16

# THE PERCEPTUAL AND ACOUSTIC CORRELATES OF "PLEASENTNESS" IN SPEECH

*Ramesh Kumar Mishra*

## Introduction

Voice is a multidimensional series of measurable events and it is necessary that various dimensions of voice be measured to obtain an accurate knowledge of the vocal function. Voice can be normal or pathological. Abnormal voice includes disorders of pitch, disorders of loudness and disorders of quality. Under the general heading of voice quality, several perceptual categories like, harsh, horse, breathy, hyper nasality is considered. A whole range of acoustic parameters and perceptual ratings scales are used to classify pathological voices. Of these, the Long Term Average Spectra (LTAS) provides information on the spectral distribution of the speech signal over a period of time. These spectra can be used to study human voice source in diverse populations [1], [2] And [3].The speech signal presents the product of the sound source and the vocal tract transfer function. The latter differs for different sound segments but in the averaging process, the short-term variations due to phonetic structures are averaged out and the resulting spectrum can be used to obtain information on the sound source, if the analysis is restricted to voiced sounds , where the sound source is the vibrating glottis. In order to further minimize variations due to phonetic structure, the analysis can be made of the reading of a standard text. A further advantage of using running speech is that it provides a more natural sample than sustained phonations.

The use of Long Term Average Spectrum (LTAS) analysis of speech has been explored in several studies. Previous studies have identified the range of individual variation in speech using LTAS [4] and [5]. In the past LTAS has been used to classify abnormal voice disorders [6], [7] & [8], accounting of voice changes after surgery and therapy [9] & [10]. Lofqvist [11] has described LTAS as an important tool in voice research. Kitzing [12] and Wendler, Rauhut & Kruger [13] have shown the usefulness of LTAS as a measure of voice quality. However, in the absence of a laryngeal disorder one's voice can be often perceived as "pleasant", "harsh" or "unpleasant". These are of course often subjective psycho-acoustic experiences by the listener. The present study aims to identify and quantify the empirical basis of such wide variations in normal voice quality perceptions. So far the acoustic and perceptual correlates of these qualities in normal voices have not been understood. In this context the present study assumes its

importance and investigated the acoustic and perceptual correlates of pleasantness and harshness in the voices of normal subjects. Specifically a four point scale was used for perceptual analysis and LTAS was used for acoustic analysis.

**Method**

2.1, Subjects

Forty subjects in the age range of 17 and 26 years and with normal voices participated in the study. Subjects spoke different native languages but the reading sample was collected in English. Apart from the voice givers there were 30 speech pathologists for the perceptual judgments.

**Material**

The first paragraph of "Rainbow Passage" was the material used for reading.

2.3 Procedure

Subjects were tested individually. They were instructed to read the first paragraph of

"Rainbow Passage" 42 words into a microphone kept at a distance of 10 cm (Sony system). They were asked to read in their normal way .The samples were audio recorded in a sound proof room. Each sample was given an identification number for later use in the perceptual judgment task.

**Analysis**

**Perceptual Analysis**

Thirty speech pathologists were the judges for perceptual analysis. They were instructed to listen to each reading sample carefully and rate the samples as pleasant, very pleasant, harsh or very harsh. They were tested individually and each subject listened all the forty samples carefully. There was a pause of 1 minute between each sample. They were also asked to indicate the perceptual correlates of these qualities on a separate sheet of paper.

All the subjective perceptual correlates were coded and quantified.

3.2 Acoustic Analysis

The reading samples were digitized at 16 KHz sampling frequency using a 12-bit A/D converter. Using the "LTAS" of the SSL Software (Voice and Speech Systems,

Bangalore) α, ß and γ ratios were extracted for each sample. α ratio is the ratio of intensity in the frequencies 0-1 KHz to the intensity in the frequencies 1-5 KHz. ß ratio is the ratio of intensity in the frequencies 0-2 KHz to the intensity in the frequencies 2-8

KHz and γ ratio refers to the ratio of intensity in the frequencies 0-1 KHz to the intensity in the frequencies 5-8 KHz.

3.3 Statistical Analysis

The data was tabulated and the percent response under "pleasant", "very pleasant", "harsh" and "very harsh" categories for each subject was calculated by the following formula

**% Response = Total number of responses under each category X 100**

## RESULTS AND DISCUSSION

### Perceptual Analysis

Of the 40 recorded reading samples 17.5% were judged as "pleasant" by more than 75% of the subjects. Also 0% was judged as "very pleasant", 7.5 % as "harsh", 0% as "very harsh" by more than 75% of the subjects. Only these samples were correlated with LTAS measures. The results indicated several parameters as perceptual correlates of "pleasantness" and "harshness" in voice samples. These parameters were grouped under 5 categories as related to quality, pitch, prosody, loudness and noise. "Harsh"/ "very harsh" voices were perceived to be rough , strained , shrilled , nasal , to have pitch breaks, reduced and inappropriate pitch variations, low pitch , variations in rate of speech , inappropriate variations in prosodic features, random variations in loudness and more noisy. Of these qualities related correlates were maximally indicated and noise related correlates were minimally indicated. Table 1 shows the perceptual correlates of "harsh"/ "very harsh" voices.

