STUDIA PHONOLOGICA XVIII (1984)

Nasal Consonant Discrimination by Vowel Independent Features

Shigeyoshi KITAZAWA and Shuji DOSHITA

ABSTRACT

This paper describes nasal consonant discrimination based on vowel independent features; murmur and onset spectrum. Classification based on spectral pattern statistics by linear discriminant functions has shown to be quite successful for a specific speaker, but not so good for a non-specific speaker set consisting of 44 speakers. Several experiments were made to determine the influence of individual factors, and some tentative conclusions are drawn.

1. INTRODUCTION

The question of invariance in consonant features is a current popular topic[1]. We have shown already some evidence for invariance of stop consonant features. Whether the invariance hypothesis can be extended to other consonants is an interesting question not only for speech science but also for speech recognition technology. In this article, nasals are the next target for a test of the invariance hypothesis.

It is known that three kinds of features are effective for nasal consonant discrimination:

(1) nasal murmur spectrum;
(2) vowel onset spectrum;
(3) formant transition from onset to stationary vowel.

Among these features, (1) is approximately independent of the vowel following the consonant. The consonants investigated in this report are CV-initial nasals composed of [m], [n], [ɲ], and [ŋ].

Since murmurs can be sustained, they may contain some phoneme-specific or manner- and place- specific information. By perceptual experiments, it has been shown that murmurs are well identifiable, even if isolated from a vowel, or directly attached to a steady-state vocalic portion (typically [a] or [æ] without formant transitions[2]. If it is true that there is some invariant acoustical feature, and it is computable, then that will be the most suitable feature for automatic speech recognition.

Shigeyoshi KITAZAWA (北澤茂良) : Assistant, Department of Information Science, Kyoto University.
Shuji DOSHITA (堂下修司) : Professor, Department of Information Science, Kyoto University.

However, this feature cannot be expected to be the speaker independent. Nasals are basically speaker dependent because of the articulatory mechanism. Nasal spectra directly reflect anatomical difference which can not be compensated by intentional effort of speaker, as in vowel. Of course, there is an invariance over speaker differences in the human perceptual level, but present pattern recognition technology has no means to treat this kind of problem.

Discrimination experiments were made for both a specific speaker and a large number of speakers, for comparison.

2. Acoustic Analysis of Nasal Consonants

2.1 Acoustic Characteristics of Nasal Consonants

The descriptions of each segment are as follows:

(1) Nasal Murmur:

A murmur is a synchronous, low amplitude and low frequency waveform preceding the CV nasal syllable, lasting several tens of milliseconds. In terms of production mechanism, murmur is characterized by nasal-opening and a totally obstructed oral pathway. Usually murmur waveform is easily distinguishable from vowel waveform, by the dominant low frequency component.

Although, it is difficult to distinguish place of articulation from the waveform characteristics, there are observable differences in spectrum due to mouth cavity coupling zeros.

Analysis data on nasal murmur in several languages reveal language independent systematic differences in spectral parameters[3]: (>=higher frequency than)

N_1 frequency values: [ŋ]>[ɲ]>[n]>[m];
N_1 bandwidth values: [ŋ]>[ɲ], [m], [n];
NZ frequency values: [ŋ]>[ɲ]>[n]>[m].
where N_1: a low first formant;
NZ: an antiformant.

Information about the perceptual relevance of murmurs as place cues is provided by experiments in which murmurs were presented for indentification[2]. The results show the following trends:

[m] murmurs (English) are quite accurately identified.
[n] murmurs (English) give 50%–60% correct responses.
[ɲ] murmurs (Polish) are not recognized as [ɲ] but mainly as [ŋ].
[ŋ] murmur (English) identification results are poor.

(2) Onset of Vowel at Oral Opening:

In terms of speech production, the onset point can be defined as the oral passage opening or the changing point from stationary murmur to nasalized vowel onset. According to the production theory, oral opening causes immediate increase in

second formant intensity[2].

(3) Formant Transition:

Formants can show frequency decrease, frequency increase, or no frequency variation from vocalic steady state to onset. The effect of the nasal consonant preceding [a] in several languages was investigated[2]. With respect to other vocalics, the few existing studies do not show any relevant characteristics. Vocalic onset behaviour does not seem to provide any vowel-independent cue for consonant place.

