The Comparisons of GRBAS Perceptual Judgments according to Levels of Utterances

Hwa-Young Pyo[*] · Hyun-Sub Sim[**]

ABSTRACT

The present study was performed to investigate adequate levels of utterances which can give essential as well as useful information about the patients' voice, by examining the degrees of correlation between the levels of utterances (vowels, words, and phrase paragraph reading) and the entire utterance including all of the levels.

For this purpose, a total of 10 individual utterance samples (5 vowels, 3 words, 1 phrase, 1 paragraph reading) were collected from each of the 30 subjects with voice disorder patients, and four experienced voice therapists evaluated them using GRBAS. The results showed that four therapists highly agreed upon on 'G' parameter. The coefficient of the correlation between each level of utterance and entire utterance tended to be above 0.70. Judgements of the vowel /ε/ as well as /o/ highly correlated with the judgement of the entire utterance. Regardless of severity, the judgement of the entire utterance highly correlated with the judgements of the vowel /u/ and the paragraph reading. These results suggest that experienced voice therapists can precisely evaluate patients' voice quality with only one sustained vowel in the clinic field, as is done with the entire utterance evaluation.

Keywords : GRBAS, Utterance, Correlation, Experienced therapist

1. Introduction

When the patients with abnormal voice were referred to voice therapists, most of the therapists judge the patients' abnormalities by a perceptual evaluation of their utterances, and make it a rule to record them with audio and/or video systems (Jeong, 1997). The recorded utterances before therapy can be useful materials to judge the degree of improvement of voice during or after therapy, and can be used as referential speech samples when the patient was referred to other therapists or ENT doctors. The procedures should be performed not only for the patients with voice problems, but also those with organic problems (e.g., apraxia) (Kearns & Simmons, 1988), or those with functional problems (e.g., functional articulation disorder) (Bernthal & Bankson, 1988).

[*] Dept. of Otorhinolaryngology, The Institute of Logopedics and Phoniatrics, Yonsei University, College of Medicine
[**] Dept. of Special Education & Interdisciplinary Program of Communication Disorders, The Graduate School, Ewha Womans University

Most of the patients with voice disorders have normal intelligence and language competence, so we can obtain their various levels of utterances. Therefore, Titze (1994) and Boone & McFarlane (2000) recommend that voice therapists should record various levels of utterances from vowels to spontaneous speech, with various pitch and amplitude, and with various manners, in order to make a correct voice evaluation.

However, recording the entire voice production at various levels of utterances in different ways is very difficult with respect to time and economy. When we consider the fact that experienced voice therapists can grasp the patient's problems with very small parts of recorded speech samples even though they have never met the patients, the necessity to record all levels of the patient's utterances is decreased. Therefore, the present study was performed to investigate adequate levels of utterances which can give essential as well as useful information about the patients' voice, by the correlation study between the GRBAS evaluation on each individual level of utterances (vowels, words, phrase and paragraph reading) and the entire utterance including all levels.

2. Methods

2.1 Subjects

Speech samples were recorded from the 30 patients with voice problems for vocal hyperfunction who were referred to the Voice and Speech Therapy Clinic of Severance Hospital. Seventeen of the thirty were males, and thirteen were females. Their age levels were the following : 0~9 years (8 persons); 10~19 years (5); 20~29 years (2); 30~39 years (5); 40~49 years (10). Patients over fifty were excluded, for the senility of vocal folds begins after the fifties (Benninger et al., 1994).

