

## Chapter 5

1338

# Towards a Physical Definition of the Vowel Systems of Languages

*Laura L. Koenig*

My purpose in this chapter is to discuss a set of observations suggesting that what are sometimes superficially described as 'the same' vowels may differ when produced by speakers of different languages. The studies reviewed here straddle the traditional boundary between phonetics and phonology in that they investigate how specifics of production (phonetics) vary depending on the speaker's language, in particular the set of sound contrasts relevant in that language (phonology). Although I find some of the results of this work compelling, I believe that the conceptual framework underlying many of the research questions is misguided. Ultimately, I will argue that cross-linguistic speech research questions can only be formulated in physical terms, based on speakers' articulatory behavior and its acoustic consequences, and on the associated behavior of listeners. Phonological descriptions that assume sound categories as abstract entities selected by languages and utilized by speakers lend themselves to misunderstanding and inappropriate descriptions of human linguistic behavior. Yngve's (1996) framework, which begins with a physical description of actions or sounds, and from this proposes speaker and listener properties defined within the context of communicative interaction, offers a promising alternative perspective.

### 1. Physical description of vowels

The following is a brief tutorial on the typical ways in which vowels have been described in articulatory and acoustic terms to aid those without phonetic training in understanding the work outlined below.

As in any review, this description is simplified, and omits many details not critical for present purposes. More extensive discussions of vowel description can be found in introductory phonetics texts (e.g. Ladefoged 1993; MacKay 1987). The relationships between vowel articulation and acoustics have been explored by many researchers over the years, including Chiba and Kajiyama (1941), Fant (1960), Gay, Boé, and Perrier (1992), Lindblom and Sundberg (1971), and Stevens and House (1961).

Vowels can be defined simply as linguistic sounds produced with a relatively open vocal tract and little impedance to airflow. Three articulatory parameters that differentiate vowels in many languages of the world are (a) the degree of opening of the jaw-tongue complex; (b) the relative position of the tongue mass in the supralaryngeal vocal tract; and (c) the configuration of the lips. The first dimension is usually referred to as vowel height: high vowels have relatively high jaw-tongue positions, whereas low vowels are more open. Raising the tongue from the floor of the mouth also tends to draw the root of the tongue forward and expand the pharyngeal cavity, so some authors speak in terms of tongue root position or pharyngeal width (e.g. Lindau 1979). The second dimension has to do with whether the bulk of the tongue is shifted towards the alveopalatal region (front vowels) or the velar/uvular region (back vowels). Lip configurations usually include rounded or unrounded/spread.

Much of the variation in these three articulatory parameters is reflected acoustically in the first and second resonant (or formant) frequencies of the vocal tract (F1 and F2, respectively). Perceptual experiments have indicated that relative F1 and F2 differences in speakers' productions can account for a large portion of listeners' labeling behavior for vowels (see Hillenbrand and Gayvert 1993; Peterson and Barney 1952), although other acoustic cues, including durational information, formant change over time, and phonetic context may also play significant roles in vowel perception (see Lindblom and Studdert-Kennedy 1967; Nearey 1989; Strange, Jenkins, and Johnson 1983). The following summary gives the basic relationships between articulatory posture and formant frequencies.

High jaw-tongue positions and expanded pharyngeal cavities are associated with low F1's; low jaw-tongue positions and constricted pharyngeal cavities are associated with high F1's.

Front vowels, with anterior tongue constrictions, are associated with high F2's, whereas back vowels are associated with low F2's.

Lip rounding has the effect of lowering all formants, but its effects are often quite pronounced for F2, and there is some evidence to

suggest that speakers can trade off lip rounding and tongue backing to achieve a lower F2 (Perkell, Svirsky, Matthies, and Jordan 1991). This is one example of a well-known indeterminacy in working back to articulation from an acoustic signal: a given acoustic pattern may be associated with different underlying articulations (cf. also Stevens and House 1955). Articulatory measurements, while more invasive and labor-intensive, thus provide a valuable complement to acoustic data.

Another complication in dealing with the acoustics of vowels (indeed, speech in general) is that the actual frequencies associated with an articulatory configuration (or set of vocal tract cavity sizes) will vary inversely with overall vocal tract size. In general, formant frequencies are lowest in men, intermediate in women, and highest in children, reflecting that vocal tract sizes are typically largest in men and smallest in children. Thus, formant frequency ranges, and hence vowel perception, are scaled as a function of vocal tract size, and determining whether a given formant frequency is high or low depends on the system in question, that is, the anatomy of the actual speaker.

