TALKER IDENTIFICATION

USING NARROW-BAND SPEECH ENVELOPE CORRELATIONS

PACS REFERENCE: 43.72.Fx, 43.70.Gr, 43.72.Pf

Name/s of the author/s KAZAMA, Michiko[(1)]; TOHYAMA, Mikio[(2)]; YAMASAKI, Yoshio[(1)]
Institution/s [(1)]Waseda University; [(2)]Kogakuin University
Address/es [(1)]3-4-1, Okubo, Shinjuku-ku; [(2)]2655-1, Nakano-machi, Hachioji-shi
Town Tokyo
Country JAPAN
Tel: [(2)]+81-426-22-9291(3422)
Fax:
E-mail: [(1)]mich@ann.hi-ho.ne.jp

ABSTRACT

A talker-identification experiment utilizing an inter-frequency-band envelope-correlation matrix as a talker's signature-database was made. The envelope-correlation matrix was formed by averaging over 10 Japanese sentences, from each of which the envelope was obtained on a decibel scale for every 1/8-octave band between 2,000 to 11,313 Hz. The envelopes were extracted by using half-wave-rectifying and low-pass filtering processes. The identification scheme is based on a comparison of the correlation matrix created from a single test sentenc e for a subject to be identified by a signature database. The experimental results showed that 95% of 20 talkers could be identified and 100% was possible if second-best fits were included.

INTRODUCTION

This article describes an experiment on talker identification using narrow-band speech envelopes for high-frequency speech components. It is known that important speech information is conveyed on narrow-band envelopes. Drullman[1] reported that an intelligible speech signal could be synthesized by modulating 24 1/4-octave-band noise signals in the 100-6400-Hz range by the temporal speech envelope in the corresponding 1/4-octave band. Kazama et al.[2] used speech-signals synthesized from phase-only information to demonstrate that the narrow-band speech envelope is a significant cue for evaluating speech intelligibility. These studies have shown that the inter-frequency band relationship in the envelopes might provide a talker's individual voice-information.

Bimbot et al.[3] found that an inter-frequency covariance matrix for temporal changes in the short-term power spectrum in the 4-8-kHz band, which could be interpreted as narrow-band power envelopes, contains the talker's information. Hanson et al.[4] reported that gender differences can be detected in the relationship between the amplitudes of the first and third formants. They then used a noise judgment based on the third formant obtained by a filter with a band-width of 600 Hz around the third-formant frequency region. They found that the speech waveform and the third formant indicated the presence of a second glottal excitation within a glottal period for some male speakers.

Following the investigations above, the authors assumed that speech-signal components above the third formant frequencies contain the speech characteristics of talkers independent of spoken sentences. In the current study, we performed a talker-identification based on a simple algorithm using inter-band envelope correlation matrix for higher frequency bands than 2000 Hz.

METHOD

Test Materials

We used 20 native Japanese speakers, (ten female and ten male between 20 and 22 years old. Eleven Japanese sentences spoken by each subject were recorded in an anechoic room and digitized at a sampling rate of 48 kHz by a 16-bit A/D converter. Ten sentences from the recorded samples were used to make the envelope-correlation matrix to be referred to, and one single sentence was utilized for an identification test (described in the following sections). The ten sentences were the same for all the subjects; however, the single sentence used for the identification test, such as each one's individual name, was different for each subject. Every sentence is easily readable in daily use, and lasts about two seconds.

Signal Processing For Getting The Envelope-Correlation Matrix To Be Referred to

All the sentences were analyzed as follows. The speech signal is divided into 21 1/8-octave bands by using a filter bank (fourth-order Butterworth, Mat-Lab signal-tool box "butter") between 2,000 and 11,331 Hz. The envelope of each 1/8-octave band signal was obtained through a low-pass filter with a cut-off frequency of 40 Hz after a half-wave rectifying process. We took the decibel level of each envelope, including the silence (only very-low-level noise) portions, in the sentence. Then, the inter-band correlation matrix (21.21) of the envelopes for each talker was derived by averaging over ten sentences. This envelope-correlation matrix was used as the database of talker-identity to be referred to.

Procedure For talker identification

First, we make the envelope-correlation matrix of a talker to be identified but using only a single test sentence for the identification test. Second, the correlation coefficient between the two correlation matrices, the matrix for the test and the one from the talker-identity database, is used as a measure for identifying the talker. Then, we select the identified talker from the database so that the correlation coefficient becomes the highest.

