

# RELIABILITY CONSIDERATIONS OF A VOICE PROFILING SYSTEM \*

Clark D. Starr  
University of Minnesota  
and  
Frank B. Wilson  
University of Alberta

## ABSTRACT

The Voice Profile is a clinical tool that is used, rather widely, to describe and evaluate features of voice disorders. This report provides data on the reliability of clinicians use of the Profile. Fifteen experienced and 15 inexperienced clinicians used the Profile to describe the voices of 25 persons who had been diagnosed as having voice disorders. Analysis of data obtained is interpreted as indicating that the Profile can be used most effectively in tasks where one determines whether features are normal or deviant, and that experienced and inexperienced clinicians use it with similar reliability.

From 1966 to 1971 the United States Office of Education, Bureau of the Handicapped sponsored a research and demonstration project at The Jewish Hospital of St. Louis. The project was intended to determine the number and type of voice disorder cases in the school population. A major product of that project was a voice profiling system, which was originally designed to meet the needs of the project Wilson (1972). The profile is now used rather widely as a clinical tool Wilson & Rice (1977). Because the profile was designed to describe a clinical population no formal reports on its reliability have been published. This report presents some descriptive information and reliability data on that profile.

## THE PROFILE - A Description

Figure 1 presents the profile form. The profile was developed in response to the project staff's need to analyze voice problems in a manner useful in clinical management, and the need to develop terminology acceptable to research personnel. Voice dimensions included in the profile are a composite of the features identified by project staff and school clinicians as relevant to clinical management. Specific features were evolved from informal evaluations of clinicians' judgments.

\*This paper was presented at the 1975 convention of the American Speech and Hearing Association, WASHINGTON, D.C.

FIGURE I.

VOICE PROFILE

NAME: \_\_\_\_\_ AGE: \_\_\_\_\_ SEX: \_\_\_\_\_

[circle one]  
VOICE RATING: 1 2 3 4 5 6 7

LARYNGEAL CAVITY	RESONATING CAVITY	INTENSITY	
PITCH	NASALTY	-2	+2
high	hypnasal	soft	loud
B	C		
+3	+4		
+2	+3		
	+2		
A open -4-3-2 1 +2+3 closed	-2 1 +2	VOCAL RANGE	
-2	-2	1	+2
-3	hyponasal	monotone	variable
low			pitch

	YES	NO
INTERMITTENT DIPLOPHONIA	_____	_____
DIPLOPHONIA	_____	_____
AUDIBLE INHALATION	_____	_____
PITCH BREAKS	_____	_____
ERRATIC PHRASING	_____	_____
IMMATURE RESONANCE	_____	_____

Indicate presence or absence of acoustic feature by [ ]

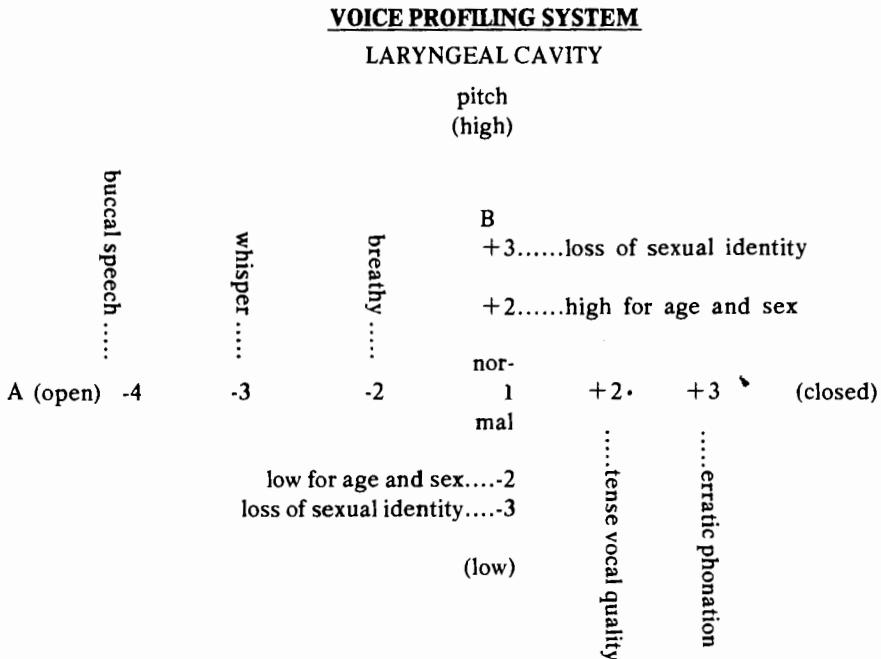
MARKING SYSTEM

Primary Feature X                      Secondary Feature /  
Intermittent Feature [hat]            Noted Feature /