Because acoustic signals can be obtained simply and non-invasively, most researchers undertaking quantitative study of vowels have relied on formant frequency measurements. To restrict formant frequency variation, many researchers have used exclusively adult male subjects. A more limited set of studies has included both male and female subjects, and a few have explored vowel articulation directly using imaging techniques or articulatory tracking.

## 2. History

The idea that languages differ in their 'phonetic substance' is nothing new. Perhaps the most fundamental example of this insight is the cardinal vowel system, which still forms the basis of vowel transcription using the International Phonetic Alphabet (IPA). Developed by Daniel Jones (see Jones 1956), the cardinal vowels were intended to define a language-independent set of articulatory reference points that phoneticians could use in describing the vowels of individual languages. Thus, one might use the [i] transcription for several languages, but the actual vowel productions represented by the transcription would be closer in some languages than others to the articulatory extreme represented by cardinal [i]. A phonetician trained in the cardinal vowel system should thus be well-equipped to apprehend and describe language-specific phonetic patterns.

Implicit in this enterprise is the realization that transcriptional labels represent a family of productions and acoustic characteristics whose precise details may not be identical across languages.

Pierre Delattre carried out pioneering cross-linguistic research using both acoustic and articulatory data. A multilingual speaker, foreign-language teacher, and perceptive impressionistic phonetician, Delattre was engaged in early studies using the sound spectrograph and pattern playback at Haskins Laboratories, and was thus uniquely equipped to carry out instrumental enquiries into the details of language-specific patterns. The work most relevant to the present discussion was an investigation of how vowels in four languages (English, French, German, and Spanish) vary depending on whether the vowel is stressed or not (Delattre 1969). The stress factor was manipulated by selecting pairs of related words in which the vowel of interest appeared in both stressed and unstressed contexts (e.g., the vowel in the second syllable of *adáp*ting vs. *adaptá*tion). Five male speakers were recorded for each language, and all data were analyzed in terms of the F1-F2 values at the most open point in the vowel. X-ray images were also obtained from a subset of speakers.

In brief, Delattre observed that the effects of stress on a vowel varied depending on the language in question. For example, in English, unstressed vowel productions showed a range of F2 values 16 per cent smaller than the stressed vowels (hence, front-back differentiation of vowels was reduced). For the other three languages, F2 ranges were only reduced by 5–7 per cent. The X-ray data suggested that variation in vowels as a function of stress also had distinct articulatory underpinnings in the different languages. Whereas an English speaker showed considerable tongue centralization and reduction of lip rounding for unstressed [o 'u], a French speaker showed reduction of lip rounding but little change in tongue position. In short, the articulatory configurations and acoustic patterns associated with a vowel category, which can, in turn, be associated with a phonetic transcription, depended on the speaker's language history. A French [u] and an English [u], although they share some physical similarities, are not identical vowels, as a transcriptional representation would suggest.

In 1972, Liljencrants and Lindblom sought to explain certain universal patterns of vowel systems by appealing to a principle of perceptual contrast. The authors reasoned that, within a language community, listeners would tend to have more success differentiating between vowels that were more acoustically different rather than those that were less so. Speakers, to avoid being misunderstood,

would accordingly adopt articulatory patterns that maximized acoustic contrast (i.e., dispersion within perceptual vowel space). Lindblom (1986) subsequently drew on data from the Stanford Project on Language Universals Archive to refine a model of perceptual contrast in vowels. Crothers (1978) had provided a phonological analysis of the Stanford data, representing each vowel in a language as the transcriptional value of the most common phonetic realization, and then comparing the vowel qualities (transcriptions) found in languages with various numbers of contrasts. Crothers found that, where three vowels were differentiated within a language, the most common inventory was [i ɑ u], the three vowels that form the extremes of F1-F2 space. As more distinctions were made, the additional vowel qualities tended to be predictable; in a five-vowel system, for example, mid front and back vowels such as [e] or [o] were usually found in addition to [i ɑ u]. Lindblom (1986) attempted to generate the Crothers data using an algorithm that began with the frequency spectrum of a vowel, adjusted for features of the human auditory system (differences in hearing threshold as a function of frequency, upward spread of masking, and transformation to the sone scale of loudness), and maximized the resulting perceptual dispersion. This effort was somewhat more successful than Lindblom and Liljencrants at predicting vowel qualities given a certain number of contrasts, but some systematic discrepancies remained (in particular, non-peripheral vowels were predicted more often than observed). Moreover, even for a given inventory size, the Crothers data indicated that the actual phonetic quality of the vowels could vary to some extent; thus, a five-vowel system might include the vowels [ɛ] and [ɔ] rather than [e] and [o]. Lindblom conceded that vowel differentiation in language may be driven by a principle of sufficient, rather than maximal, perceptual contrast. In other words, the number of vowel contrasts made in a language may explain some aspects of speakers' vowel productions for that language, but unexplained language-specific patterns remained.

