Differentiation of Harsh and Hoarse Vocal Qualities through Spectrographic Analysis

Chanut Akamanon

A survey of the literature dealing with speech and voice reveals the problems of voice quality classification systems and definitions of the terms. The most confusing term seems to be harshness. This is often equated with throatiness, strident, guttural, metallic and others (Fairbanks, 1940; Anderson, 1977). It is still unknown whether the metallic, guttural, strident, throatiness, and some others are definitive qualities distinguishable from harshness or not. The problem is more complicated when "hoarseness" is described as the combination of harshness and breathiness (Fairbanks, 1960; Zemlin, 1968; Morris and Spriestersbach, 1978) when it is not really known what harshness is.

For harshness, it is believed that this quality is characterized by the presence of vocal fry and hard glottal attack (Van Riper, 1972). However, it was found that vocal fry can be distinguished from harshness perceptually (Michel and Hollien, 1968). The physiologic correlates of vocal fry have been reported which indicate the modification of vocal fold behavior, subglottic pressure, and air flow rates during its production, when compared with modal register (McGlone, 1967; Murry, 1971). In contrast, physiologic correlates of harshness have not been thoroughly investigated. The physiological mechanisms of harshness have been explained mostly through subjective experience. Only data on acoustic features are obtained which suggest below average fundamental frequency, aperiodicity, pitch and/or amplitude perturbations as basic characteristics of harsh vocal quality (Wendahl, 1963; Bowler, 1964; Michel, 1964). These same acoustic features have been reported to exist in hoarseness (Lieberman, 1963; Moore and Thompson, 1965; Cooper, 1974). In addition to these, the noisy characteristic heard in hoarseness is speculated to be due, in part, to the turbulent air flow arising from an incomplete closure of the glottis (Isshiki, Yanagihara and Morimoto, 1966).

It can be concluded that harshness and hoarseness, at least, in part, possess the same basic acoustic features that most researchers refer to as "roughness" perceptually. And this "roughness" perception is reported to arise from amplitude perturbation (shimmer), pitch perturbation (jitter), aperiodicity, and variations in waveform in successive cycles. The distinction between these two qualities may lie in the "presence of noisy characteristics of turbulent air flow" in hoarseness.

It is hypothesized that spectrographic analysis can bear out the distinction between these hoarseness and harshness.

Fairbanks (1960) and Perkins (1971) illustrated the spectrograms of one speaker simulating four different voice qualities including hoarseness and harshness. Visual inspection of these spectrograms lead to the speculation that hoarseness and harshness can be differentiated through the use of sound-spectrography. However, it is still unknown whether the spectrographic data characterizing simulated hoarse and harsh vocal qualities presented by those two authors can be used as the basis for identifying such qualities in the patients with voice disorders. So the criteria for differentiating hoarseness and harshness through spectrographic analysis were specifically constructed (see Table 1) for this study. The criteria for spectrographic analysis of hoarseness based on various findings are summarized as follows:

1. Indistinct striations, irregular spacing of striations: This characteristic is described by the pattern of indistinct, uneven spacing of the vertical lines indicative of weak, aperiodic glottal pulses. The lack of, or decrease in, pulse characteristics observed on the spectrograms is speculated to be due to some lack of glottal pulse energy. The aperiodicity or loss of regular vocal pulses on the spectrograms has been reported by many investigators (Reich and Lerman, 1978).

2. Regular spacing of striations : This is directly opposite to the one mentioned above. Under this situation, the spacing of the striations sometimes appears to be regular although some lack of glottal pulse energy is expected. This type of spectrographic characteristic is mentioned here because not every case of hoarse voice speaker has aperiodic fundamental periods. This was also observed by Yanagihara (cited in Isshiki, et al., 1966) through the use of vibralizer.

3. Scattered energy in high frequency region above 4000 Hz: The spread of energy on the spectrograms of hoarse voice speakers has been reported in various studies (Rontal et al., 1975; Reich and Lerman, 1978). The presence of scattered energy in the high frequency region is said to result from breathiness (Rontal, et al., 1975)

4. Formant breakdowns: Discontinuities in formant patterns have been observed in various pathological larynges. Rontal, et al. (1975) emphasize the presence of formant breakdowns as an essential spectrographic feature of hoarseness.

5. Aphonic voice breaks: This means the intermittent absence or discontinuity of phonation for a brief period. The presence of aphonic voice breaks as one characteristic of acute laryngitic hoarseness has been reported by Shipp and Huntington (1965).

