

A Study on the Word Selection for Intelligibility Assessment of Synthesized Speech by Rule

(규칙 합성음성의 이해성 평가를 위한
단어표 구성에 관한 연구)

Jin-Woo Hong*, Soon-Hyob Kim*, Seong-Hoon Kang**

홍진우*, 김순협*, 강성훈**

ABSTRACT

As a result of recent progress in speech synthesis techniques, the those new services using new techniques are going to introduce into the telephone communication system. In setting standards, voice quality is obviously an important criterion. It is very important to develop a quality evaluation method of synthesized speech for the diagnostic assessment of system algorithm, and fair comparison of assessment values. This paper has described several basic concepts and criterions for intelligibility assessment of synthesized speech by rule based on assessment item, assessment unit, and requirements to word selection. Finally, a method of word selection and word list to be used in word intelligibility tests was proposed.

요 약

최근에 음성합성 기술이 발전됨에 따라 이러한 기술을 이용한 새로운 통신 서비스가 등장하고 있으며, 그 표준 설정에 있어서 음성 품질은 아주 중요한 요인이 된다. 따라서, 시스템 알고리즘의 진단적 평가 및 평가치의 상호 비교를 위해 품질 평가 방법을 개발하는 것은 매우 중요하다. 본 논문에서는 규칙 합성음성의 이해성 평가를 위한 기본적 사항들과 개념을 기술하고, 이해성 평가에 사용되는 단어표를 구성하는 방법과 단어표를 제안하였다.

I. Introduction

Recently, synthesized speeches are developed using various methods according to technical advances and expansions of their application. Therefore, it

is very important to assess the quality of synthesized speech for optimum design and effective communication network planning.

But articulation and naturalness of synthesized speech by rule are not enough to compare with that of natural speech. In this situation, it is necessary to establish the assessment method counting the quality of synthesized speech by rule,

*Dept. of Computer Eng., Kwang Woon Univ.

**Signal Processing Sec., ETRI.

and then this assessment method can be offered to useful index for investigation and development of synthesized speech through result analysis.

In the case of quality assessment of synthesized speech, word intelligibility is used as a measure, especially in synthesized speech by rule because it has some possibility of quality impairment owed to the sound connection.

This paper described several basic concepts for intelligibility assessment of synthesized speech by rule and proposed the method of word selection and word list to be used in intelligibility tests.

II. Establishment of Criterion

A. Assessment Item

It is noticed that synthesized speech has the different characteristics as contrasted with natural speech because of synthesized unit and connection rule. Therefore, when listen to synthesized speech by rule repeatedly, intelligibility score is increased. That is, learning effect is reflected on intelligibility of synthesized speech.

Also, because the learning effect according to person is different, the measurements for intelligibility assessment have to consider following items.

- (1) learning effect
- (2) personal differences of listening capability

B. Assessment Unit

Assessment unit is very important factor to implement test and is classified into "syllable", "word", and "sentence". Each assessment unit is used according to its target and object. Generally, syllable is used for sound articulation, word or sentence is used for intelligibility efficiency. So, we used word for assessment unit.

C. Necessity of Assessment Word

In the case of intelligibility assessment, it is

evident that measurement results have great differences according to selected test words. However if the selected test words contain word number too much, it can be generated that test environment and test time can be very big. The appropriate word selection and word list, therefore, are required for a reasonable test of synthesized speech by rule and that point lead to one of goals in this paper.

III. Analysis of the Mother Group for Word List Construction

A. Choice of the Mother Group

The mother group must be chosen to construct word lists. In this paper, the data base (total 17,532 words) of "Korean vocabulary frequency inquiry" was chosen as a mother group, inquired into appearance frequency of Korean vocabulary in primary school text, issued to Korea Psychological Society.

The statistical analysis was made in valid 13,883 words except derivative words by a suffix, person's names, and names of a place from the mother group.

B. Applicable Attributes to Word List

The criterions to select words in the mother group are introduced to identify word list with a distribution of the mother group's attributes. In this paper, we considered five articles as word attributes, i. e.,

- (1) word class
- (2) word length
- (3) frequency number
- (4) anomalous pronunciation
- (5) kinds of phoneme in first syllable

C. Statistical Analysis of the Mother Group's Attributes

- (1) Classification and Distribution of Attributes

Each word attribute is divided again, using basic concept to construct word lists that they have to maintain balanced frequency distribution against the mother group for five attributes proposed. Namely, each word attribute is classified into several detailed regions by granting characteristics and words are selected from each detailed region so as to identify with frequency distribution of the mother group.

From the viewpoint of this selection method, we decided region numbers of each attribute like Table 1 so that each region can include word number as equally as possible.

Table 2 shows the analysis results of the mother group to each attribute and each region. Total region numbers are 864 ($4 \times 4 \times 3 \times 2 \times 9$ regions).