The general idea that the acoustic characteristics of vowels can be predicted based on the system of vowel contrasts in a language has continued to drive cross-linguistic work. Keating and Huffman (1984), in a study of Japanese, focused on acoustic variability over repeated productions of a vowel. Speakers of Japanese differentiate among five vowel sounds, typically transcribed as [i e ɑ o u]. Keating and Huffman predicted that [i] and [e] would show relatively little variation in F2, because they lie close together in acoustic space, whereas the back unrounded vowel [u] and the low vowel [ɑ], which

do not have close acoustic neighbors, would vary more. Seven speakers of Tokyo Japanese read passages and produced words in citation form. Results provided partial support in that F1 variation was considerable in [ɑ] productions, and F2 variation was extensive for [w]. A principle of contrast did not entirely explain the data, however, because the acoustic spaces for contrastive vowels showed some overlap of formant values (suggesting that contrasts were not always maintained).

Perhaps the strongest evidence for cross-language differences in vowel production comes from studies that explicitly compare vowels produced by speakers of different languages under controlled conditions. A pilot study by Manuel and Krakow (1984) compared formant frequencies in two men, one speaker of Swahili and one of English. In Swahili, five vowels are contrasted, whereas in English 11–12 monophthongal vowels are typically differentiated, depending on dialect. (Some speakers of English, for example, do not differentiate between the words *cot* and *caught*, whereas for other speakers these words have distinct pronunciations, i.e., different vowels). Manuel and Krakow hypothesized that the English speaker, with more vowel distinctions to make, would show relatively less variability for a given vowel sound than the Swahili speaker. The specific type of variability under investigation was coarticulation, that is, variations in the pronunciation of a sound that arise as a function of neighboring sounds in the word or syllable. The speakers read nonsense two-syllable words constructed to yield all possible combinations of the vowels [i e ɑ o u] in the two syllables, and formant frequency measurements were carried out to determine whether the speaker of Swahili showed greater acoustic variation in his vowels as the neighboring vowel changed. Results indicated significant vowel context effects on both F1 and F2 in the Swahili speaker's productions, but only on F2 for the English speaker. The authors tentatively concluded that speakers of languages with fewer vowels may show more extensive coarticulatory effects, and that, in general, production details for speech sounds are constrained by the system of contrasts relevant for the language in question.

Manuel (1990) followed up on this idea with a more extensive study of three African languages: Shona and Ndebele, with five vowel contrasts, and Sotho, with seven vowels. Three men were recorded for each language. Analysis focused on how the mid-to-low unrounded vowels [ɑ e] varied depending on whether the following vowel was rounded ([u]) or not ([i]). A complex pattern of results emerged. For [ɑ], statistically significant F2 context effects were

found at vowel midpoints and offsets for both Ndebele and Shona, but the formant frequency variation was more extreme in Ndebele. The Sotho speakers had significant F2 effects at vowel endpoints, but not midpoints, and the frequency differences (in Hertz) were smaller than in the other two languages. The results for F1 did not clearly differentiate languages, however. Further, no language showed much variation in [e] as a function of context, and there were no clear language differences. To a limited extent, then, Manuel's results did support the premise that the vowel inventory of a language may be related to the amount of coarticulatory influence speakers show during vowel production. At the same time, the differences between Ndebele and Shona were at odds with the strong hypothesis that vowel inventory size alone can predict coarticulatory extent as measured by formant frequencies.

