The acoustic properties of voice timbre types and the importance of these properties in the determination of voice classification in male singers

Cleveland, T.

journal: STL-QPSR
volume: 17
number: 1
year: 1976
pages: 017-029

http://www.speech.kth.se/qpsr
C. THE ACOUSTIC PROPERTIES OF VOICE TIMBRE TYPES AND THE IMPORTANT OF THESE PROPERTIES IN THE DETERMINATION OF VOICE CLASSIFICATION IN MALE SINGERS*

T. F. Cleveland**

Abstract

An investigation to determine the significance of formant frequencies, pitch, and source spectrum slope on voice classification is reported. Eight professional singers sang five common vowels on four common pitches, and, in a forced-choice test, vocal pedagogues classified the vocalizations as tenor, baritone, or bass. Source spectrum slopes of 12 dB per octave approximated the vocalizations of the three voice classes, while formant frequencies and pitch accounted for most of the difference in jury evaluation results. To evaluate more accurately the significance of formant frequencies, pitch, and source slope on voice classification, vowel synthesis was employed. Representative voice class formant frequency combinations, four different pitches, and source spectrum slope were systematically altered while, vibrato, and vowels were stabilized. The original jury classified the synthesis in a forced-choice test, and the dependence of the jury evaluation upon formant frequencies and pitch was evidenced in analysis of variance tests. Additional experiments based on the above findings suggested the following results and conclusions:

1) Measurements of spoken vowel vocalizations revealed higher formant frequencies in tenor timbre type and lower formant frequencies in bass timbre type.

2) The "average formant frequency" seems to be a better description of timbre type than separate, individual formant frequencies.

3) Long-time-average-spectra seem to afford a good estimate of voice classification.

4) Formant frequency percentage differences between basses and tenors were similar to those found between males and females suggesting that the mouth and pharynx lengths known to differentiate females and males, probably also differentiate tenors and basses.

* This paper is a comprehensive overview of the author's doctoral dissertation presented to the University of Southern California. The research and writing was accomplished during 1973-1975 while the author was a guest researcher at KTH, Dept. of Speech Communication under the sponsorship of the Fulbright-Hays Foundation.

** Author's present address: Laryngeal Dynamics and Vocal Acoustics Laboratory, University of Southern California, School of Medicine, Research Division, Department of Otolaryngology, 903 Hoffman, 2025 Zonal Ave., Los Angeles, California 90033, USA.
Introduction

Voice timbre is often defined as that particular attribute of a given voice which distinguishes that voice from another when the vowel and the pitch are the same. Individual voice timbres may be said to exist along a continuum of a myriad of various timbres. Vocal pedagogues have empirically divided the continuum of various timbres into at least three main timbre divisions: bass, baritone and tenor. Acoustically, individual timbres would correspond to characteristic acoustic signals of the laryngeal source and the vocal tract resonances, while voice timbre types, i.e., bass, bariton and tenor, would be groups of individual timbres which possess similar laryngeal source and vocal tract resonance characteristics.

Because no cataloging of acoustic characteristics of timbre types exists, the realization of such a catalog might assist in the understanding of certain required acoustic properties of timbre types and, more importantly, the ascertainment of the importance of timbre type in the determination of voice classification might assist in the development of a theory of voice classification for both speaking and singing voices. The present work is intended to contribute to such a theory.

Scope of this study

The present investigation encompasses a series of experiments involving voice timbre and voice classification in male singers. First, the study seeks to ascertain the acoustic properties which are influential in the perceptually recognized timbre types of bass, baritone and tenor. Second, the study probes the importance of timbre type in the determination of voice classification. Third, the study presents rough calculations of "typical" timbre type vocal tract lengths computer from formant frequency measurements.

Experiment I - Real Sounds

Method

In order to determine the spectral characteristics of timbre types, eight professional Swedish male singers sang the vowels [i], [e], [a], [o], and [u] in a prearranged scheme on the pitches C3, F3, A3 and E4. The source spectra and formant frequencies of these vocalizations were determined by analysis by synthesis on a terminal analogue of the vocal tract. The formant frequencies of lower pitches were also determined by sonogram measurement.
A tape of commonly spliced (common onset/decay), randomly ordered, vowel vocalizations was prepared for the perceptual evaluation of eighteen vocal pedagogues. The individual vocalizations were presented three consecutive times with six seconds of silence following the last presentation affording the jurors an opportunity to classify the sounds as bass, baritone or tenor.

The listening test which was played to the jury on a tape recorder (Revox #77) was heard by the jury via headsets (Sennheiser HD 44). Because of the large number of sample sounds (5 vowels x 4 pitches x 8 subjects) the test was divided into two sessions of approximately 25 minutes each. An interval "break" of 30 minutes separated the two sessions. The jury was instructed to not discuss the test during the "break".

Conversion of jury response to jury number

A number representative of the jury evaluation (JN) of each sound stimulus was computed from the formula \(2_T + 1_{Br} - 2_B = JN\), where \(T\), \(Br\), and \(B\) are the voice timbre types Tenor, Bariton and Bass, respectively.*

Results

a. Formant frequencies

Correlation of the average formant frequency to the mean jury number of each subject may be seen in the scattergram of Fig. 1-C-1. The average formant frequency is an average of the four lowest formant frequencies of the combined vowels \([i], [e], [a], [o], [u]\) of each singer-subject and the mean jury number is the mean of the jury numbers of all the vocalizations by an individual singer-subject.