(2) Distribution of Word Numbers Included in One Region

On analyzing the mother group into five attributes, word numbers belonging to classified regions

consist of certain frequency distribution. Figure 1 shows how many word numbers in one region are included. The word numbers included in one region are represented by x axis, y axis represents the accumulation of region numbers that word numbers are below the value coping with x axis. In Figure 1, it is shown that the ratio of regions that word numbers of one region are below 30 to total region is 83%.

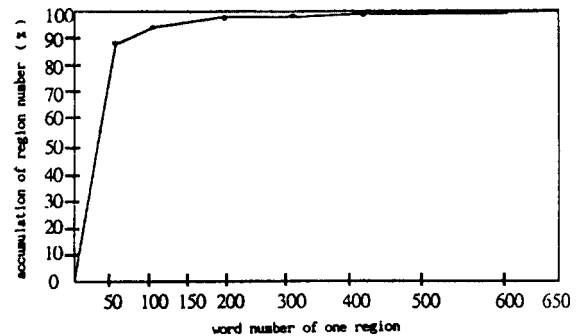


Fig. 1. Relation of the ratio of word number to region number

Table 1. Classified region number of each attribute

attributes	region number	kinds of region
word class	4	noun, verb, adjective, adverb
word length	4	1 syllable, 2 syllable, 3 syllable, above 4 syllable
frequency number	3	frequency no. above 20 : high frequency no. 6-9 : middle frequency no. below 5 : low
anomalous pronunciation	2	exist or not exist of anomalous pronunciation
kinds of phoneme in first syllable	9	single vowel stop sound glottalized sound strong sound fricative sound nasal sound diphthong liquid sound affricate sound

Table 2. Frequency distribution to region of each attribute

attributes	classified region	region no.	word no.	rate(%)
word class	noun	195	9162	66
	verb	136	2967	21
	adjective	131	808	6
	adverb	141	946	7
word length	1 syllable	67	422	3
	2 syllable	166	5396	39
	3 syllable:	190	4199	30
	above 4 syllable	180	3866	28
frequency number	high	198	2319	17
	middle	197	3200	23
	low	208	8364	60
anomalous pronunciation	exist	273	3957	29
	not exist	330	9926	71
kinds of phoneme in first syllable	single vowel	78	1452	10
	stop sound	79	4448	32
	glottalized sound	71	498	4
	strong sound	63	749	5
	fricative sound	75	2153	16
	nasal sound	76	1660	12
	diphthong	70	820	6
	liquid sound	15	34	0
	affricate sound	76	2069	15

Additionally, the valid region numbers are 603 because the regions not to be included ever a word are 261.

IV. Construction of Word List

A. Requirement Condition to Word List

On constructing word list based on the contents as mentioned above, following three items must be considered as requirement condition.

- (1) To reflect directly the characteristics of Korean language being used usually.
- (2) To reduce a learning effect to remember words by repeating test if possible.
- (3) To make a word list including small size of words, maintaining balance with the mother

group and considering test environment and test condition.

We applied three conditions to word list construction like lower part. First is to construct word list having same features as the mother group for five attributes so that it has the balance as equally as possible. Second is to make variable word lists not to fix words within the selected word list. Third is to utilize word selection method deliberating frequency distribution of the mother group and efficient test.

B. Word Selection Method

Because words for assessment objective are all Korean vocabularies to exist really, words as many as possible must be used for intelligibility test. To

take into account of efficient measurement, however, it is desirable that global intelligibility can be grasped by the small word number if possible.

A antinomy is arised from this problem.

And then we implemented a process of applying to tradeoff between both issues optimally. Now, think about a method of word selection utilizing frequency distribution of the mother group for five attributes. When total word number is AN, word number to desire selection is SN, and word number belonging to certain region is n, the condition for a word to be at least selected in that region can be represented by equation(1).

$$(SN / AN) \times n \geq 1 \quad (1)$$

That is, words are selected into regions to satisfy equation (2).

$$n \geq (AN / SN) \quad (2)$$

The smaller SN is, in this time, the more the ratio of regions not to be able to select even a word increases. So, approximation goes down from the distribution of the mother group.

From now, "cover rate of region (BRN/ERN $\times 100[\%]$)" and "cover rate of word (RAN/AN $\times 100[\%]$)" are defined, which express the ratio of regions number (BRN) for each word within a word list to the valid region numbers (ERN=603) and the ratio of entire word number (RAN) belonging to this region to total words (AN) each other. It is a nice idea to bring up SN so as to select words from many regions, but the cover rate of region and word makes a saturation state beyond some level of SN. To bring up SN aimlessly, that is, generates a matter of large scale test and a subject not to increase in approximation to the mother group so much as that. This appearance is demonstrated from Figure 2 that depicts

the relation between cover rate of word and cover rate of region to SN.

