Articulatory differences between spoken and sung vowels in singers

Sundberg, J.

journal: STL-QPSR
volume: 10
number: 1
year: 1969
pages: 033-046

http://www.speech.kth.se/qpsr
III. SINGING AND SPEECH

A. ARTICULATORY DIFFERENCES BETWEEN SPOKEN AND SUNG VOWELS IN SINGERS

J. Sundberg

In a previous report the acoustical differences between spoken and sung vowels, articulated by four trained bass singers, were investigated. (1) Certain differences were observed in the formant frequencies of the vowels. These acoustical discrepancies obviously have some articulatory background.

In the present study we will investigate such articulatory differences. Articulatory positions in spoken and sung vowels were analyzed by means of X-ray pictures and, as regards the lips, photographies. The acoustical implications of the articulatory differences were determined by means of an electrical line analogue (LEA) simulation technique.

Material

Three of the subjects presented in the earlier report served as subjects. One of them, D2, had no possibility to submit himself to the radiological investigations. This, however, was of minor importance, since his formant frequency changes were of the same nature as those of the other subjects.

The same nine Swedish long vowels as in the preceding report were investigated: [u], [o], [a], [æ], [e], [i], [y], [u], and [ø]. These vowels were spoken in the context /en dV:t/, and lateral X-ray pictures were taken during the articulation of the sustained vowel. The sound produced was simultaneously registered on a tape recorder. The same procedure was repeated for the sung vowels produced on a constant note.

The X-ray pictures included the entire vocal tract with lips, glottis and the frontal part of the cervical vertebrae (Fig. III-A-1). Neither filter nor contrast were used.

Frontal X-ray pictures of the larynx, taken previously when the subjects spoke or sang an [æ], were also utilized in this part of the investigation.

Formant frequency changes

The formant frequencies of the vowels, spoken and sung during the X-ray photographing, were determined from sonagrams. The differences
Fig. III-A-1. Tracings of lateral X-ray pictures for the vowel /u/.
Solid contour: spoken vowel.
Dashed contour: sung vowel.
in the formant frequencies between the spoken and sung vowels were found to be of the same type as in the previous investigation. Thus, the sung vowels as compared to the spoken vowels showed the following characteristics: a) the second formant frequency, F2, is lowered in all vowels except in the back vowels [u], [o], [a], and [e]; b) F3 is raised in the back vowels and lowered in the other vowels; c) F4 is lowered in all vowels; d) the frequency distance between F3 and F4 is reduced in all vowels.

One of the subjects, B4, did not produce these differences as clearly as in the previous investigation. Also his articulation differences between the spoken and sung vowels were much smaller than for the other subjects. A formant frequency analysis was therefore made of vowels produced by this subject in some Swedish songs with piano accompaniment. Here, the formant frequencies differed clearly from those of the spoken vowels and in the same manner as for the other subjects. Thus, the articulation of the sung vowels that this subject produced when the X-ray pictures were taken was not fully representative of his articulation in actual singing.

Area functions

A study of the acoustical implications of articulatory differences requires knowledge of the area function of the vowel. Therefore, area functions were worked out for the sung and spoken vowels of one of the subjects, (B1). The procedure used for deriving area functions from the measures available on lateral X-ray pictures of the vocal tract is described in the Appendix.

The agreement between the formant frequencies of the subject and those measured on the analogue provides an indication that the procedure for deriving area functions and computing formant frequencies is reasonably accurate and essentially correct. Moreover, the results support the assumption that the vowels were articulated without nasality also in singing.

