A quantitative theory of cardinal vowels and the teaching of pronunciation

Lindblom, B. and Sundberg, J.

journal: STL-QPSR
volume: 10
number: 2-3
year: 1969
pages: 019-025

http://www.speech.kth.se/qpsr
B. A QUANTITATIVE THEORY OF CARDINAL VOWELS AND THE TEACHING OF PRONUNCIATION*

B. Lindblom and J. Sundberg

Abstract

The problem of devising a reference system for specifying the phonetic value of vowels is discussed. The classical theory of Cardinal Vowels is examined as well as previous quantitative frameworks for describing vowel pronunciation. An attempt is made to construct a model of vowel production that combines the idea of a reference system with an objective numerical method of specification. A set of vowels is generated that represent the most extreme vowels in terms of the total acoustic vowel space characteristic of the model. These sounds are selected so as to be also approximately equidistant acoustically. These model-based "cardinal vowels" are compared with a set of true cardinal vowels pronounced by Daniel Jones. The two sets display many qualitative similarities both acoustically and auditorily. The implications of a vowel reference system that could indeed be produced "from written descriptions", is touched upon in particular with regard to the teaching of pronunciation to hard of hearing as well as normal students. An analysis-by-synthesis application is suggested in which the model is used to supplement visual spectral displays of vowel sounds with articulatory interpretations. Some of the problems that would have to be overcome in such a method of "automatic articulation instruction" are mentioned, for instance, those of normalization and compensatory articulation.

The Cardinal Vowel System

In the teaching of pronunciation one principle is to describe the unknown sounds to be learned in relation to sounds that are known to the student. Phoneticians have pointed out that the sounds of a given language cannot be used for this purpose since there are large variations among speakers of the same language owing to dialectal, sociallogical, and other factors (1). Instead of teaching, for instance, vowel pronunciation in terms of the vowel sounds of a particular language a reference system that is independent of any given language has been devised. A famous example of such a system is the Cardinal Vowels. These sounds are said to be "a set of fixed vowel-sounds having known acoustic qualities and known tongue and lip positions"(2). There are a primary set and a secondary set of cardinal vowels each set comprising eight vowels.

* This paper was presented at the Second International Congress of Applied Linguistics, Cambridge, England, 8-12, Sept. 1969.
It is customary to refer to a given cardinal vowel by its number and to describe its articulation in terms of three dimensions: tongue height, front-back position of the tongue and degree of rounding. In Table I-B-1 below we show the articulatory specifications and symbols used to describe the primary and secondary sets.

**TABLE I-B-1**

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>close</td>
<td></td>
<td></td>
</tr>
<tr>
<td>half-close</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>half-open</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>open</td>
<td>a</td>
<td>a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>y</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>close</td>
<td></td>
<td></td>
</tr>
<tr>
<td>half-close</td>
<td>θ</td>
<td>γ</td>
</tr>
<tr>
<td>half-open</td>
<td>æ</td>
<td>θ</td>
</tr>
<tr>
<td>open</td>
<td>Ω</td>
<td>Ω</td>
</tr>
</tbody>
</table>

The following features are said to characterize these sounds:

1. They are independent of the vowels of any language.
2. They are fixed reference points of "exactly determined and invariable quality".
3. They are peripheral vowels. Thus in principle it should be possible to describe an arbitrary vowel quality of any language by interpolating between the reference points.
4. They are auditorily equidistant.
5. Moreover, "the values of cardinal vowels cannot be learnt from written descriptions; they should be learnt by oral instruction from a teacher who knows them".

The last property indicates that the system is not an objective and quantitative one but relies heavily on the motor skills and perceptual acuity of the student. It is passed on by oral tradition.

**Quantitative Frameworks of Vowel Specification**

From the early fifties onwards acoustic phonetics has made rapid progress. Among the achievements in this field are the schemes devised by Fant (4) and Stevens and House (5) to study the relation between vowel articulations and their acoustic results. We are referring to the three-parameter models of these investigators. The following three parameters...
are controlled in these models:

1. length and opening area of lip section,
2. position of maximal tongue constriction,
3. the magnitude of this constriction.

On the basis of these three numbers the cross-sectional areas along the vocal tract are derived with the aid of rules that differ somewhat between the two versions of the three-parameter model. Given the distribution of cross-sectional areas along the tract, or the area function, the acoustic determinants of vowel quality, the formant frequencies, are computed. According to this type of model possible vowel articulation is defined as any permissible combination of parameter values the parameters being dimensions of the area function.