| S.No | Perceptual Correlates | VeryHarsh (%) | Harsh (%) |
|---|---|---|---|
| 1 | Quality related (Roughness, strain, shrill, nasality) | 30.43 | 52.94 |
| 2 | Pitch related (Pitch breaks, reduced and inappropriate variations in pitch, low pitch) | 23.91 | 23.52 |
| 3 | Prosody related (variations in rate of speech Inappropriate variations in prosodic | 17.39 | 8.82 |

| | | | |
|---|---|---|---|
| | features) | | |
| 4 | Loudness related<br>(Random variations in loudness) | 15.86 | 8.82 |
| 5 | Noise related<br>(More noisy speech) | 10.86 | 5.88 |

**Table 1 Perceptual correlates of "harsh"/ "very harsh" voices. Percent indicates the percent of subjects who indicated the correlates**

"Pleasant" and "very pleasant" voices correlated with softness, smooth onset, and absence of nasality, absence of pitch breaks, adequate pitch and loudness, appropriate prosodies

, little variations in rate of speech, perfect rhythm, appropriate pausing, melodious, absence of abnormal fluctuations in loudness and absence of noise. More than 33% of the judges considered quality related parameters and less than 11% considered loudness related parameters as perceptual correlates of "pleasant"/ "very pleasant" voice. Table 2 shows the perceptual correlates of "pleasant" / "very present" voice.

| S.No. | Perceptual Correlates | Very Pleasant (%) | Pleasant (%) |
|---|---|---|---|
| 1 | Quality related<br>(Good quality, softness, smooth onset No strain, no nasality) | 34.04 | 33.33 |
| 2 | Pitch related<br>(No pitch breaks, adequate pitch Minimal variation in pitch) | 4.89 | 17.94 |

| | | | |
|---|---|---|---|
| 3 | Prosody related<br>(Appropriate suprasegmentals, Good prosody, little variation in speech Rate, perfect use of rhythm and pause, Melodious) | 2.65 | 23.07 |
| 4 | Loudness related<br>(No fluctuations in loudness) | 10.63 | 5.12 |
| 5 | Noise related<br>(No or very little noise in speech) | 12.76 | 20.51 |

**Table 2 Perceptual correlates of "pleasant"/ "very pleasant" voices. Percent indicates the percent of subjects who indicated the correlates**

**Acoustic Analysis**

The α, ß and γ ratios of voices which were judged by more than 75% of judges to be pleasant or harsh were examined. Table 3 shows the α,ß and γ ratios of voices judged as "pleasant". The results indicated low ß and γ ratios for "pleasant" voice samples. This indicated presence of high energy in high frequencies. Further α: ß was 1: 1.8, α: γ was

1:7.05 and ß: γ was 1.4.0.

Table 4 shows the α,ß and γ ratios of " harsh" voices. Compared to α,ß and γ ratios of " pleasant" voices all the three ratios were increased in samples judged as " harsh" , indicating low energy in high frequencies. Further α: ß was 1.8, α: γ was 1: 12.2 and ß: γ was 1: 7.2. It was observed that α: γ and ß: γ were higher in "harsh" voices. Figures 1-2 show LTAS for a "pleasant" and a "harsh" voice. The average α, ß and γ ratios in other samples (which were classified as pleasant / very pleasant/ harsh/ very harsh by <75% of subjects) wer α = 147, ß=351 and γ= 2812 e. It was noticed that α: ß were 1:2.3, α: γ was

1:19.1 and ß: γ was 1:8.1. All ratios were higher compared to voice samples judged as "pleasant" and "harsh".

| Sl. No. | α | ß | γ |
|---|---|---|---|
| 1 | 39 | 80 | 278 |
| 2 | 43 | 74 | 274 |
| 3 | 37 | 69 | 247 |
| 4 | 35 | 58 | 414 |
| 5 | 39 | 71 | 255 |
| 6 | 50 | 68 | 265 |
| 7 | 40 | 81 | 247 |
| **Average** | **40** | **71** | **282** |

**Table 3 α, ß and γ ratios of "pleasant" voices**

| Sl.No. | α | ß | γ |
|---|---|---|---|
| 1 | 168 | 297 | 1131 |
| 2 | 94 | 136 | 974 |
| 3 | 152 | 215 | 3289 |
| **Average** | **138** | **249** | **1798** |

