PITCH-SYNCHRONOUS RESPONSE OF CAT COCHLEAR NERVE FIBERS TO SPEECH SOUNDS

Toru Hashimoto, Yoshifumi Katayama*, Keiichi Murata*, and Ikuo Taniguchi*

Institute for Medical and Dental Engineering, Tokyo Medical and Dental University, Kanda, Chiyoda-ku, Tokyo 101, Japan
*Department of Neurophysiology, Medical Research Institute, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo 113, Japan

Abstract Responses of individual fibers of the cat cochlear nerve to Japanese monosyllabic speech sounds were described. Responses were analyzed by making post-stimulus time histograms and "pitch-synchronous" period histograms. Almost all neurons responsive to speech sounds showed repetitive discharges synchronous with the pitch period irrespective of their characteristic frequencies (CF). The frequency relation between the tuning characteristics of single neurons and the spectral distribution of the stimulus was systematically investigated to find out stimulus parameters to produce the "pitch-synchronous" impulse discharges. It was found that the neurons responded to vowels if the latter contained formants with frequencies near the neurons' CFs.

The aim of the authors is to investigate the response characteristics of single neurons at lower levels of the auditory pathway of various animals to human speech sounds and to disclose the neural mechanism of feature extraction preceding the recognition of the acoustic patterns from the comparative physiological point of view. The present report describes responses of individual fibers of the cat cochlear nerve. The temporal characteristics of the response to speech sounds were systematically investigated and the distribution of responsive neurons in the cochlear nerve was analyzed. The primary process of the feature extraction was discussed on the basis of the experimental result. A plausible system of "pitch-synchronous" analysis at the higher stations of the auditory pathway was also suggested. The outline of the present research was preliminarily reported (Hashimoto et al., 1973).

Received for publication May 6, 1975
METHODS

Experiments were carried out on 17 adult cats. Each animal was anesthetized with pentobarbital sodium (30 mg/kg) given intraperitoneally. A constant level of anesthesia was maintained by a frequent administration of a small dose of the same drug through the cannulated right femoral vein. The trachea was cannulated for artificial respiration. The scalp was opened caudolaterally and a portion of the cerebellum was displaced or carefully sucked out to expose the cochlear nerve under a binocular dissecting microscope.

Unitary discharges of the cochlear nerve fibers were recorded with metal microelectrodes (tungsten wire) which were electrolytically sharpened and insulated except for the fine tip. The electrode was inserted along the cochlear nerve with the aid of a micromanipulator under a dissecting microscope. The microelectrode was connected to a conventional high input impedance preamplifier. The impulse discharges were converted into a normalized pulse train through an amplitude discriminator and displayed on a CRO (Tektronix: Model 565) with the stimulus waveform. The pulse train was fed to a digital computer (DEC: LINC-8) for on-line data processings.

A free sound field was produced in a sound-quieted room by an electrostatic loudspeaker (Quad & Sony: 3120F) which was placed 1.0 m in front of the animal's head. A pure tone, either continuous or in burst, a combination of pure tones and a human speech sound were the stimuli used. The human speech sound was a monosyllabic Japanese vowel or a consonant followed by a vowel. A series of five separate monosyllabic speech sounds with an interval of 2 sec was recorded. It was repeatedly copied on endless magnetic tape, so that when reproduced, the series of speech sounds was generated at a rate of 4 per minute (TEAC: A-350). On copying, an onset mark of each voice sound was also recorded on another track of the endless magnetic tape. This mark was used to trigger a sweep for compilation of the post-stimulus time (PST) histogram of the neural response.

During an insertion of a microelectrode into the cochlear nerve a pure tone stimulus was used to search for sound-sensitive neurons. The frequency of the sound was sequentially changed by a voltage controlled generator (Wavetek: VCG 116). When a responding auditory unit was encountered, the tuning characteristic was automatically measured under computer control and displayed on a computer scope. The tuning characteristic of an individual fiber of the cochlear nerve was specified in terms of the characteristic frequency (CF).

