These characteristics are assumed to be different from those used in "we usually use." These characteristics are assumed to be different from those of the auditory pathway represented by a spatio-temporal masking pattern and to clarify adaptation, saturation, combination tones generation, masking from the physical spectrum. Also, the remarkable abilities of speech sounds are assumed to be performed using these spectral pattern: speech onsets and the formant structure, in particular, whole period of each masker. Spatio-temporal masking patterns thus as maskees. simultaneous and temporal masking are measured for the masking patterns. These spectral emphases in the auditory pathway are signal processing by computers.

INTRODUCTION
Spectrum analysis in the human auditory system is performed by cochlear function and neural network processing. These characteristics are assumed to be different from those of speech analysis techniques based upon digital signal processing which we usually use. A number of psychoacoustical and neuro-physiological studies have been carried out to date to obtain knowledge on this auditory spectral analysis characteristics[1,2,3]. These studies indicate that the auditory system has its own signal processing functions such as, critical band filtering, lateral inhibition, adaptation, synchronization, combination tones generation, masking, and so on. Therefore, the inter-aural spectrum, i.e. sound spectrum representation in the auditory pathway, is different from the physical spectrum. Also, the remarkable abilities of the human auditory system to detect, separate, and recognize sound spectra are assumed to be performed by using these inter-aural spectral input data for higher level signal processing. Therefore, inter-aural spectrum is superior to the physical spectrum representation when discussing perceptual costs of speech sounds.

From this standpoint, recent efforts have been made to investigate how the time axis, inverse of the masker, affected the experimental procedure remained the same throughout the experiments. Measurements were repeated at least 3 times for each subject. Two well trained male subjects participated in the experiments. The measurements were repeated at least 3 times for each subject.

RESULTS
The sound spectrum and speech waveform of the monosyllabic masker /de/ are shown in Fig.2(a) and (b), respectively. In Fig.2(c), the spatio-temporal masking pattern is depicted. The sound spectrogram and speech waveform of the monosyllabic masker /de/ as a function of time. When compared with the power spectral pattern, three distinctive characteristics are observed in the masking pattern. (1) Masking does not take the value of 0 dB at the time before the beginning (t = -25 msec.) and after the end (t = 400 msec.) of the masker sound. (2) Masking pattern increases remarkably at speech onset (t = 25 msec.) and at the transitional part of the formant. (3) Masking value decreases gradually in the vowel part. These characteristics were commonly observed in each masking pattern measured with respect to other monosyllabic maskers.

The sound spectrum and speech waveform of the continuous speech masker are shown in Fig.5(a) and (b). A spatio-temporal masking pattern measured every 25 msec. for this masker is depicted. The fine spectral structures of the masker sound, in particular, the first formant transition (e.g. at t = 175 to 300 msec.), and the vowel formant structure (e.g. at t = 175 to 300 msec.), are clearly observed in the masking pattern. Three experiments were carried out. Two well trained male subjects participated in the experiments. The measurements were repeated at least 3 times for each subject.

Fig.1 Experimental setups and the time chart of the stimuli. Both masker and masker signal are D/A converted (20kHz, 12bit), low-pass-filtered (3kHz, 96dB/oct), individually attenuated, mixed together, then presented to a subject monauraly.

Fig.2 (a) Wide band spectrogram. (b) speech waveform and (c) the spatio-temporal masking pattern measured at the lower frequency region. Time in msec.

Fig.3 The measuring masks and the time chart of the stimuli. Both masker sound and masker signal are D/A converted (20kHz, 12bit), low-pass-filtered (3kHz, 96dB/oct), individually attenuated, mixed together, then presented to a subject monauraly.

Fig.4 Masking patterns and 1/3 octave band power spectral patterns for /de/ as a function of time at three frequency bands (a) 0.006kHz, (b) 0.10kHz and (c) 0.2kHz.

Fig.5 The sound spectrum and speech waveform of the continuous speech masker are shown in Fig.5(a) and (b). A spatio-temporal masking pattern measured every 25 msec. for this masker is depicted in Fig.5(c). Formant structures and formant transitions are clearly represented in Fig.5(a) and well shown in Fig.5(c).

Fig.6 The sound spectrum and speech waveform of the continuous speech masker are shown in Fig.5(a) and (b). A spatio-temporal masking pattern measured every 25 msec. for this masker is depicted in Fig.5(c). Formant structures and formant transitions are clearly represented in Fig.5(a) and well shown in Fig.5(c).
Physical to inter-aural spectrum transformation. 3) The direction of AMFM component movements in speech sounds is of great importance in producing the inter-aural spectrum pattern. 4) Taking into account the considerable differences between inter-aural and physical representation of speech spectrum, the inter-aural spectrum can be implemented as a better representation of speech spectrum in feature extraction and speech signal processing by computers, particularly in automatic speech recognition systems.

ACKNOWLEDGEMENTS
This work was mainly done at NTT Integrated Communications Laboratories. The author is grateful to Dr. Ryuhei Nakatsu for his continuing guidance and encouragement, and to Dr. Noboru Sugamura for his fruitful discussions. The author also wishes to thank Dr. Yutaka Tokuhara for his helpful advice in writing this paper.

References