The effects of articulatory differences were analyzed in the following way. On the area functions, simulated on the analogue, the area was reduced by 26% over a length of 1.5 cm. The effect of this modification was systematically evaluated at ten different places along the vocal tract. Thus, the first four formant frequencies were measured after each perturbation. The percentual changes in formant frequency plotted as a...
function of the length of the vocal tract inform us about the acoustical sensitivity to small articulatory perturbations at different places in the vocal tract. The mathematical expression of such perturbation sensitivity curves is

$$\frac{\Delta F}{F} = \frac{1}{2} \int_0^\lambda \left[ \frac{\rho}{2A(x)} U^2(x) - \frac{1}{2} \frac{A(x)}{\rho c^2} P^2(x) \right] \frac{\Delta A(x)}{A(x)} \, dx$$

where $F$ is the formant frequency

$\Delta F$ is the formant frequency difference due to the perturbation

$A(x)$ is the area function

$\Delta A(x)$ is the area difference between the unperturbed and perturbed area function

$\lambda$ is the length of the vocal tract

$x$ is the length coordinate of the perturbation

$U(x)$ is the volume velocity at the length $l=x$

$P(x)$ is the sound pressure at the length $l=x$

$\rho$ is the density of air

$c$ is the sound velocity

Experiments with area functions obtained from other subjects than B1 showed that, between subjects, the perturbation sensitivity curves differ only slightly, provided that the place of maximum constriction is located at the same place.

Perturbation sensitivity curves were worked out for the area functions pertaining to the spoken vowels of subject B1. In order to obtain a general view of the effects of perturbing the vocal tract at different places, the perturbation sensitivity curves for each formant of all nine vowels were plotted on the same graph. To make such plots feasible the length was normalized from $\lambda=0$ at the lips and $\lambda=1$ at the glottis. These plots, Fig. III-A-2a-d, yield the information needed for a study of the acoustical implications of differences in the articulation of a given vowel.

Fig. III-A-2d shows, that the subject has to increase his mouth cavity or decrease the volume in the lower pharynx region, if he wants to rise his first formant frequency. According to Fig. III-A-2c a narrowing in the middle of the pharynx will make $F2$ to sink substantially in all vowels,
Fig. III-A-2. (a-d) Percentual changes in the first four formant frequencies due to a perturbation inserted at ten different places along the vocal tract simulated on the electrical line analogue LEA. The perturbation involved a narrowing of the area by 26% over a length of 1.5 cm. The area functions of the Swedish vowels [u], [o], [a], [e], [i], [y], [u], and [β] give values within the dashed lines.
but a narrowing of the cross-sectional area in the mouth will give a rise to F2 only if the area is not constricted before as is the case in [i]. Fig. III-A-2b shows, that F3 can easily be altered by articulation differences in the mouth: a reduced volume immediately behind the incisors and an expanded volume in the soft palate region will rise F3. Finally, Fig. III-A-2a illustrates, that F4 is practically unsensitive to articulation differences outside the larynx tube. An expansion of the sinus morgagni and a narrowing of the entrance to the larynx tube lowers F4 essentially. We will next proceed to study how the subject uses these possibilities to alter his formant frequencies in singing.

Articulatory differences

From a physiological point of view there are several elements in the articulatory organs that are independently variable. In some of these dimensions, notably (1) larynx position, (2) the jaw and the lip openings, (3) the cross-sectional area of the vocal tract in the velum region, and (4) the tongue shape in the back vowels, typical differences were found between spoken and sung vowels. We will now describe these articulatory differences and discuss their consequences as regards the formant frequencies,

(1) The position of the sagittal mid point of the rima glottis related to the lower frontal corner of the second cervical vertebra is shown in Fig. III-A-3a. In speech the larynx assumes a high position in [e] and [i] whereas the rounded front vowels [y], [u], and [ø] are articulated with a lower larynx. A high larynx position in vowels where the tongue is pulled upwards seems natural, since the tongue root and the larynx are attached to each other via the hyoid bone and the epiglottis. Our acoustic analysis indicates that larynx depression would be specially effective in [y]. In singing, however, the big differences in larynx positions are smoothed out and the average position is lower (Fig. III-A-3b). This is in accordance with the findings of Faaborg-Andersen & Vennard, that muscles lowering the larynx are active in singing.(4)

Singing teachers often recommend their pupils to sing [i] as [y]. This might be a simple way to make the pupil lower his larynx, since the normal speech position of the larynx in [y] is much lower than in [i].