Although these frameworks of vowel specification are objective and quantitative which the cardinal vowel system is not they sometimes specify "possible vowel articulation" in too generous a fashion and in a manner which is not always easy to interpret in intuitively meaningful articulatory terms such as open-close, front-back, etc.

A Model of Vowel Production

In the present paper we shall report on an attempt to construct a model that combines the idea of an articulatory and perceptual reference system inherent in the cardinal vowel theory. Our aim has been to build into the model all that we know at present about the natural degrees of freedom of the vocal tract. In so doing we hope that we might arrive at an improved definition of the notion of "possible vowel articulation".

A. Articulatory Properties

Our model is controlled by means of the following independent components:

- **the mandible** whose movements we restrict to a single fixed path.
- **the tongue** whose shape can be varied continuously by linear interpolation between three basic configurations of tongue contours corresponding to [i], [a] and [u], respectively. A certain mixture of palatalization, velarization and pharyngealization ("[i]-ness", "[u]-ness" and "[a]-ness", respectively) corresponds to certain numerical values of these parameters.
This choice can be justified partly on the basis of data obtained from lateral X-ray profiles of Swedish vowels. It turns out that three main families of tongue contours are obtained provided that these contours are plotted with the lower jaw as reference (Fig. 1-B-1). An explanation of this rather restricted set of contours is readily apparent when we think of the arrangement of the major extrinsic muscles of the tongue. We find the genioglossus, styloglossus and the hyoglossus muscles which seem mechanically capable of participating in the contraction patterns underlying the production of [i], [u], and [a], respectively (7).

labio-muscular activity (rounding-spreading)

larynx height

which is independent of jaw position.

All of these parameters lend themselves naturally to an interpretation in terms of "muscle lengths". To compute a sound wave from such specifications the procedure is the following:

1. Choose parametric values (jaw opening, tongue shape, rounding-spreading, larynx height).

2. Compute the associated articulatory profile, that is, the contours of the vocal tract and its length in a lateral projection.

3. Translate the result of 2 into an area function (the variation of cross-sectional area along the tract).

4. Compute the formant frequencies corresponding to this area function.

In Fig. 1-B-2 the three basic tongue shapes are shown. A coordinate system anchored on the mandible is also depicted. This system is used to compute interpolated tongue shapes. At the top left we see the parameters of width and height of lip separation with the aid of which the opening area is derived. Below a lateral profile tracing is shown with another coordinate system. This system is used to compute the area function.

B. Acoustic Properties

Now assume that like the child learning to talk, we combine different mandible positions, tongue shapes, lip states and larynx heights in all possible ways and listen to the acoustic result in each individual case. Whatever we do with our articulatory components it is clear that the human speech organs are constrained in such a way so as to permit only certain vowel qualities, or combinations of formant frequency values.
Fig. I-B-1. Midsagittal tongue contours for Swedish vowels in relation to outline of mandible. Top left: [i, e, ɛ, æ]. Top right: [u]. Below: [ɔ, o, ø].
In the upper left-hand part the parameters determining the mouth opening area $A$ are shown: $h =$ vertical separation between lips; $w =$ distance between mouth corners; $p$ is a number that specifies the curvature of the lip contours. These contours when projected on a frontal plane are assumed to be given by

$$A = h \cdot w \cdot \frac{p}{p+1}$$

In the upper right part the basic tongue shapes of the model are shown. A polar coordinate system defined in relation to the mandible is also indicated. With the aid of this coordinate system interpolated tongue shapes associated with $[i, u]$ and $[a]$ were computed.

In the lower part of the figure a lateral X-ray tracing can be seen. Superimposed on the profile is a coordinate system defined in relation to fixed structures such as the maxilla. This system was used in the determination of area functions.
Certain other $F_3-F_2-F_1$ combinations characterize vowels that we could produce only with the aid of a terminal analogue synthesizer - not with our mouths. From the point of view of the human speech mechanism such combinations would be impossible. The space that characterizes the acoustic possibilities of our model is shown in Fig. I-B-3. In this figure we have separated the rounded and spread subspaces. The lower fields refer to all possible combinations of $F_2$ and $F_1$ values. The top areas pertain to the corresponding $F_3$ values. When we explore the contours of these spaces auditorily we find the qualities indicated by the vowel symbols. These points have been selected at approximately equidistant $F_1$ steps.

True and Model-Based Cardinal Vowels

Clearly the first four features mentioned on p. 20 as characterizing cardinal vowels apply also to the vowels generated by the model. They are independent of any language, they are peripheral and fixed reference points and they are acoustically (if not auditorily) equidistant.