A correlation coefficient of 0.90 was calculated showing that a change in the mean jury number corresponded closely to a change in the average formant frequencies of the singer-subject. The statistical significance of the correlation coefficient was tested by transforming it to a "t" statistic. The correlation coefficient of 0.90 corresponds to a "t" statistic of

* The formula \(1_T + 0_{Br} - 1_B\) was felt to reflect the linear relationship which probably exists between voice timbre types in the remaining experiments of this investigation so it was employed. However, results from this new formula did little to change the results from the previous formula except lower all jury numbers by 7 jury numbers with a standard deviation of \(\pm 1\) jury number.
MEAN JURY NUMBER
AS A FUNCTION OF AVERAGE
FORMANT FREQUENCY
[i, e, a, o, u]
6.94, which, with six degrees of freedom, is significant at the .001 level of confidence. From these results it seems reasonable to assume that one of the acoustic properties on which the jury evaluation was dependent was formant frequencies.

It is of significant interest to note the grouping of the singer-subjects as shown in Fig. 1-C-1. It appears that the jury placed the singer-subjects into three different groups. The groups may be readily distinguished in the figure. The lowest group is comprised of HH and BE; the second group is made up of JA, SS and UB; and the third group is comprised of SEA, LS and CS. Furthermore, because the groups are so explicitly delineated, it may be assumed that the jury used a criterion which is related to the average formant frequency. * Average formant frequency boundaries between the groups might be appropriately drawn at 1.69 kHz (bass and baritone) and at 1.79 kHz (baritone and tenor).

b. Pitch

In order to determine if the different pitches elicited a different response from the jury, the mean jury number was correlated to the various pitches. In this case, computation of the mean jury number was derived from the average jury numbers of all vowels from a single pitch by a given singer-subject. For purposes of correlation, pitches were defined by their fundamental frequency. The results of this correlation may be seen in Fig. 1-C-2. The scattergram shows that the jury votes changed with a concurrent change in pitch. A correlation coefficient of .81 was computed for these data, and it seems reasonable to infer that pitch was also an influential acoustical property in the jury's determination of voice timbre type.

c. Source

As the timbre type may be dependent on the source spectrum, the source spectra of the singer-subjects in this investigation were studied. Resulting measurements from the study determined little consistent deviation from a source falling at -12 dB per octave for the various singers.

* An extensive examination of several representative values of formant frequencies suggested that the "average formant frequency" as utilized in this experiment was a representative of the formant frequencies as any number that was explored. A similar number has been used in other investigations, e.g. Coleman, 1973.
JURY NUMBER AS A FUNCTION OF PITCH
[i, e, a, o, u]

Fig. I-C-2.
on pitches F₃ and A₃. However, slight deviations from -12 dB per octave were found in two tenors and a bass. The bass deviation appeared on the highest pitch the bass sang, E₄, and the tenor deviation appeared on the lowest pitch sung by the tenors, C₃. Graphs representing the mean source spectrum deviations from a -12 dB octave slope for the vowels [i], [e], [a], [o], [u], are shown in Fig. I-C-3. A source falling at -12 dB per octave does not deviate from the zero dB line as shown in letter a. The mean deviations of the tenor and bass may be seen in letters b and c, respectively.

Both pitches E₄ and C₃ represent extremes in the pitch range of both voice types. Consequently, it may be hypothesized from this finding that a source spectrum slope of less than -12 dB per octave, as seen in Fig. I-C-3c, may be associated with extreme "vocal effort". Such may be the rational for the source slope deviation of the bass singer subject on the pitch E₄, which is the top extreme of his range. The deviation close to the tenor's lowest pitch, C₃, as shown in Fig. I-C-3b, may perhaps be explained as a consequence of a lack of vocal efficiency, which may be typical of the lowest pitches singers produce in singing. For the most part, however, the overall results of the source spectrum analysis suggest that the source spectrum of the three voice timbre types is not consistently different from the -12 dB per octave source spectrum slope frequently utilized in synthesis.

Discussion

The results regarding the source spectrum seem to support findings of Sundberg (1973). His investigation showed that the source spectrum differed only slightly between "dark" and "light" voices, and, in addition, Sundberg stated that "the development of voice timbre in voice training would be a matter of learning a special articulation rather than having the vocal chords to vibrate in a very special way".

Conclusions

Summarily, the results of this investigation suggest that both the formant frequencies and pitch may be influential in timbre type determination. Of course it may be assumed that the source spectrum contributes to the individual timbre of a given voice, but the results indicate that a special source spectrum does not appear to belong to the typical features of a given voice category. The results of this section may be
**SOURCE SPECTRUM SUMMARY**

<table>
<thead>
<tr>
<th></th>
<th>Harmonic number</th>
<th>measured</th>
<th>synthesis (Tenor)</th>
<th>synthesis (Bass)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dB</td>
<td></td>
<td>Δ</td>
<td>x</td>
</tr>
<tr>
<td>&quot;NORMAL&quot;</td>
<td>&quot;NORMAL&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOURCE</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Harmonic number</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>329</td>
<td>1645</td>
<td>3290</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Harmonic number</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>650</td>
<td>1300</td>
<td>1950 Hz</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Harmonic number</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>1645</td>
<td>3290</td>
<td>4935 Hz</td>
</tr>
</tbody>
</table>

Fig. I-C-3.
viewed somewhat tentatively, however, as there are many uncontrolled variables in real sounds. More definite conclusions may be drawn if the acoustic properties which seem important in the generation of timbre types are synthesized and aurally evaluated by the same group of vocal pedagogues.