A lowering of the larynx implies a lengthening of the pharynx part of the vocal tract. The X-ray pictures showed that this length expansion
Fig. III-A-3. (a) Larynx position in different Swedish vowels. Solid lines: spoken vowels, dashed lines: sung vowels. Subjects: • B1, × B3, ▲ B4.

(b) Difference in larynx position between spoken and sung Swedish vowels. Subjects: • B1, × B3, ▲ B4. The solid line shows the mean values.

(c) Changes in the formant frequencies due to a 1 cm lowering of the larynx as measured on the electrical line analogue LEA.
takes place immediately above the epiglottis. A lowering of the larynx by 1 cm could therefore be simulated by simply inserting two extra length sections at the corresponding place on the analogue. The changes in the formant frequencies obtained are shown on Fig. III-A-3c. F3 is the least sensitive measured formant irrespective of vowel. F1 is the most sensitive formant in the back vowels, and F2 is the most sensitive in the front vowels. F4 is rather sensitive for all vowels.

Thus, the lowered F2 in the sung vowels may be an effect of the lowered larynx, since the degree of larynx lowering as well as the sensitivity of this formant to a lowered larynx both culminate in the front vowel series. Also the reduced frequency distance in the sung vowels between F3 and F4 may be partly explained with reference to the larynx lowering, since F4 is much more sensitive than F3 in all vowels.

(2) The lip and the jaw openings were examined not only on the X-ray pictures, but also in a series of frontal and lateral photos, at least two for every spoken and sung vowel. From these photos it is possible to deduce the relations between the width of the mouth opening and its height and the position of the mouth corners. These relations are required for the estimation of the lip opening area.

The lip opening seems to be strongly dependent on the vocal effort. This circumstance appear likely when we examine Fig. III-A-4 which shows a comparison of the degree of jaw opening at various efforts. As effort increases, jaw opening - and thus, by implication, lip opening - also tends to increase. The depression of the mandible was observed to be larger in the sung vowels. The sound levels of the spoken vowels were lower than those of the sung vowels. The vocal effort was therefore presumably bigger when the subjects sung than when they spoke the vowels. Thus, the larger lip opening in the sung vowels, evident on Fig. III-A-5a, and the lowered jaw could be effects of a difference in the vocal effort rather than characteristics of the sung vowels.

If, however, the lip opening in the sung vowels tends to be of the same size as in the spoken vowels or smaller, in spite of an increased jaw opening, as is the case with [e] and [i], this is likely to be a significant feature of the articulation of these vowels in singing. Another characteristic of the vowels [e] and [i] in singing is a considerably increased protrusion of the lips. This effect is also observed in [y] but is negligible in the
Figure III-A-4. The effect of vocal effort on the depression of mandible. Measurements made by P. Lindblom.

Depression of Mandible
(normal effort)

Strong effort

Weak effort
other vowels, as shown in Fig. III-A-5b. It is interesting that the pro-
trusion as well as the larynx lowering culminates on the two spread front
vowels. This observation is in agreement with the suggestion of Perkell,
that there seem to be some interrelations between the larynx position and
the lip protrusion, together governing the total length of the vocal tract. (7)

An acoustically relevant variable of the lip conditions is the ratio be-
tween its effective length and area, $\frac{l_0}{A_0}$. (8) This ratio was increased
by 50% in the area functions, and the formant frequency changes thereby
obtained are shown in Fig. III-A-5c. F1 and F2 of the most rounded vowels
are most sensitive. F3 is unsensitive in the back vowels but very sensitive
in the front vowels [e], [i], and [y]. Thus, the lowered F3 in the sung
front vowels is probably an effect of the changed lip articulation. F4 ap-
ppeared to be practically unsensitive to the lip perturbation introduced on
the area functions: the maximum change amounted to 5°/oo.