Table 1 example of inter-band correlation matrix

		1/8 octave band center frequency (Hz)																				
		2000	2181	2378	2593	2828	3084	3363	3668	4000	4362	4756	5187	5656	6168	6727	7336	8000	8724	9513	10374	11313
1/8 octave band center frequency (Hz)	2000	1	.94	.90	.88	.84	.75	.75	.79	.71	.60	.49	.43	.35	.31	.44	.45	.37	.37	.50	.48	.45
	2181	.94	1	.95	.91	.89	.82	.81	.79	.69	.62	.49	.38	.31	.27	.40	.41	.33	.32	.45	.46	.41
	2378	.90	.95	1	.92	.87	.82	.82	.76	.65	.60	.44	.33	.26	.23	.36	.38	.27	.27	.40	.40	.34
	2593	.88	.91	.92	1	.90	.83	.85	.83	.71	.63	.49	.37	.32	.27	.44	.43	.28	.29	.42	.42	.34
	2828	.84	.89	.87	.90	1	.90	.86	.85	.76	.73	.62	.49	.41	.37	.50	.53	.38	.38	.51	.52	.46
	3084	.75	.82	.82	.83	.90	1	.94	.89	.82	.75	.64	.52	.49	.46	.60	.59	.46	.46	.55	.57	.54
	3363	.75	.81	.82	.85	.86	.94	1	.92	.83	.80	.65	.52	.51	.46	.60	.60	.47	.45	.55	.57	.54
	3668	.79	.79	.76	.83	.85	.89	.92	1	.93	.85	.76	.66	.63	.57	.70	.70	.59	.60	.69	.70	.64
	4000	.71	.69	.65	.71	.76	.82	.83	.93	1	.88	.84	.78	.76	.69	.79	.80	.73	.74	.81	.81	.76
	4362	.60	.62	.60	.63	.73	.75	.80	.85	.88	1	.91	.81	.76	.70	.74	.83	.73	.70	.79	.80	.77
	4756	.49	.49	.44	.49	.62	.64	.65	.76	.84	.91	1	.91	.87	.77	.80	.88	.80	.76	.84	.84	.79
	5187	.43	.38	.33	.37	.49	.52	.52	.66	.78	.81	.91	1	.94	.90	.89	.93	.88	.88	.93	.88	.87
	5656	.35	.31	.26	.32	.41	.49	.51	.63	.76	.76	.87	.94	1	.88	.90	.92	.89	.89	.90	.90	.86
	6168	.31	.27	.23	.27	.37	.46	.46	.57	.69	.70	.77	.90	.88	1	.94	.92	.90	.90	.87	.83	.84
	6727	.44	.40	.36	.44	.50	.60	.60	.70	.79	.74	.80	.89	.90	.94	1	.93	.86	.88	.88	.84	.83
	7336	.45	.41	.38	.43	.53	.59	.60	.70	.80	.83	.88	.93	.92	.92	.93	1	.92	.89	.92	.90	.88
	8000	.37	.33	.27	.28	.38	.46	.47	.59	.73	.73	.80	.88	.89	.90	.86	.92	1	.96	.93	.93	.93
	8724	.37	.32	.27	.29	.38	.46	.45	.60	.74	.70	.76	.88	.89	.90	.88	.89	.96	1	.94	.92	.91
	9513	.50	.45	.40	.42	.51	.55	.55	.69	.81	.79	.84	.93	.90	.87	.88	.92	.93	.94	1	.95	.93
	10374	.48	.46	.40	.42	.52	.57	.57	.70	.81	.80	.84	.88	.90	.83	.84	.90	.93	.92	.95	1	.94
	11313	.45	.41	.34	.34	.46	.54	.54	.64	.76	.77	.79	.87	.86	.84	.83	.88	.93	.91	.93	.94	1