In order to investigate temporal structures of the responses to voice stimulation PST histograms were made. The time intervals between the onset mark of a stimulus and responsive impulse discharges were measured by a crystal-controlled clock. For a precise analysis of the response a time bin of 200 μsec was chosen. For a rough measurement a time bin of 1 msec was preferred in order to save the computer memory. For obtaining one PST histogram the same stimulation was applied 100 times.
In order to make a precise analysis of the periodic activity of neurons in response to vowel stimulation pitch period histograms were calculated (Brugge et al., 1969). The pitch period was defined as a time interval of the repetition of an identical speech waveform (Mathews et al., 1961). A PST histogram was divided into segments with an interval equal to the pitch period with reference to the stimulus waveform. The tallies of corresponding bins in all the segments were added bin by bin to make up one “pitch-synchronous” period histogram.

The concept of the “pitch-synchronous” period histogram is based on the assumption that the sound wave of the vowel is a true periodic phenomenon. However, vowels are inherently quasiperiodic. Therefore preliminary analysis of the recorded vowels was made and it was ascertained that the variation of the pitch period and the error of segmentation were less than the bin width, 200 µsec.

The sound intensity of stimulation was measured near the animal’s ear. It was varied in a range below 84 dB SPL. It was expressed as a dB value above the threshold intensity for the frequency corresponding to the CF of each neuron.

Spectral analyses of the stimulus speech sounds were carried out with a sound spectrograph (Rion).

RESULTS

Responses to vowels

Fig. 1. PST histograms of the responses of a single unit of the cochlear nerve of a cat to Japanese vowels, /a/, /i/, /u/, /e/ and /o/. Unit’s CF: 6.6 kHz. The stimulus intensity: 70 dB. The bin width: 1 msec. In the histograms for /a/, /e/ and /o/ the periodic rise and fall of the firing probability are observed, occurring synchronously with the pitch period of the vowel. No response is indicated to /i/ nor /u/. A time histogram of spontaneous discharges is shown in the right lowest.
In Fig. 1 PST histograms to five kinds of Japanese vowels are shown for a single unit of the cochlear nerve. This unit had the CF at 6.6 kHz and responded remarkably to /a/, /e/ and /o/, while neither to /i/ nor to /u/. The difference of response patterns was possibly dependent upon how the neuron’s tuning characteristic covered the spectral distribution of the stimulus. In response to /a/, /e/ or /o/ sharp peaks and dips were produced in the histogram with a constant interval which coincided with the pitch period of the stimulus vowel. This means that the neuron tended to fire “pitch-synchronously”, i.e. each impulse was phase-locked to the fundamental frequency of the vowel. This type of a discharge pattern will be termed the “pitch-synchronous” response.

The discharge pattern of a neuron with the CF of 0.5 kHz is shown in Fig. 2. The unit responded clearly to /i/ and /u/ as well as to /a/, /e/ and /o/. In this case effective formants of all the vowels were located within the neuron’s tuning curve.

In Fig. 3 is shown a relation between the CF’s of single neurons and the spectrograms of five Japanese vowels used for stimulation. Since the sound spectrograph used in the present experiment had a limited frequency coverage, the illustrated spectrograms lack components higher than 4 kHz. The units (total number, 51) are plotted on the frequency axis of each spectrogram with a distinction according to whether they responded to a given vowel (closed circles) or not (open circles). The responsive units showed the discharge pattern of the “pitch-synchronous” response. One can see that the responsive units are clustered in
COCHLEAR NERVE RESPONSE TO SPEECH SOUNDS

Fig. 3. Responsiveness of the cat cochlear nerve fibers to Japanese vowels. Single units are plotted according to their CF’s on the frequency axis of the spectra of the stimulus sounds. Solid circles: responsive units. Open circles: unresponsive units. The sound intensity is about 50 dB SPL.

Fig. 4. PST histograms and “pitch-synchronous” period histograms of the responses of a single cochlear neuron to vowels. Unit’s CF: 0.3 kHz. The bin width: 200 μsec in both histograms. The intensity: 30 dB. The trace above each histogram is the stimulus waveform with the same time scale as that for the histogram. The “pitch-synchronous” period histograms of impulse discharges are shown for two pitch periods.
the frequency range where the vowels had potent components. It seems likely that whether units respond to a given vowel is determined solely by their tuning characteristics and the spectral distribution of the stimulus.

In assembling the units plotted in Fig. 3 the sound intensity was fixed at about 50 dB SPL. This intensity is the one which is used for ordinary conversation. At this moderate level the sound spectrograms were well represented by responsive neurons as in Fig. 3. The aggregate of responsive neurons is a duplicate of the short-time sound spectrogram of the stimulus. The "pitch-synchronous" excitation of cochlear nerve fibers is a basic feature of the response to voiced sounds.