2.2 Recording of Speech Samples and Perceptual Judgment

Subjects' utterances were recorded by a JVC RC-QW33 cassette tape recorder with a Sony FV-310 microphone placed 5 cm from the mouth. Subjects were asked to read the experimental materials, that is, 3 word-level utterances, 1 phrase, 5 vowels, 1 paragraph reading, and 1 entire utterance including all of the 10 individual utterances (see Table 1), at their comfortable pitch and amplitude. Their recorded speech samples were divided into 9 individual utterances and 1 entire utterance, according to the levels of speech utterances. A different level utterance has a different order of speech samples. This means that, for example, under the title of Phrase level,

Table 1. Speech Samples

Word levels - Gender / Age / Year

              (e.g.: nam-ĵa / i-šip-sa-sɛ / čɔn-gu-bæk-gu-šip-gu-ɲɔn)

                  (Male / 24 years old / The year of 1999)

Phrase levels - Date (e. g. ku-wɔl-ku-il)(The ninth of September)

Vowel levels (sustained for 2 seconds) - /a/ /ɛ/ /i/ /o/ /u/

Paragraph levels - Subjects read a part of 'San Chaek (Taking a Walk)' paragraph:

     '바닷가에 나가 조개를 주우며 넓게 펼쳐있는 바다를 바라보면 내 마음 역시 넓어

     지는 것 같다. 가로수 길게 뻗어있는 곧은 길을 따라 걸어가면서, 마치 쭉쭉 뻗어

     있는 나무들처럼, 그리고 반듯하게 놓여있는 길처럼, 바른 마음으로 자연을 벗하며

     살아야겠다는 생각을 한다.'

     (Pa-dat-k'a-ɛ na-ga čo-gæ-ɾil ču-i-myɔ nɔl-k'ɛ pʰyɔl-čʰyɔ-in-nin pa-da-ɾil

     pa-ɾa-bo-myɔn næ ma-im yɔk-š'i nɔl-bɔ-ĵi-nin kɔt kat-t'a. Ka-ɾo-su kil-gɛ p'

     ɔ-dɔ-in-nin ko-tin ki-ɾil t'a-ɾa kɔ-ɾɔ-ga-myɔn-sɔ ma-čʰi čuk-čuk-p'ɔ-dɔ-in-ni

     n na-mu-dil-čʰɔ-ɾɔm ki-ɾi-go pan-di-tʰa-gɛ no-yɔ-in-nin kil-čʰɔ-ɾɔm pa-ɾin

     ma-im-i-ɾo ča-yɔ-nil pɔ-tʰa-myɔ sa-ɾa-ya-gɛt-t'a-nin sæŋ-ga-kil han-da)

---

30 subjects' phrase utterances were randomly ordered, and the order was different from that of the paragraph reading.

Four experienced voice therapists, who treated voice disorder patients for 2.5 years to 4 years (mean, 3.5 years), participated as judges. They evaluated the subjects' voice quality with GRBAS scales, listening to 10 individual utterances. In assigning GRBAS scores, they used 4 scales, i. e., 0 (normal) 1, 2, 3 (severe). If necessary, the use of .5 score (e.g., G 1.5) was allowed. Before the experiment began, the information about the subject's age and gender was given to the judges. When the judges evaluated the samples, they were allowed to listen to one voice sample three times maximum. An evaluation of the entire utterance was asked for 3 days later, since they had evaluated individual utterances, for minimizing the influence of individual utterances on the entire one.

2.3 Analysis

The mean score of each voice was calculated by averaging the four different scores from four judges. In each voice, when the difference between the lowest score and the highest one is over 1.5, these two scores were excluded in the data analysis. Instead, the remaining two scores were analyzed. But if the difference was over 1.5, it was not excluded if the two judges assigned the same score. The correlation between the values of each individual utterance and the entire utterance was analyzed by the Pearson r Correlation Coefficient.

The intra-judge reliability was assessed, and the mean score was 0.94 (range : 0.87~ 0.98), meaning that the results of the present experiment were proved to be very reliable.

3. Results

Table 2 indicates the numbers of subjects showing the difference between highest and lowest score across judges was within 1.0 for each GRBAS parameter.

