Clinical applications of acoustic voice analysis. Part I: Background and perceptual factors

Fritzell, B. and Hammarberg, B. and Wedin, L.

journal: STL-QPSR
volume: 18
number: 2-3
year: 1977
pages: 031-038

http://www.speech.kth.se/qpsr
B. CLINICAL APPLICATIONS OF ACOUSTIC VOICE ANALYSIS

Abstract

A set of 17 voices representing various voice disorders were presented to a jury of voice experts who rated them on a 5-point scale for each of 27 terms frequently used for describing voices. A statistical treatment of these ratings yielded 5 bipolar factors accounting for 85% of the total variance. Significant correlations were found between 4 of these factors and certain acoustic data extracted from long-time-average spectra and fundamental frequency histograms of the same voice samples. Also, a new preliminary strategy for measuring vocal aperiodicity in sustained vowels is presented and informally evaluated.

Part I. BACKGROUND AND PERCEPTUAL FACTORS

B. Fritzell, B. Hammarberg*, and L. Wedin**

1:1 Introduction

Who has a normal voice? And, who has an abnormal voice? Can we make a definition? - This appears to be a most difficult problem. We certainly have a feeling for what is normal and abnormal, but our feelings are subjective, and our judgements differ. So, we want to make measurements of some kind, in order to be more objective.

A large part of our clinical case load in phoniatrics and logopedics is voice patients. During the three-year period 1974-1976, 47% of the new patients admitted to the Department of Phoniatrics of Huddinge University Hospital had voice problems. We give these patients various kinds of voice therapy, we try to improve the quality of their voices. Could we measure the change brought about by our voice therapy?

It must be recognized that what we hear is not measurable in a simple way. Sound is elusive, it disappears. By means of a tape recorder we can repeat it, but in order to be truly scientific we must transfer the acoustic signal into some kind of permanent optical recording, in the form of figures or a tracing on a paper and use a ruler.

The present study, reported in this paper and the following one, is an attempt to use acoustic voice analysis for clinical purposes, to supplement our clinical observations by acoustic data. The ultimate goal is a handy little electronic box, into which we plug the voice of our patient, and out come valid and reliable figures of relevant acoustic parameters.

* Department of Phoniatrics, Huddinge University Hospital, Sweden

** Institute of Psychology, Stockholm University
And these will enable us to study, communicate, and report diligently in the area of voice disorders, and voice therapy in particular, a deplorably neglected subject in our scientific journals.

Acoustic measurements do not make sense on their own, however, they must be related to perceptual characteristics in order to be clinically useful. The first part of our presentation here is a report on a perceptual analysis of patients' voices. This will be followed by a description of the acoustical methods used, namely fundamental frequency analysis, long-time-average spectrum analysis, and a new type of aperiodicity analysis. In the near future, we will also start making use of acoustic glottography by means of inverse filtering procedures ad modum Rothenberg. Finally, some preliminary results of the comparison of our perceptual and acoustic data will be presented.

The study is a joint project between the Department of Phoniatrics of the Huddinge Hospital and the Department of Speech Communication of the Royal Institute of Technology in Stockholm.

1:2 Perceptual factors

In our search for acoustic measures of voice quality it seemed important to get a starting point in a perceptually based description of voice quality.

In phoniatric-logopedic clinics in Sweden a more or less firmly established set of terms are used to describe voice qualities. These terms, however, are based on impressionistic, perceptual judgements, which make them problematic in at least two respects: (1) they are esoteric, i.e. they are difficult to handle in external communication, and (2) it is uncertain whether they are precise enough to serve as a basis for reliable measurements.

Can we make use of these varying descriptive terms to obtain a frame of reference, which is more precisely defined?

Factor analysis

In our effort to answer this question we have applied factor analysis. The perceptual qualities of human voices can be interpreted as bundles of various traits or characteristics, which are more or less closely connected with each other. By means of factor analysis we can obtain a
limited number of clusters of variables with high intercorrelations. Such clusters are then treated as factors, i.e. scales along which measurements can be made.

Previous studies

Some previous studies of voice perception by means of factor analysis can be found in the literature, such as the ones by Isshiki et al (1969), Isshiki & Takeuchi (1970), and Takahashi & Koike (1975). In these studies the stimulus material consisted of sustained vowels. In all three studies Osgood's semantic scaling method was used. In this case the input terms consist of bipolar verbal scales such as 'awful-nice, sweet-sour, soft-hard', which are commonly used adjectives with no special reference to voice perception.