**Experiment 2 - Synthesis**

**Method**

A series of vowel sounds were synthesized which reflected the acoustic properties found to be influential in timbre type determination of real sounds. The synthesis was accomplished utilizing the source-filter network shown in Fig. I-C-4. Formant frequencies and pitch were systematically varied while the source spectrum and vibrato were kept constant. A synthesis evaluation tape of the synthesized sounds, which was similar to the evaluation tape of the real singer-subject vocalizations, was made and exposed to the original jury of vocal pedagogues, who classified the sounds as bass, baritone or tenor. The formant frequencies were derived from coordinates of a best fit line through formant frequencies of individual formants and equally spaced points along a continuum of jury numbers. The best fit line was drawn through a scattergram of formant frequencies of a given formant from all singer subjects on all pitches and the equally spaced jury numbers along a continuum were chosen by the investigator as representative of the various timbre types. A scattergram example from which data points were derived are given in Fig. I-C-5. Pitches were the same as those used by the singer-subjects and the vibrato was $\pm 2\%$ of the fundamental. This vibrato generation is consistent with human vibrato sounds (Wolfe et al., 1965).

**Results**

The results of the jury evaluation were statistically measured by univariant analysis of variance (Finn, 1974) on the representative vowels [i] [a] [u]. The results of this test are shown in Table I-C-I. The statistical F values (F) for the variable formant frequencies, here referred to as formant type, and pitch are both significant at the .001 level of confidence.

As can be seen in Table I-C-I an additional test of orthogonal polynomial contrast was also accomplished. Such a test determines how the
VOWEL SYNTHESIS
SOURCE-FILTER DIAGRAM

Fig. I-C-4.
RELATIONSHIP OF JURY NUMBER TO FORMANT FREQUENCY

Fig. I-C-5.
### TABLE I-C-I. Univariant analysis of variance results

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Formant type</td>
<td>2</td>
<td>1621.75</td>
<td>36.56</td>
<td>.0001</td>
</tr>
<tr>
<td>(lin.)</td>
<td>(1)</td>
<td>(3243.37)</td>
<td>(73.11)</td>
<td>.0001</td>
</tr>
<tr>
<td>(quad.)</td>
<td>(1)</td>
<td>(.125)</td>
<td>(.003)</td>
<td>.9582</td>
</tr>
<tr>
<td>Pitch</td>
<td>3</td>
<td>3881.29</td>
<td>87.49</td>
<td>.0001</td>
</tr>
<tr>
<td>Interaction</td>
<td>6</td>
<td>175.45</td>
<td>3.96</td>
<td>.0069</td>
</tr>
<tr>
<td><strong>Within</strong></td>
<td>24</td>
<td>44.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Univariant analysis of variance test results showing highly significant dependence of jury note on the synthesis treatments, formant type, pitch and the interaction of the two. Furthermore, orthogonal polynomial contrasts show formant type treatments to be linearly related with a statistical significance beyond the .01 level of confidence.
means are related. For example, it asks the question, Do the means of the different treatment groups, as in this case, formant type, have a linear, quadratic or cubic relationship? As can be seen in the table the statistical F value is significant at the .01 level of confidence for the linear (lin) contrast of means. This result indicates that the means of the formant type treatments-bass, baritone and tenor are different, and their differences fall on a straight line. Fig. I-C-6 illustrates the linear relationship between mean jury number and formant type for the vowels [i, a, u].

Table I-C-I also shows the interaction of the two variables, formant type and pitch, as tested by the univariant analysis of variance. The interaction is significant at the .006 level of confidence.

Discussion

It is interesting to compare the results obtained from analysis of the real sounds and the synthesized sounds. Where in the real sounds the formant frequencies seemed to be a more important acoustic property than pitch, the reverse seems to be the case with the synthesized sounds. Possible explanations for this difference may be as follows: 1) For analysis, the jury numbers from the synthesized results were "normalized" with respect to pitch and formant type, whereas "normalized" results in the real sounds were not feasible. Thus, "normalized" results were not compared to "normalized" results. In addition, 2) The crude method utilized to derive the formant frequencies for the synthesis may be responsible for some of the difference.

Conclusions

Regardless of the difference between the real and synthesized statistical results, the results are compatible and suggest that it is reasonable to conclude that timbre type is dependent on formant frequencies and pitch. However, we may regard the influence of pitch on timbre type determination as a property more of theoretical interest rather than practical import, and, as such, from a practical and pedagogical point of view, it may be concluded that timbre type is most practically influenced by formant frequencies.

Experiment 3 - Practical import of timbre type

Introduction and method

Since voice timbre type seems quantifiable through the utilization of formant frequencies or its counterpart the average formant frequency,
JURY NUMBER AS A FUNCTION OF FORMANT TYPE
(IPA SYMBOLS)

Fig. 1-C-6.
as shown in Fig. I-C-1, the second purpose of this investigation—determining the practical import of timbre type in voice classification—was explored. To do so, it was necessary to ascertain the importance of other properties important in voice class determination. Range is such a property. The center range was determined by finding the range of a given singer-subject on a piano with 81 keys and assigning the most central pitch in that range the number of that key as they were counted from bass to treble. This logarithmic number was preferred over the linear number which would have prevailed if the pitch frequency had been employed. Fig. I-C-7 shows the correlation of the center range with autoclassification (the professional singing classification of the singers in this investigation).

