Acoust. Sci. & Tech. **28**, 2 (2007)

ACOUSTICAL LETTER

Effects of stimulus contents and speaker familiarity on perceptual speaker identification

Kanae Amino* and Takayuki Arai

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

(*Received 20 September 2006, Accepted for publication 23 October 2006*)

Keywords: Nasals, Speaker identification, Speaker familiarity
PACS number: 43.71.Bp [doi:10.1250/ast.28.128]

1. Introduction

In daily conversation, listeners use both phonological and speaker information of the utterance in order to achieve successful communication [1]. It is reported that the processing of the phonological content and that of the speaker identity occur separately, although they interact with each other [1,2]. One example of this interaction is that certain sounds are effective for perceptual speaker identification [3]. Conducting a perceptual speaker identification test enables us to know which sounds are effective for the judgment of speaker identity [4], and those effective sounds must convey acoustic properties that indicate the speaker's identity. Using these acoustic properties, we can achieve higher accuracy in speech technologies, such as automatic speaker recognition and/or automatic speech recognition.

In our previous studies [5–7], we carried out familiar-speaker identification experiments, where differential effects of the phonological content on perceptual speaker identification were tested, and we found that stimuli containing nasal sounds were more effective for conveying identity than stimuli containing only oral sounds. However, it has been pointed out that the processes of identifying familiar speakers and unknown speakers are different [8–11]. Previous research [12] has shown that familiar listeners performed significantly better than naive listeners, or listeners who did not know the speakers previously, in identifying the same speakers.

In this study, we carried out a perceptual speaker identification test in order to examine the effective sounds for perceptual speaker identification in terms of speaker familiarity, and to determine whether the idiosyncrasy of the nasal sounds that was observed in familiar-speaker identification [5–7] can also be seen in unknown-speaker identification.

2. Method

2.1. Participants and speech materials

In speaker identification tests conducted with unknown speakers, the size and the composition of the speaker ensemble have a great effect on identification performance [13,14]. It is suggested that the desirable speaker ensemble for an experiment has a relatively small size and that the speakers' ages and genders should be consistent [3]. Thus, in this study, we used the speech materials of 4 male speakers of similar ages. These materials were previously recorded for a familiar-speaker identification test [6]. The 4 speakers were selected out of 10 male speakers on the basis of average fundamental frequencies, because it had been pointed out that fundamental frequency has a large effect on unknown-speaker identification [11,15], and in this study we aimed to determine the effects of the articulatory properties rather than the voice properties.

Nine monosyllables of the 4 speakers were used. The stimulus syllables are shown in Table 1. The speech samples are identical to those used in [6]. Each stimulus syllable was manually excerpted from four-moraic words that were read out within carrier sentences. The vowel in the CV syllables was always /a/ in order to keep the experiment simple.

Sixteen volunteers unfamiliar with any of the speakers participated in the listening speaker identification test. They were all native speakers of Japanese, and none of them had known hearing problems.

2.2. Procedures

All the sessions were carried out in a soundproof room. In order to make the test comparable to our previous experiments [5,6], we decided to conduct a speaker identification test rather than discrimination or matching tasks used in other similar studies [16–18].

Familiarisation sessions preceded the test sessions. In the familiarisation phase, the subjects listened to the recorded utterances of each speaker saying a common sentence, "/hoɴdʒitsɯ wa seiteɴ nari/ (It is fine today)," three times. They heard the utterances as many times as they wished, until they were confident that they could recognise the speakers. To avoid confusion, the speakers were introduced to the listeners by an ID number from 1 to 4, and the utterances were also always presented in order from 1 to 4 throughout the familiarisation sessions.

Familiarisation was followed by a practice session of 8 sentences, 2 for each speaker, presented in a random order with feedback after each trial. The sentences used here were identical to those used in familiarisation sessions. Familiarisation and practice were then repeated until the subject could identify the speaker with more than 90% accuracy. The average time subjects spent learning the voices was about 15 minutes, and the subjects listened to each speaker's utterances 3 to 12 times.

Then, we moved on to the test session without taking a break. The test session had 180 trials, that is corresponding to

*e-mail: amino-k@sophia.ac.jp

Table 1 Stimulus syllables.

Sonorants	Nasals	/ma/ /na/ /nja/
	Approximant	/ja/
Obstruents	Fricatives	/sa/ /za/
	Flap/Tap	/ra/
	Plosives	/ta/ /da/

Table 2 Results of the identification test with naive listeners ($N = 320$).

Stimulus	Percent correct
/na/	55.31
/nja/	50.31
/ja/	48.75
/sa/	46.88
/ma/	46.56
/za/	43.75
/da/	43.44
/ta/	42.81
/ra/	41.88

9 stimuli, 5 tokens and 4 speakers. The stimuli were presented in a balanced random order. The subjects were not allowed to listen to the reference sentences during the test session. They took a break after 90 trials.