The Voice Rating Scale, upper right, is used to record ratings of overall severity of a vocal deviation. It is a scale of apparently equal interval, with one indicating normal speech and seven, disorder of the greatest severity. It may be used in conjunction with a prerecorded calibration tape prepared by the project staff, which delineates points one through seven.

This is the only section of the profile that directly addresses severity. The Laryngeal Cavity section, shown in Figure II, includes the perceptual features of pitch, open and closed. Each subdivision is designed to be a semi-discrete category, which has the potential to be rated for severity. The open and closed concept is based on observations of laryngeal function attributed to Moore and VonLeden (1953). The concept rests on the premise that in voicing the vocal folds progress through three positions during the vibratory cycle. The cycle starts closed, during which time infraglottal air pressure builds. The vocal folds then open to release air pressure. Reduction of air pressure and presence of myoelastic tension results in a return to the closed position. Each complete vibratory cycle occurs approximately 180 to 320 times per second during voicing. If any part of this cycle is altered, the voice will be affected. For example, if the folds do not close, the end product will be a whisper. Conversely, if the fold closure is longer than normal, the voice will tend to be tense.

FIGURE II.



The Resonance section is an array of semi-discrete subdivisions. Hypernasality and hyponasality are traditional resonance dimensions. However, the inclusion of throaty or guttural and effeminate speech in this section, as shown in Figure III is unique. The Intensity section provides ratings of inappropriately soft, normal and inappropriately loud voice. The Vocal Range section provides ratings on pitch variability. Monotone is a traditional concept. Excessive variability is used less frequently, and refers to the excessive and inappropriate use of inflections and pitch shifts.

The middle section of the form presents a list of vocal features to be marked as present or absent. Some features, such as diplophonia and intermittent diplophonia, are related, but most features in this section are independent. This Marking System provides an opportunity to relate ratings on each feature to the overall perception of vocal deviancy. This appears to be a unique profile characteristic.

**FIGURE III**

**RESONATING CAVITY**

NASALITY

hypernasal

C

+4....nasalized vowels and consonants  
(with or without audible nasal air  
emission)

+3....nasalized vowels

+2....assimilated nasality

nor-

throaty or guttural..... -2

1

+2.....effeminate speech

mal

-2....denasal speech

hyponasal

**FUNCTIONS OF THE SCALE**

**Quantification.** The voice profile offers the user four opportunities to quantify observations. First, one may count the number of deviant features identified. Second, on the Voice Rating Scale one may rate the extent to which overall deviancy is perceived. Third, the subdivisions in the Laryngeal Cavity, Resonance Cavity, Intensity and Vocal Range sections allow one to make some judgments of relative severity of individual features, specifically on the Open and Closed features. It may be assumed that buccal speech is more deviant than whispered speech or, on the nasality feature, that nasalized vowels constitute a more severe deviation than assimilated nasality. However, no attempt was made to specify the interval between subdivisions. The profile constructors did not assume that the interval between buccal and whispered speech was the same as the interval between whispered and breathy speech. Therefore, it appears that the subdivisions should be

considered as rank order or nominal scales. Fourth, the extent to which a deviant feature contributes to the overall perception of deviancy may be quantified through the use of the Marking System.

**Comparison.** The profile offers users a way of comparing one speaker with another or with himself over a period of time. Also, users can compare speakers with their own norms for variables such as Intensity or Vocal Range. There are no data accompanying the profile that allow users to compare profile patterns with any normative group, such as spastic dysphonics or speakers with unilateral vocal fold paralysis. Hopefully, data of this type will be available in the future.