The studies discussed thus far suggest that one way in which vowels may differ across languages is in the degree of variability, including coarticulatory variation induced by surrounding speech sounds or movements. Another body of work suggests, in contrast, that speakers across languages may differ in the overall range of articulatory configurations they use. For example, an English speaker, who needs to differentiate a large number of vowels, may utilize more extreme articulations for some vowels, and hence produce a wider range of formant frequencies, than a speaker of a language in which fewer vowel distinctions are required. Jongman, Fourakis, and Sereno (1989) measured formant frequencies of Greek (a five-vowel system) and German (a 14-vowel system). Four speakers of Greek and three speakers of German were recorded saying real words that included all the distinctive vowels of their language, with neighboring sounds kept similar to reduce coarticulatory effects. Results indicated that the Greek speakers produced a narrower range of formant frequency values in both F1 and F2 than the speakers of German. Bradlow (1995) similarly compared formant frequencies of vowels produced by speakers of English and Spanish (a five-vowel system). The Greek data from Jongman *et al.* were used as another example of a five-vowel language. Results showed that the overall formant frequency ranges, in Hertz, were 13 per cent larger for English speakers than Spanish, and 17 per cent larger than Greek. Formant frequency standard deviations of individual vowels did not differ between English and Spanish. Coefficients of variation (a relative measure, computed as the standard deviation divided by the mean value) did approach significance for F1, however, with more relative variability in Spanish than English. Bradlow also found that F2 values for English tended to be higher

overall than in Spanish, suggesting that the English speakers were, in general, using more anterior tongue positions. The latter finding, that speakers of a language may adopt a characteristic vocal tract posture or setting, has been termed a language-specific setting or base of articulation (see Laver 1980).

Flege (1989) asked whether within-category variability or the overall size of vowel space would best account for language-related vowel differences in the speech of bilingual English-Spanish speakers. The subjects, Spanish-dominant immigrants to the USA, were recorded under normal speaking conditions and in a bite-block experiment, in which speakers' jaws were fixed in a slightly open position. Flege hypothesized that, if speaking English required greater constraint on the range of tongue postures (or formant frequencies) so as to maintain necessary sound distinctions, subjects might compensate more effectively for the bite block when speaking English than Spanish. The speech production measure was the distance between the tongue and the hard palate, with tongue position tracked by means of sensors on the tongue. Results indicated that the highest vowels, [i] and [u], were produced with smaller tongue-palate distances in English than in Spanish. Standard deviations of sensor positions did not differ between English and Spanish. Significant differences were not seen for the normal vs. the bite-block condition. Finally, average tongue position, calculated over all sensors, did not differ between the two languages, but Flege noted that the average tongue contour (obtained by fitting a polynomial to the four sensors) was somewhat more convex for English. Again, then, some language differences were found, but the pattern of results was more complex than what would have been predicted simply on the basis of phonetic inventory size.

Recently, a colleague and I have undertaken a comparison of formant frequencies in vowels produced by monolingual speakers of Greek and English. This work is in its early stages, and presently only one male and one female speaker of each language have been analyzed fully. Our pilot results (Okalidou and Koenig 1999), however, seem to support the findings of Jongman *et al.* (1989) and Bradlow (1995) in that the English speakers produced a wider range of formant frequencies than the Greek speakers, suggesting that more extreme articulatory postures may be used when speakers need to differentiate a larger number of vowel sounds in their native language.

The study of acoustic and articulatory differences among languages is in its infancy. The preliminary hypotheses of some authors have been partially supported, but much more data from



more languages needs to be collected to refine these hypotheses. For those languages that have been investigated, only a handful of speakers have been recorded, and only a limited range of contextual variation has been explored. The available data do, however, appear to indicate that 'the same' vowel, produced by speakers of different languages, may differ in several respects, including the range of token-to-token variability, extent of contextual effects, extremity of articulatory postures or formant frequencies, and base of articulation. Some of these differences may be related to the number of vowel contrasts a speaker needs to maintain in order to be understood in his or her speech community, but other findings currently lack a general explanatory theory.

### 3. Critique

The basic logic of the studies reviewed above has been to start with the system of vowel contrasts in a language, and then to correlate acoustic or articulatory features of actual (phonetic) vowel productions with aspects of the phonological system. The data tend to support the very general hypothesis that vowel categories have language-specific acoustic and/or articulatory patterns. Thus, although we may use the IPA symbol [i] when transcribing English as well as Greek, it is not really accurate to say that English and Greek share 'the vowel [i]'.

