

# Evaluation for speech signal based on human sense and signal quality

Yoshito MEKADA, Hiroshi HASEGAWA,

Takeshi KUMAGAI, and Masao KASUGA

Faculty of Engineering, Utsunomiya University

Ishii, Utsunomiya, 321 Japan

{mekada, hasegawa, kmg, kasuga}@infor.utsunomiya-u.ac.jp

**Abstract :** Each reproducing speech signal has each particular signal property, because of the processing of encoding and decoding for communications through various media. In this paper, we examine the correlation between speech signal quality and sensory pleasure for the sensory improvement of that signal. In experiments, we evaluate the quality of speech signals through various media by psychological auditory test and physical features of these signals.

## 1 Introduction

Wide spreading of digital audio technology enables us to get easily various high quality signals in our daily life. In the communication system, especially in communication for presence, a transmission of high quality sound signals is necessarily required[1]. These recent advances in such communication systems depend on the spreading of infrastructure, such as ISDN, and the remarkable progress of digital signal processing technology. Therefore, lots of researches make use of that processing technology and many studies are carried out for compression of information by bit reduction and bandwidth limitation for highly efficient transmission. In Japan, two types of digital data compression technology are ap-

plied to carrying personal phones in public line for practical use. One is PDC(Personal Digital Cellular) and the other is PHS(Personal Handy Phone)[2]. PDC uses VSELP(Vector Sum Excited Linear Prediction) coding and PHS employs ADPCM(Adaptive Differential Pulse Code Modulation) coding as compression technology. These technologies enable us to have the communication with relatively high quality signal at low bit rate. However, such speech quality became far from original one in comparison with a facing conversation and a communication for presence. That is because, in these communication systems, their main purpose is to understand what speaker says. In this case, transmitted signals lose high frequency component and have much

quantization noise because of small number of quantizing bits. Therefore, procedure for restoring the essential elements are needed for comfortable communication[3].

In this paper, we examine the correlation between speech signal quality and sensory pleasure for the sensory improvement of the signal. First, We transform this original speech signal into sample signals that features are well-known by using digital signal processing, and transmit through various media. Second, we perform psychologically auditory test and compare that results with physical features of speech signals. In psychologically auditory test, we refer to the evaluation procedure for speech and audio signals[4].

## 2 Survey of speech signal

In this section, we mention brief survey of speech signals in Japan. In general, the objectives of information transmission by speech are as follows.

- 1) Transmitting the information about facts and states.
- 2) Transmuting the emotional information with feeling or atmospheres of speakers.

The transmitting bandwidth in Japanese public phone line is from about 300Hz to 3400Hz, however, bandwidth of original speech signal is from about 100Hz to 10kHz. From this point of view, quality of public phone line is sufficient for the first objective mentioned above. But, the second objective is not sufficient by reason of the lack of high frequency components and individuality. So, we should analyze inferior speech signal from the

viewpoint of both human sense and physical characteristic.

## 3 Analysis of speech signals

For analyzing speech signals, there are two types (subjective estimation and objective one). In the former one, speech signals are evaluated by person within a given framework. In the latter one, many physical features of speech signal are utilized to evaluate it through digital signal processing. Our purpose is to examine the correlation between speech signal quality and its sensory pleasure. Therefore, we examine subjective evaluation and objective features of speech signals. From the beginning of the first step of this research, we employed basic well-known physical features[4]. As experimental sample, we applied band-limiting filter and reduced the number of bit from 16 bits to 10 and 8 bits per sample. Furthermore, we transmitted that reproducing signals through PHS and PDC.

In this paper, we use speech signals of twenties man and woman for experimental samples(original speech signals). These samples are Japanese and include three word categories(monosyllable, number and city name) to evaluate total feeling of each word category. Table 1 shows experimental samples of speech signals in this examination.

Specifications of each original sample data are as follows.

Sampling frequency	:	48kHz
Quantization	:	16bit stereo

All original speech signals are recorded on digital audio tape(DAT) and transferred into

computer. To make processing signals(band-limited signal and bit reduced signal), we used 3rd order Butterworth digital filter for making the former signals, and omitted least significant bits to generate the latter signals. Furthermore, we get the transmitted signals from PHS to PHS and PDC to PDC . These all signals have the same specifications as original speech signals and are recorded on DAT except that transmission signals are two channel datum with monaural.