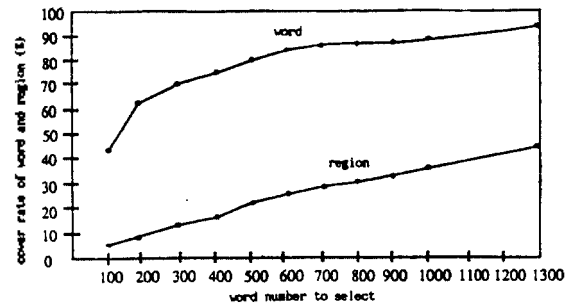


Fig. 2. Cover rate of word and region to SN

C. Word List Construction

First of all, note that some degree of rate creates a good efficiency of selection. Figure 3 shows the relation between cover rate of region and cover rate of word to word number selected. If each region has the same number of words, the distribution draws a dotted line A as in figure 3. But on comparing word distribution with line A distribution, we can know that the relation of both is inclined toward one side. In the case of word selection considering of a tradeoff to cover rate of word and region, best fine efficiency is obtained by a P point that straight line B to move parallel with the dotted line A is contacted with distribution curve.

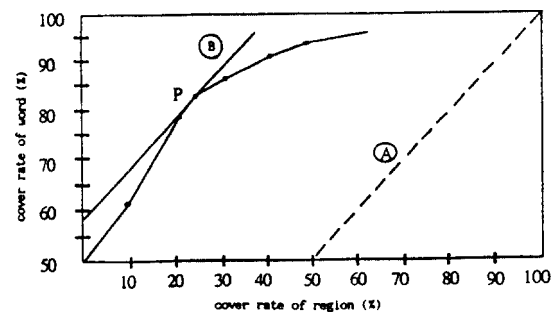


Fig. 3. Relation of cover rate between region and word.

In the result, word numbers of a word list are

540 from Figure 2 because cover rate of word is about 84% in this P point. This is optimal value of word numbers within a word list to test intelligibility assessment.

V. Conclusion

The necessity of the quality assessment for synthesized speech by rule cannot emphasize too much, but there are not standard evaluation methods and word lists for assessment in Korea, so far.

This paper has described several basic concepts and criterions for intelligibility assessment of synthesized speech by rule, based on assessment item, assessment unit, and requirements to word selection. So, it was implemented to do following : to choice the mother group, to give word attributes, to analyze the frequency distribution of the mother group, and to present word selecting conditions.

Finally, we proposed the method of word selection and word list construction for intelligibility measurement.

REFERENCES

1. N. Kitawaki, K. Itoh, and K. Kakehi, "Speech Quality Measurement Methods for Synthesized Speech", Review of ECL, Vol. 29, No. 9-10, Sep. 1981.
2. T. Watanabe, "Intelligibility Assessment of Synthesized Speech Using Word Intelligibility Score", NTT R & D, Vol. 38, No. 10, pp. 1,143~1,152, 1989.
3. T. Watanabe, H. Nagabuchi, and N. Kitawaki, "A Study on Quality Evaluation Method of Speech Synthesized by Rule", NTT ECL, SP86-11, 1988.
4. N. Higuchi, S. Yamamoto, and T. Shimizu, "Evaluation of Intelligibility and Naturalness of the Synthetic Speech Generated with a Japanese Speech Synthesizer by Rule", Japan EIC D-II, Vol. J71-D-II, No. 8, pp. 1,133~1,140, Aug. 1989.
5. S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, Objective Measurement of Speech Quality, Prentice Hall, 1988.

▲Jin Woo Hong was born in YuJoo, Korea, on April 15, 1959. He received the B.S., M.S. degree in electronics engineering from the Kwang-Woon University, Seoul, in 1982 and 1984, respectively. Since 1984 he has been with the Transmission

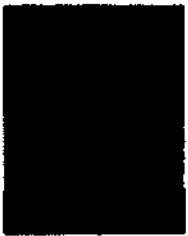


System Section and the ISDN Technology Department in ETRI, where he is now a senior member of technical staff. He is currently working toward the Ph.D. degree at Kwang-Woon University from March 1990. His research interests include the ISDN, speech quality assessment, and speech synthesis.

▲Soon Hyob Kim was born in JeJu, Korea, on Dec. 28, 1947. He received the B.S. degree in electrical engineering from the Wool-San University in 1974, and the M.S., Ph.D. degree in electronics engineering from the Yonsei University, Seoul, in 1976 and 1983, respectively. Since 1979 he has been a professor in the Kwang-Woon University. From 1986 to 1987 he worked for the University of Texas at Austin as a visiting professor. His research interests include the digital signal processing, speech recognition and synthesis, AI, and neural network.



▲Seong Hoon Kang was born in Kwangju, Korea, on June 15, 1956. He received the B.S degree in electronic engineering from the Kwangju University, in 1981, the M.S degree in electronic engineering from the Yonsei University, in 1983, and the



Ph.D. degree in Acoustics from the Kobe University, Japan, in 1987. During the period of April 1987 to August 1987, he was a visiting researcher at the Acoustic Laboratory of YAMAHA Corporation, where he was involved with acoustic designs, measurements, and subjective assessment of concert halls.

Since February 1988 he has been with the Korean Electronics and Telecommunications Research Institute, as a Senior Researcher. His current interests are the transmission quality of telephone network, Electroacoustics, Room Acoustics, and Psychoacoustics.

Dr. Kang is a member of the Acoustical Society of Japan.