**Time structure of responses to vowels**

Further details of time structures of the "pitch-synchronous" response were investigated with a finer time resolution of 200 μsec. In the left part of Fig. 4 PST histograms are shown for the vowel stimulation of a neuron with the CF of 0.3 kHz. The discharge pattern of the "pitch-synchronous" response was fairly simple. Single or dual peaks were repeated periodically. Neural firing occurred at a fixed phase in one pitch period. Figure 5 shows another example of the

![Fig. 5. PST histograms and "pitch-synchronous" period histograms of the responses of a single cochlear neuron. Unit's CF: 0.7 kHz. The sound intensity: 10 dB.](image-url)
neuronal responses to the same utterance as in Fig. 4. The CF was 0.7 kHz and about twice as high as that of the neuron in Fig. 4. PST histograms had many peaks and dips during one pitch period. The firing probability of the neuron rose more than twice in one pitch period.

The "pitch-synchronous" period histograms are shown in the right part of Figs. 4 and 5. The temporal distribution of the firing probability during two pitch periods is demonstrated. Impulse discharges were concentrated at specific phases of the pitch period. Comparing the histograms with the waveforms of the two pitch periods, it is seen that the profile of the firing probability is not always similar to the stimulus waveform.

In Fig. 6 the response pattern of the same unit as in Fig. 5 to the utterance by a different person is shown. The response pattern had an invariant property of the "pitch-synchronous" impulse discharges, but peaks to vowels /a/, /u/ and /e/ were less numerous and much broader in Fig. 6 than in Fig. 5. Differences of the response pattern between Figs. 5 and 6 would reflect a subtle difference in the spectral composition of voices between uttering individuals. This fact may imply that the response pattern of cochlear nerve fibers could play a role in identification and verification of the voice of individual persons.

Fig. 6. PST histograms and "pitch-synchronous" period histograms of the same neuron as that in Fig. 5 to a series of vowel stimuli of a different person's utterance. The stimulus intensity: 20 dB.
In the unit of Fig. 6 a comparison was made between responses to the original vowels and those deprived of high frequency components through a band-pass filter. The band-pass filter was tuned to 160 Hz. Figure 7 shows the response to the filtered /e/ and, in addition, the response to the original /e/ in Fig. 6 is presented for comparison. The sound spectrogram of the original /e/ is shown in the top part of Fig. 7. Through the band-pass filter the frequency components other than the fundamental frequency as shown by the leftmost bar in the spectrum were thoroughly eliminated. In the PST histograms as well as in the “pitch-synchronous” period histograms it is seen that the neuron’s response to the filtered /e/ was a single peak in a one pitch period and its height was much lower than in the response to the original /e/. From the experimental result shown in Fig. 7 it may be pointed out that the stimulus parameter to elicit the “pitch-synchronous” response pattern would not be the component of the fundamental frequency but the “pitch-synchronous” repetition of favorable formants for the neuron. The discharge pattern is the neuronal representation of the temporal change of effective formants for the neuron.

Responses to consonants

PST histograms of responses to Japanese consonants followed by vowels were investigated. In Fig. 8 are shown histograms to /ka/ and /ke/. In each case a marked response to the consonant /k/ is observed preceding the response to the vowel. In stimulations with /ki/, /ku/ and /ko/ the responses to /k/ were
not so clear as those to /ka/ and /ke/. This was probably because intensities of
the frequency components favorable for the neuron’s tuning characteristic were

**POST-STIMULUS TIME HISTOGRAMS**

**SOUND INTENSITY**

25 dB ABOVE THRESHOLD

Fig. 8. PST histograms of responses to Japanese consonant /k/ followed by vowels. The CF
of the neuron: 5.2 kHz. In the above two PST histograms the discharges from 0 to
around 30 msec are responses to /k/.

**CONTROL**

**NO STIMULUS**

CF 5.2 kHz

**POST-STIMULUS TIME HISTOGRAM**

**SOUND INTENSITY**

55 dB ABOVE THRESHOLD

Fig. 9. PST histogram of the response to a fricative consonant combined with a vowel,
/sa/. At the top is shown the stimulus waveform. The unit’s CF: 10 kHz. In the
lower figure there is an increase of impulse discharges in the time interval between 0 and
70 msec, which is the response to the consonant /s/.
different among the phonemes with the same /k/.