2.2 Segmentation of Nasal Waveform

We use a statistical discrimination method, therefore the training waveform should be segmented as precisely as possible. This is done by visual observation. A characteristics description in other terms than spectral is necessary since the waveform is used for segmentation.

Sometimes it is difficult to identify the onset point from visual observation of speech waveform because of the small and gradual increase in amplitude. The following two kinds of waveform filtering are utilized for onset point detection. One is high-pass filtering and the other is inverse filtering by a linear prediction filter adapted to murmur.

> Highpass filter: This is a 55 point FIR filter designed using a Kaiser window. The minimum stopband attenuation is 60 dB, and the ideal cut off frequency of the filter is 5.8 kHz (0.35).
>
> Inverse murmur filter: This is a 23rd order linear prediction filter obtained as a best fit to the stationary murmur waveform.

The characteristics of these two filters for onset detection are similar in essence. Since nasal murmur is dominated by low frequency components, output from the highpass filter eliminates almost the entire signal. When the oral cavity opens, the high frequency component appears immediately and this is visible in the waveform.

When murmur waveform is fed to the inverse murmur filter, the output is a pulse train synchronous with the murmur driving source, but this pattern changes as soon as the second formant amplitude builds up significantly at the vowel onset.

Fig. 1. shows an example of onset detection for nasal /no/. Observing the original waveform it is difficult to determine the onset point of time. The highpass

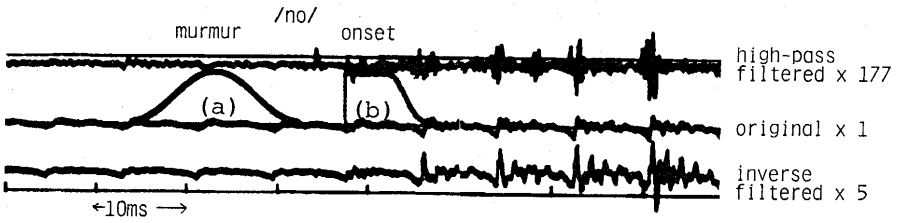

Fig. 1. Waveform and time window for (a) murmur and (b) onset.

filtered waveform shows a significant impulse at the onset, high frequency components follow. The inverse filter eliminates the murmur waveform, leaving only the residue pulse train. The vocalic waveform, even for a small amplitude, produces a recognizable amount of residue signal besides the driving source. Combining these three waveforms, the onset point can be identified unambiguously.

The stationary part of murmur waveform is weighted with a 20 ms Hamming window and 23rd order linear prediction analysis is applied to obtain a smoothed spectrum.

For an onset spectrum calculation, the waveform is excised at the beginning of a vocalic excitation with a forward rectangular and backward Hamming complex window of approximately 8 ms duration so that the analysis interval falls within a pitch interval. A linear prediction analysis similar to that for murmur is applied to compute spectrum.

The stationary vowel spectrum about 60 ms after onset is necessary as a reference for transition measurement from the vocalic onset spectrum. The 60 ms seems to be sufficient to reach a stationary state of vowel.

2.3 Spectrum Estimation and Representation for Discrimination

Nasal murmur spectrum can be described with nasal tract poles and oral cavity zeros, so a pole-zero model should be applied to estimate the murmur spectrum. However, human perception is more sensitive to spectral peaks than spectral zeros, and the approximation by an all-pole model is used in practice because of stability and simplicity in computation. The vowel onset spectrum is also better described by the all-pole model, eventhough the vowel onset is still nasalized. The stationary vowel spectrum, of course, is well represented by the all-pole model.