## 4 Psychological evaluation on speech signal

To evaluate speech signal quality, many types of techniques are proposed. So that, assessment processes are different in each research field.

Our purpose in this paper is to make sure the correlation between speech signal quality and its sensory pleasure by psychological auditory test and feature analysis. Therefore, we propose SD(Semantic Differential) scaling method with new assessment words. Evaluation method and experimental procedure are shown in next two subsections.

### 4.1 Assessment words and evaluation method

From the beginning of the first step of research, we select some assessment word pairs that used in evaluating speech signals[4~7]. Table 2 shows the relation between assessment 15 word

pairs we chosen and their representative assessment word[7,8]. Using these assessment word pairs, we organized seven grades of SD scale and made performed psychological auditory test. A assessment sheet for psychological auditory test is shown in table 3. Test subjects check the difference between original signals and processing or transmitted signals subjectively.

### 4.2 Experimental procedure

Psychological auditory test performed in the

Table 1 Speech signals for examination.

category	speech signal
monosyllable	ha, hyo, a, myu, ga
number	0287, 5732
place name	hachinohe, kesen-numa, yukuhashi, sapporo

Table 2 Assessment words for psychological auditory test.

representative assessment word	assessment word pair		
clarity	clear	—	indefinite
sharpness	sharp	—	uncrisp
vividness	vivid	—	blur
tenderness	soft	—	hard
beauty	pretty	—	dirty
showiness	showy	—	plain
strength	strong	—	weak
spreadness	spread	—	restricted
brightness	bright	—	dark
transparency	transparent	—	dim
liveliness	bouncy	—	depressed
fineness	fine	—	rough
stretchness	stretched	—	confined
glossiness	glossy	—	rough
naturality	natural	—	artificial

Table 3 Assessment sheet for psychological auditory test

	exceedingly	quite	middle	quite	exceedingly	
indefinite	-----				-----	clearly
uncrisp	-----				-----	good sharpness
vivid	-----				-----	blur
hard	-----				-----	soft
dirty	-----				-----	pretty
plain	-----				-----	showy
weak	-----				-----	strong
spread	-----				-----	restricted
dark	-----				-----	light
transparent	-----				-----	dims
bouncy	-----				-----	depressed
fine	-----				-----	rough
stretch	-----				-----	confine
glaze	-----				-----	rough
natural	-----				-----	artificial

soundproofing room. All speech signals on DAT are reproduces with two speakers (Fig. 2).

We employed pair test method[9]. All subjects listened three times in comparing original speech signal(A) to processing signal(B) or (A) to transmitted signal(B) at adequate time intervals to fill in the assessment sheet. Psychological auditory tests were executed two times for keeping reliability of tests. Furthermore, we set up the test time no longer than 30 minute per one psychological auditory test because of much fatigue of subjects in a test. Total number of subjects are 13(speech signal of male) and 12(speech signal of female). Fig.3 shows the average profile of psychological auditory tests.

## 5 Discussion

According to results of the psychological auditory test by male speech signals and female speech signals, the difference of valuation between female speech signals and these of original one is larger than that of male speech signals and their original's. However, the tendency of that drawing in Fig.3-1 and Fig.3-2 are almost the same each other, except that variations of the test results by female speech signal are larger than these of male's. This shows as follows. If we want to transmit the speech signals of female containing its emotional information, we should keep their physical features as possible as we can.

In about half of assessment words, transmitted signal by PHS located between band-limited signal(3kHz) and bit reduced signal(8bits). In qualitatively speaking, our feeling of band-limited signal may be expressed as "to be vague", and that of bit reduced signal may also be expressed as "to be distorted".