Results

The correlation coefficient computed for these data was .97 suggesting an almost perfect correlation between center range and autoclassification. From this result it seems reasonable to assume that, practically speaking, the center range might be regarded as indicative of the autoclassification of a professional singer. Consequently, a high correlation between center range and average formant frequency would suggest that an estimate of the autoclassification might also be derived from the average formant frequency alone.

Fig. I-C-8 is a scattergram of center range and average formant frequency (the number utilized in this experiment to quantify timbre). The plot supports the conclusion—the higher the center pitch of the range, the higher the average formant frequency—though there is one case in which this does not hold—singer subject HH does not conform to the trend in the remaining data. In spite of this result, however, the finding seems to suggest that, as a rule, the formant frequencies and center range tend to "match", and since center range seems to be a good indicator of professional voice classification, formant frequencies or the average formant frequency of a singer-subject might, therefore, be a good indicator of professional singing classification.

Further assistance in classifying voices might be gained if it could be determined that formant frequencies in speech correlated to autoclassification. Results from an extensive study of speech sounds of the professional singers in this experiment reflect that formant frequencies of speech sounds correlate highly to autoclassification. Measured
AUTOCCLASSIFICATION AS A FUNCTION OF CENTER RANGE

Fig. I-C-7.
CENTER RANGE AS A FUNCTION OF AVERAGE FORMANT FREQUENCY

Fig. I-C-8.
formant frequencies from several vowels were correlated to the auto-
classification of the singer-subjects and the resulting correlation coef-
ficients are given in Fig. I-C-9. As seen in the figure, the high corre-
lation coefficients seem to indicate that measurements of formant fre-
quencies of speech vocalizations may be helpful in voice class deter-
mination.

Discussion

The results of these experiments seem to be valuable for singers
and pedagogues alike for the findings may offer explanations to the peda-
gogical observations heretofore unexplained. Pedagogues have often
made the observation with reference to untrained singing voices, "If the
quality is there, the range will come". It seems possible to suggest an
explanation from the evidence in this study as to what a teacher means
when he makes such a statement.

Range is usually a product of vocal maturity and is normally fully
extended only in trained singers. On the other hand, this experiment
has shown that voice timbre is related to formant frequencies which
depend on vocal tract size and dimensions (Fant, 1960). If timbre is a
result of vocal tract size, it may be suggested that timbre would exist
at an earlier age than range in the beginning singer, as it would be ex-
pected that a singer would reach a more physically mature state than
that stage at which his range would be more fully developed. Therefore,
timbre may be a better early estimator of voice class in the beginning
singer than range. The pedagogue who has made this observation, em-
pirically, has possibly recognized through years of observant voice
teaching that timbre in the untrained voice is a good indicator of the
future voice class of the singer because, in professional singers, as
shown in Fig. I-C-8, the timbre and the range tend to "match". The
same explanation may even apply for the more mature singer who pos-
sesses a wide, ambiguous range, encompassing more than that required
for single voice class. In determining the voice class of this singer,
the pedagogue may be seeking the best "match" between timbre and range
utilizing timbre as an empirical cue to assign the appropriate voice class.

For the pedagogue who has witnessed a singer who possesses baritone
range, but tenor quality, a similar explanation may hold. Such an im-
pression may be evoked by the empirical observation that a given voice
CORRELATION COEFFICIENT SUMMARY of the RELATIONSHIP of FORMANT FREQUENCIES in SPOKEN VOWELS to SINGING VOICE CLASS

Fig. I-C-9.
is a "mismatch" voice. As this experiment shows that, generally, if the formant frequencies are high, the range may also be high, the pedagogue may be expressing empirically the same point that timbre and range should "match".

Conclusion
---
It seems in order to conclude that in the absence of an extended range of pitches, and in the presence of an ambiguous, extensive range of pitches, timbre may be a good estimate of the future voice class of a given singer.

Experiment 4 - Long-Time-Average-Spectra

Introduction and method
---
Since the finding of this experiment seems to have much practical value, a more expedient method than that of sonogram measurement and spectrum matching should be realized. Such a method may exist in long-time-average-spectra. LTAS is a single spectrum of the averaged acoustical components of short or extended signals as a function of time.

In speech LTAS peaks are located around the average formant frequencies of the three lowest formants (Müller, 1973). In singing, three peaks generally appear, but the third peak, the singing formant peak, is more pronounced (Jansson and Sundberg, 1972). It would be anticipated that singers singing the same task on the same pitch would reflect characteristic "timbre" spectra.

Results
---
Such is the case in Fig. I-C-10 which reflects a comparison of LTAS of two professional singers singing the same vowel scheme on the same fundamental frequency. Singer-subject BE's autoclassification is bass, and he was classified by the vocal pedagogues as possessing bass timbre. Singer-subject LS has the professional singing classification of a tenor and was classified by the jury as a singer with tenor timbre. The figure clearly delineates the timbre differences.

Conclusion
---
The results from this LTAS, and a systematic study of different LTAS on other tasks allows the conclusion that LTAS holds much potential in the ascertainment of timbre type and voice classification.
Superimposed LTAS of bass (BE) and tenor (LS) showing consequential spectral differences resulting from the singing of the same vowel scheme on the common pitch $A_3$.

