Auditory Neural Information Processing Modeling for Speech Recognition

음성인식을 위한 청각신경 정보처리 모델링

Hee-Kyu Lee*, Kwang-Hyung Lee**

(이 희 규, 이 광 형)

요 약

음성처리 및 인식기기의 기능을 향상시키기 위해서는 생체공학적인 방법을 이용한 인체의 청각신경 정보처리 시스템의 연구가 중요하다. 그래서 본 논문에서는 와우각의 메카니즘을 분석한 기저막의 IIR 디지털 필터 모델링이 연구되었다. 또한 음소검출필터와 특성 추출을 위한 변별기능을 이용한 자음인식의 다층신경 모델을 구성한다.

이 모델은 자음인식에 있어서 90% 이상의 높은 감지율을 나타내고 있다.

ABSTRACT

A neural auditory system is studied for the aim of making better speech recognition systems. The cochlear mechanics is described. A IIR digital filter modeling of basilar membrane is discussed for the speech recognition.

A multi-layer model of consonant recognition using phoneme detection filters and discriminant functions for feature estimation is constructed.

This model shows more then 90% recognition rate in consonants.

I. Introduction

It is impotant to study human auditory neural information processing system using the method of bionics for improving the function of speech processing and recognizer.

In this article we shall do this to explain how the information is carried and transformed on its way to the brain of the listener. We can discuss the recognizable function of frequency feature with different methods in auditory system.

Especially, we can build the modeling theory

*Bucheon Tech College Assistant prof.
부천공업전문대학 교수
**Sungsil Univ. Assosiate prof.
숭실대학교 교수

about the exact feature extration of required response in pitch discrimination.

Though it has any indistinctness to auditory nerve system consists of external ear, middle ear, basilar membrane in the cochlea and sensory cells nerve fibers, brain-stem centers and auditory cortex, the auditory neural information processing can take more quick effect than the original pitch discrimination. Also when we build the hardware, we are able to make target of hardware in the near A. I.

In this paper, we discussed the recognition function of the frequency feature of the basilar membrane. And, as we build the exact feature extraction modeling theory of the required sound level in pitch discrimination. We want to take method for designing of the advanced speech recognizer.

Ⅱ. Mechanism of feature extraction in auditory system.

The moving function of a cochlear duct shows that vibration sensed by eardrum, propagates at the fluid in the cochlea, when the oval window is pushed by a sound wave, the fluid in the cochlea moves, this vibration makes traveling wave at basilar membrane in cochlea, and detects sound feature parameter from the hair cell, transfers the central nerve, all this, processes are seperated into macro mechanics, micro mechanics and transduction about the cochlea moving characteristics. The macro mechanics describes the motion of the fluid and motion of the basilar membrane in the scali.

Micro mechanics shows the motion of the organ of the tectorial membrane. By transduction, it means a description of the inner hair cell response to basilar membrane hair cell synapse.

In the basilar membrane, distinctioin between the scala vestibuli and scala tympani, it's magnitude and solidity is increasing as it's onwarding to the inner. As mechanical impedance is different according to the position, the resonant frequency is the lower, the onwarder in the inner. So that, when the basilar membrane's partition vibrates in the near basilar membrane, the frequency becoming high sound and, vibrating partition nearing the top of cochlea, it becomes low sound (showen in fig. 1). By a large and small of amplitude moving in the basilar membrane, the sound amplitude is distinguished, high tone or low tone.

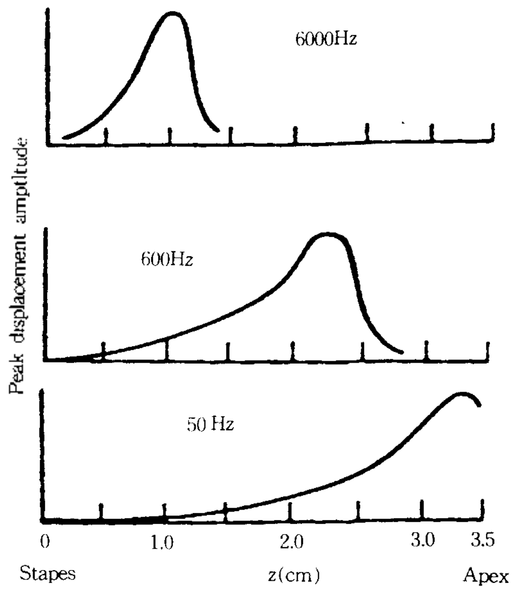


Fig. 1 Peak displacement amplitude of the basilar membrane for a pure tone.