**Prediction.** The profile was developed with the hope that it could be used to predict appropriate clinical management procedures and clients' responses to them. To date, no data have been presented concerning this aspect of its use.

**Analysis.** The profile is designed to help listeners identify vocal features that contribute to the overall perception of vocal deviancy and to identify features whose modification will lead to changes in these perceptions. Its value is dependent upon the extent to which its authors selected relevant variable that can be subjected to acoustical or perceptual analysis. In the Open and Closed section the authors have gone one step further by suggesting that perceptual features are related to specific characteristics of laryngeal function. This information is considered to be of importance in clinical management.

To date, no formal validity studies of the profile have been published. As previously noted, the developers' primary goals were to identify vocal features that account for listeners' perceptions of overall vocal deviancy, and that are related to clinical management programs. No formal attempts have been made to test the ability of the profile to serve these goals. At the present time, profile users must accept the developers' report that experienced clinicians who have worked with a large number and a wide range of voice problems have found that the features present in the profile serve their needs. The fact that there are no published validity studies of this profile does not mean that individual features represented in the profile have not been discussed elsewhere. Clinical literature is replete with observations on the significance of pitch, breathiness, tension, nasality and many other features in the perception of vocal deviations and to their clinical management. What we lack is evidence that this profile is a collection of relevant features arranged appropriately to allow users to accomplish specified goals.

If one accepts the potential validity of this tool, the question remains whether or not a clinician can use it in a reliable manner. Again, reliability must be defined in relation to the users one intends to make of the profile. The developers provide only informal comments to support their contention that the profile can be used in a sufficiently reliable manner to accomplish the goals they set for it. While it is not possible to determine all the uses a clinician may make of the profile, it seems appropriate to provide potential users with descriptive data on the reliability of clinicians who do use it. The following sections describe a procedure used to measure reliability and presents the data obtained.

## THE STUDY

**Purpose.** This study has two general purposes. The first is to describe the reliability with which speech clinicians profile recorded samples of children with voice disorders. The second is to compare the ratings of clinicians who have extensive clinical experience with voice disorders to those of clinicians with limited experience.

### **Procedure**

**Voice Samples.** Using tapes available from clinic files, we selected 25 recordings of children, ages six through fifteen, judged to have voice problems. We attempted to select tapes that were of reasonable quality and represented a range of severity.

**Experimental Tapes.** Thirty second samples of speech were extracted from each clinic tape and arranged randomly on a master tape. Five samples were presented twice on the tape. A number preceded each sample and ten seconds of silence followed each sample. Eight copies were made from the master tape and were used as experimental tapes in the study.

**Listeners.** Two groups of listeners were used. One group consisted of 15 undergraduate and graduate students who had just completed a voice course that included three hours of analyzing recordings of children and adults with voice problems. The voice profile was one of the tools used by these students. This group was labeled Inexperienced Listeners.

The other group contained 15 speech clinicians each of who had at least three years of clinical experience and had worked with 15 or more voice cases. All members of this group had participated in at least one voice workshop and had used the Voice Profile in their clinical practice. This group was labeled Experienced Listeners.

**Listening Task.** Each listener was provided with a tape, a tablet of voice profile blanks and a set of directions. The latter included a statement of the general purpose of the study and the nature of the task. This was followed by a brief review of the features of the voice profile and how it is scored. Next, listeners were given an opportunity to profile three sample voice recordings, and compare their profile scores with those most commonly used by a group of speech clinicians. Group scores were available from a previous study. Finally, listeners were told that they could listen to each sample as often as they needed to and that they must score every feature on the profile.

Inexperienced Listeners were provided with a high quality tape recorder with earphones and were instructed to adjust the volume to a comfortable level. Experienced Listeners were asked to listen on the best quality tape recorder available to them and to use earphones if they were available.

**Results.** The purposes of the study were to describe listeners' reliability and compare experienced and inexperienced listeners' performance. In fulfilling these purposes, listener groups and voice features were analyzed separately. Several definitions of listener agreement were used and several indices of reliability computed.