The main effect of a lowered jaw is a rise in F1, provided that the
tongue shape is preserved. (9) If the lowered jaw really is a characteristic
of the sung vowels, it can be seen as a compensation of the effect of the
lowered larynx.

(3) The area in the velum region was expanded in nearly all sung vowels.
Exceptions to this rule were only [u] and [a]. As regards the front vow-
els this expansion may be a physiological consequence of the lowering of
the larynx and the jaw. If the frontal position of the tongue is preserved
and the tongue root and the jaw at the same time is pulled downwards by a
lowered larynx, the result will be an expanded cross-sectional area in the
velum region. In the back vowels the same effect is obtained by means
of an altered tongue shape, as we shall see below.

In another investigation reported by Sovijärvi (10) a lowering of the soft
palate was observed in singing. This does not apply in the case of the sub-
jects of this investigation, eventhough one of the subjects showed a slight
tendency on the back vowels. This confirms once more the assumption
that the sung vowels were articulated without nasatization.

Singing teachers frequently speak of the necessity of "widening the throat"
in singing. Probably, it is the mentioned expansion of the area in the velum
region that gives this sensation. If so, this sensation would mainly be a
sign of a lowered larynx.
Fig. III-A-5. (a) Ratio between the lip opening area of spoken ($A_{Sp}$) and sung ($A_{Su}$) vowels. Subjects: B1, B3, B4. The solid line shows the mean values.

(b) The difference in mouth corner position between spoken and sung vowels. Subjects: B1, B3, B4. The solid line shows the mean values.

(c) Changes in the formant frequencies due to an increase in $L/A$ by 50% as measured on the electric line analogue LER.
The velum is situated at about .35 to .45 in the normalized vocal tract. The acoustic consequences of an area expansion in this region are illustrated in Fig. III-A-2a-d. F1 will rise in the front vowels and be rather unaffected in the back vowels. F2 will drop a little in all vowels. F3 will tend to rise, especially in the back vowels. Thus, the lowering of the larynx and the area expansion in the velum region both contribute to the lowering of F2 in the sung vowels. The rise in F3 of the back vowels may to a certain extent be an effect of the same articulatory difference.

(4) The tongue shape is changed considerably only in the back vowels. In [u] and [o] the tongue tip is given a more frontal position in singing than in speech. At the same time the tongue root is pulled downwards by the slightly lowered larynx. The velar tongue shape in [u] and [o] is thus replaced by a pharyngeal shape, more resembling the tongue shape of an [a]. This means, that the entire form of the frontal cavity is altered in the way demonstrated in Fig. III-A-1. For the spoken back vowels the largest area is found immediately behind the incisors, and for the sung vowels the maximal values are situated in the velar region of the mouth, since the tongue tip fills out a good deal of the frontal part of it. These changes mainly affect F3, as is seen on Fig. III-A-2a-d. The filled out frontal part of the mouth cavity and the expanded posterior part of it both tend to rise F3. The effect on other formant frequencies is small. Here we have a probable explanation of the raised F3 of the sung back vowels.

Discussion

Above we have found support for ascribing certain formant frequency differences between spoken and sung vowels to certain articulatory differences: a) The lower F2 in the front vowels is an effect of the lowered larynx and its physiological consequence in the velum region. b) The rise in F3 in the back vowels is an effect of an altered tongue shape involving a decrease of the cavity immediately behind the incisors and an increase of the posterior part of the mouth cavity. In the front vowels the lowered F3 is an effect of the lowered larynx and, as regards the spread front vowels, also of an increased lip protrusion. c) The lowered F4 is an effect of the lowered larynx. d) The reduced frequency distance between F3 and F4 is a consequence of articulatory differences mentioned under d) and c).
Let us now discuss why the spoken and sung vowels are articulated differently. Is it in order to achieve certain esthetical, or better, acoustical characteristics of the sung note, or is it for physiological reasons?

