

## On the Interpretation of Vowel "Quality": The Dimension of Rounding

LEIGH LISKER

*University of Pennsylvania  
and  
Haskins Laboratories*

The usual description of vowels in respect to their "phonetic quality" requires the linguist to locate them within a so-called "vowel space," apparently articulatory in nature, and having three dimensions labeled high-low (or close-open), front-back, and unrounded-rounded. The first two are coordinates of tongue with associated jaw position, while the third specifies the posture of the lips. It is recognized that vowels can vary qualitatively in ways that this three-dimensional space does not account for. So, for example, vowels may differ in degree of nasalization, and they may be rhotacized or r-colored. Moreover, it is recognized that while this vowel space serves important functions within the community of linguists, both the two measures of tongue position and the one for the lips inadequately identify those aspects of vocal tract shapes that are primarily responsible for the distinctive phonetic qualities of vowels (Ladefoged 1971). With all this said, it remains true enough that almost any vowel pair of different qualities can be described as occupying different positions with the space. Someone hearing two vowels in sequence and detecting a quality difference will presumably also be able to diagnose the nature of the articulatory shift executed in going from one vowel to the other. Asked to reproduce the vowel quality ([i] and then [a]), a subject is likely to change tongue position along both the high-low and the front-back dimensions of the vowel space. If familiar with phonetic jargon, the subject is likely to call the first vowel high and front, the second low and back. In other words, the change will be factored into a shift along two dimensions of the articulatory space. A shift from [i] to [u] also involves two dimensions of the vowel space, this time the front-back and the unrounded-rounded. The speaker will possibly be equally accurate in diagnosing the articulatory basis of this vowel quality change.

It is often said that there is a relatively simple relation between tongue position and the lowest resonances of the vocal tract, so that a given talker's vowel quality, insofar as it is interpreted as tongue position, can be inferred from a sound spectrogram (Potter, Kopp and Green 1947; Joos 1948; Ladefoged 1975). As we tell students beginning the study of phonetics: the "higher" a vowel is, the lower is its first formant frequency; the "backer" it is, the lower the second formant. In other words, tongue position "maps" quite well into formant frequencies; and, it might seem, vice versa.

Complicating this picture, however, is the fact that the rounding feature also leaves its trace in the resonance pattern, and that its effect is somewhat like that of tongue backing. The simple rules relating vowel height and degree of fronting to the frequencies of the lowest two formants work pretty well for most of the vowels of English, and even better for languages with a smaller vowel inventory, for example Spanish and Japanese. But if backing the tongue and pursing the lips have similar acoustic effects, then we are entitled to wonder just how well the standard articulatory vowel space, with its three mutually independent dimensions, corresponds to the listener's perceptual vowel space. In English and a good number of other languages the vowel inventories include only front unrounded and back rounded vowels. Thus for English speakers generally there is no need, and may

not be the competence to disentangle the acoustic consequences of tongue backing and lip rounding, and quite possibly in perceiving speech they take advantage of the fact that they are not required to distinguish **both** front from back **and** unrounded from rounded qualities. But speakers of certain other languages, and linguists in their descriptive exercises, **do** have to do so. The question arises: How do they factor out the effects of tongue backing and lip rounding, given the assumption that both map into the same two or three formants that are generally said to be sufficient cues to vowel identity?

Before we address this question, it would be nice to have some assurance that it is not premature. That is, we first want to know how well linguistically sophisticated listeners do in fact separate out variations in lip and tongue position. There are some hints in the literature to suggest that lip position is **not** so readily inferred from the acoustic signal (Abercrombie 1967). In any vowel transcription exercise we are well advised to look at as well as listen to the speaker. (After all, the practice of speech perception purely by ear is for the sighted a luxury of rather recent invention.) On the other hand, it could be argued that if the phonetic quality of vowels is in fact entirely a matter of auditory perception, then any uncertainty as to the precise contributions of tongue and lips should be taken as evidence for the linguistic irrelevance of so specific an articulatory description. This view, of course, runs counter to the strong claim of some current theorizing that what we perceive when engaged in speech interpretation is precisely a human vocal tract engaged in linguistically relevant signaling.

To get some answer to the question of how unambiguously the rounding feature is conveyed by the acoustic signal, I turned to the set of so-called cardinal vowels that Daniel Jones recorded some time in the '50s. These are eighteen isolated vowel qualities meant to serve as standards for users of the alphabet of the International Phonetic Association. The IPA letters representing those qualities are shown in Figure 1 in the positions assigned them with a simple representation of the three-dimensional vowel space.