**Table 4 α, ß and γ ratios of "harsh" voices**

**Discussion**

Voice perception includes many events in the speech signal. It is very difficult to explain in any objective way what one exactly means when he or she describes any voice sample as " pleasant" or " harsh" when the sample is of normal humans , without any pathology. This study aimed at quantifying the different

acoustic and perceptual parameters that may be present in any speech samples which help us to objectively decide any speech as good or bad as far as their aesthetic perception is concerned, The Long Term Average Spectra was used as a tool in the acoustic measurements of samples and α ß and γ ratios were computed for each speech sample. LTAS measures generally indicate the sound source spectrum and are easily calculated from running speech. LTAS averages spectral features of the speech signal over duration of time, from which voice quality is inferred. Any sudden change in the energy content of the speech signal at different frequencies could perceptually affect the perception of "pleasantness" or "harshness"" in speech. . LTAS is averaged over time, and it is suitable for analysis of longer utterances of speech and indicates this shift of energy content at different frequency bands. The results of this indicated several points of interest. First of all, quality related parameters were the most often used perceptual correlates of "pleasant" / "harsh" voices. Second, "pleasant" voices were correlated with more energy in high frequencies. It was a very concrete acoustic observation that the α ß and γ ratios of speech samples judged as "pleasant" were much low compared to the samples judged "Harsh". This might indicate precision in the production of highly frequent phonemes in "pleasant" voices. This could be explained as the general pattern of energy distribution in a "pleasant speech signal compared to a "harsh" speech sample.

When the subjects judged the samples they precisely perceived these acoustic events and then could decide the perceptual correlates. This study also quantified several subjective remarks about the qualities of the speech. It is an important observation that speech samples that are considered generally good and soothing have no pitch break or rapid shift in their energy modulation. And speech samples which are perceived as " harsh" have rapid and sudden changes in these acoustics structures.LTAS was a good measure of these acoustic parameters.

There could be one important limitation of this study as far as the subjects are concerned.All the forty subjects who listened the samples and gave their perceptual correlates were trained speech pathologists. It could be the case that since they had training in these subjects and are aware of the technicalities of acoustic events in a speech, this information could have colored their opinions. If the subjects would have been normal lay people then it is would have been difficult to elicit perceptual correlates. Then this being a preliminary study into the nature of speech quality perception, further studies can take into account these limitations. Further studies relating the perception of "pleasantness" or "harshness" to specific languages should prove very interesting.

**Acknowledgements**

The study was completed while the author was at the All India Institute of Speech and Hearing. Thanks to Prof. Savithri for her help in design of the study and the many students who participated and lent their voices for this research.

**References**

Hammarberg, B, B. Fritzell, H.Schiratzki.1984. Teflon Injection in 16 patients with Paralytic Dysphonia: Perceptual and Acoustic Evaluation. *Journal of Speech and Hearing Disorders* . 49. 72-82.

Hartman E, D. van Cramon .1984. Acoustic Measurements of Voice Qualities in Dysphonia after Severe Closed Head Trauma: A follow Up Study. *British Journal of Disorders of Communication*. 19. 40-68.

Hurme, P. 1980. Auto-monitored Speech Level and Average Speech Spectrum. *papers in Speech Research in University of Jyvaskyla* . 2. 121-127.

Hurme P, M. Pirinen. 1984. Accessing Variations in LTAS of Speech. U Ikonen & T Tikka(eds.) *P*apers from the Twelfth Meeting of Finish Phoneticians-Joensuu .Finland:University of Joensuu. *Studies in Language* . 23-36.

Kitzing, P. 1986. LTAS Criteria Pertinent to the Measurement of Voice Quality. *Journal of Phonetics*. 14. 477-482.

Linville, SE, J.Rens.2001. Vocal Tract Resonance Analysis of Aging Voice Using Long –term average spectra. *Journal of Voice*. 15(3).323-30.

Linville, SE.2002. Source Characteristics of Aged Voice Assessed from Long-term Average Spectra. *Journal of Voice* . 16(4).472-7

Lofqvist, A. The Long –time-average- spectrum as a Tool in Voice Research. *Journal of Phonetics* . 14. 471- 7

Prytz, S.1977. Long-Term -average-spectra (LTAS) Analysis of Normal and Pathological Voices. *Proceedings of the 17th International Congress of Logopedics and Phoniatrics.Copenhagen* 1. 459-475.

Sonneinen, A., P, Hurme.1982. Clinical and Acoustic Observations of Normal and Hoarse Voices .Papers in Speech Research.University of Jyvaskyla.4.2-16. 475.

Wenderl, J, A.Rauhut , H. Kruger . Classification of Voice Qualities. *Journal of Phonetics* . 14.483-488.

Wendler, J, E. Doherty, H.Hollien. 1980.Voice Classification by Means of Long-term-average- spectra. *Folia Phoniatrica* . 32.51-60.

White, P.A . 1998. Study of the Effects of Vocal Intensity Variation on Children's Voices Using Long Term Average Spectrum (LTAS) Analysis. *Logopedics, Phoniatrics* . 23.(3).111-120.