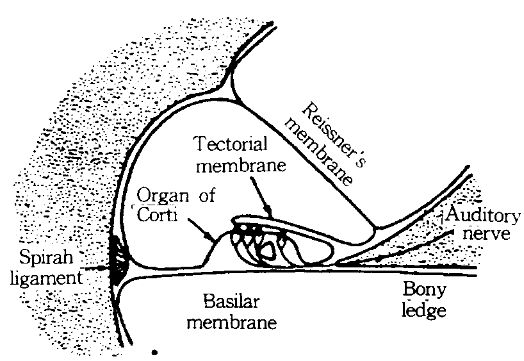


Fig. 2 Cross Section of the Cochlea duct.

1) The vibration phase is slowly decaded onwarding in the inner when it's amplitude is maximum. This decading tendancy is maximum.

The basilar membrane vibration following by the corti organ's vibration arranged in aline upon it. [1] There are some 30,000 hair cells in cochlea tube and hair cell vibrates with them by the 4 colum arrage in the basilar membrane. The edge of hair cell is touched the tectorial membrane of a unvibrated, and moving the basilar membrane followings by moving hair cell. When mechanical energy converts to electrical energy by the Vibration, the voltage of hair cell is rised. The cochlear duct (showen in fig. 2), defined as the space between Reissner's membrane and the basilar membrane, is at an 90 mV potential. This potential is improtant in the transduction process. It is important imformation impulse's generating frequency and timing. The voltage generated by hair cell accordings to the sound strength. Impulse generating operation synchronized a vibration phase of the basilar membrane and being frequency higher, impulse generating is disappeared. It is wonder that, in the over 500Hz, impulse generates not by sound's onwording, this generation is sequential. In voiced sound, impulse is by a peripheral nerve whether this sound is a voiced or not unvoiced.

Ⅲ. Basilar membrane modeling

Considering a steady and trensient vibration of the basilar membrane, frequency characteristic is shown by the modeling D.F(shown in fig. 3(a) [3] IIR D.F. having a characteristic of LPF is consisted in fig. 3(b) considered the envelop and traveling wave of basilar membrane.

Audiable frequency range is about 10 oct. of 20~20,000Hz, but the sound of less than 30 Hz is equal to the forward of the cochlear amplitude, so pattern frequency range of the basilar membrane vibration in this model is 39.0625Hz~ 20,000Hz, that is 9 oct so that model consists of 54 stage being dependent conjunction 6 stage per 1 oct.

The basilar membrane shows LPF characteristics from its vibration feature. Model of the basilar membrane is consists of dependent conjuctional D.F. as a fig.3(b). Here, higher frequency is sensored by the haircell, later is the attenuate by onward in the inner.[2] The other side, lower frequency becomes several times onwarding in the inner, and sensored in the edge of basilar membrane (shown in fig.4(b)).

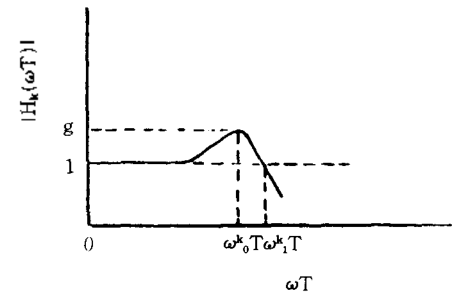


Fig. 3(A) Amplitude Characteristic of D.F.