It is a common observation that the sound level in the frequency range 2.5-3 kHz is abnormally high in sung vowels. This increase is often referred to as the "singing formant". When F2 is low in frequency, as in the back vowels, the spectral level in this frequency range is low. However, it can be raised if the higher formants are brought closer in frequency. This is illustrated in Fig. III-A-6a, showing the spectral envelope of an [u] with the formant values normal for this vowel according to Fant, and with the source characteristic -10 dB/oct.

The graph also shows what happens to the spectral level if F3, F4, and F5 are brought closer in frequency. It is seen on Fig. III-A-6b that this spectral envelope matches an actually sung vowel spectrum envelope of an [u]. Experiments showed that the "singing formant" can be explained in the same manner in other vowels.

The "singing formant" appears especially in the back vowels to give the timbre "brilliance". The singer seems to bring about this spectrum characteristic by a skilled control of the higher formant complex. He does not nasalize and the source spectrum appears to be normal. However, needless to say, other singers may employ different methods.

So far, we have found a plausible reason for the articulatory modifications of the back vowels. In the front vowels, the big difference appears in the larynx position and in the lip opening. Probably, the larynx lowering is made for physiological reasons. It is well known that singing with a high larynx position may damage the vocal cords. A reason for this might be that a low larynx position makes the tissues between the thyroid and cricoid cartilages more lax. The shifts in the relative position of these two cartilages, required for the big pitch variations in singing, could be made with less muscular activity, if the tissues joining them are lax. However, this is a question for future research. The differences in the lip opening might be due to habits associated with the adjustment of the larynx position.

Finally, it shall be pointed out, that this investigation is based on observations from a restricted number of bass singers singing a rather low note. It is of course possible that the observed articulation in singing is dependent on the type of voice, the pitch, and the shape of the vocal tract.
Fig. III-A-6. (a) Spectral envelopes of the vowel [u] simulated on a terminal analogue and demonstrating the effects of a reduced frequency distance between the third, fourth, and fifth formants. Source characteristic: -10 dB/oct.

(b) Spectrum of a sung [u]. The dashed curve is the same as the chain dashed curve on the upper figure.
Acknowledgments

This work was made possible by support from the Tri-Centennial Fund of the Bank of Sweden and by valuable discussions with Prof. G. Fant and Dr. B. Lindblom at the Dept. of Speech Communication, KTH. Dr. K. Lagergren and his staff at the Radiological Dept. D at Karolinska Sjukhuset, Stockholm, and Mr. G. Garpendahl of the Dept. of Speech Communication, KTH, are gratefully acknowledged for their kind assistance in taking the X-ray pictures viz., the lip photos. The author also wishes to thank the singers for their kind cooperation.

References:


(6) Personal communication with Dr. B. Lindblom, who also offered Fig. III-A-4a.

(7) Perkell, J.S.: "Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study", The Massachusetts Institute of Technology (1968).


On the problem of obtaining area functions from lateral X-ray pictures of the vocal tract

The problem of deriving an area function from the measures available on lateral X-ray pictures of the human vocal tract has been dealt with by several authors. \(1,2,3\) A method for obtaining measures from X-ray pictures is described by Heinz & Stevens and Öhman \(3,4\). A polar coordinate system is used the center of which is the center of a circle tangent to the horizontal portion of the hard palate and the back pharynx wall.

It has been stated, that the cross-sectional area in the mouth region, \(A_M\), is a power function of the distance \(a_M\) between the hard palate and the tongue.

\[
A_M = k \cdot a_M^\alpha,
\]

where \(k\) and \(\alpha\) are constants. According to Heinz & Stevens, these constants have different values in different parts of the mouth. The constants were determined from plaster casts of the mouth by measuring the areas for different positions of the tongue.