Speech waveform is first differenciated with coefficient 0.95, then a smoothed spectrum is obtained by the linear prediction method. The optimal prediction order

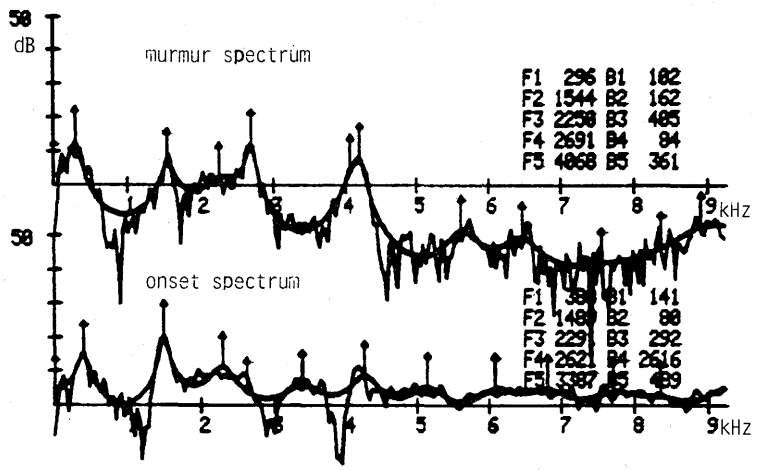

Fig. 2. Spectra for the parts of waveform in Fig. 1.

was estimated in a preliminary experiment to be 23. Zero mean log spectra are used for the nasal consonant discrimination described in the following paragraphs.

In Fig. 2. examples of LPC murmur and onset spectra are compared to Fourier spectra.

For spectral observation, linear prediction spectra are typically represented on a linear frequency scale. However, for research in connection with auditory performance, a bark or modified log frequency scale is usually more appropriate. We therefore chose a 28 channel critical-band representation with decibel amplitudes. Because these features have psycho-physical meanings, interpretation remains possible. Although other parameters, e.g., LPC cepstrum coefficients provide nearly equal results, further interpretation is impossible. This representation was found to be useful in stop consonant discrimination[4,5,6].

The frequency range 0 to 9.25 kHz was split as follows. Up to 1 kHz, the spectrum is covered by about ten 100 Hz bands. Above 1 kHz, the bands broaden exponentially. A total of 28 sections covers the whole frequency range. Average spectral energy of each section was calculated, and the whole spectra information was represented by a resulting 28 variable set: a 28 dimensional vector for each spectral observation.

3. Discrimination Experiment of Nasal Consonants

3.1 Materials for Experiment

Syllables examined are combinations of [m], [n], [ŋ] and five Japanese vowels, and [ɲ] and [ja], [ju], and [jo], a total of 18 different syllables. Utterances for each syllable were made separately, with a small pause between syllables, once for each speaker. Speech was recorded in a sound-proofed booth via an electret-condenser microphone and directly digitized at 18.5 kHz (every 54 μsec) with 12 bits, after going through an 8.9 kHz 70 dB/oct low-pass filter to avoid aliasing. The speakers are 45 young male native Japanese. One of them performed 20 repetitions of utterances used for specific speaker experiments, and the other 44 performed each utterance only once, for non-specific speaker experiments.

3.2 Consonant Discrimination

Consonant discrimination is performed based on statistical decision theory, because data is obviously noisy.

If d is the number of features to be used in making the classification, then a pattern $(f_1, f_2, \ldots, f_d)$ is a point in E_d (d-dimensional Euclidian space) with f_i being the value of the i-th feature. If patterns are drawn from R classes, then the classifier would assign an appropriate class to an input pattern based on a training set of patterns. The division of the feature space into R regions can be described by a set of R discriminant functions. A minimum distance classifier of this kind corresponds to a Bayes classifier in which a multivariate Gaussian distribution has been assumed

as the joint probability density function of the feature space for each class. This classification method has the advantage of simplicity, though the Gaussian assumption has to be checked. This was checked indirectly on our data set by testing each marginal distribution. Even with a Gaussian assumption, we must estimate the many parameters of the density function from a finite number of sample observation vectors. A simple and common choice for a pattern classifier is the linear classifier using the assumption that the covariance matrices are identical for each class. This assumption was also checked on our data set by a test of hypothesis, and proved to be within an allowable difference. The advantage of the linearity hypothesis resides not only in simplicity but also in robustness against non-normality and unequal covariance matrices[7].