Fig. I-C-10.
Experiment 5 - Morphology

Introduction

Since this study has shown characteristic formant frequency patterns for bass, baritone and tenor timbre, morphology of the vocal tract must be related to voice timbre type and voice classification. Studies of various investigators have shown that, in most vowels, the formant frequencies of women and children are considerably higher than those of males (Fant, 1960; Peterson and Barney, 1952) and that these differences are due to the relatively greater pharyngeal length and more highly developed laryngeal cavities of adult males over that of females, and that same physical maturity of females over children. However, within these groups no systematical subdivisions seem to have been established.

Results

Characteristic formant frequencies typical of the voice timbre type (as determined by the jury of vocal pedagogues) and the autoclassification (professional singing classification) of the singers in this study are given in Table I-C-IIa and b, respectively. The characteristic frequencies are presented by vowel and represent an average of measured formant frequencies of all singer-subjects who possess a like timbre type as a result of jury classification or a similar voice class as a consequence of autoclassification.

Scattergrams of the correlation of the formant frequency difference (measured in per cent) between males and females and between autoclassified basses and tenors, and jury classified bass timbre and tenor timbre is shown in Fig. I-C-11. The results were obtained using the formula \[ F_n^\% = \frac{F_{it} - F_{ib}}{F_{ib}} \times 100\% \], where \( F \) equals the formant frequency in question, "i" is the formant number, and "t" and "b" equals tenor and bass, respectively. Comparison of the percentage differences between jury classification (dashed line) and autoclassification (solid line) with female/male (line-dot) of Fant’s study (Fant, 1966) shows a remarkable correspondence in most cases. Especially close relationships are seen in \( F_1 \) of [\( a \)] and [\( o \)] and \( F_2 \) of [\( i, e, a, u \)]. As can be seen in the Figure, the largest percentage differences between females and males are also the largest percentage differences between tenors and basses. Since it is known that the female/male differences are due to relatively greater pharyngeal length and more highly developed laryngeal cavities
<table>
<thead>
<tr>
<th></th>
<th>F₁</th>
<th>F₂</th>
<th>F₃</th>
<th>F₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>312</td>
<td>1996</td>
<td>2602</td>
<td>3116</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Bar</td>
<td>287</td>
<td>1864</td>
<td>2500</td>
<td>2973</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>3</td>
<td>20</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Bass</td>
<td>278</td>
<td>1557</td>
<td>2312</td>
<td>2800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>F₁</th>
<th>F₂</th>
<th>F₃</th>
<th>F₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>352</td>
<td>1942</td>
<td>2424</td>
<td>3041</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>11</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Bar</td>
<td>348</td>
<td>1744</td>
<td>2339</td>
<td>2950</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-1</td>
<td>14</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Bass</td>
<td>350</td>
<td>1533</td>
<td>2075</td>
<td>2822</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>F₁</th>
<th>F₂</th>
<th>F₃</th>
<th>F₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>623</td>
<td>1005</td>
<td>2620</td>
<td>2919</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Bar</td>
<td>522</td>
<td>942</td>
<td>2478</td>
<td>2823</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-1.5</td>
<td>1</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Bass</td>
<td>530</td>
<td>934</td>
<td>2298</td>
<td>2721</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>F₁</th>
<th>F₂</th>
<th>F₃</th>
<th>F₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>389</td>
<td>698</td>
<td>2757</td>
<td>3010</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-6</td>
<td>-7</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Bar</td>
<td>414</td>
<td>750</td>
<td>2569</td>
<td>2938</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>13</td>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Bass</td>
<td>366</td>
<td>708</td>
<td>2557</td>
<td>2599</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>F₁</th>
<th>F₂</th>
<th>F₃</th>
<th>F₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>339</td>
<td>683</td>
<td>2538</td>
<td>2944</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-4</td>
<td>-2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Bar</td>
<td>326</td>
<td>700</td>
<td>2522</td>
<td>2943</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-3</td>
<td>-6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Bass</td>
<td>336</td>
<td>742</td>
<td>2404</td>
<td>2700</td>
</tr>
</tbody>
</table>

TABLE I-C-II.a Average formant frequencies of timbre types
TABLE I-C-II. b Average formant frequencies of voice classes

[i ]

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>304</td>
<td>1969</td>
<td>2567</td>
<td>3105</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>9</td>
<td>13</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Bar</td>
<td>278</td>
<td>1744</td>
<td>2482</td>
<td>2897</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-7</td>
<td>10</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Bass</td>
<td>300</td>
<td>1587</td>
<td>2214</td>
<td>2752</td>
</tr>
</tbody>
</table>

[e ]

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>350</td>
<td>1942</td>
<td>2414</td>
<td>3061</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>0</td>
<td>17</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Bar</td>
<td>350</td>
<td>1662</td>
<td>2247</td>
<td>2873</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-2</td>
<td>8</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Bass</td>
<td>356</td>
<td>1539</td>
<td>2041</td>
<td>2754</td>
</tr>
</tbody>
</table>

[a ]

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>609</td>
<td>994</td>
<td>2576</td>
<td>2909</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>15</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Bar</td>
<td>530</td>
<td>944</td>
<td>2400</td>
<td>2849</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Bass</td>
<td>503</td>
<td>900</td>
<td>2386</td>
<td>2527</td>
</tr>
</tbody>
</table>