6. Formant structures characterized by noise elements (Sections and Conventional spectrograms): The presence of noise within and between the formant structures is observed. This scattered energy within or between the formant structures of hoarse speakers, even with a mild degree, has been reported by Yanagihara (1967).

7. Lack of higher harmonics: The decrease of energy in the higher formants or replacement by noise is observed. The reduction of harmonic components in the spectrum of hoarseness has been reported by Isshiki, et al. (1966). There is the possibility that the investigation of the noise patterns alone as has been used by Yanagihara (1967) and many others, may not be sufficient in every case of hoarse voice speakers since the energy of high frequency noise may be too weak to be present on the spectrograms. The investigation of the lack of higher harmonics is appropriate in this study in order to objectify every possible acoustic characteristic of hoarseness.

Criteria for spectrographic analysis of harshness are summarized as follows:

1. Distinct striations, irregular spacing of striations: This characteristic of distinct striations is speculated to be the result of high glottal pulse energy. The irregularity

of spacing pattern of the striations is also noticeable on the spectrograms. The presence of this spectrographic feature corresponds with various findings that aperiodicity in the vocal signals is responsible for the perception of harshness (Wendahl, 1963; Bowler, 1964; Michel, 1964).

2. Frequency breaks equal to or more than one octave: The gross measurement of the fundamental frequency was performed by counting the number of striations in the period of one-fourth to one inch and transformed into number of cycles per second. In this way any dramatic changes of the fundamental frequency of equal to or more than one octave can be objectified. The observation of frequency breaks of about one octave or more in harshness has been observed by Bowler (1964).

3. Frequency breaks of less than one octave: Wendahl (1963) demonstrated that the abrupt frequency changes in successive cycles could be of less than one octave and still give the impression of harshness to the listeners.

4. Definitive formant structures: Generally, harsh vocal quality occurs only on certain portions of speech (Bowler, 1964; Anderson, 1977), especially during downward inflection at the end of sentences (Bowler, 1964; Anderson, 1977). The remaining normal vocalization will display the formant patterns that are comparatively normal.

5. Extraformants (below 4,000 Hz): This means that there may be formant-like energy patterns in places where the formants of the vowels are not expected. It is speculated that the glottal pulse energy in the cases of tense vocal productions is too high to be damped out by the vocal tract and is present in the spectrum in the form of extraformant patterns.

6. Occurrence of fundamental frequencies in the range of vocal fry (2-78 Hz) and hard glottal attack: As was stated by Van Riper (1972), vocal fry and hard glottal attack are two main characteristics of harshness. The presence of hard glottal attack during harshness perception has been mentioned by Fairbanks (1960). The spectrographic characteristics of hard glottal attack is expected to be in the form of a series of sharp, pulse-like vertical striations occurring at the initiation of phonation.

The purpose of this study is to investigate whether the criteria for spectrographic analysis mentioned above can reliably differentiate the voice samples of known qualities into hoarseness and harshness.

Procedures

Voice samples were selected from the available tapes from the Phonetics Evaluation Clinic at the University of

Colorado which included five hoarse (voice ≠ 1 - 5) and four harsh voice samples (voice ≠ 6 - 9). The recorded voice samples were analyzed using a sound spectrograph (Sona-Graph 661A, Kay Elemetrics Corp., Pine Brook, N.J.). Two judges, the present investigator and the experienced phonetician, examined the spectrograms of each voice sample and recorded the presence of certain spectrographic features on the spectrographic evaluation form (see Table 1).

Results

The results of the spectrographic analysis are shown in Table 1. It is apparent that voice number one to five, which are voice samples of hoarseness, have certain acoustic characteristics of hoarseness as shown in the upper half of the chart and none in the lower half of the chart. In contrast, the voice samples of harshness (voice ≠ 6 - 9) have most of their acoustic features appearing in the lower half of the the chart. Voice number 7 and 8 which are samples of moderate and severe harshness have some acoustic features similar to hoarseness in addition to their spectrographic features of harshness. The examples of the spectrograms of normal, hoarse and harsh voice samples are shown in figure 1 and 2 respectively.

Discussion

This study attempted to investigate if spectrographic data can differentiate hoarseness from harshness. The criteria for spectrogrophic analysis were tested through voice samples of known qualities and were found to be adequate for differentiating hoarseness from harshness. Perceptually, these two qualities were reported to be difficult to be differentiated (Akamanon, 1980). However, it is not claimed that these criteria are definitive ones since this is still a preliminary study of spectrographic characteristics of hoarseness and harshness. Because these criteria identified the voice samples reliably in this study, it was further believed that this method could be used to identify hoarse and harsh vocal qualities in the voice-disordered patients.