There are some important techniques in applying linear discriminant functions. One is feature selection and the other is bias compensation. The program BMDP-7M that is used in the following analysis permits stepwise discriminant analysis[8]. The variables used in compution of the linear classification functions are chosen in a stepwise manner. At each step the variable that adds most to the separation of the groups is entered into the discriminant function and the variable that adds the least is removed. Those variables which are highly correlated or do not contribute much to separation are removed. Jackknife-validation procedure, also referred to as the one-leaving-out method, is used to reduce the bias due to the finite number of samples in group classification.

3.3 Discrimination of Nasal Consonant of a Specific Speaker

One male native Japanese speaker uttered 18 different nasal initial CV syllables 20 times, resulting in a data set of 360, where nasal consonants consist of [m], [n], [ɲ], and [ŋ] and vowels consist of [a, i, u, e, o]. Consonant discrimination was performed irrespective of the following vowel. The result is shown in Table 1 in terms of correct recognition rate, after the Jackknife-validation procedure.

The murmur spectrum (28 variables) discriminates better than onset spectrum (28 variables). This result reflects the fact that the murmur spectrum possesses consonant specific features that are independent from the succeeding vowel, while the onset spectrum does not. The two spectra have however mutually independent

Table 1. Jackknifed classification for data set of a specific speaker.
(Consonants in the leftmost column are classified as one of the upper row.)

group	with murmur & onset					with murmur					with onset				
	%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]
[m]	94.0	94	3	0	3	88.0	88	2	5	5	69.0	69	12	8	11
[n]	96.2	0	77	0	3	96.2	1	77	0	2	83.7	5	67	3	5
[ɲ]	100.0	0	0	80	0	98.7	1	0	79	0	97.5	0	0	78	2
[ŋ]	97.0	1	2	0	97	96.0	1	3	0	96	67.0	12	16	5	67
total	96.7					94.4					78.1				

features. When murmur and onset spectra are combined together (56 variables) and reduced by selection to 25 variables the resulting 96.7% score shows better discrimination than the 94.4% obtained with the murmur spectrum only. Half the variables entered to the discriminant function came from the onset spectrum, which shows the effectiveness of the onset features. The murmur F_1 is known to be one effective feature for nasal place discrimination, the addition of the first formant value to the spectrum set, as a useful feature resulted in an improved 97.2% correct classification.

Since the articulation point of /ni/ in Japanese is the same as that of /ɲja/, /n/'s in /ni/ are grouped into [ɲ] in the discrimination experiment.

3.4 Discrimination of Nasal Consonants of Non-specific Speakers

It can be predicted that nasal consonant place discrimination may be difficult when the speaker is not restricted to a specific speaker, because of the so called speaker factor. The apparent anatomical variability of the width of the nasal passages and the amount of mucous filling in cavities and constrictions are reflected in the variability of spectrographic details when data from different subjects are compared.

We prepared utterances of 18 different CV syllables in the same way as in 3.3 for 44 male speakers. These samples were processed as described in section 2 to form a data set consisting of 792(=44×18) samples, then the same discrimination procedure was applied to this data set with the assumption that a linear discriminant function is applicable.

Table 2. Jackknifed classification for data set of 20 speakers.

group	with murmur & onset					with murmur					with onset				
	%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]
[m]	75.0	75	9	5	11	62.0	62	18	13	7	48.0	48	15	9	28
[n]	82.5	3	66	1	10	63.7	13	51	5	11	67.5	9	54	6	11
[ɲ]	78.7	5	3	63	9	50.0	12	7	40	21	67.5	15	3	54	3
[ŋ]	56.0	14	9	21	56	57.0	17	12	14	57	35.0	29	15	21	35
total	72.2					58.3					53.1				

Table 3. Classification on nasal consonant from 44 speakers with murmur and onset spectrum.

group	samples	Apparent					Jackknifed				
		%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]
[m]	220	71.8	158	17	13	32	68.6	151	19	15	35
[n]	176	72.2	34	127	3	26	69.9	22	123	4	27
[ɲ]	176	79.5	10	7	82	10	77.3	11	7	136	22
[ŋ]	220	59.5	34	21	34	131	57.3	37	21	36	126
total	792	70.2					67.7				

For a subset of 20 speakers, discrimination based on murmur spectrum selecting 8 variables resulted in 58.3% correct discrimination, while onset spectrum selecting 11 variables resulted in 53.1% correct discrimination. A murmur-onset combination selecting 15 variables improved the score to 72.2% as shown in Table 2. Increasing the number of speakers to 44 degraded the score by a few percent (Table 3). The tendency of murmur discrimination error is similar to human identification error for isolated murmur as sketched in 2.1, but here the score is better.