According to Jones's description, sixteen of the eighteen qualities are distributed evenly along the eight vertical edges of the volume, with four levels of height and two of rounding. Two additional qualities allow for three degrees of fronting for the highest vowels.

Linguists trained in the British tradition are likely to have undergone a rigorous drilling in the production and recognition of these eighteen vowel qualities as recorded by Daniel Jones. Indeed, persons who have not enjoyed such training have no right, according to Peter Ladefoged, to claim competence in judging vowel quality. Competent observers are, in his words, "phoneticians who are thoroughly conversant with the exact quality of the original cardinal vowels as a result of prolonged training and instruction by Daniel Jones, or by one of his pupils" (Ladefoged 1967). American linguists, so far as I know, neither give nor receive this kind of drill. Despite this deficit, over the past several months I have arranged to have a number of linguists and other speech researchers in Philadelphia and New Haven participate in an exercise that involved labeling Jones's eighteen vowels. Instead of having to identify the eighteen presumably distinct qualities, a really formidable task without the intensive training that Ladefoged prescribes, participants were asked to decide which of them had been produced with lip rounding—in other words, to make a simple binary choice. For this task it seemed appropriate to exclude two kinds of persons: phoneticians rigorously schooled to recognize Jones's cardinal vowels and those

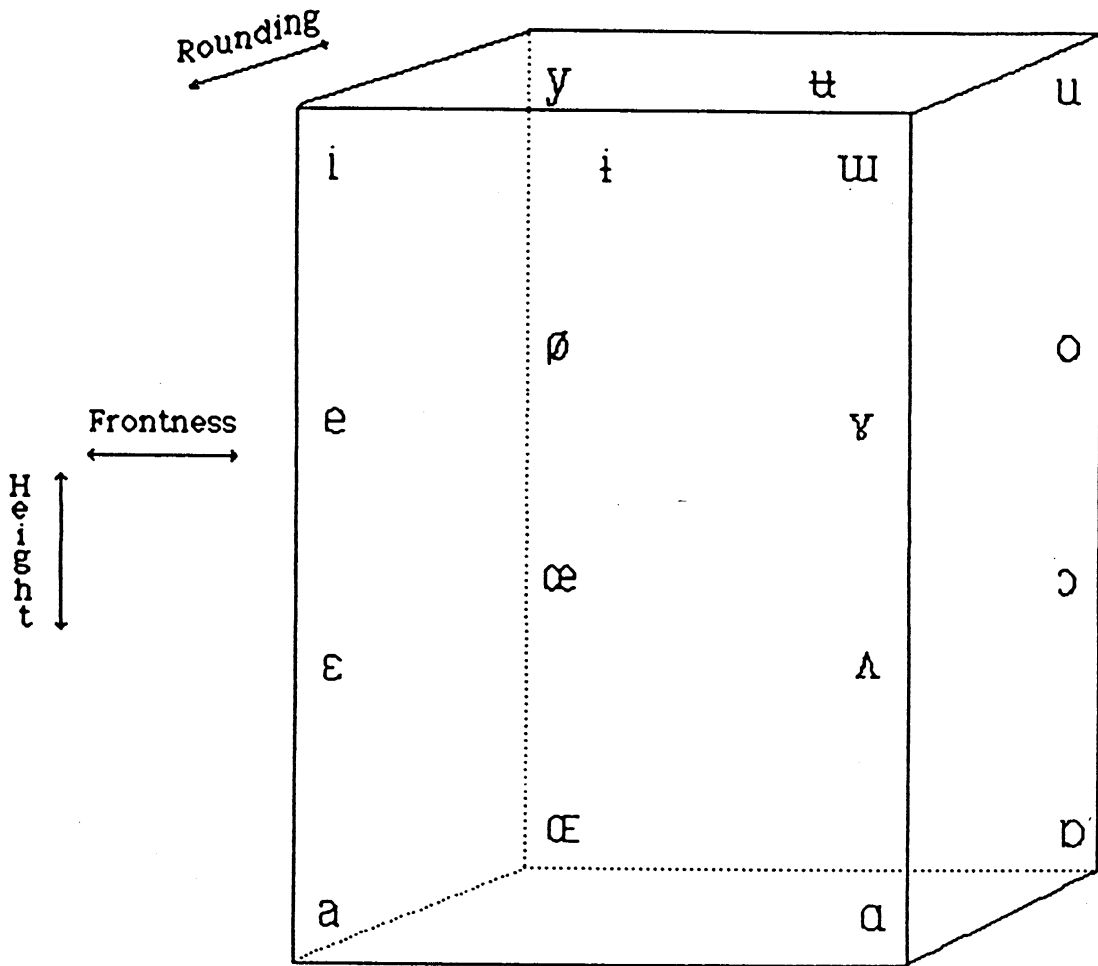


Figure 1. The eighteen cardinal vowels of Daniel Jones in a simple model of the "vowel space."