References

Akamanon, Chanut. 1980. Comparison of Perceptual Judgments and Spectrographic Features in the Differentiation of Harsh and Hoarse Vocal Qualities. Unpublished Master thesis, University of Colorado.

Anderson, V.A. 1977. Training the Speaking Voice. Third Edition, New York: Oxford University Press.

Bowler, N.W. 1964. A fundamental frequency analysis of harsh voice quality. In Speech Monograph, 31 : 128-134.

Cooper, M. 1974. Spectrographic analysis of fundamental frequency and hoarseness before and after rehabilitation. In Journal of Speech and Hearing Disorders, 39 : 286-297.

Fairbanks, G. 1940. Voice and Articulation Drillbook. First edition, New York : Harper and Brothers.

Fairbanks, G. 1960. Voice and Articulation Drillbook. Second edition, New York: Harper and Row.

Isshiki, N., N. Yanagihara and M. Morimoto. 1966. Approach to the objective diagnosis of hoarseness. In Folia Phoniat., 18 : 393 - 400.

Lieberman, P. 1963. Some acoustic measures of the fundamental periodicity of normal and pathologie larynges. In Journal of Acoustic Society of America, 35 : 344-353.

McGlone, R.E. 1967. Airflow during vocal fry phonation. In Journal of Speech and Hearing Research, 10 : 299-304.

Michel, J.F. 1964. Vocal Fry and Harshness. Unpublished doctoral dissertation, University of Florida.

Michel, J.F. and J. Hollien. 1968. Perceptual differentiation of vocal fry and harshness. In Journal of Spooch and Hearing Research, 11 : 439-443.

Moore, G.P. and C.L. Thompson. 1965. Comments on physiology of hoarseness. In Arch. Otolaryng., 81 : 97-102.

Morris, H.L. and D.C. Spriestersbach. 1978. Appraisal of respiration and phonation. In Darley and Spriestersbach, Diaghostic Methods in Speech Pathology. New York: Harper and Row.

Murry, T. 1971. Subglottal pressure and airflow measures during vocal fry phonation. In Journal of Speech and Hearing Research, 14: 544-551.

Perkins, H.W. 1971. Vocal function: A behavioral analysis. In Travis (ed.), Handbook of Speech Pathology and Audiology. New York: Appleton-Century-Crofts.

Reich, A.R. and J.W. Lerman. 1978. Teflon laryngoglasty: An acoustic and perceptual study. In Journal of Speech and Hearing Disorders, 43 : 496-505.

Rontal, E., Rontal, M. and M.I. Rolnik 1975. Objective evaluation of vocal pathology using voice spectrography In Annals of Otology, Rhinology and Lanyngology, 84 : 662-671.

Shipp, T. and D.A. Huntington 1965. Some acoustic and perceptual factors in acute-laryngitic hoarseness. In Journal of Speech and Hearing Disorders, 30 : 350-359.

Van Riper, C. 1972. Speech Correction. Fifth Edition, Englewood Cliffs, New Jersey: Prentice-Hall.

Wendahl, R.W. 1963. Laryngeal analog synthesis of harsh voice Quality. In Folia Phoniat., 15 : 241-250.

Yanagihara, N. 1967. Significance of harmonie changes and noise components in hoarseness. In Journal of Speech and Hearing Research, 10 : 531-541.

Zemlin, W.R. 1968. Speech and Hearing Science. Englewood Cliffs, New Jersey : Prentice-Hall.