3. Results and discussion

The results of the test session are shown in Table 2. Each monosyllable was evaluated 320 times, that is, 5 tokens, 4 speakers and 16 subjects. We can see that all the stimuli in this study gained scores of more than chance level (25%). As was in the experiment conducted with familiar listeners [6], the nasals /na/ and /nja/ obtained high scores, and the bilabial nasal /ma/ ranked in a middle position.

Figure 1 shows the results of the test session in comparison with the results of the previous experiment [6] conducted with familiar listeners. The stimuli are placed in ranking order of this study. The results of the previous experiment shown here are recalculated for the 4 speakers whose speech materials were used in this study.

As can be seen, the tendencies among the stimulus syllables were similar. Analysis of variance revealed significant main effects of both speaker familiarity (F (1, 17) = 803.4, $p < 0.001$) and stimulus content (F (8, 17) = 10.9, $p < 0.001$), with overall values in the familiar-listener test (Mean 75.2%, $S.E. = 0.019$) being higher than those in the naive-listener test (Mean 46.6%, $S.E. = 0.014$), and with the result for nasal /na/ being significantly higher than those for any other syllables and the result for /nja/ being significantly higher than those for /da/, /ta/ and /ra/.

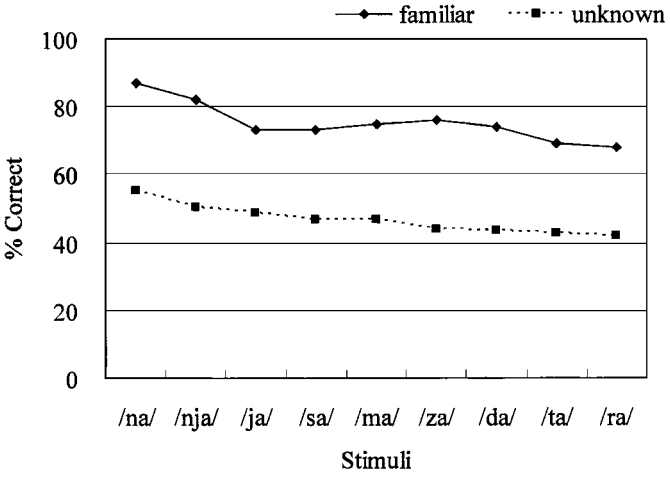

Fig. 1 Results of speaker identification test conducted with naive listeners in comparison with the previous test conducted with familiar listeners [5,6]: percentages of correct identification ($N = 100$ for familiar listeners and $N = 320$ for naive listeners). Data for familiar-speaker identification [5,6] are recalculated for the 4 speakers involved in this study. From left to right, the stimuli are placed in ranking order of the results of the unknown-speaker identification.

As pointed out in a previous study [12], speaker familiarity has a great effect on identification performance. The listeners in the previous experiment [6] knew the speakers very well; they had lived in the same dormitory as the speakers for at least 4 years. The results of this study in comparison with the results of [6] show that speakers are identified more accurately when the listeners have known them for a long time and when the listeners have had more chance to hear their utterances.

The difference in speaker identification performance of familiar and naive listeners may be explained by different cognitive processes. It is suggested that the identification of familiar speakers depends on pattern recognition, whereas that of unknown speakers is more based on feature analysis [8,9]. Because feature analysis is more difficult to execute, high familiarity generally yields more accurate identification [11], although there are differences among listeners in the ability to identify people by speech sounds alone [3]. Another study [17] that investigated the relationship between perceptual speaker identification and acoustic features indicated that the contributions of the spectral properties on speaker identification performance were greater in familiar-speaker identification than in unknown-speaker identification.

In spite of these differences, the nasals /na/ and /nja/ were found to be effective for both familiar- and unknown-speaker identification. This implies that nasality is one of the important features that represent speaker identity. It was confirmed that, regardless of familiarity with the speakers, listeners use nasal features contained in utterances in order to recognise speakers.

In this study, we compared the results of the identification of 4 out of 10 familiar speakers and of 4 unknown speakers. However, even if the speech materials used in the experiments are identical, the results will be different when the speaker

ensembles are different. As mentioned in a previous study [19] (reviewed in [3]), speaker identification depends not only on the individual characteristics of each of the speakers, but also on the characteristics of the other speakers with whom they are being compared. We drafted 4 speakers who had similar fundamental frequencies, and this should have had an effect on the difficulty of the identification task. We should examine the effects of the difference in speaker ensembles in our next study.