of unspoiled phonetic naiveté. The responses of the first group would not be interesting because the associations between the stimuli and their articulatory interpretations have been memorized. Utterly naive listeners were disqualified on the ground that the question asked called for judgments not generally pronounced by listeners other than linguists. Moreover, the question would pose a considerable threat to the purity of their phonetic naiveté. The members of the group chosen for testing (there were 22 of them) ranged from undergraduates taking an introductory phonetics course to senior linguist/phoneticians with years of experience in linguistic field work. By and larger there were no serious differences between the responses of the least and the most senior groups of participants; if anything, the first group scored higher, that is, in better agreement with Jones's description.

The vowels, long ago copied onto tape from a commercially available disc recording (Linguaphone ENG 254/255, 1955), were presented in a random order, with each occurring ten times. Each vowel quality was represented by a single token of the several in the recording, and some of these were edited to eliminate noticeable quavers and to reduce gross differences in stimulus duration. Each of the test stimuli was also subjected to spectral analysis, LPC and spectrographic. The data on first and second formant frequencies are shown in Figure 2. Note that four vowels are absent, those of most "open" quality. They were omitted because no plausible second formant frequencies were determined by either of the two methods of analysis used. (I would like to believe that they were where we have been taught to expect them, but were too close to the first formants for resolution.) For the other vowels there is the expected relation between vowel height and F1 frequency. The relation between F2 frequency and both backing and rounding is also not surprising. F2 rises progressively as we go from the back rounded to the back unrounded, then to the front rounded, and finally to the front unrounded. The two central vowels at the highest level fit into the general picture very nicely.

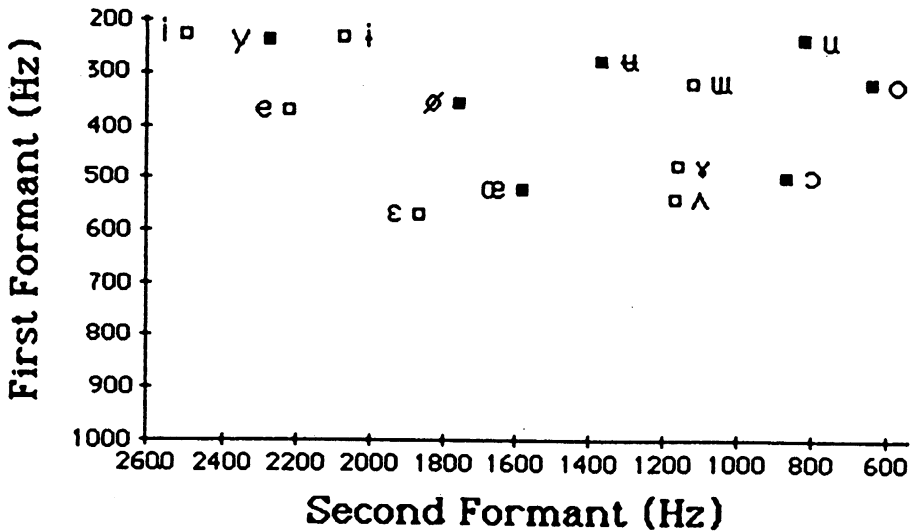


Figure 2. Plotted values of formant 1 and formant 2 for fourteen of Daniel Jones's cardinal vowels. (filled squares = rounded; unfilled squares = unrounded).



In Figure 3 are shown the pooled responses of the 22 participating linguists to Daniel Jones's vowels, which are arranged in order of increasing percent "rounded" judgments. The four leftmost vowels, [i e ε a], least often judged rounded, are those Jones described as front and unrounded; the four judged 90% or more as rounded are the back rounded vowels [u o ɔ α]. The ten remaining qualities, comprising the front rounded, the two central, and the back unrounded vowels, were not judged in close conformity with Jones's description, nor does there seem to be any orderly relation among them. All but one of these vowels have a quite unamerican quality, the exception being the back lowermid unrounded [ʌ], which surprisingly was judged rounded in 38% of the responses. The two high central vowels were both judged mostly rounded; whatever the degree of rounding difference as Jones pronounced them, it did not significantly affect the spectral pattern so far as the listeners could hear.