[o ]

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>401</td>
<td>724</td>
<td>2706</td>
<td>2989</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Bar</td>
<td>391</td>
<td>711</td>
<td>2554</td>
<td>2906</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>7</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>Bass</td>
<td>365</td>
<td>729</td>
<td>2605</td>
<td>2969</td>
</tr>
</tbody>
</table>

[u ]

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>330</td>
<td>682</td>
<td>2548</td>
<td>2957</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-1</td>
<td>-5</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Bar</td>
<td>333</td>
<td>719</td>
<td>2420</td>
<td>2716</td>
</tr>
<tr>
<td>Δ(%)</td>
<td>-4</td>
<td>-4</td>
<td>-5</td>
<td>-2</td>
</tr>
<tr>
<td>Bass</td>
<td>348</td>
<td>749</td>
<td>2536</td>
<td>2784</td>
</tr>
</tbody>
</table>
A COMPARISON OF TENOR/BASS AND FEMALE/MALE FORMANT FREQUENCY DIFFERENCES

Fig. I-C-11.
of adult males over those of females, it is likely that similar morphological distinctions exist between tenors and basses.

Since these vocal tract differences are likely to exist between tenors and basses, it is of interest to approximate what the differences in vocal tract length might be. Fant (1960) has shown that F2 in the vowel [i] is clearly a half wave-length resonance of the back cavity. Furthermore, he has linked the dependence of the front cavity to F3 in the vowel [i]. Utilizing Fant's table of percentage dependence of given formant frequencies on various cavities (Fant, 1960) hypothetical vocal tract lengths of both tenors and basses can be approximated. Calculations show that overall vocal tract lengths for basses and tenors is approximately 19 cm and 15.5 cm, respectively, when singing the vowel [i].

Discussion

It is significant to note that these findings are not inconsistent with the empirical observations of various laryngologists. Brodnitz (1961) observed that, as a rule, high voices are found in persons with round faces and short necks, while long faces and long narrow necks seem frequent in singers with deep voices. Similar observations have been made by Arnold (1965) and van Deinse (1974).

Conclusions

As a consequence of the various experiments presented in this paper the following summary of conclusions seems to be in order:

1. Voice timbre type seems significantly dependent on formant frequencies.
2. Voice classification may be estimated from voice timbre acoustic cues - specifically the average of the four lowest formant frequencies.
3. Measured formant frequencies from speech sounds seem indicative of voice classification.
4. Long-time-average-spectra seems to afford a good estimate of voice classification.
5. The mouth and pharynx lengths known to differentiate females and males may also differentiate tenors and basses.

Acknowledgments

The author is deeply indebted to Dr. Johan Sundberg for his scholarly guidance in the research and writing of this paper. The encouragement of Professor Gunnar Fant, and the stimulating atmosphere provided
by the members of the Department of Speech Communication, KTH is gratefully acknowledged. Appreciation is also expressed to the singer-subjects for their kind co-operation.

References


SYMPOSIA PUBLICATIONS

1. The Proceedings of the Symposium on Auditory Analysis and Perception of Speech, which was held in Leningrad in August 1973, have been published in two parts. One part, containing the neurophysiological and psychoacoustic basis and auditory models, was published in a special issue of Acustica, Vol. 31, No. 6, December 1974. It can be obtained from S. Hirzel Verlag, Stuttgart, BRD.


2. The Proceedings of the Speech Communication Seminar, which was held in Stockholm in August 1974, have been published by Almqvist & Wiksell International, Stockholm 1975 and jointly by Halsted Press, a division of John Wiley and Sons, Inc., New York. These Proceedings are bound in four volumes, one for each session of the seminar, and the title is Speech Communication, Vols. 1-4. G. Fant has edited.

Vol. 1: Speech Wave Processing and Transmission
Vol. 2: Speech Production and Synthesis by Rule
Vol. 3: Speech Perception and Automatic Recognition
Vol. 4: Speech and Hearing, Defects and Aids. Language Acquisition

These final volumes contain 28 articles not included in the preprint version as well as comments on the sessions. The preprint version was distributed to the participants.

The address of Almqvist & Wiksell International is:
P.O. Box 62
S-101 20 Stockholm, Sweden

Symposium on Auditory Analysis and Perception of Speech.
Leningrad, USSR, August 21-24, 1973

INTRODUCTION, by G. Fant

REVIEW OF SESSION I by B. L. Cardozo (co-chairman)

MODEL OF THE PERIPHERAL AUDITORY SYSTEM AS A TOOL IN SPEECH PERCEPTION RESEARCH by L. A. Chistovich (chairwoman)

papers presented in SESSION I: "Neurophysiological and psychophysical basis for designing models of speech processing"

A. R. Möller: CODING OF AMPLITUDE AND FREQUENCY MODULATED SOUNDS IN THE COCHLEAR NUCLEUS

G. V. Garšuni: SOUNDS EMITTED BY BIOLOGICAL SOURCES AND THE FUNCTIONAL CHARACTERISTICS OF CENTRAL AUDITORY NEURONS

I. A. Vartanian: ON MECHANISMS OF SPECIALIZED REACTIONS OF CENTRAL AUDITORY NEURONS TO FREQUENCY-MODULATED SOUNDS

N. G. Bibikov: ENCODING OF THE STIMULUS ENVELOPE IN PERIPHERAL AND CENTRAL REGIONS OF THE AUDITORY SYSTEM OF THE FROG

A. V. Popov and V. F. Shuvalov: TIME-CHARACTERISTICS OF COMMUNICATIVE SOUNDS AND THEIR ANALYSIS IN THE AUDITORY SYSTEM OF INSECTS