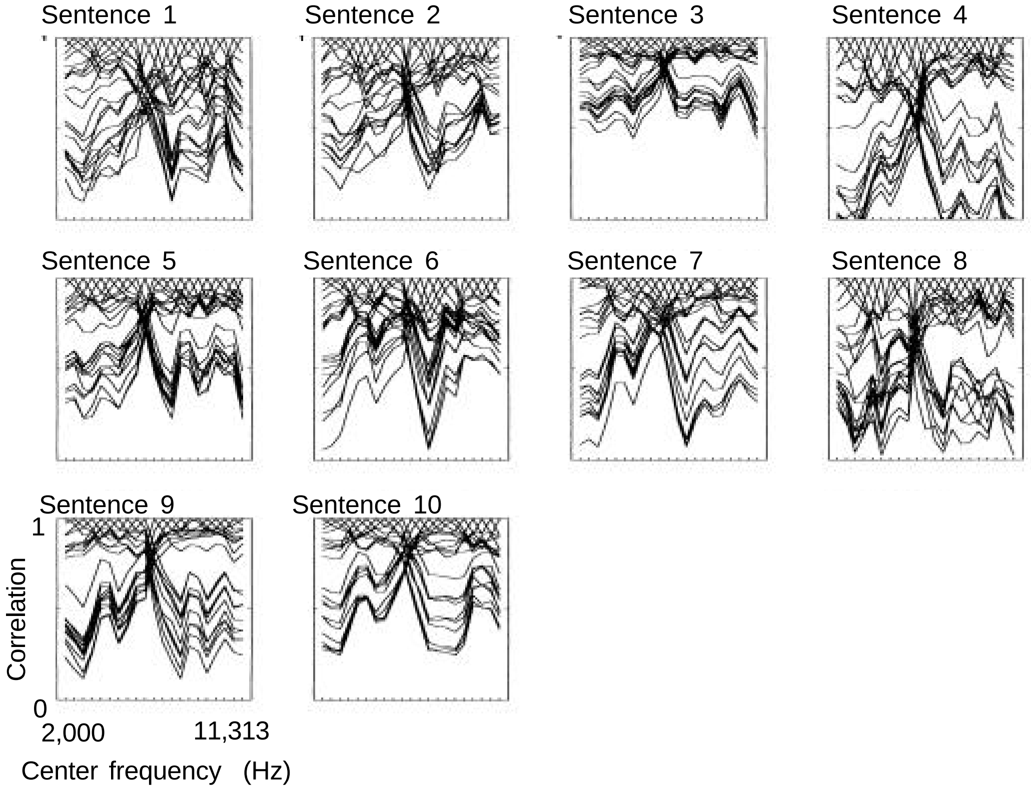

Fig 1. example of inter-band correlation matrix for subject A

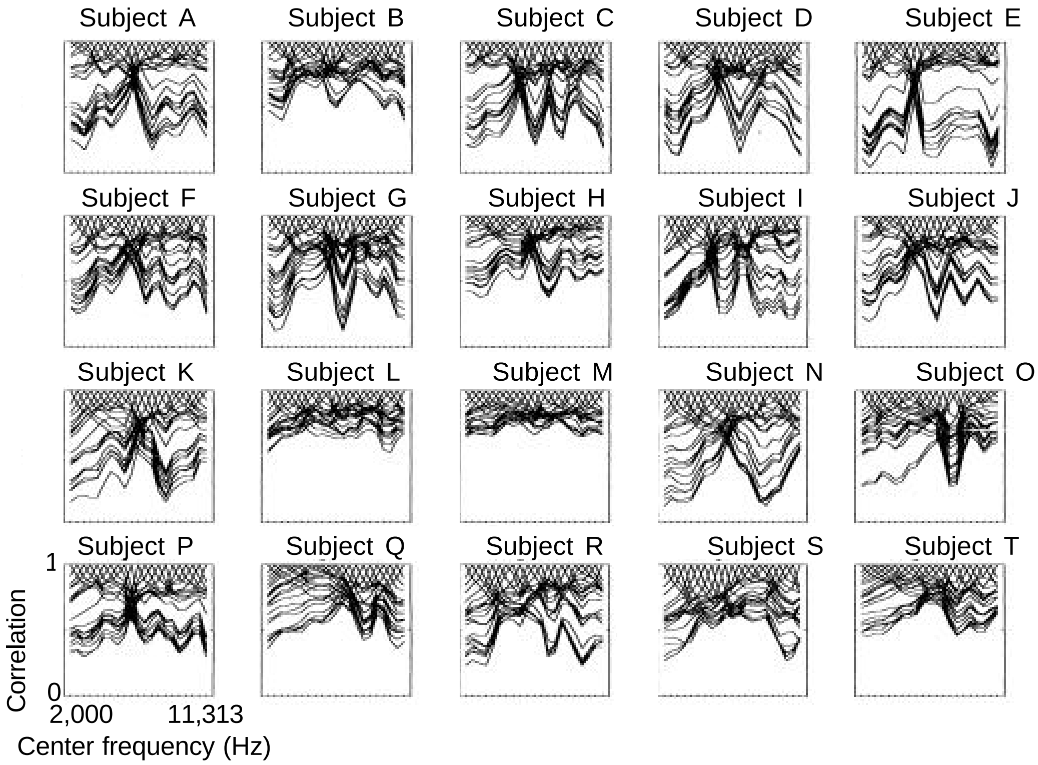

Fig. 2 Reference inter-band correlation matrix

Results

Table 1 shows an example of the inter-band correlation matrix obtained by using sentence 1 for subject A. Figure 1 shows the correlation matrix using all the sentences (1-10) for subject A. Sentence 1 in Fig. 1 shows all the entries of Table 1. The variations in the correlation matrices are due to the sentences. Figure 2 shows the correlation matrices used as the talker-identification database for 20 subjects. All the correlation matrices were obtained by averaging the 10 sentences in Fig. 2. It is clear that the variances in Fig. 2 are greater than those for Fig. 1. This variance makes the talker-identification possible.

Figure 3 shows the results of the talker-identification experiments. The solid vertical line shows the subject to be identified and the circle gives the identified talker-candidate whose reference matrix produces the highest correlation coefficient with the test matrix of the subject to be identified. If the circle is located on the solid line, the talker can be correctly identified. In the experiment, 95% of talkers could be identified and 100% could be if the second-best candidate was included.