Such measurements have recently been made by the author on plaster casts of three different subjects' mouths in order to determine the variations of the constants. These constants were found to have about the same values regardless of the place in the mouth, as is evident from the graphs on Fig. 1a-c. Fig. 1b shows the function obtained from the described polar coordinate system as well as measures obtained when the areas were measured at an angle normal to the upper teeth contour. The figure indicates that only small errors are made, if the constants \(k\) and \(b\) are assumed to have the same values in the entire mouth region, and that the angle to the mid line of the vocal tract is of minor importance.

Heinz & Stevens obtained good results estimating the cross-sectional areas in the pharynx region, \(A_{Ph}\), as the area of an ellipse. The distance between tongue contour and back pharynx wall, \(a_{Ph}\), is taken as one axis and the other axis varies as follows: 3 cm in the upper pharynx region, 2.5 cm behind the epiglottis, 2.0 cm between the epiglottis and the entrance to the larynx tube, and 1.5 cm in the larynx tube.
Fig. 1 (Appendix). (a-c) The cross-sectional area in the mouth region plotted as a function of the palate-tongue distance for three subjects.
On the other hand, Fant has published tomograms* of the pharynx, showing that the major axis of the equivalent ellipse decreases substantially with $a_{ph}$ in narrow passages as found for example in back vowels. (5) An examination of Fant's tomographic data showed that the side walls of the pharynx cavity tend to approach each other in the manner indicated on Fig. 2a, when the distance between the tongue and the back pharynx wall exceeds a certain value (1.9 cm). This value is exceeded in the vowel [i]. Fig. 2b shows the area as a function of this distance.

For the estimation of the lip opening the measures available from a lateral X-ray picture are unsufficient as long as the relation between the width and height of the opening and the position of the mouth corners are not known. These relations seem to differ substantially from one subject to the other.

The area values in the larynx tube can be estimated as the area of an ellipse with the anterior-posterior measure as one axis and the left-right measure from a frontal X-ray picture as the other.

**Measurements**

The above-mentioned principles were used for deriving area functions for nine Swedish long vowels, spoken and sung by a bass singer. The measured $a_{ph}$ and $a_{M}$ were obtained in the following way. First, a midline in the sagittal plane was established by means of a grid of the type described by Heinz & Stevens. Thereafter the values of $a_{ph}$ and $a_{M}$ were measured at every half cm along this line and perpendicular to it, and defined as the distance from the palate viz, back pharynx wall contours to the midline of the tongue contours on the X-ray pictures. The lip opening area was estimated by deducing the width of the mouth from its height and from the position of the mouth corners. (6) These relations were determined from a series of frontal and lateral photographs of the subject speaking and singing the vowels several times.

As stated in ref. (2) the area function of the larynx tube is extremely important to the fourth formant frequency which is almost entirely governed by the shape of the larynx tube and the length of the vocal tract.

* made by Dr. P. Edholm, Karolinska Institutet, Stockholm
Fig. 2 (Appendix). (a) The major axis of the equivalent ellipse giving the cross-sectional area in the pharynx region, if the anterior-posterior measure is used as the minor axis. The major axis is plotted as a function of the back pharynx wall-tongue distance.

(b) The cross-sectional area in the pharynx region plotted as a function of the back pharynx wall-tongue distance.
of the subject. Furthermore, the quantization steps of the analogue were too big to achieve the required accuracy of the area. The area function of the larynx tube was therefore given a form that gave a correct value of the fourth formant frequency.

The differences between the formant frequencies of the subject and those measured on the analogue are given in Table I. The differences are rather small in spite of the rather crude method.

Big percentual differences appear only for the first formant, especially for the front vowels. These negative differences might be due to the formant frequency raising effect caused by the wall inductances as discussed by Fant. (5)

References:
The formant frequencies of the subject B1 and those of the derived area functions, measured on the electrical line analogue LEA. Sp refers to spoken vowels and Si to the sung vowels.