T. Houtgast: AUDITORY ANALYSIS OF VOWEL-LIKE SOUNDS

A. Rakowski and A. Jaroszewski: ON SOME SECONDARY MASKING EFFECT

B. L. Cardozo: SOME NOTES ON FREQUENCY DISCRIMINATION AND MASKING

J. Mártony: CHANCE OF INTENSITY DISCRIMINATION LIMEN IN CHILDREN

papers presented in SESSION II: "Modelling the peripheral processing of the speech wave"

V. K. Labutin: ANALOG OF AUDITORY NERVE RESPONSE TO ACOUSTIC SIGNALS AND ITS RELATION TO SOME PSYCHO-ACOUSTIC PHENOMENA

N. G. Zagoruiko and V. G. Lebedev: MODELS FOR SPEECH SIGNALS ANALYSIS TAKING INTO ACCOUNT THE EFFECT OF MASKING

I. A. Chistovich et al: A FUNCTIONAL MODEL OF SIGNAL PROCESSING IN THE PERIPHERAL AUDITORY SYSTEM

J. J. Zwislocki: A POSSIBLE NEURO-MECHANICAL SOUND ANALYSIS IN THE COCHLEA

R. Carlson, G. Fant, and B. Granström: TWO-FORMANT MODELS, PITCH, AND VOWEL PERCEPTION (summary only)

M. R. Schroeder and J. L. Hall: "A MODEL FOR MECHANICAL TO NEURAL TRANSDUCTION IN THE AUDITORY RECEPTOR"
Session III. Vowel perception

R. Plomp: "Auditory analysis and timbre perception"

L. C. W. Pols: "Analysis and synthesis of speech using a broad-band spectral representation"


R. Carlson, G. Fant, B. Granström: "Two-formant models, pitch, and vowel perception"

R. Carré, R. Lancia: "Perception of vowels amplitude transients"

A. V. Baru: "Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration and number of formants) in dog" (a summary)

W. A. Ainsworth: "Intrinsic and extrinsic factors in vowel judgments"

A. Q. Summerfield, M. P. Haggard: "Vocal tract normalization as demonstrated by reaction times"

K. Centmayer: "Interrelations of vowel perception and linguistic context"

E. Fischer-Jørgensen: "Perception of German and Danish vowels with special reference to the German lax vowels"

L. V. Bondarko, L. A. Verbitskaya: "Factors underlying phonemic interpretation of phonetically non-defined sounds"

Session IV. Feature detection and auditory segmentation.

Consonant perception

H. Fujisaki, K. Nakamura, T. Imoto: "Auditory perception of duration of speech and non-speech stimuli"


V. V. Lublinskaja, W. Mikiel: "Scaling of pitch intervals between vowel-like sounds with constant and variable fundamental frequency"


J. M. Pickett, E. M. Danaher: "On discrimination of formant transitions by persons with severe sensori-neural hearing loss"

D. H. Klatt, S. R. Shattuck: "Perception of brief stimuli that resemble rapid formant transitions"

K. N. Stevens: "The potential role of property detectors in the perception of consonants"

G. I. Tsemel: "Application in speech recognition, some data on auditory segmentation and speech wave parameters' perception"

Session IV, cont.

M. Derkach: "Acoustical cues of softness in Russian syllables and their application in the automatic speech recognition"

W. Tscheschner: "A contribution to the investigation into the partial functions of human speech processing"

M. Schönfeld: discussion on Tscheschner's paper


Session V. Temporal organization of connected speech. Prosody

B. Lindblom: "Some temporal regularities of spoken Swedish"

I.H. Slis: "Consequences of articulatory effort on articulatory timing"

S.G. Nooteboom: "On the internal, auditory representation of syllable nucleus durations"

J.J. Ohala: "The temporal regulation of speech"

A.W.F. Huggins: "On isochrony and syntax"

R.K. Potapova: "Auditory estimate of syllable and vowel duration in sentences"

L.V. Zlatoustova: "Rhythmic structure types in Russian speech"

P. Janota, J. Ondráčková: "Some experiments on the perception of prosodic features in Czech"

O.F. Krivnova: "Some notes on melody component of intonation as a cue in text segmentation (a summary)"

N.D. Svetozarova: "The inner structure of intonation contours in Russian"

R. Haavel: "On the significance of some parameters of a pitch contour"

L.P. Blochina: "On the question of auditory analysis of frequency parameters in sentences (a summary)"

W. Jassem: "Normalization of F0 curves"

J. Baštecký, B. Borovičková, Vl. Maláč, O. Vinař: "Analysis of speech errors under delayed auditory feedback"

G.H. Coker, N. Umeda: "Sub-phonemic variations in American English"

Session III: Chairman review by G. Fart

Session IV: Chairman review by K.N. Stevens

Session V: Chairman review by S.G. Nooteboom
Table of contents
Vol. 1 SPEECH WAVE PROCESSING AND TRANSMISSION