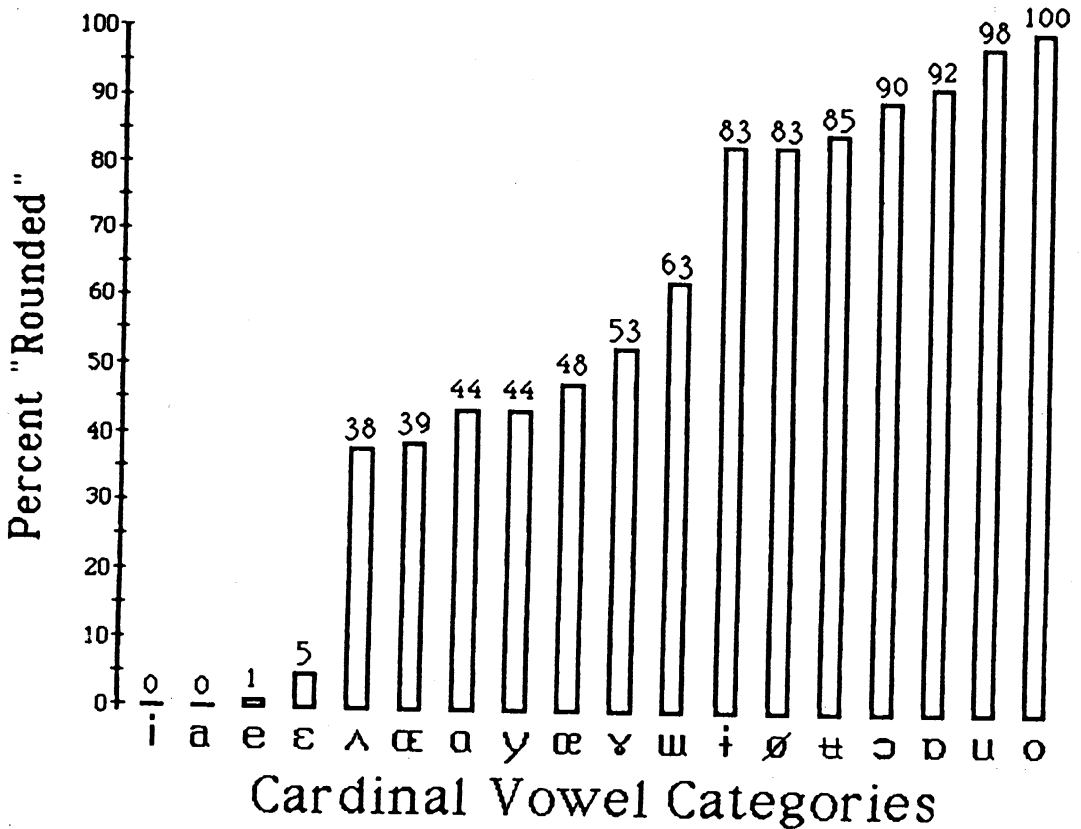


Figure 3. Daniel Jones's cardinal vowels, ordered by degree of perceived rounding. Judgments by 22 listeners, all with some linguistic training.

In Figure 4 the distribution of "rounded" and "unrounded" judgments for the two vowel types as per Jones's description are given separately, and for each type the front, central and back vowels are scored separately. The front unrounded and back rounded show inconsequential mean error rates. Aside from the two poorly discriminated central vowels, there was a general failure to decide whether the remaining vowels had been produced with or without lip rounding.