Table 2. The comparisons of judgement consistency across GRBAS parameters

		G	R	B	A	S
	GENDER	19	14	17	30	20
WORD	YEAR	23	7	19	30	14
	AGE	20	9	16	28	16
PHRASE	DATE	20	9	15	30	11
	a	28	11	24	30	15
	ε	25	7	25	30	13
VOWEL	i	25	15	24	29	13
	o	25	7	20	29	19
	u	20	7	23	30	10
PARAGRAPH		26	8	23	30	8
ENTIRE		26	6	20	30	8

As we see in the table, the 'A' parameter showed the highest degree of consistency across judges. But, it does not suggest that judgments on the 'A' parameter were more consistent compared with other parameters, because voices of patients with vocal hyperfunction revealed few 'A' parameter(Pyo et al., 1999), Therefore, the 'A' parameter will not be mentioned further in the current study. The 'G' and 'B' parameters showed higher consistency among the judges, compared with other parameters.

Figure 1 showed minimal, maximal and mean coefficients of the correlation between the individual utterances and entire utterance. As we see in Figure 1, the minimal coefficients across the levels of utterance fluctuated from 0.186 to 0.684. For example, 4 of 10 individual utterances, i. e., age, vowel /ε/, /o/, and paragraph, showed even their minimal values over 0.60, and vowel /a/ and /u/ showed the coefficient over 0.40. The highest correlation in minimal scores was found in the vowel /o/. On the other hand, the range of mean values (0.781 to 0.915) was smaller than that of the minimal value. Paragraph reading showed the highest correlation in mean scores, and vowel /o/ showed a high correlation of over 0.90, too. Finally, the maximum scores ranged from 0.98 to 1.00, and the highest coefficients in

the maximum score was found in the vowel /ɛ/.
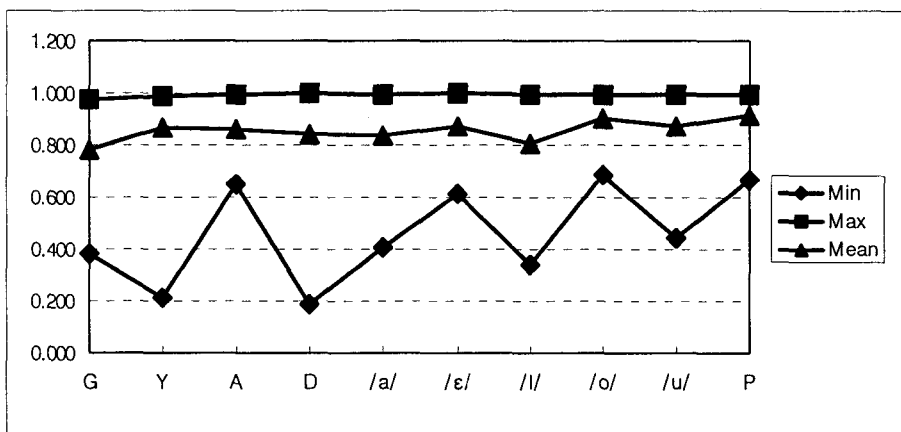
* G = Gender / Y = Year / A = Age / D = Date / P = Paragraph
/a/, /ɛ/, /i/, /o/, /u/ = five vowels

Figure 1. The comparison of the coefficient of correlation between individual utterances and entire utterance

We also compared the coefficient of correlation according to the severity. On the basis of the values of the 'G' parameter, we ordered the speech samples from the highest score, meaning the most severe voice, to the lowest score meaning the least severe voice. We selected fifteen patients from these, five of the most severe voices (Severe group), five middle of the severity order (Moderate group), and five of the least severe voices (Mild group) (see Figure 2).