The variables

Our approach to this difficult problem of finding appropriate terms for perceptual impressions is to base the perceptual analysis on the terminology that is actually used in a phoniatric-logopedic clinic. Therefore the variables for the listening test were chosen among voice descriptive terms found in a corpus of 115 case histories of voice patients in our clinic. The 28 most frequently used terms were picked out for the listening test out of a total of some 50 terms. 23 of the terms refer to voice quality, 2 to pitch, and 3 to register. Some examples of these terms are given in Table II-B-I.

Voice sample. Listeners. Listening test

The set of stimuli consisted of a short story (of 92 words) tape recorded from 17 subjects with various functional and organic voice disorders, such as vocal nodules (2), chronic laryngitis (2), unilateral paralysis (2), polyps and polypoid thickening of the vocal cords (4), contact ulcer (1), laryngeal cancer (1), incomplete mutation (1), and functional dysphonia (4). Three of the voices were duplicated, in order to provide an estimation of the consistency of the listeners, and consequently, the listening tape included 20 voice examples. There were 8 males and 9 females ranging in age from 17 to 70. For the listening test a test form was prepared with the 28 variables. The listeners were the staff of the phoniatric department of Huddinge Hospital, i.e. 1 male phoniatrician and 13 female logopedes. Every voice example was repeated 4 times.
on the listening tape. Listeners rated their auditory impression of the voices on a 5-point scale for each of the 28 variables.

**TABLE II-B-I.**
Examples of the voice describing variables in Swedish and their approximate equivalents in English.

<table>
<thead>
<tr>
<th>Swedish professional term</th>
<th>English equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>läckande</td>
<td>breathy</td>
</tr>
<tr>
<td>skrovlig</td>
<td>rough, raucous</td>
</tr>
<tr>
<td>diplofoni</td>
<td>diplophonia</td>
</tr>
<tr>
<td>knarr</td>
<td>creaky, vocal fry</td>
</tr>
<tr>
<td>press</td>
<td>strained</td>
</tr>
<tr>
<td>väsande</td>
<td>wheezing, hissing</td>
</tr>
<tr>
<td>förskott</td>
<td>instant</td>
</tr>
<tr>
<td>registerbrott</td>
<td>voice breaks</td>
</tr>
<tr>
<td>instabil klang</td>
<td>unstable quality</td>
</tr>
<tr>
<td>instabil röstläge</td>
<td>unstable pitch</td>
</tr>
</tbody>
</table>

The factors

Dr. Lage Wedin at the Institute of Psychology at the University of Stockholm carried out the statistical calculations.

The so-called "Principal Component Analysis" was used. To summarize the results of this factor analysis, calculated on the intercorrelations of the mean ratings of all the 28 variables over the 20 voice examples, 5 factors were extracted representing 85% of the total variance.

The factors all turned out to be bipolar:

**Factor 1** (see Table II-B-II) included the variables bitonalitet "bitonality", diplofoni "diplophonia", instabil röstläge "unstable pitch", fladder "flutter", skrap "harshness", registerbrott "voice breaks" and instabil klang "unstable quality" in the negative pole. The only variable with significant positive loading was återhållen "repressed". We tentatively label factor 1 stabil-instabil "steady-unstable".
FACTOR 1 STEADY - UNSTABLE

† + .535 restrained
- .669 changes of voice quality
- .774 unstable quality
- .774 voice breaks
- .795 harsh (sw. 'skrap')
- .810 flutter
- .827 unstable pitch
- .831 diplophonia
- .872 bitonality

TABLE II-B-II.

FACTOR 2 BREATHY - OVERTIGHT

† + .787 creaky, vocal fry
- .730 husky, veiled
- .773 intermittent aphonia
- .832 lack of timbre
- .849 wheezing
- .899 breathy

TABLE II-B-III.

FACTOR 3 HYPO - HYPERKINETIC

† + .852 strangled phrase endings
+ .831 strained
+ .812 hyperkinetic
+ .698 guttural, throaty
+ .540 glottal stroke
† - .675 monotonous

TABLE II-B-IV.

