanese listeners all perceive it as /u/, the nearest quality for them. This phenomenon is a result of Japanese listeners' categorical perception of vowels.

5. CONCLUSION

Japanese has no oral-nasal phonemic contrast in vowels whereas Bangla has oral-nasal phonemic contrast for all its seven vowels. The question of interest of our work is how oral-nasal distinction of Bangla vowels is perceived by Japanese listeners in their own language background. We observed that, instead of perceiving nasal /˜i/ as phoneme /i/ which is within same vowel category, Japanese listeners perceive most of the data as oral /u/ which is across vowel category. For further investigation of the language dependent perception in discriminating oral-nasal contrast, we constructed speech corpus using synthetic stimuli with varying nasality. As the nasality of /˜i/ is increased in synthetic stimuli, Japanese listeners tend to perceive /˜i/ as /u/ (i.e. across vowel category) abruptly with an average slope of 19%/dB. Bangla listeners perform this change gradually from category /i/ to /˜i/ at an average slope 11.8 %/dB, at threshold. From our study, we may conclude that, due to Japanese listeners' categorical perception of vowels, similar spectral location of nasal formant of /˜i/ and F2 of /u/ results in perceptual illusion of perceiving /u/ which is the nearest vowel quality for them.

REFERENCES

[1] Hai, A., "*Dhvani-Vignan O Bangla Dhvani Tattwa*", June, 1985. (In Bangla)

[2] Deller, J.R., and Hansen, J.H.L., "*Discrete-Time Processing of Speech Signals*", pp. 137, IEEE Press, 2000.

[3] Furui, S., "*Digital Speech Processing, Synthesis, and Recognition*", Second Edition, pp. 30–31, Marcel Dekker, Inc, 2001.

[4] Imai, S., "Log Magnitude Approximation (LMA) filter", *Trans. of IECE Japan, J63-A*, 12, pp. 886–893 (1980). (In Japanese).

[5] Oppenheim, A.V., "*Applications of Digital Signal Processing*", pp. 156-159, Prentice-Hall, Inc, 1978.

[6] Takeuchi, S., Kasuya, H., and Kido, K. , "On the acoustic correlate of Nasality," *J.Acoust. Soc. Jpn.*, Vol.31, pp.298-309 (1975). (In Japanese)

[7] Hawkins, S. and Stevens, K.N., "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels", *J.Acoust.Soc.Am.*, 77(4), pp. 1560-1575, April, 1985.

[8] Haque, S. and Takara, T., "Rule Based Speech Synthesis by Cepstral Method for Standard Bangla", *18th International Congress on Acoustics, ICA 2004*, 4-9 April, 2004.

[9] Harnad, S., "Categorical Perception", *Encyclopedia of Cognitive Science, Nature Publishing Group/Macmillan*, (2003).http://cogprints.ecs.soton.ac.uk/archive/00003017/01/catperc.html.

[10] Fujisaki, H. and Kawashima, T., "A Model of the Mechanisms for Speech Perception Based on Discrimination of Synthetic Speech Sounds," *J.Acoust. Soc. Jpn.*, Vol.27, No.9, pp.453-462 (1971). (In Japanese)