The recognition rate is quite low compared to the result with a specific speaker. This reflects nasal variability between speakers.

Since the articulation point of /ni/ in Japanese is the same as that of /ɲja/, /n/'s in /ni/ are grouped into [ɲ] in the same manner as in 3.3.

3.5 Interactive Effects between Murmur and Onset

Murmurs are approximately independent of their following vowel, though, in fact, the murmur spectrum is slightly affected when the vocal cavity changes its form between vowels.

Onset, being the beginning of the vowel, is directly dependent on the target vowel, its spectrum is therefore vowel-dependent more than that of murmur, nevertheless the common articulation place seems to be preserved irrespective of the target vowel.

Quantitative measurements on coarticulation have been given by Tabata from a detailed study of V_1CV_2 syllables[9]. His method is based on multivariate statistical analysis of spectrum, instead of formant frequencies which are the common material in consonant coarticulation study. He discriminated four factors, two vowel-factors for V_1 and V_2, a consonant-factor, and a speaker-factor using an analysis of variance technique. He found the following results by analysis of VCV-type words uttered by five male adults:

(1) At the stationary part of nasal consonants, the effects of the speaker-factor is the largest among four factors;

(2) At the stationary part of the vowel, the effect of vowel-factor is the largest among four factors. The influence of the nasal consonant continuous to vowel is smaller than that of the speaker;

(3) The effect of speaker is generally considerable while that of consonant is not so strong. But the directions of the main axes, seen in the distributions of the three factors of vowel-effect, consonant-effect and speaker-effect, meet nearly at right angles with each other.

Combination of these two spectra should provide vowel independent features for nasal place discrimination, because it is known empirically and theoretically that there are some vowel independent patterns such as in first formant frequency and amplitude, and zero frequency change. This assumption was examined by the above discriminant test and proved to be true by the improved accuracy both in 3.3

for a specific speaker and in 3.4 for non-specific speakers.

4. STATIONARY VOWEL INTERACTION

4.1 Coarticulation Effect on Stationary Vowel

The amount of consonant place information residing in the stationary part of the vowel following the consonant was measured by discrimination of the nasal place based on the stationary vowel spectrum. The data set consists of 64 samples for each vowel: one complete set of utterances by 44 speakers and 20 complete sets of utterances by one speaker.

(1) Examination of stationary [a] spectral pattern following nasal:

The stationary [a] spectrum was calculated with a 20 ms Hamming window placed at about 60 ms later from the beginning of transition of vowel after nasal murmur. The result shows quite good discrimination for [ɲ] because of high coarticulation between [j] and [a], while most frequent misclassification was observed between [n] and [ŋ] as shown in Table 4. The coarticulation might be explained from measurement of articulator movement, however, we have no alternative to intuitive interpretation now.

(2) Examination of stationary [u] spectral pattern following nasal:

Correct discrimination was worse than vowel [a] as shown in Table 5.

Table 4. Classification of stationary vowel [a] according to nasal place. (once each for 44 speakers and 20 times for one other speaker)

group	samples	Apparent					Jackknifed				
		%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]
[m]	64	65.6	42	7	2	13	65.6	42	7	2	13
[n]	64	53.1	5	34	5	20	46.9	7	30	6	21
[ɲ]	64	85.9	0	7	55	2	82.8	1	7	53	3
[ŋ]	64	57.8	7	13	7	37	54.7	7	15	7	35
total	256	65.6					62.5				

Table 5. Classification of stationary vowel [u] according to nasal place. (once each for 44 speakers and 20 times for one other speaker)

group	samples	Apparent					Jackknifed				
		%	[m]	[n]	[ɲ]	[ŋ]	%	[m]	[n]	[ɲ]	[ŋ]
[m]	64	45.3	29	20	1	14	42.2	27	21	1	15
[n]	64	45.3	17	29	6	12	42.2	18	27	6	13
[ɲ]	64	84.4	0	7	54	3	82.8	0	8	53	3
[ŋ]	64	53.1	13	14	3	34	48.4	15	15	3	31
total	256	57.0					53.9				