In tabulating results, it was found that the Marking System (i.e. primary feature, secondary feature, etc.) was used by only six listeners and that the features Erratic Phrasing, Immature Resonance and Frontal Resonance were used by five or less listeners. Therefore, these features were not used to tabulate agreement scores and information relating to them is not on any of the Tables.

In tabulating these data, two definitions of listener agreement were used. Complete Agreement indicates that two listeners agree exactly on the rating they used for a voice feature. Partial Agreement indicates that on the Voice Rating Scale two listeners' ratings were identical or within one scale number of each other and on the Laryngeal Cavity, Resonating Cavity, Intensity and Ranges Scales two listeners agreed that the feature was normal (i.e. a rating of 1) or deviant (i.e. a rating of other than 1). Partial Agreement for features rated "yes" or "no" could not be determined.

Using these definitions, a listener's ratings on the 25 voice samples were compared with every other listener's ratings and the mean number of agreements determined. Next a mean of the means was computed for the 15 Experienced Listeners and for the 15 Inexperienced Listeners. Finally, these means were divided by the number of opportunities for agreement (i.e. 25). This yielded Complete Agreement Scores and Partial Agreement Scores. Procedures used to compute these scores are referred to as Total Decision Analysis.

Table I provides a summary of complete agreement. Information in Table I shows that some features were infrequently noted as being deviant. Under these circumstances, the Total Decision Analysis procedure may at times be misleading. An attempt was made to compensate for these circumstances. When a listener's ratings on a feature were compared with those of another listener for the 25 samples we determined the number of samples judged by both, or either one, to be deviant and we divided complete agreements and partial agreements by this number. This is referred to as the Deviation Decision Analysis.

A second purpose of the study was to compare experienced with inexperienced listeners. This purpose was accomplished, in part, by comparing agreement scores, as they appear in Tables I and II using a "t" for independent samples. None of the "t" tests were significant at the .05 level. Inspection of Tables I and II indicates that the largest group differences were on the voice features, Range and Intensity. The smallest were on Severity, Open, Closed, and Nasal Resonance. When large differences occurred, Experienced Listeners had high agreement scores. When the differences were small, no trend was observable.

Table I includes information on the frequency with which features were noted as being deviant by two or more listeners. The "t" tests computed were used to evaluate differences between the Experienced and Inexperienced Listeners. Inexperienced Listeners noted less deviancy on eight of the ten features tabulated.

Five voice samples were rated twice by all listeners. Table III contains data on the percent of repeated judgements on which Complete and Partial Agreement occurred. Intra-judge reliability was similar for both groups.

In summary, the analysis revealed some numerically large differences between ratings made by Experienced and Inexperienced Listeners. However, rating behavior on the total profile did not show group difference to be statistically significant.

Table IV shows the distribution of mean group ratings on the Voice Rating Scale for the voice samples used in the study. The authors interpret these data as indicating that their original selection of subjects did succeed, in that the group selected contained a wide range of severity.

**TABLE I. Complete Agreement. Summary of 15 Experienced [Exp.] and 15 Inexperienced [Inex.] Listeners' Complete Agreement Scores for 25 Voice Samples, based on Total Decision and Deviation Decision Analysis procedures. Summary of frequency with which vocal features were noted as being deviant by two or more Listeners.**

<u>Vocal Features</u>	<u>Total Decision Analysis</u>		<u>Deviation Decision Analysis</u>		<u>Frequency Noted Deviant</u>	
	Exp.	Inexp.	Exp.	Inexp.	Exp.	Inexp.
Severity	$\bar{X}$ 48 SD 12	49 15	*	*	*	*
Pitch	$\bar{X}$ 33 SD 3	47 4	21 6	19 3	52	32
Open	$\bar{X}$ 62 SD 5	65 5	51 6	45 5	88	64
Closed	$\bar{X}$ 66 SD 5	68 5	54 6	51 6	84	68
Nasal	$\bar{X}$ 97 SD 2	95 1	68 4	80 5	16	16
Intensity	$\bar{X}$ 90 SD 2	48 4	20 10	15 8	24	20
Range	$\bar{X}$ 78 SD 6	19 4	16 5	4 5	28	20
Intermittent Diplophonia	$\bar{X}$ 71 SD 9	85 7	25 11	21 8	40	24
Diplophonia	$\bar{X}$ 65 SD 10	50 7	32 10	18 5	36	12
Audible Inhalation	$\bar{X}$ 80 SD 7	65 8	20 13	12 8	24	8
Pitch Breaks	$\bar{X}$ 78 SD 11	64 10	42 11	29 8	36	48

\*Deviation analysis and Frequency noted are not appropriate for severity.