FACTOR 4 LIGHT - COARSE

† + .930 coarse
+ .858 rough
+ .701 harsh (sw. 'ströv')
- .618 restrained
- .671 middle register
- .760 high pitch

TABLE II-B-V.

FACTOR 5 CHEST - HEAD REGISTER

† + .577 chest register
† - .747 head register

TABLE II-B-VI.

TABLES II-B-II - II-B-VI. Ranking of the significant factor loadings of the variables in each of the 5 factors.
Factor 2 (cf. Table II-B-III) included at the negative pole läckande "breathy", väsande "wheezing", klangfattig "lack of timbre", afoniska inslag "moments of aphonía" and beslöjad "husky" opposed to knarr "creaky/vocal fry" at the positive. Factor 2 therefore was tentatively labeled otät-övertät or läcka-knarr "breathy-overtight".

Factor 3 (cf. Table II-B-IV) has at one pole variables such as monotön "monotonous" and beslöjad "husky", at the other strypa frassut "strangled phrase endings", pressad "strained", hyperkinetisk "hyperkinetic", halsig "guttural, throaty", and hårdansatser "glottal stroke", which made us tentatively interpret factor 3 as hypokinetisk-hyperkinetisk "hypokinetic-hyperkinetic".

Factor 4 (cf. Table II-B-V) consists of högt röstlage "high pitch", mellanregister "middle register" and återhållet "restrained" at one pole contrasted to grov "coarse", skrovlig "rough" and sträv "harsh" at the other. This factor seems to reflect a high-light quality towards a harsh-dark, which makes us tentatively label factor 4 ljus-grov "light-coarse".

Factor 5, finally (cf. Table II-B-VI), is easily described as it turned out to consist of only one variable in each pole: "chest register" versus "head register" (or in Hollien’s terms "modal" versus "loft register").

As a final step in the factor analysis, factor scores were calculated for each of the 20 voice examples (including the three duplicated ones). The factor scores give you the position of each voice on each of the 5 factors; thus for instance voice no. 9 had a high negative score in factor 2, which means a high degree of läckage "breathiness" and voice no. 17 had a high positive score in factor 4, that is a high degree of grov/skrovlig "coarse/rough", cf. Table II-B-VII.

Reliability and validity
Test-retest reliability

In order to test the consistency of the listener’s ratings the retest reliability was calculated for the first and second presentations of the three duplicated voices (i.e. voices no. 3/11, 6/14, and 12/18). Pearson’s correlation coefficient was calculated on the mean ratings of the 28 variables for each paired voice. The correlation coefficients were all very high: .93 - .97. This means that the consistency of the listeners is significantly reliable.
TABLE II-B-VII. The factor scores of the 20 voice examples. The items within squares are commented in the text.

Validity

To test the validity of the factors, i.e. to see if the factor analysis has really produced such factors as are relevant to the perception of voice quality, a second listening test with the same 20 voice examples was given. This time the same listeners as before (except for one drop-out) rated the voices on a 9-point scale for each of the 5 factors. The contents of the factors, as in the Tables II-B-II - II-B-VI just shown, had been explained to the listeners.

The mean values over all listeners per factor were correlated with the factor scores of the voices from the first listening test. The correlations (Pearson's r) were all significant at the 1% level: for the first 4 factors varying from .80 to .91, for factor 5 "registers" .60. The validity of the factors, in the sense in which it has been tested, can therefore be said to be very good.
Discussion of results

By the techniques of factor analysis we have shown in a sample of pathological voices how 28 clinically used voice describing variables can be analyzed as to their interrelationships and reduced to a simpler structure of 5 bipolar factors.

What conclusions can be drawn from this study?

Firstly, the reliability of the listener group, as far as the consistency of the ratings concerns, turned out to be very good, which means that the 28 voice describing variables are highly reliably handled within this group of listeners. It should be borne in mind that these persons have all been in the same clinic for at least 2 years, forming a closely working team.

Secondly, the 5 factors turned out to be valid for voice perception of this sample of pathological voices. This makes us optimistic as to further research. It seems worthwhile to continue along these lines. The extension of the investigation to larger voice samples and to other listener groups might eventually yield a useful perceptual instrument for voice analysis.

Acknowledgments:

This work was supported by the Swedish Medical Research Council (project No. B77-17X-04783-02) and the Bank of Sweden Tercentenary Foundation, contract No. 67/48:4.

References