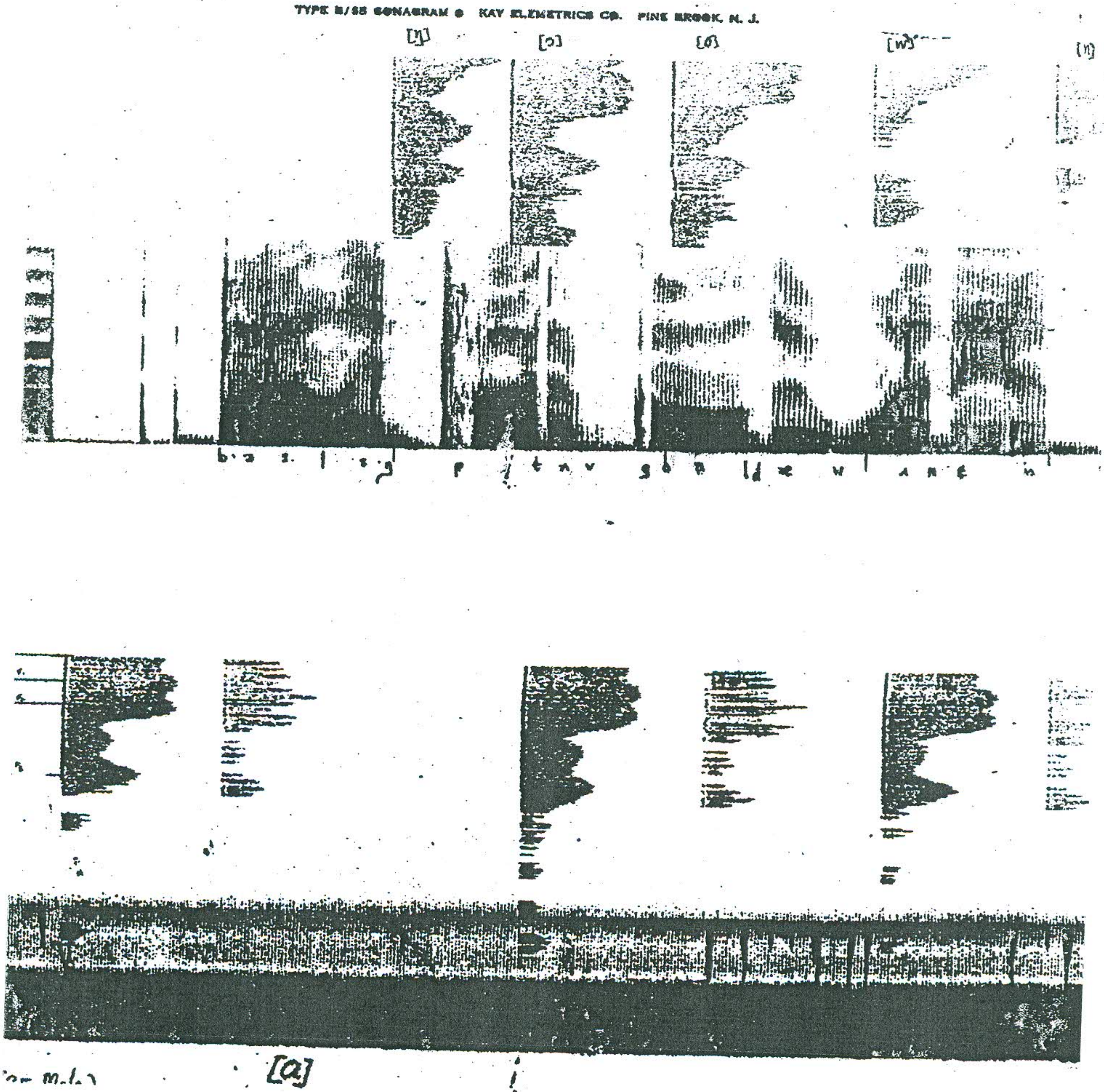

Figure 1 Spectrographic data of normal vocal quality

[illegible]

[illegible]

[illegible]

[illegible]