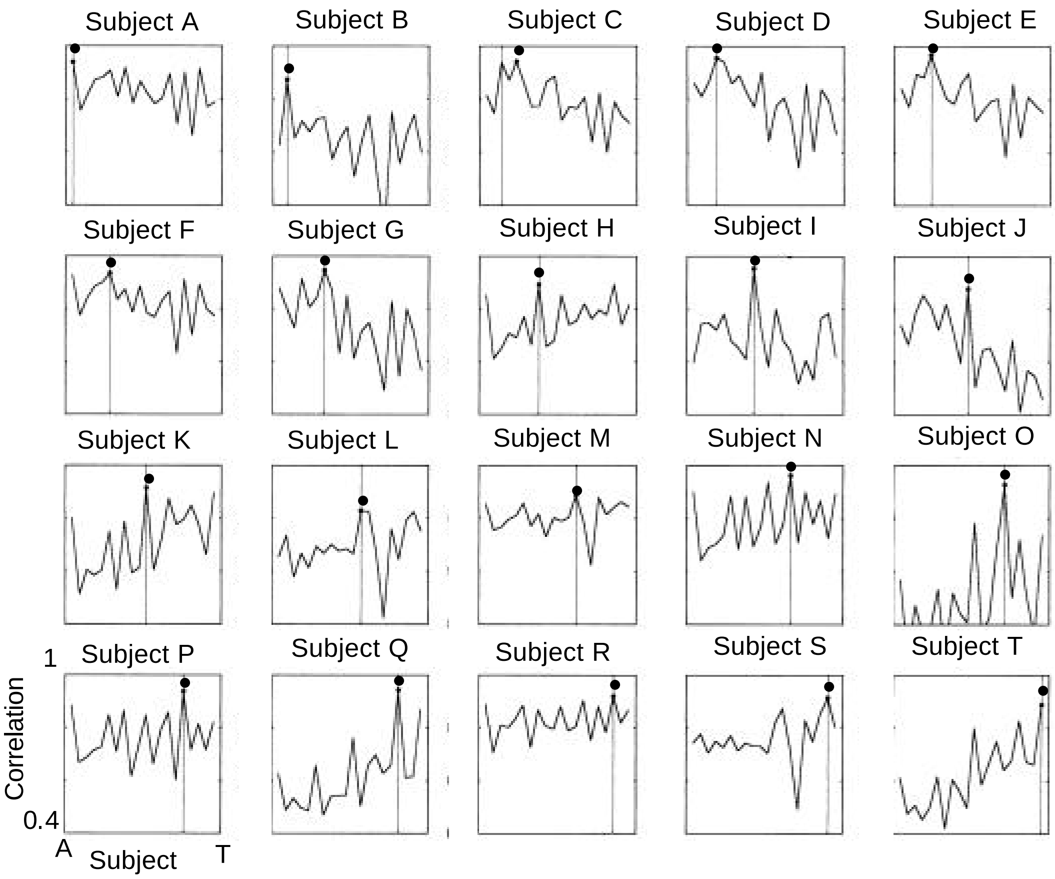

Fig. 3 Text-independent speaker identification

Subject A-J: male speaker

Subject K-T: female speaker

Line: correct identification

● : result

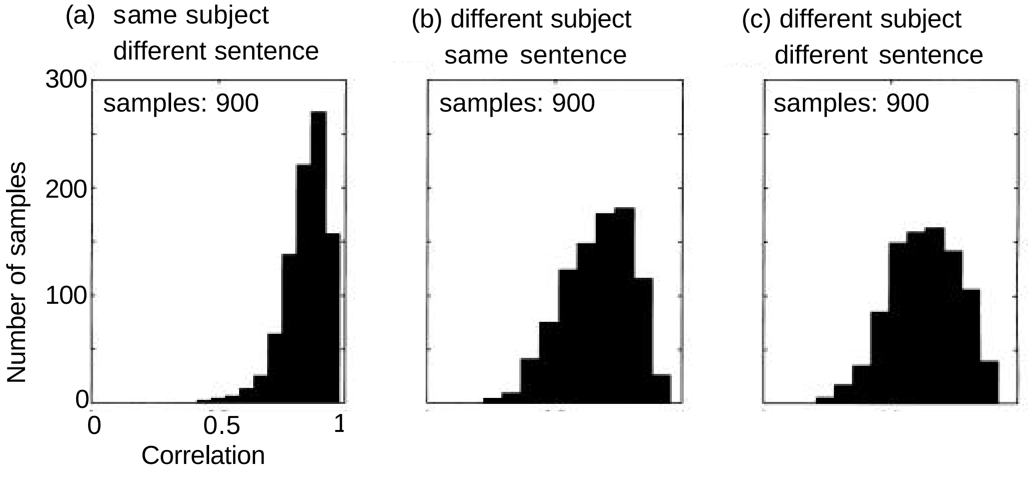

Fig. 4 Histograms of the correlation between two sentences inter-band correlation matrix

Figure 4 shows an example of analysis of identification errors. We made both the reference and test correlation matrices by a single sentence. Figure 4 presents the distributions of the correlation coefficients between the two matrices. Figures 4(a) shows the case for 900 different-sentence pairs with the same subject, Fig. 4(b) shows the distributions for the same sentence with 900 different subject pairs, and Fig. 4(c) shows the results for 900 pairs with different sentences and subjects. We confirmed high correlations for the same subject pairs even when the sentences are different. On the other hand, the correlations are low for the different-talkers pairs with the same sentence pairs. However, the variance in the case of Fig. 4(a) must be reduced in order to get high identification scores.

SUMMARY