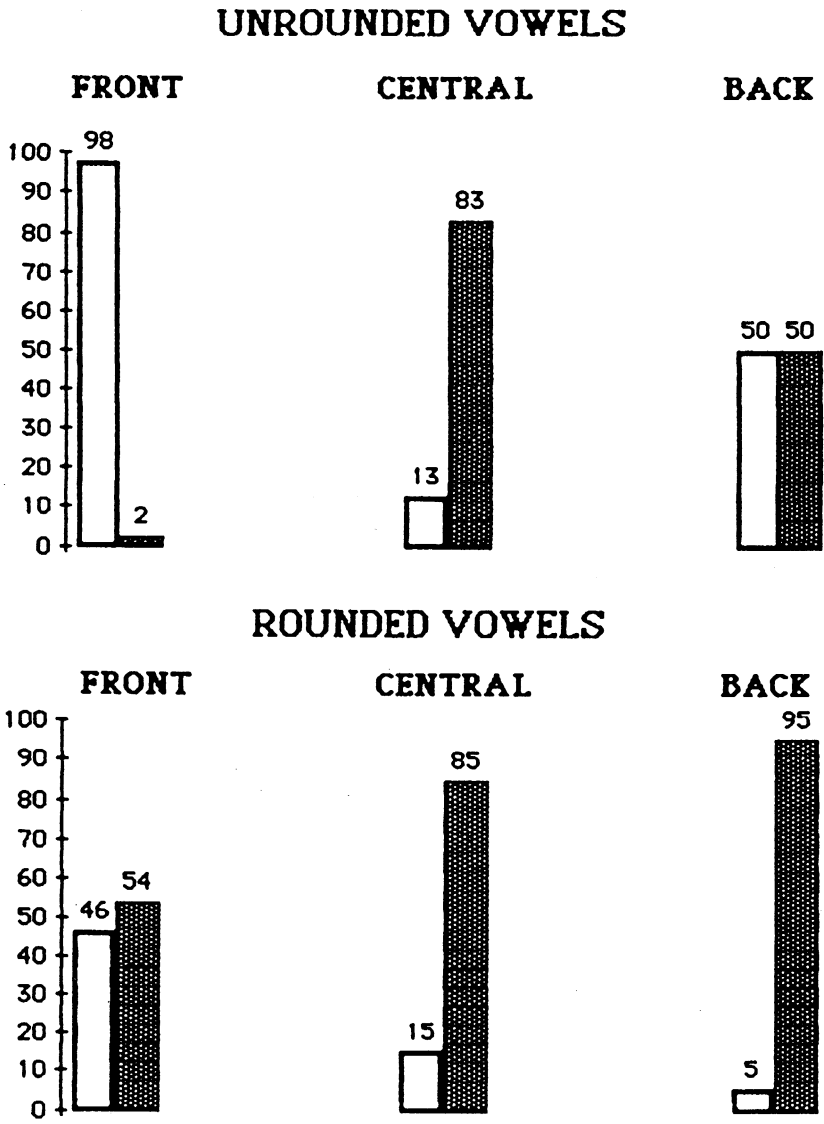


Figure 4. Rounding of eighteen Daniel Jones cardinal vowels; judgments of 22 listeners (shaded = % "rounded"; unshaded = % "unrounded.")

These results may seem surprising, and they cannot be taken to imply the impossibility of correctly inferring lip position on a purely auditory basis. Indeed, one test participant, with an overall error rate of just 14%, managed to achieve scores of about 80% correct for both the front rounded and back unrounded vowels. He is, however, a phonetician with years of field and teaching experience who has, moreover, been exposed to the Daniel Jones recordings. (But an even more senior linguist showed an overall error rate of 30%.) The group data generally bear out the recent advice from David Abercrombie, that "if one wishes to relate a speaker's vowels to the Cardinal Vowels, it is necessary to be able to *see* him; one needs to know what his lips are doing" (Abercrombie 1985). If this is so, that we cannot decide by ear whether a vowel quality was produced with or without lip rounding, we might ask whether this fact, important as it is for the phonetician concerned with speech production, argues for the significance or the irrelevance of visual information for the perception of vowel quality by persons with normal hearing who happen not to be professional linguists.

A recently expressed version of the long advocated motor theory of speech perception (Lieberman and Mattingly 1985) claims that the process involves the direct apprehension of the signal-generating human vocal tract. I am uncertain whether this view constitutes an insight into the nature of speech perception, or whether, instead, it is merely the definition of what we must mean when we say that an auditory signal is perceived to be speech. But in any case, if speech perception is crucially a matter of the articulatory interpretation of an acoustic signal, that interpretation need not be correct.

### References

- ABERCROMBIE, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- ABERCROMBIE, D. (1985). Daniel Jones's teaching. In *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, 15-24. Orlando, FL: Academic Press.
- JOOS, M. (1948). *Acoustic Phonetics*. (Linguistic Society of America Language Monograph no. 23.) Baltimore: Waverly Press.
- LADEFOGED, P. (1967). The nature of vowel quality. In *Three Areas of Experimental Phonetics*, 50-142. London: Oxford University Press.
- LADEFOGED, P. (1971). *Preliminaries to Linguistic Phonetics*. New York: Harcourt Brace Jovanovich.
- LADEFOGED, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- LIBERMAN, A. M. AND I. G. MATTINGLY (1985). The motor theory of speech perception revised. *Cognition* 21, 1-36.
- POTTER, R. K., G. A. KIPP, AND H.C. GREEN (1947). *Visible Speech*. New York: Van Nostrand.

[This work was presented at the 115th meeting of the Acoustical Society of America in Seattle, Washington, May 16-22, 1988. Research support was provided by NIH Grant HD-01994 to Haskins Laboratories.]