Table 6. Vowel dependent discrimination of nasal consonant followed by [u]. (once each for 44 speakers and 20 times for one other speaker)

group	samples	with murmur and onset					with murmur, onset, and vowel				
		apparent score %	[m]	[n]	[ɲ]	[ŋ]	apparent score %	[m]	[n]	[ɲ]	[ŋ]
[m]	64	87.5	56	4	0	4	90.6	58	2	0	4
[n]	64	79.7	1	51	5	7	89.1	1	57	2	4
[ɲ]	64	90.6	0	3	58	3	98.4	0	1	63	0
[ŋ]	64	70.3	10	7	2	45	87.5	5	3	0	56
total	256	82.0					91.4				

Among consonants [ɲ] was best discriminated probably for a similar reason as [ɲ] before [a]. On the other hand [m, n, ŋ]'s are likely to be misclassified.

The results are only preliminary and need further study on with other vowels, however, they show some coarticulation effects between nasal consonant and vowel. In other words, the stationary vowel spectrum after the nasal consonant still contains some consonant place information.

4.2 Vowel Dependent Discrimination of Nasal Consonants for a Set of Speakers

Nasal consonants in 256 CV syllables (where V is fixed to a specific vowel), by 45 speakers (one additional speaker repeated 20 times the utterances) were investigated.

First, results of a discrimination experiment concerning murmur and onset spectrum is shown in the left half of Table 6 for discrimination between [m, n, ɲ, ŋ] followed by the vowel [u] (a total of 256 CV's). Better discrimination is visible than in the vowel independent case (compare with Table 2 and 3).

In comparison to the above result, the next shows the effect of an additional feature, stationary vowel spectrum, used with murmur and onset spectrum for discrimination. The right half of Table 6 shows the resultant significantly better discrimination, showing the importance of the vowel transitional feature.

With Jackknife-validation the correct discrimination rate was reduced by several percent from the apparent score, the resultant total score was 79.7% in the with murmur and onset case, and 85.2% with stationary vowel and murmur and onset.

From the results shown above, it seems possible to conclude that the stationary

Table 7. Nasal consonant classification rate depending on vowel.

following vowel	[a]	[i]	[u]	[e]	[o]
murmur and onset	91.0	73.4	82.0	88.0	92.2
murmur, onset and vowel	90.6	75.0	91.4	88.0	93.4

Table 8. Stationary vowel contribution to vowel independent nasal classification for a set of 44 speakers.
(note: unlike to Table 2 or Table 3, [n] group includes [ni].)

group	samples	murmur and onset					murmur, onset and stationary vowel				
		apparent score %	[m]	[n]	[ɲ]	[ŋ]	apparent score %	[m]	[n]	[ɲ]	[ŋ]
[m]	220	71.8	158	18	12	32	80.5	177	13	1	29
[n]	220	63.2	22	139	27	32	71.8	26	158	7	29
[ɲ]	132	77.3	5	9	102	16	79.5	4	7	105	16
[ŋ]	220	60.9	34	23	29	134	65.5	29	29	20	142
total	792	67.3					73.5				

Table 9. Nasal consonant discrimination including syllabic nasal.

group	samples	with murmur and onset					
		Jackknife score %	[m]	[n]	[ɲ]	[ŋ]	[N]
[m]	220	59.5	131	11	7	25	46
[n]	176	72.2	13	127	6	22	8
[ɲ]	176	41.8	9	6	125	26	10
[ŋ]	220	71.0	40	25	20	92	43
[N]	220	54.1	29	7	4	61	119
total	1012	58.7					

vowel [u] spectrum contributes to discrimination, but this conclusion cannot necessarily be extended to other vowels. Table 7 shows results in respect to different vowels. Generally, front vowels are less coarticulated than back or middle vowels.

The reason is not yet explained, but differences in nasal coarticulation with vowel could be one explanation. The narrower tract in front vowel results in less coupling with the nasal tract, producing therefore less change in spectral pattern, for example.