FOREWORD by G. Fant, A. Risberg, and S. Felicetti, p. I.
CHAIRMAN'S COMMENTS ON SESSION 1 by J. N. Holmes, pp. II-IV.
CHAIRMAN'S COMMENTS ON SESSION 1 by M. R. Schroeder, pp.
V-IX.
SOME COMMENTS ON LINEAR PREDICTIVE CODING by B. Gold,
pp. X-XXI.
ATAL, B. S.: "Towards determining articulator positions from the
NAKAJIMA, T., OMURA, H., TANAKA, K., and ISHIZAKI, S.: 
"Estimation of vocal tract area functions by adaptive inverse
filtering methods and identification of articulatory model",
pp. 11-20.
RICE, D. L.: "Articulatory tracking of the acoustic speech signal",
op. 21-26.
ATAL, B. S. and SCHROEDER, M. R.: "Recent advances in predictive
coding - applications to speech synthesis", pp. 27-31.
33-34.
MAKHOUJ, J.: "Linear prediction vs. analysis-by-synthesis",
pp. 35-43.
WAKITA, H. and GRAY Jr., A. H.: "Some theoretical considerations
for linear prediction of speech and applications", pp. 45-50.
ROZSYPAL, A. J.: "Computer supported gating of speech signals",
pp. 51-55.
CARTIER, M. and GRAILLOT, P.: "Reduction and reconstitution
of spectral data", pp. 57-64.
JANSEN, L. P. C.: "Relations between some of the parameters of
vowel sounds", pp. 65-70.
McLARNON, E., HOLMES, J. N., and JUDD, M. W.: "Experiments
with a variable-frame-rate coding scheme applied to formant
synthesizer control signals", pp. 71-79.

LÄNGLE, D. and PAULUS, E.: "Efficiency and limitations of
linear transformations in digital speech transmission", pp. 81-86.
POLS, L. C. W.: "Intelligibility of speech resynthesised by using
a dimensional spectral representation", pp. 87-95.
SUZUKI, H., OYOYAMA, G., and KIDO, K.: "Analysis-convrsion-
synthesis system for improving naturalness and intelligibility
of speech at high-pressure helium gas mixture", pp. 97-105.
SUZUKI, J. and NAKATSUI, M.: "Perception of speech uttered
under high ambient pressure", pp. 107-114.
* GARDINI, B. and SERRA, A.: "Identification of speech parameters
using a recursive method", pp. 119-128.
* JAYANT, N. S.: "Spectral analysis of speech based on linear
delta modulation", pp. 129-137.
* LUKATELA, G., TOMIĆ, T., and DRAJIĆ, D.: "Efficiency
and accuracy in vocal pitch determination by some methods
using digital computer", pp. 139-147.
* HOLLIEN, H., FEINSTEIN, S. H., ROTHMAN, H., and HOLLIEN, P.: 
"Underwater SONAR systems (passive) in humans", pp.
149-157.

* these papers were not published in the preprint version
Table of contents

Vol. 2 SPEECH PRODUCTION AND SYNTHESIS BY RULES

CHAIRMAN'S COMMENTS ON SESSION 2 by J. L. Flanagan, pp. I-II.

CHAIRMAN'S COMMENTS ON SESSION 2 by B. Lindblom, pp. III-XIX.


MANSELL, P.: "Some EMG data on the articulation of [f] and [p]", pp. 23-32.


van der GIET, G. and LANCE, D.: "Influence of time on formant frequencies estimated by a time dependent horn equation", pp. 113-119.


FANT, G., STÅLHAMMAR, U., and KARLSSON, I.: "Swedish vowels in speech material of various complexity", pp. 139-147.

NORD, L.: "Vowel reduction-centralization or contextual assimilation?", pp. 149-154.


SELL, K.: "A digital speech synthesiser for simultaneous voice output on several independent channels", pp. 327-331.

* these papers were not published in the preprint version
Vo13 SPEECH PERCEPTION AND AUTOMATIC RECOGNITION

CHAIRMAN’S COMMENTS ON SESSION 3 by M. Haggard, pp. I-VIII.
CHAIRMAN’S COMMENTS ON SESSION 3 by D. H. Klatt, pp. IX-XXVIII.


SUMMERFIELD, A. Q.: "Processing of cues and contexts in the perception of voicing contrasts", pp. 77-86.


NOOTEBOOM, S. G.: "Contextual variation and the perception of phonemic vowel length", pp. 149-153.

GARNES, S.: "Implications of the perception of duration", pp. 155-164.


MIYUKA, M. and HENARD, J. S.: "Word recognition based either on stationary items or on transitions", pp. 257-263.


These papers are not published in the preprint version.
Table of contents
Vol. 4 SPEECH AND HEARING, DEFECTS AND AIDS.

LANGUAGE ACQUISITION

CHAIRMAN'S COMMENTS ON SESSION 4 by I. J. Hirsh, pp. I-IV.

CHAIRMAN'S COMMENTS ON SESSION 4 by I. Lehiste, p. V.


MARTONY, J., and AGELFORS, E.: "Two psychoacoustic tests with severely hard of hearing children", pp. 69-76.


HIRSH, I. J.: "Recent developments in psychoacoustic tools and hearing aids", pp. 79-82.


MILLER, J. D.: "Preliminary research with a three-channel vibrotactile speech-reception aid for the deaf", pp. 97-103.


*these papers were not published in the preprint version