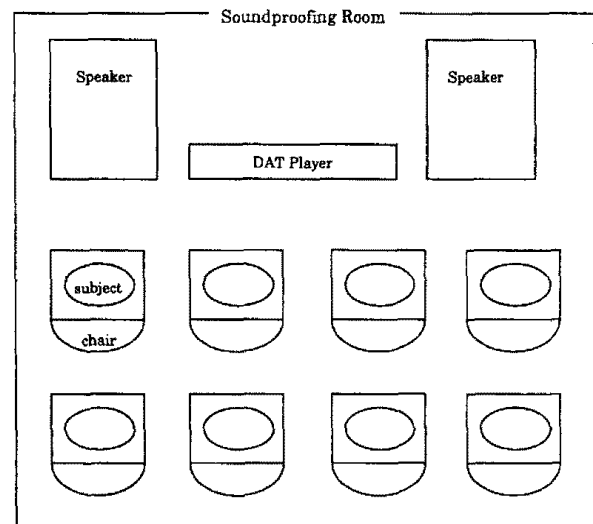


Fig.2 Arrangement of listening room for auditory test.

Transmitted signal by PHS has both characteristics by reason of using the technique of ADPCM. In view of this psychological auditory test, it is also known that test results in transmitted signal by PHS is more similar to band-limited signal (3kHz) than that of bit reduced signal(8bits). Transmitted signal by PDC has poor characteristics in the sense of human sense. So, this results is easily understood for us from the reason that PDC uses VSELP coding.

In these experiments, we used only basic features (variation of spectrum, e.t.c.) as physical features. These were used for having confirmation of speech signal characteristics. We would like to have next study for creating the other new features.

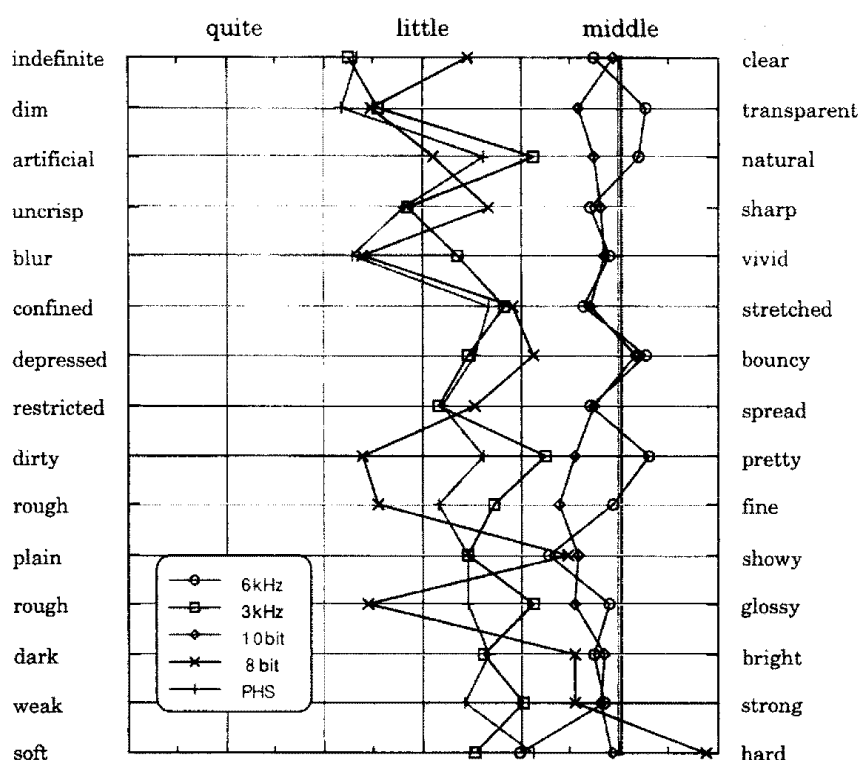


Fig.3-1 Results of psychological auditory test for speech signals of male.