4.3 Stationary Vowel Contribution to Vowel Independent Discrimination in Non-specific Speaker Case

Here, in contrast to 4.2, vowel context knowledge is not used. However, the stationary vowel spectrum is used as a reference to compensate vowel dependency of onset, in contrast to 3.4. In Table 8, by comparison to Table 3, the stationary vowel contributes to discrimination score by 6 percent. Note that [ni] contrarily to 3.3 and 3.4, is not grouped into the [ɲ] class, but into the [n] class.

4.4 Syllabic Nasals

In Japanese, nasals can follow a vowel (in contrast to other consonants that can only occupy the initial position of a CV syllable.) Syllabic nasals were discriminated like other nasals by a similar method, the results are shown in Table 9.

Since the articulation point of /ni/ in Japanese is the same as that of /ɲja/, /n/'s

in /ni/ are grouped into [ɲ] in this discriminant experiment.

5. Summary

5.1 Idiosyncrasy

Previous research has shown that the nasal spectrum is highly dependent on individual person. Tabata[9] for example, has shown that the most significant factor at the nasal stationary part is the speaker factor. However, concerning back vowels, if the dependency on the followed vowel is compensated, a fairly high discrimination score is achievable, suggesting the presence of some invariant features for place of articulation, sufficient to overcome idiosynchrasy.

5.2 Vowel Dependency

Since murmur and onset spectrum is dependent on the following vowel, the discrimination of consonants sharing the same vowel improves correct classification rate significantly for back vowels and a little for front vowels.

5.3 Stationary Cues

The stationary nasal murmur posseses invariant cues that discriminate consonant well. Even the stationary vowel possesses some invariants cue for consonants.

5.4 Transitive Cue

The transitive cue has been best studied by synthetic speech psycho-acoustic tests, and is known to be an effective place cue, though vowel dependent.

In statistical discrimination of nasal spectra, a transition cue is implicitly used when murmur, onset and stationary spectra are combined as a vector from which linear discriminant functions are generated, although no attempt is made to extract explicitly formant frequency increase/decrease. It seems evident that the transitive cue has contributed in these discriminations. Of course, the transitive cue contributes most to vowel dependent discrimination, but it is also seen to be effective in some extent in speaker and vowel independent discrimination as a semi-invariant cue.

6. Conclusion

Vowel independency of nasal consonant place feature for discrimination has shown up quite well for a specific speaker in terms of classification based on spectral pattern statistics by linear discriminant function, but it has been difficult to show the same for non-specific speaker recognition, probably because of idiosyncrasy.

Acknowledgement

The authors would like to express their gratitude to students who contributed to the experiment described in here, especially Mr. Naoki Yokoi and Mr. Ken-ichi Arakawa, and Mr. Yoshihiro Nishimura.

REFERENCES

1) R. A. Cole and B. Scott, "Toward a theory of speech perception," *Psychol. Rev.*, 81, pp. 348–374, (1974).
2) D. Recasens "Place cues for nasal consonants with special reference to Catalan," *J. ASA* 73 1346–1353 (1983).
3) G. Fant, *Acoustic Theory of Speech Production*, (Mouton & Co., 'S-Gravenhage, 1960).
4) S. Kitazawa and S. Doshita, "Discriminant analysis of burst spectrum for Japanese initial voiceless stops," *Studia Phonologica*, vol. 16, pp. 48–70, 1982.
5) S. Kitazawa and S. Doshita, "Plosive discrimination by running spectra in 40-ms initial segment," *Studia Phonologica*, vol. 17, pp. 27–38, 1983.
6) S. Kitazawa and S. Doshita, "Speaker and vowel independent recognition of CV initial plosives by reduction and integration of running spectra," *Proc. of Seventh International Conference on Pattern Recognition*, Montreal, Canada, July 30—August 2, 1984.
7) P. A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York, 1975.
8) W. J. Dixon and M. B. Brown, *BMDP-77*, University of California Press, 1977.
9) K. Tabata, *Multivariate Statistical Analysis of Speech Sounds*, Doctoral Thesis, Kyoto Univ. 1973.

(Aug. 31, 1984, received)