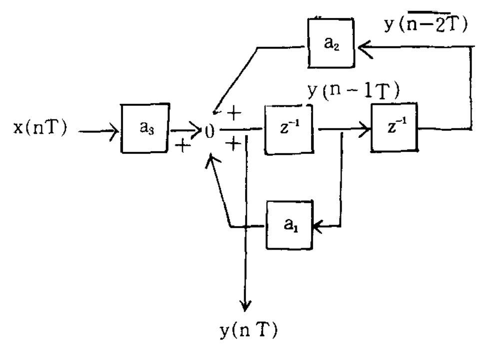


Fig. 3(B) Digital filter of 1 stage.

Ⅳ. Neural model for speech recognition using discriminant analysis.

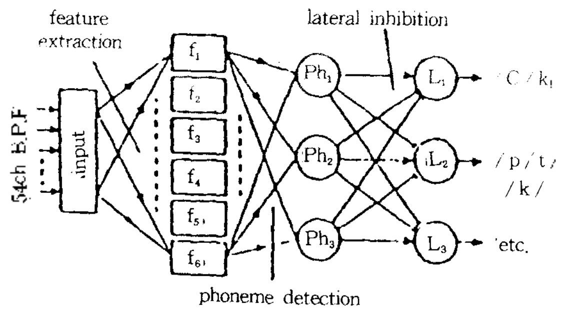


Fig. 4(A) 4-layer model of consonant recognition.

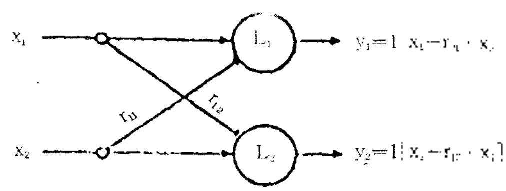


Fig. 4(C) basic model of lateral inhibition.

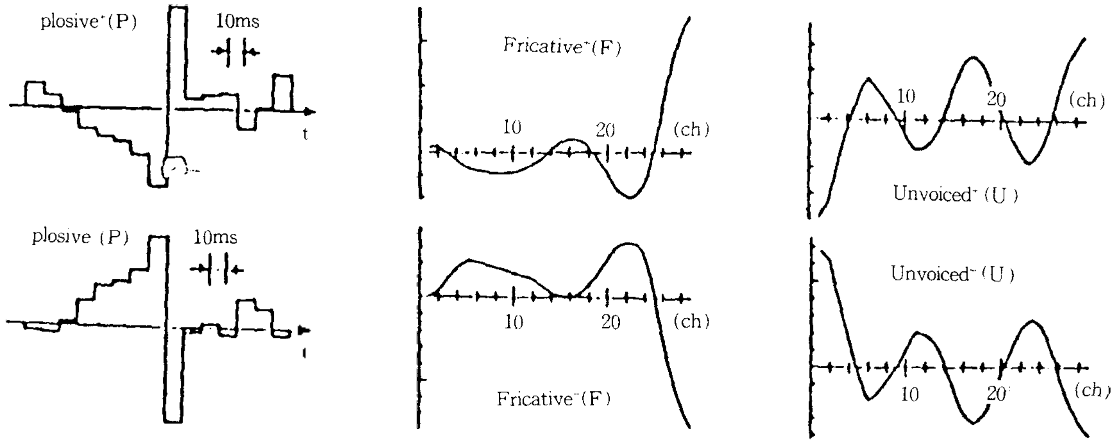


Fig. 4(B) discriminant functions for 6 features.

Fig.4(a) shows the multi-layer neural model for consonant recognition. The input speech signals are analized by using BPF bank and transformed into a few of features by discriminant coefficients found by perceptron learning, are computed, by the value of which the detection of specific phonemes or phonemic groups are performed. In this operation, the correlation modeling of the feature is included.

Conducting the output unification by the lateral inhibition between those detected outputs, the unique consonant recognition output are determined. The first input layer gets the frequency analysis outputs of the BPF banks due to cochlear basilar membrane operation.

The input layer computes the logarithm spectrum per unit frame(10ms). The second layer makes the feature extraction.

This layer estimates kinds of features and those complementary featurs. (shown in fig. 4(b)).