This discussion begs the question, however, of what one means by the vowel [i] (or any other vowel). On what basis would one ever suggest that English and Greek share a vowel? Such a view is promoted by transcriptional representations of speech which provide a finite set of symbols for vocal tract configurations and sound patterns from which languages are assumed to select. At the same time, traditional phonemic analyzes of speech (e.g. Bloomfield 1933; Swadesh 1934) argued that a sound category (phoneme) can only be defined within a specific language, i.e., in the context of a specific set of sound contrasts maintained among a set of speakers and listeners involved in regular communicative exchange. According to this view, an IPA representation is a shorthand notation for a class of sounds that speakers of a language produce under certain circumstances, and to which listeners respond in a particular way. The data presented above suggest that one insight of traditional phonemic analysis was essentially correct: sound categorization within a language can really only be defined for a specific linguistic community, that is, a particular group of communicating individuals.

Explicitly formulating the issue this way, *viz.*, in terms of the communicative participants, offers straightforward explanations for phenomena in speech perception as well as production. For example, studies of infants' responses to speech have documented numerous instances of perceptual attunement to the sound system of the ambient language, including loss of response to sound differences that are irrelevant in the infant's speech community (e.g. Best 1992; Best, McRoberts, and Sithole 1988; Burnham 1986; Jusczyk 1993; Werker 1989). Refinement of the details of speech sound categorization may persist into childhood (see Bernstein 1983; Nittrouer 1992; Nittrouer and Studdert-Kennedy 1987; Ohde 1994; Ohde and Haley 1997; Zlatin and Koenigsnecht 1975). Although some aspects of speech perception abilities may well be innate (see, e.g., Grieser and Kuhl 1989), any discussion of perceptual attunement to the surrounding language must begin with a description of the physical signals the infant receives, and proceed to describe how the infant responds to those signals.

When category formation is conceived of as being an active process of establishing behavioral responses to physical signals, it follows that we should expect children brought up in different language communities to develop different response criteria for specific acoustic patterns. Language-specific patterns of speech discrimination and labeling behavior in adults also follow as a direct consequence. Decades of cross-linguistic work on speech perception have established that some details of listener response depend on his or her language background (e.g. Abramson and Lisker 1970; Flege and Eefting 1986; Lisker and Abramson 1970; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura 1975; Polka 1991; Schmidt 1996; Sekiyama and Tohkura 1991; Terbeek 1977), for both vowel and consonant distinctions. If we recognize that Spanish speakers and English speakers have developed linguistic-phonetic properties appropriate to their language communities, it is immediately clear why a native English speaker differentiates easily between the words *beet* and *bit*, whereas a native Spanish speaker might protest that these words sound the same.

The experience of teaching introductory phonetics courses to a linguistically diverse student population has provided me with many examples of the conceptual quandries one encounters by adopting language-universal notions of sound categories as a starting point for discussion. One particularly memorable incident for me involved a Jamaican student who, when faced with the IPA representations for the vowels of 'Standard American English' (SAE) and a sample list of

words in which such vowels typically occur, asked (quite appropriately) why her productions of those words seemed to involve a completely different set of sounds – indeed, hardly one item of our word list would have included the same transcriptional symbol for her speech as what I had presented for SAE. Answering such a question requires that the definition of ‘a vowel of a language’ be utterly reworded and reconceptualized as being a set of sounds that a certain set of speakers produces to yield a particular kind of response (or understanding) in their listeners. The focus of emphasis shifts, rightly, to the speakers and listeners involved in communicative exchange. With this kind of definition, there is little reason to expect that the physical production and acoustic structure of sounds be constant across dialects of English, or across languages, except as may occur because all humans share essentially the same sound-producing and sound-perceiving mechanisms. What is constant is simply that communication is possible and largely error-free within a single speech community; across communities, misunderstandings may occur. With this redefinition, we stop thinking about a set of ‘universal’ vowels (or rather, vowel symbols) and concentrate instead on the human beings who produce and interpret these sounds.

I do not give this example in order to argue that phonetic transcription is a useless tool. I do, however, find it illustrative of the kinds of difficulties encountered when one takes too seriously the notion that linguistic interchange involves a set of objects or entities that are defined independently of actual speakers and listeners. For purposes of an introductory course in which the students share a considerable amount of linguistic knowledge (or have shared linguistic-phonetic properties), there is probably some utility in presenting a set of examples and their simplified transcriptional representations for a dialect all or most students in the class may be expected to share. But any real understanding of phonetics as a linguistic pursuit must acknowledge that the reality is not a set of pre-determined categories, as defined by the International Phonetic Association or anyone else, but rather classes of behaviors that occur in the context of human communication. The categorization of these classes of behaviors into vowel qualities that can be represented transcriptionally is quite secondary to the inquiry.