The final goal of this study is to explain the mechanism of speaker identification, and to understand its relationship to the linguistic aspects of speech sounds. Our next task is to verify the potentiality of the nasal sounds that indicate speaker's individual characteristics. We should investigate the nasality feature in other phonetic circumstances, such as in the coda position, with other vowels and nasalised vowels themselves, and under other conditions, such as at different fundamental frequencies and in different speaker ensembles.

In addition, the identification score differences among the coronal nasals, /na/ and /nja/, and the bilabial nasal, /ma/, is still unexplained. This tendency was also observed in our previous studies [6,7], and the same tendency was true with the oral sounds, the result for /da/ being higher than that for /ba/ [7]. Phonetic and/or acoustical explanation will be necessary for this.

4. Summary

In this study, we conducted a perceptual speaker identification experiment in order to examine the effects of speaker-listener familiarity and of the stimulus content. We used the same materials as those used in our previous study [6], where familiar listeners identified the speakers.

The results showed that familiar listeners performed significantly better than naive listeners; however, the overall effects of the stimulus content were similar between familiar and naive listeners. The nasals /na/ and /nja/ were particularly effective for speaker identification, and the identification score differences among the coronal nasals and the labial nasal was again observed in this study.

Acknowledgement

This work was supported by a Grant-in-Aid for JSPS Fellows (17·6901).

References

[1] L. Nygaard, "Perceptual integration of linguistic and non-linguistic properties of speech," in *The Handbook of Speech Perception*, D. Pisoni and R. Remez, Eds., (Blackwell Publishing, Oxford, 2005), pp. 390–413.

[2] J. Goggin, C. Thompson, G. Strube and L. Simental, "The role of language familiarity in voice identification," *Mem. Cognit.*, **19**, 448–458 (1991).

[3] P. Bricker and S. Pruzansky, "Speaker recognition," in *Experimental Phonetics*, N. Lass, Ed. (Academic Press, London, 1976), pp. 295–326.

[4] D. O'Shaughnessy, *Speech Communications: Human and Machine* (Addison-Wesley Publishing Company, New York, 2000).

[5] K. Amino, T. Sugawara and T. Arai, "Correspondences between the perception of the speaker individualities contained in speech sounds and their acoustic properties," *Proc. Interspeech*, pp. 2025–2028 (2005).

[6] K. Amino, T. Sugawara and T. Arai, "Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties," *Acoust. Sci. & Tech.*, **27**, 233–235 (2006).

[7] K. Amino, T. Sugawara and T. Arai, "Effects of the syllable structure on perceptual speaker identification," *IEICE Tech. Rep.*, **105**, 109–114 (2006).

[8] D. Van Lacker, J. Kreiman and K. Emmorey, "Familiar voice recognition: Patterns and parameters. Part 1: Recognition of backward voices," *J. Phonet.*, **13**, 19–38 (1985).

[9] D. Van Lacker, J. Kreiman and T. D. Wickens, "Familiar voice recognition: Patterns and parameters. Part 2: Recognition of rate-altered voices," *J. Phonet.*, **13**, 39–52 (1985).

[10] A. Schmidt-Nielsen and K. R. Stern, "Identification of known voices as a function of familiarity and narrow-band coding," *J. Acoust. Soc. Am.*, **77**, 658–663 (1985).

[11] A. D. Yarmey, A. L. Yarmey, M. J. Yarmey and L. Parliament, "Commonsense beliefs and the identification of familiar voices," *Appl. Cognit. Psychol.*, **15**, 283–299 (2001).

[12] H. Hollien, *The Acoustics of Crime* (Plenum, New York, 1990).

[13] I. Pollack, J. M. Pickett and W. H. Sumby, "On the identification of speakers by voice," *J. Acoust. Soc. Am.*, **26**, 403–406 (1954).

[14] K. Stevens, C. Williams, J. Carbonell and B. Woods, "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material," *J. Acoust. Soc. Am.*, **44**, 1596–1607 (1968).

[15] T. Kitamura and P. Mokhtari, "Effects of intra-speaker variation of speech sounds on perception of speaker characteristics," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 525–526 (2005).

[16] R. O. Coleman, "Speaker identification in the absence of inter-subject differences in glottal source characteristics," *J. Acoust. Soc. Am.*, **53**, 1741–1743 (1973).

[17] M. Hashimoto, S. Kitagawa and N. Higuchi, "Quantitative analysis of acoustic features affecting speaker identification," *J. Acoust. Soc. Jpn.* (*J*), **54**, 169–178 (1998).

[18] M. Owren and G. Cardillo, "The relative roles of vowels and consonants in discriminating talker identity versus word meaning," *J. Acoust. Soc. Am.*, **119**, 1727–1739 (2006).

[19] C. Williams, "The effects of selected factors on the aural identification of speakers," *Air Force Systems Command*, Electronics Systems Division, ESD-TDR-65-153 (1964).