The third layer becomes the phoneme detection filter (shown in fig. 3(b)). In this layer, by performing the modeling of the temporary correlation or of the intensity between each features, the detection filter is constructed to truncate the phoneme from the continuous speech.

The fourth layer is for the output unification. As the phonemes detection filter are independently designed for each phoneme. The different phonemes are used to be detected simultaneously in a region.

In fig. 4(c), as we introduce the lateral inhibition scheme to the detected output of the third layer, the recognition output is unified and is gotten uniquely. Because these four layers are built step by step, this neural network model becomes the learning model.

V. Experiment results and discussion

Fig. 5 shows the confusion table resulted from the phoneme detection experiment for the learning data, 212 wards uttered by five audlt males and five females. In fig. 5, we can find high phoneme detection rates ; 93.1%, 94.0%, 98.4% and 90.4% for unvoiced affricate /c/, /ki/, unvoiced plosive /p/, /t/, /k/, unvoiced fricative /s/ and voiced fricative /z/ respectively.

in \ out	/c/ /ki/	/p/ /t/ /k/	/s/	/z/	total
/c/ /ki/	90.6%	1.6%	1.6%	2.8%	320
/p/ /t/ /k/	3.2	94.3	0.0	0.5	793
/s/	0.0	0.0	98.4	0.8	494
/z/	1.9	0.0	0.5	90.4	209
vowel	0.02	0.3	1.1	1.6	5,718
the others	0.3	1.9	1.6	3.3	2,962

Fig. 5. experiment results

VI. Conclusion

In this paper, we have presented the multi-layer auditory neural information modeling for consonant recognition. This model is considered to make the four layers the organization such as the count matching of phonology concept by linear transform and nonlinear transform responded to the property of that's concept. The results of recognition experiment marks the high point more than 90% for the four phoneme groups with very different propertys. Cochlear basilar membrane is modeled as cascaded IIR digital filters which constitute the BPF bank's frequency characteristics agreed with the biological data by Bekesy. The silicon cochlea implant as a hearing-aid is reported successfully operated by an ENT doctor in recent days.

REFERENCES

1. J. B. Allen, Cochlear modeling, JAN, 1985, IEEE, ASSP, Magazine pp. 3~29.
2. S. Seneff, "A computational model for the peripheral auditory system", Int Conf. Acoustic and Speech Signal Processing, Tokyo, 1986, pp. 1980~1986.
3. Lee, H. K. and Lee, K. H., "Frequency discriminative modeling of cochlear in neural auditory information processing", BMF conf. paper, KAIST, 1988. (Korea).
4. T, Kawabata at, al. "Feature extraction of phoneme based on the discriminant analysis", A-151, paper no, 591s 57 IEICE (Japanese)
5. G. Von, Bekesy, "Experiments in Hearing", Mc Graw-Hill, 1960.
6. L. R. Rabiner and B. Gold, "Theory and Application of Digital Signal Processing", Prentice-Hall, 1975, P218.
7. Lee, H. K. "A Digital Filter Modeling of Basilar Membrane in Neural Auditory Information Processing"MS. Sungsil univ. 1988.

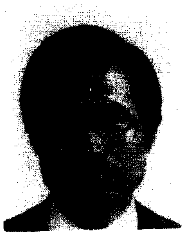
Hee Kyu Lee

1978. 2 : Department of Electronics Engineering, Inha University (B. S)

1981. 8 : Department of Industrial Engineering, Yonsei University (M. S)

1988. 2 : Department of Electronics Engineering, Sungsil University (M. S)

1982. 2~ : Assistant professor, Dept. of Electronics Engineering, Bucheon Tech. College,



Kwang Hyung Lee

1968. 2 : Department of Electronics Engineering, Seoul National University (B. S)

1972. 8 : Department of Electronics Engineering, Seoul National University (M. S)

1982. 8 : Dept. of Electronics Eng., Gradute school of Tokyo National University 연구원

1987. 2 : Department of Electronics Engineering, Chung-Ang University (Ph. D)

1983. 3~ : Assosiate professor, Dept. of Electronics Engineering, Sungsil University.