In the Severe group, the lowest correlation was shown in gender (2 of 5), and the highest was in vowel /ɛ/ and /u/ (2 of 5, each). Vowel /a/ was proved the least correlated with entire utterance in the Mild group (3 of 5), and vowel /u/ and paragraph was the most correlated (2 of five, each). The moderate group showed similar results with the Severe group, though paragraph rather than vowel /u/ had a higher correlation with the entire utterance.
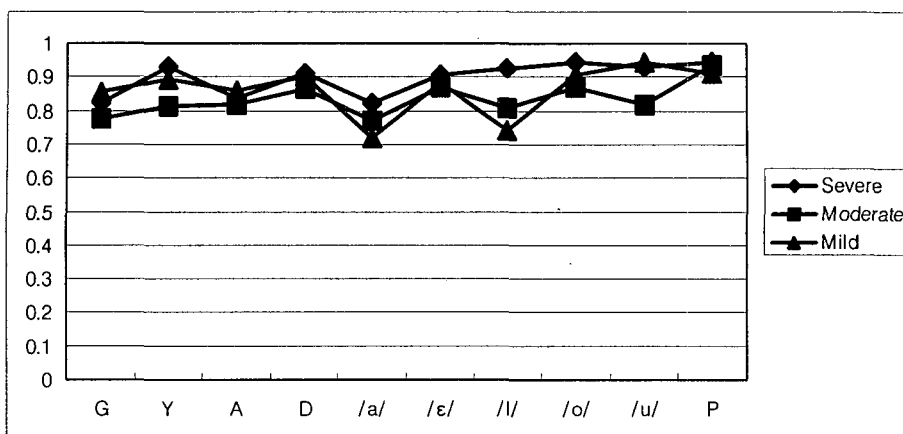
* G = Gender / Y = Year / A = Age / D = Date / P = Paragraph
  /a/, /ɛ/, /i/, /o/, /u/ = five vowels
* y-axix means the coefficient of correlation.

Figure 2. The comparisons of correlation between the individual and entire utterances according to their severity

4. Discussion and Conclusion

We studied the coefficient of correlation between each individual utterance and the entire utterance including all levels of individual utterances. As a result, we found that one sustained vowel of a patient is enough for experienced therapists to grasp the patient's general voice quality. The correlation between the entire utterance and the vowel /a/, frequently used in the clinical field, was lower than that of /ɛ/ and /o/. Of five parameters of GRBAS, the 'G' parameter showed the most consistent results among judges, and the 'B' parameter was more consistent than the 'R' parameter.

As we can see, in perceptual GRBAS judgments, experienced voice therapist's judgment of a patient's one sustained vowel sample showed similar results as the judgment of all of the patient's speech samples. This is consistent with Fex (1992), which reported the more experienced judges show the higher reliability in GRBAS evaluation. These results suggest that if we do not have to consider unexperienced judges or naive listeners in speech samples recording, or if we will not use the samples for further research purposes, recording only one sustained vowel is enough to judge the patient's voice quality.

The results that the vowel /a/ showed relatively lower correlation with the entire utterance than the vowel /ɛ/ and /o/ did suggest the differential effect of vowels on the GRBAS judgment. Vowel /a/ has been used more frequently than /ɛ/ and /o/ in the clinical field; however, there was no scientific evidence about the reason that /a/ is preferred in the

clinical field. /a/, /ɛ/ and /o/ of the five vowels used in this experiment are open vowels. Generally, when you want someone to open the mouth, you will ask him or her to say 'ah'. This means that most people know that they should open their mouth wide when they pronounce the vowel /a/. Considering this, vowel /ɛ/ and /o/ are not widely known as open vowels, so most patients do not open the mouth wide. In the voice therapy approaches of Boone & McFarlane (2000), 'Open the mouth' and 'Chewing' approaches are based on the fact that speaking with a wide open mouth makes the laryngeal muscles and vocal folds relax, which can produce a more comfortable and acceptable voice. Therefore, speaking /a/ with an open mouth can produce a clearer voice than the patient's usual voice. This means that vowel /a/ might not represent the ordinary voice of the patients. Consequently its correlation with entire utterance can be lower than other vowels.

According to Table 1, the number of patients who showed a difference among judges lower than 1.0 was the highest in the 'G' parameter. The number was higher in the 'B' parameter than the 'R' parameter. These results are consistent with previous studies which proved that, regardless of experience, the 'G' parameter showed relatively high intra-judge reliability (Kreiman et al., 1992; de Bodt et al., 1997), and the parameter 'G' is highly correlated with the 'B' parameter (Dejonckere et al., 1993).