A talker-identification experiment was performed under the assumption that talkers individual voice signatures independent of sentences are contained in the inter-frequency relationship between envelopes in the frequency band higher than 2000 Hz. The experimental results confirmed 95% of 20 talkers could be identified by using correlation matrices for the narrow-band envelopes in every 1/8-octave-band between 2,000 and 11,313 Hz.

ACKNOWLEDGEMENTS

The authors would greatly thank Prof. Tammo Houtgast (Amsterdam Free University, The Netherlands) for his valuable discussions, suggestions, and able guidance. This research was partly supported by TAO Fellowship, Japan.

REFERENCES

[1] Drullman, R., 1995. Temporal envelope and fine structure cues for speech intelligibility. J. Acoust. Soc. Am. 97(1), pp. 585-592.

[2] Kazama, M., Tohyama, m. and Houtgast, T., Speech reconstruction by using only its magnitude spectrum or only its phase, 17th ICA, 2001, 7P.51

[3] Bimbot, F., I. Magrin-Chagnolleau and L. Mathan, 1995. Second-Order statistical measures for text-independent speaker identification. Speech Communication, 17(1995), pp. 177-192.

[4] Hanson, Helen M., Chuang, Erika S., 1999, Glottal characteristics of male speakers: Acoustic correlates and comparison with female data, J. Acoust. Soc. Am. 106(2), pp. 1064-1077.