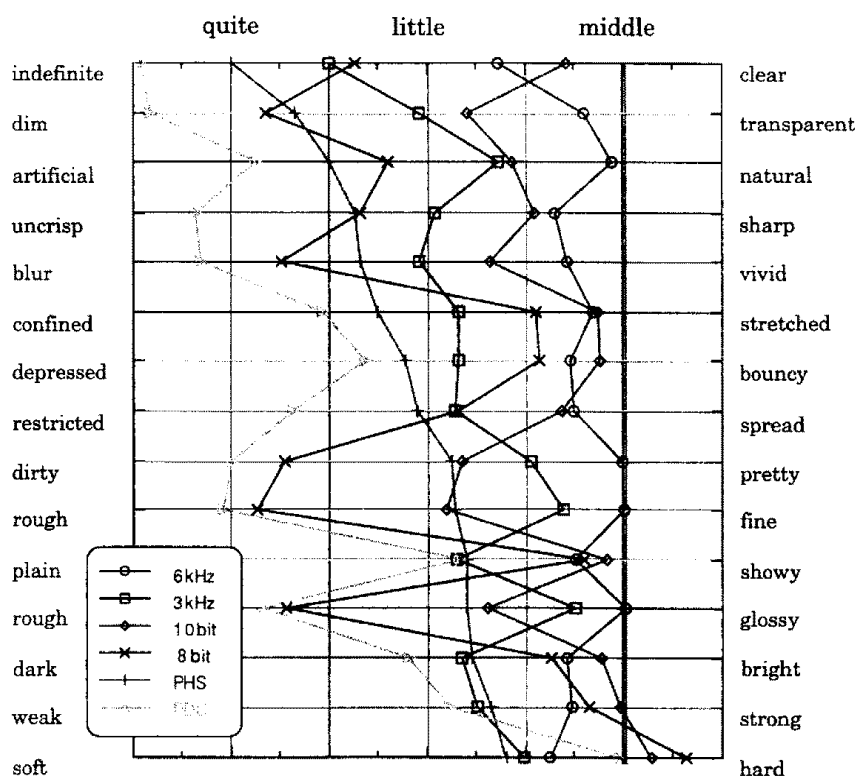


Fig.3-2 Results of psychological auditory test for speech signals of female.

## 6 Conclusions

In this paper, we analyzed speech signals of male and female to improve their quality from viewpoint of human sense. We paid attention to qualities of the transmitted speech signals through various media. Then, we analyzed them in comparison the results of psychological auditory test with their physical features. On our these experiments, it concludes as follows. As for the difference between speech signal of male and female, it is known that the emotional quality of compressed speech signals of female are inferior to these of male's. And, the variances of female's graph are a little lager than male's one.

In closing this paper, we say following things for our futureworks.

1. Considering for other physical features, such as time-mean of envelope, and so.
2. Analysis for factors[9].

This work was supported in part by grants from TAF(Telecommunications Advancement Foundation). We thank of many participants in our psychological auditory test and we also gratefully acknowledge Ms. Kumiko Okahisa, Mr. Hiroshi Inoue, and Ms. Michiko Hososhima for many helpful assistance.

## REFERENCES

- [1]Y. Yoshida, M. Abe : "Generation of Wideband Speech from Narrowband Speech by Codebook Mapping", Trans. of IEICE, J78-DII, 3, pp.391-399(1995-03).
- [2]F. Watanabe : "Technologies of Personal Handyp-phone System", Journal of ITE, Vol.50, No2, pp.207-215(1996).
- [3]T. Nakayama, T. Koshikawa, and T. Miura : "Fundamental Consideration on the Evaluation of the Quality of the Reproduces Sound", Journal of Acoust. Soc. Jpn., Vol.52, No.7, pp.209-215(1965).
- [4]T. Nakayama, K. Kodama, T. Futagami et al. : "Evaluation and Analysis of Coded Speech Quality", Trans. of IEICE, J75-A, 1, pp.1-9(1992-01).
- [5]E. Masuyama, S. Kobayashi : "Sensory Evaluation", Kakiuchi press, pp.7-49(1993).
- [6]S. Iwamura : "The Effect of the Matching of Sound and Video on the Interaction between Auditory and Visual Processing in Communication via Audio-visual Media", Journal of Acoust. Soc. Jpn., Vol.48, No.9, pp.649-657(1992).
- [7]M. Miyahara, Y. Morita : "Investigation and Analyses of Words for Assessment of Tone Quality", Journal of Acoust. Soc. Jpn., Vol.52, No.7, pp.516-522(1996).
- [8] K.Okahisa, H. Inoue, M. Hososhima et al. : "Examination on the correlation between speech signal quality and sensory pleasure", Proc. of 12th symposium on Human Interface, pp.275-280(1996)
- [9]T. Nakayama, T. Miura, F. Hayama : "Factor Analysis of Sound Quality", Technical Report of IEICE, EA75-13(1975-05).