One could, of course, argue that this conceptual leap is properly the stuff of an introduction to phonetics, and that anyone working in the field of speech understands this. The work reviewed above suggests that this may not be as true as one might hope, however. The formulation of questions in the cross-language vowel research

implies that a set of vowel categories is assumed for human language in general, and the burden of proof is on the researcher to establish that these categories are manifested in phonetically different ways across languages. I believe that this is a backwards way of thinking about the issue. Rather than assuming linguistic objects (represented by phonemic, transcriptional notation) *a priori* and determining their physical realizations, we should first be assessing the physical signals produced by speakers of a specific language, and observe how listeners in that language community respond to the physical variations in these productions. Despite considerable evidence for language-specific sound patterns, and some compelling arguments from speech researchers to build phonetic science up from speech signals (or movements) rather than assuming phonological categories (e.g. Liljencrants and Lindblom 1972), traditional notions of linguistic objects have generally persisted within phonology, and are borrowed into phonetic research when researchers attempt to deal with the issue of language-specific sound systems. At present, making the leap across the traditional phonetics-phonology divide typically involves a drastic change in the nature of the phenomena under investigation. It is certainly true that some phonologists over the years have attempted to incorporate aspects of physical speakers into their phonological descriptions. Theories of feature geometry (e.g. Broe 1992; Clements 1985; Sagey 1986) essentially acknowledge that speech is produced with a vocal tract that has a variety of coupled components, and articulatory phonology (Browman and Goldstein 1992) proposes that speakers' articulatory activities form the basis of phonological contrasts. Nevertheless, the field of phonology still remains largely a study of abstractly defined units, schematized as sound types which occur across languages. Only at low, peripheral levels are these types 'converted' into the physical signals that serve as the communicative channel between participants. In this kind of conceptualization, the details of language-specific sound patterns revealed by cross-language work are difficult to predict or explain.

I do not believe that the difficulty lies in simply explicating the details of such a conversion; speech researchers are quite accustomed to dealing with continuous scales upon which individual speaker-listeners impose some kind of categorization. Rather, I believe that there is a larger problem, *viz.*, having any notion of linguistic sounds as a set of abstract objects that can be selected by languages or speakers of a language. Instead of a process of selection, we ought to be able to say, simply (or not so simply), that human linguistic communication involves making interpretations of physical signals, and that those

interpretations are made by listeners in accord with habits they have learned by growing up as speaker-listeners of their particular language. If we subsequently find that speakers' productions or listeners' interpretations and categorizations share some similarities across languages, that may tell us something interesting about limitations of the human vocal tract or auditory-perceptual system. Such an alternative conception, where we begin with physical signals and listeners' responses to them, and work back to linguistic theory from there, is essentially the perspective that Yngve's work embodies.

#### 4. An alternative formulation

Let us outline briefly how this research enterprise might be reformulated. A phonetician interested in studying Greek and English might begin by noting that a Greek speaker utters a sound (or combination of sounds) to communicate something about a flat piece of bread. An IPA transcription [pitə] or even the Greek spelling  $\pi\acute{\iota}\tau\tau\alpha$  might yield some information about the vocal tract activities the speaker uses and his/her categorization of those actions. Alternatively, the phonetician might simply observe that the speaker begins with a bilabial closure; initiates vocal-fold vibration on release of the closure; moves into a vocal tract configuration in which the jaw is fairly high, the tongue is bunched toward the hard palate, and the lips are spread; produces a tongue-tip closure; and then ends with a fairly open vocal tract position. The latter description has the advantage of making no assumptions that the specific vocal-tract activities and corresponding acoustic output of a Greek speaker are the same as those used by an English speaker requesting a similar piece of flat bread at a Middle Eastern restaurant. The phonetician might go on to have the speaker repeat how s/he would request such a piece of bread. A native Greek listener could then be recruited to listen to this set of productions and verify that s/he would respond by supplying the bread and not some other item, indicating that these vocal productions yielded a consistent effect in the listener. The phonetician could then go about making systematic measurements of the speaker's vocal tract activities and/or the corresponding acoustic record. The results of such an investigation would yield a range of activities and/or acoustic patterns that constitute a functionally equivalent class in Greek – what, in traditional terms, we would call the word *pita* and the Greek vowel /i/.