Consequently it would be of some interest to compare a set of model-based vowels with a set of true cardinal vowels. Fig. I-B-4 demonstrates the results of an acoustic comparison of this type. The first three formant frequencies were measured in a set of cardinal vowels as spoken by Daniel Jones (1). The left plot in this figure shows the extreme vowels generated with the model (also shown in Fig. I-B-3). It is seen that the true cardinal vowels span a somewhat larger range. This difference is probably due to the fact that, among other factors, Daniel Jones has a shorter overall tract length than that of our model and that he also deliberately shortened his tract by elevating his larynx to an extreme position when producing [i] and probably [a]. Qualitative similarities do exist between the sets, however. Table I-B-2 contains a specification of the articulatory parameters that underlie the model generated sounds.

Table I-B-2. The articulatory dimensions of peripheral vowel types

<table>
<thead>
<tr>
<th>Tongue shape</th>
<th>Palatal</th>
<th>Palato-pharyngeal</th>
<th>Pharyngeal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td>i</td>
<td></td>
<td>e</td>
<td>u</td>
</tr>
<tr>
<td>JAW</td>
<td>e</td>
<td></td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>€</td>
<td></td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

Fig. I-B-3. The maximal rounded and spread vowel spaces that the model is capable of generating.
Fig. 1.B-4. Acoustic comparison of a set of true and a set of model-based cardinal vowels.
Implications of a Quantitative Theory of Cardinal Vowels for the Teaching of Pronunciation

In theory, a physiological model of vowel production should be a useful tool in the teaching of pronunciation to second language learners and hard of hearing children. In practice, such an application would require solutions to some technical problems which, by no means, however, appear insurmountable. Imagine that the formant pattern of a vowel in a given language or dialect could be measured automatically and represented as a dot on an oscilloscope screen. $F_2$ and $F_3$, or some measure combining these two frequencies, could be plotted along the ordinate and $F_1$ along the abscissa. Technologically this is not wishful thinking. Attempts have already been made along these lines (6). Suppose that we first plot a target vowel as indicated in Fig. I-B-5. Next we ask our subject who might for instance be a deaf child to produce a vowel as close to the target as possible. In all probability our pupil will miss the target perhaps in the manner indicated in Fig. I-B-5. We might at this stage have the child try to improve his pronunciation by a trial-and-error procedure. An even better method would be to supplement the visual display with an indication of how the child should change his articulation in order to reach the target. Such information could be given in terms of trajectories depicting the formant frequency shift associated with "isolingual" and "isomandibular" conditions. A set of such curves is given in Fig. I-B-5. This instrument would serve as a sort of "automatic articulation instructor".

This goal might also sound utopian but it should be possible given a computer and an acceptable theory of vowel production.

The present model of vowel production is obviously unsatisfactory as such a theory. There are a number of modifications that are clearly needed to improve it*. There is for instance the problem of normalizing F-pattern data for talkers whose vocal tract lengths differ often non-uniformly in size (8). There is also the problem of compensatory articulations that our model at present often incorrectly allows in the case of non-peripheral vowels. Although it might be possible to produce an [ø]

* e.g., independent control of tongue blade in retroflexion and coronal consonant articulation; independent control of pharynx width in connection with the tense-lax distinction on which the vowel harmony of the Akan languages seems to be based; hyperpalatalization, velarization and pharyngealization to produce compensatory articulations and different types of dorsal constrictions larger than those for vowels; control of the form of the cross-sectional area (lateralis, fricatives etc.).
MOVE TONGUE FORWARD!

PUPI~S VOWEL FIRST FORMANT FREQUENCY

Fig. I-B-5. Stylized visual display of spectral properties of vowels intended to be used in improving the pronunciation of vowels by hard of hearing subjects. The target vowel and the pupil's vowel are represented by dots. The trajectories indicate the formant frequency shifts that would result if the pupil changed his articulation as indicated, that is, by lowering his jaw and by moving his tongue forward. These trajectories are assumed to be based on an analysis-by-synthesis of the pupil's vowel using a quantitative model of vowel production of the type described in the paper.
with spread lips by compensating elsewhere along the tract it is extremely rare to find a Swedish native talker who consistently spreads his [s] sounds. Among other things there seems to be a principle of disambiguation that is yet to be discovered in future research.

References:


(3) D. Abercrombie: Elements of General Phonetics (Edinburgh 1967).


Acknowledgments

This work was supported by the National Institutes of Health Research Grant No. NB 04003-07 and the Tri-Centennial Fund of the Bank of Sweden Contract No. 67/48.