In Fig. 9 a PST histogram of the response to the fricative consonant /s/ combined with the vowel /a/ is shown. Since the fricative /s/ has a wide range of spectral components of higher frequencies, only units with high CF's were able to detect the high frequency spectral component of the stimulus. This was the case in Fig. 9. In the lower part of the figure an increase of impulse discharges in the time interval between 0 and 70 msec is the response to /s/. Units with low CF's did not respond to /s/ at all. The responses to other consonant-vowel combinations such as /n/, /m/, /r/ and /t/ followed by five vowels were investigated systematically. The responsiveness depended primarily upon the intensities of the frequency components of the short-time spectrum, which were extracted by the tuning characteristic of the neuron.

DISCUSSION

The frequency analysis of a sound stimulus was made using a travelling wave in the cochlear partition and the neuronal tuning characteristic along the auditory pathway (Békésy, 1960; Katsuki et al., 1958; Watanabe and Katsuki, 1974). The auditory neuron is assumed to be a tuning filter whose center frequency was the characteristic frequency (CF). Single units of the cochlear nerve showed a response pattern of repetitive discharges synchronous with the pitch period of the speech stimulus. The occurrence of the neural discharge at a given instant indicates the existence of a favorable spectral component for the neuron at that instant. The "pitch-synchronous" periodic discharges represent the periodic repetition of the spectral component in the voiced sound. This is due to the fundamental process of speech production. The aggregation of responsive neurons as shown in Fig. 3 has a close relation to the respective stimulus spectrogram. Therefore the spectral component near the CF of the neuron was more effective than the fundamental which is sensed by the low-frequency tail of the tuning curve. This is not consistent with the idea of Kiang and Moxon (1974).

Along the auditory pathway the "pitch-synchronous" neural information is transferred to higher stations, where the information might be integrated or processed possibly with a "pitch-synchronous" logic. The "pitch-synchronous" activity to speech sounds was observed at the cochlear nucleus and the inferior colliculus (Moore and Cashin, 1974; Watanabe and Sakai, 1973). The sharpening of the response to the sustained vowel in the temporal domain, which was suggested to be the characteristic of the secondary neuron by Moore and Cashin (1974), was also observed at the primary neuron of the cochlear nerve. The "pitch-synchronous" impulse discharges might be a basic feature of the auditory response to vowels. This will be inherently related to the acoustic characteristic of vocalization. The "pitch-synchronous" striation observed in the sound spectrogram (Fig. 3) is one of essential characteristics of the short-time spectrum of the vowel. Further study on the "pitch-synchronous" response is now in progress,
using much simpler synthetic vowel-like sounds. The key factor that elicits periodic discharges synchronous with the pitch period will be extracted in the near future from the study with synthetic vowels. The mathematical "pitch-synchronous" analysis of vowels was found to have a great utility and it was suggested that the "pitch-synchronous" representation of the sound spectrum was precisely related to the vocal characteristic of speech production (MATHEWS et al., 1961). Further speculation on the utility of the neuronal "pitch-synchronous" analysis of vowels in the higher centers awaits detailed analysis at various nuclei along the auditory pathway.

On the feature extraction of the consonant response further experimental studies will be necessary at the neuronal relay station. It is essential to elucidate how a few phasic discharges responding to transient and nonstationary stimuli such as consonants could be discriminated from the background spontaneous discharges and processed as a valid information.

The primary aim of the present study is to know the neural mechanism of the feature extraction of auditory signal. For this purpose we used the cats as experimental material. This may be justified from the comparative physiological viewpoint that the primary process of feature extraction in the lower auditory pathway might not have so much difference inherent in the animal species. The vocal communication may fully depend upon the evolutionary progress of the central nervous system of the animal species. The functioning of the primary and secondary neurons is not different among different animal species as far as conventional sound stimuli are concerned. The lower auditory pathway of the mammal will be a good living model of the human auditory system for the experimental analysis of the auditory feature extraction.

This research was supported by grants from the Ministry of Education of Japan. The authors wish to thank Dr. Yasuji Katsuki for his valuable comments and advice.

REFERENCES