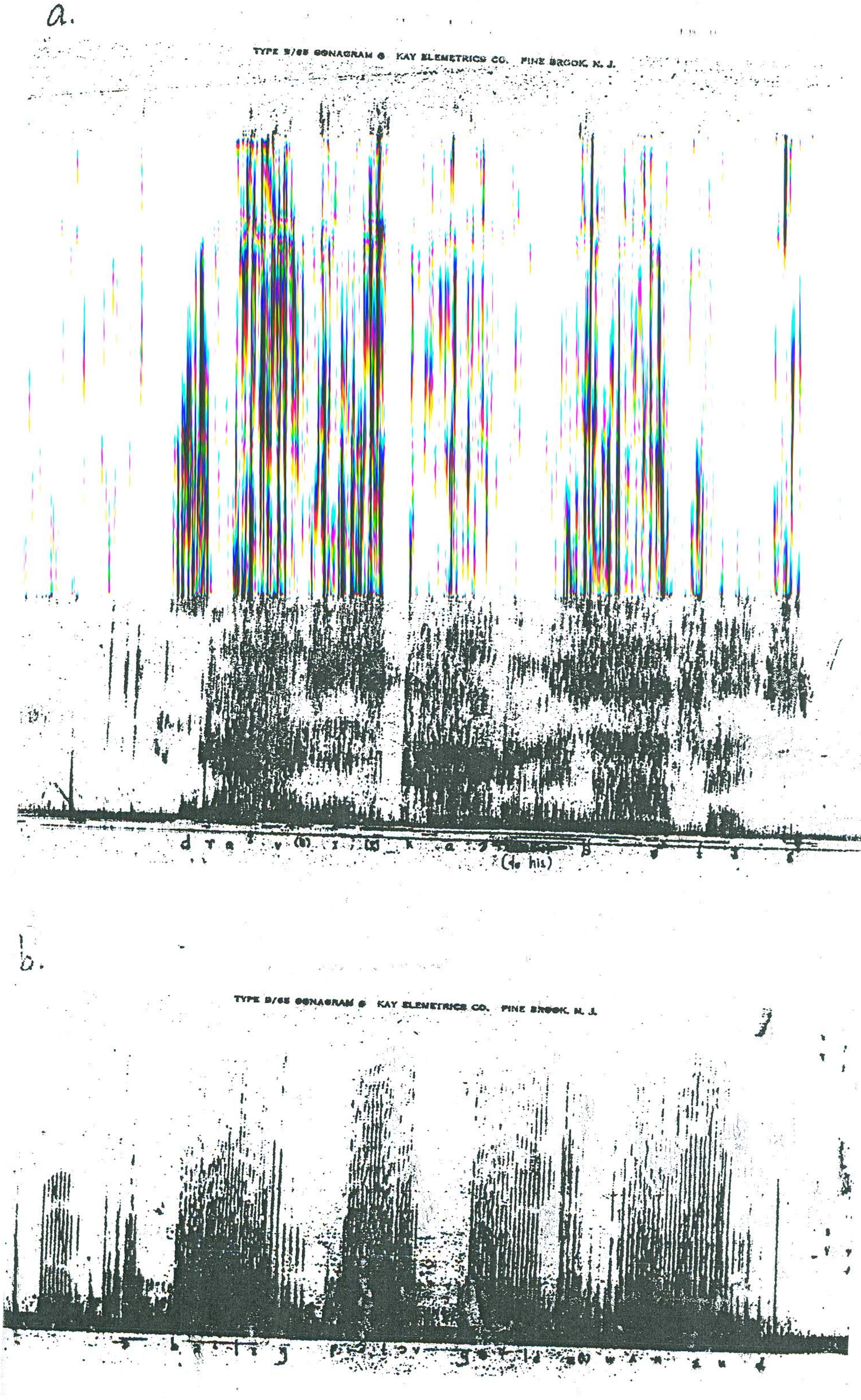

Figure 2 Spectrographic data of the hoarse voice sample of (a) and harsh voice sample (b)

Spectrographic Evaluation Form

Voice#	Indistinct striations, irregular spacing of striations.		Regular spacing of striations.		Scattered energy above 4000 Hz.		Formant breakdowns.		Aphonic voice breaks.		Noise interfering formant structures.		Lack of higher harmonics.	
	C*	V*	C	V	C	V	C	V	C	V	C	V	C	V
1	x o		o		x o		x o				x o		o	
2			x		o		x o		x		x o		x o	
3	x o		o				x o		x		x		x o	
4	x		o						x		x		x o	
5			x o										x o	
6														
7					o						x		o	

Voice#	Distinct striations, irregular spacing of striations.		Frequency breaks ≥ 1 octave.		Frequency breaks < 1 octave.		Definitive formant structures.		Extraformants (below 4000Hz)		Occurrence of F_o in the range of vocal fry (2-78 Hz).		Strong, sharp striations at beginning of phonation.	
1														
2														
3														
4														
5														
6	x o		x o		x		x o				x o			
7	x o		x		x o		x o		x		x o			

C* = Connected speech V* = Vowel prolongation x = investigator o = another judge

Table 1 . Spectrographic data of the voice samples (Voice # 1-5 = hoarseness; # 6-9 = harshness).

Spectrographic Evaluation Form

ce#	Indistinct striations, irregular spacing of striations.		Regular spacing of striations.		Scattered energy above 4000 Hz.		Formant breakdowns.		Aphonic voice breaks.		Noise interfering formant structures.		Lack of higher harmonics.	
	C*	V*	C	V	C	V	C	V	C	V	C	V	C	V
8	o		o		x		o				x			
9														

ce#	Distinct striations, irregular spacing of striations.		Frequency breaks ≥ 1 octave.		Frequency breaks < 1 octave.		Definitive formant structures.		Extraformants (below 4000Hz)		Occurrences of F_o in the range of vocal fry (2-78 Hz).		Strong, sharp striations at beginning of phonation.	
8	x o		x o				•				x o			
9	x o				x o		x o		x		x			

C* = Connected speech V* = Vowel prolongation x = investigator o = another judge

Table 1 Spectrographic data (Continued)