**TABLE II. Summary of 15 Experienced [Exp.] and 15 Inexperienced [Inexp.] Listeners. Partial Agreement Scores for 25 voice samples, based on Total Decision and Deviation Decision analysis procedures.**

<u>Vocal Feature</u>		<u>Total Decision Analysis</u>		<u>Deviation Decision Analysis</u>	
		Exp.	Inexp.	Exp.	Inexp.
Severity	M	85	82	*	*
	SD	5	6		
Pitch	M	61	53	33	40
	SD	4	4	5	3
Open	M	81	77	75	76
	SD	5	3	7	7
Closed	M	79	81	71	73
	SD	3	3	3	6
Nasal	M	98	96	82	85
	SD	2	1	5	5
Intensity	M	91	59	28	59
	SD	3	3	21	7
Range	M	78	63	16	50
	SD	4	3	8	8

\*Deviation Analysis is not appropriate.

**TABLE III. Intralistener agreement on five voice samples rated twice.**

<u>Total Decision Analysis</u>			<u>Listeners</u>	
			<u>Experienced</u>	<u>Inexperienced</u>
Complete Agreement	M	85	81	
	SD	7	7	
Partial Agreement*	M	90	90	
	SD	7	8	
<u>Deviation Decision Analysis</u>				
Complete Agreement	M	46	40	
	SD	13	9	
Partial Agreement*	M	62	65	
	SD	12	14	

\*Partial agreement tabulated for Severity, Pitch, Open, Closed, Nasal Intensity and Range. Complete Agreements tabulated for all vocal features.

TABLE IV. Distribution of Mean Severity Ratings for 25 voice samples.

<u>Severity Rating Scale</u>	<u>Experienced Listeners</u> <u>Ratings</u>	<u>Inexperienced Listeners</u> <u>Ratings</u>
Normal...1	2 samples	1 sample
2	0	1
3	7	7
4	7	6
5	5	5
6	4	5
Extremely Severe...7	<u>0</u>	<u>0</u>
	N=25	N=25

### CONCLUSIONS

Within the limits of the procedures used in this study, it seems reasonable to conclude that this Voice Profile system can be used most reliably by individuals in tasks that require one to determine whether a voice feature is normal or deviant. When tasks require analysis of feature in terms of the subcategories contained in the profile (i.e. +3 Closed, -2 Pitch, etc.), reliability is reduced. Interlistener and intralister performances for experienced and inexperienced listeners appear to be similar. Clinicians who use or contemplate using the Profile, should interpret these findings with caution. Data were collected from recordings of short speech samples; listeners had no prior knowledge of clients to help focus their perceptions; listening conditions were not closely controlled; and there were no specific benefits provided to listeners who agreed to participate. The authors interpret the data obtained, as suggesting that most clinicians who wish to analyze voice deviations in terms of the features included in the profile will find its reliability to be acceptable. Clinicians who wish to analyze features in terms of the subcategories included in the profile may not find it to be reliable in all situations. Some clinicians may find that their perceptions are too variable. Others may find that their clients' behavior varies along dimensions not included in the profile or that the increments of change that interest them are too small or too large to be registered on the profile. Some will find the profile to be useful in a wide range of clinical activities.

### REFERENCES

- Moore, G.P., and von Leden, H., *The Larynx and Voice, Laryngeal Physiology under Daily Stress. Motion picture produced at the Voice Research Laboratory, Northwestern University Medical School, Chicago, 1953.*
- Wilson, F.B., *The Voice Disordered Child: A descriptive approach. Language Speech and Hearing Services in Schools., #4, 1972.*
- Wilson, Frank B., and Rice, Mabel, *A Programmed Approach to Voice Therapy, (Austin Texan: Learning Concept 1977).*