Therefore, when the patient's voice samples are needed for voice therapy, experienced therapists can evaluate their voice quality with only one vowel, with enough reliability, and this result can be used very efficiently in the clinical field.

But, this study is limited because spontaneous speech was not included in this experiment. Our 'entire utterance' is a kind of 'intentional' speech which is different from 'unintentional' spontaneous speech, and it means that the results of the present study cannot be also used in spontaneous speech. Spontaneous speech is not the speech sample obtained by the experimenter, so many conditions can influence the voice quality such as the subject's feelings, state of health, and environmental and emotional situations of the time. Also, it is very hard to answer the question of which part of the sample will be used for the final analysis, because it is very controversial which part of spontaneous speech can explain the subject's speech behaviors representatively (Bernthal & Bankson, 1988). Therefore, for further study with spontaneous speech, we should obtain more detailed information about the speech samples.

References

Benninger, M. S., B. H. Jacobson, & A. F. Johnson. 1994. *Vocal Arts Medicine.*New York, NY: Thieme Medical Publishers, Inc.

Bernthal, J. E. & N. W. Bankson. 1988. *Articulation and Phonological Disorders.* Englewood Cliffs, NJ: Prentice-Hall.

Boone, D. R. & S. C. McFarlane. 2000. *The Voice and Voice Therapy*. Needham Heights, MA:
    Allyn & Bacon.

De Bodt, M. S., F. S. Wuyts, P. H. Van de Heyning, & C. Croux. 1997. "Test-Retest Study
    of GRBAS Scale: Influence of Experience and Professional Background on Percep-
    tual Rating of Voice Quality". *J Voice*, 11(1), 74-80.

Dejonckere, P. H., C. Obbens, G. M. de Moor, & G. H. Wieneke. 1993. "Perceptual Eval-
    uation of Dysphonia: Reliability and Relevance". *Folia Phoniatr*, 45, 76-83.

Fex, S. 1992. Perceptual Evaluation. *J Voice*, 6(2), 155-158.

Jeong, O. R. 1997. "Psychoacoustic Evaluation (II): Voice Screening Test". *J Korean Logo
    Phon*, 8(1), 101-109.

Kearns, K. P. & N. N. Simmons. 1988. "Interobserver Reliability and PerceptualRatings :
    More than Meets the Ear". *J Speech Hear Res*, 31, 131-136.

Kreiman, J., B. R. Gerratt, K. Precoda & G. S. Berke. 1992. "Individual Differences in Voice
    Quality Perception". *J Speech Hear Res*, 35, 512-520.

Pyo, H. Y., Choi, S. H., Lim, S. E., Sim, H. S., Choi, H. S., Kim, K. M. 1999. "The Corre-
    lation between GRBAS Scales and MDVP Parameters on the Pathologic Voices of
    the Patients with Vocal Polyps." *J Korean Logo Phon*, 10(2), 154-163.

Titze, I. R. 1994. "Towards Standards in Acoustic Analysis of Voice". *J Voice*, 8(1), 1-7.

▲ Hwa-Young Pyo
    The Institute of Logopedics and Phoniatrics
    Department of Otorhinolaryngology
    Yonsei University, College of Medicine
    134, Shinchon-dong, Sudaemun-gu
    Seoul, 120-752, Korea
    Tel: +82-2-361-8599 (O), Fax: +82-2-393-0580
    HP: 016-295-9788
    E-mail: vtpyo@yumc.yonsei.ac.kr

▲ Hyun-Sub Sim
    Department of Special Education
    Interdisciplinary Program of Communication Disorders
    The Graduate School
    Ewha Womans University
    11-1, Daehyun-dong, Sudaemun-gu
    Seoul, 120-750, Korea
    Tel: +82-2-3277-3538 (O), Fax: +82-2-3277-3110
    HP: 018-244-4354
    E-mail: simhs@mm.ewha.ac.kr