The researcher might then carry out a study asking an English speaker how s/he would communicate something about a kind of red

root vegetable, and obtain a set of productions that traditionally we would call repetitions of the word beet. Comparison of the Greek and English speakers' productions might indicate certain articulatory or acoustic similarities. Such a comparison would make no assumptions that English and Greek both use a vowel [i]; it might, however, suggest that Greek and English speakers hear certain similarities in each others' languages. We might further suspect that the English speaker would use his set of beet vowels to speak about pita bread, regardless of whether he were speaking to an English or a Greek speaker. A Greek speaker, listening to the English speaker say pita, might note that the English speaker produced a similar set of movements/sounds to that of a Greek speaker, but further recognize that the English speaker's activities don't exactly match those of a native Greek speaker. In other words, the English speaker might speak Greek with an English accent. On the other hand, if an English listener, hearing a Greek speaker, thinks that some of the Greek words include vowels more similar to those in the word bit than the word beet, that tells us that the linguistic-phonetic properties of the English listener differ from those of the Greek listener.

At face value, the above formulation may appear to be little more than a rewording of the existing cross-language work. There is one very important difference, however: at no point do we assume – are we even led to assume – that there exists a generic linguistic object [i] which is used by speakers of English and Greek. Given that linguists have historically used the same symbol in transcriptions of English and Greek, we may expect to find certain physical similarities between an English speaker's production of beet and the Greek speaker's production of pita, but the extent and nature of those similarities remain empirical questions. To adopt predefined transcriptional categories is effectively to assume similarity (indeed, sameness) across languages rather than treating it as a fundamental question for research.

Although instrumental analyzes of speech have been carried out for many years, only during the past couple of decades has it been possible for an individual speech researcher to contemplate carrying out the analyzes necessary to determine the full acoustic space for the vowels of a language, taking into account contextual variations, for even a few speakers. As such data become more widely available, we will begin to define precisely how vowels differ across languages. I believe that these data will ultimately lead to definitions of individual vowels which are inherently specific to a given speech community. Taking as our starting place the speech signal and listeners'

judgments, we have every reason to expect cross-linguistic differences in both speech production and speech perception. At the same time, after a considerable body of data have been amassed on many languages, we may well discover that some characteristics of speech sound categories or categorization occur frequently or universally across languages. These will form the basis of quantitative, empirically-testable human phonetic universals.

### Note

This chapter is dedicated to the memory of Karen Landahl of the University of Chicago who first introduced me to the field of phonetics and has always been an inspiration to me.

### References

- Abramson, A. S. and Lisker, L. (1970), 'Discriminability along the voicing continuum: Cross-language tests', in B. Hála, M. Romportl, and P. Janota, *Proceedings of the 6th International Congress of Phonetic Sciences, Prague 1967*. Prague: Academia Publishing House, Czechoslovak Academy of Sciences, pp. 569-573.
- Bernstein, L. E. (1983), 'Perceptual development for labeling words varying in voice onset time and fundamental frequency'. *Journal of Phonetics*, 11, 383-393.
- Best, C. T. (1992), 'The emergence of language-specific phonemic influences in infant speech perception', in H. C. Nusbaum and J. Goodman, *The Transition from Speech Sounds to Spoken Words: The Development of Speech Perception*. Cambridge, MA: MIT Press.
- Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988), 'Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants'. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.
- Bloomfield, L. (1933), *Language*. New York: Holt.
- Bradlow, A. R. (1995), 'A comparative acoustic study of English and Spanish vowels'. *Journal of the Acoustical Society of America*, 97, 1916-1924.
- Broe, M. (1992), 'An introduction to feature geometry', in G. J. Docherty and R. D. Ladd, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, pp. 149-165.
- Browman, C. P. and Goldstein, L. M. (1992), 'Articulatory phonology: An overview'. *Phonetica*, 49, 155-180.
- Burnham, D. K. (1986), 'Developmental loss of speech perception: Exposure to and experience with a first language'. *Applied Psycholinguistics*, 7, 207-240.

- Chiba, T. and Kajiyama, M. (1941), *The Vowel: Its Nature and Structure*. Tokyo: Tokyo-Kaiseikan Publishing Company Ltd.
- Clements, G. N. (1985), 'The geometry of phonological features'. *Phonology Yearbook*, 2, 225-252.
- Crothers, J. (1978), 'Typology and universals of vowel systems', in J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik, *Universals of Human Language, Vol. 2: Phonology*. Stanford: Stanford University Press, pp. 93-152.
- Delattre, P. C. (1969), 'An acoustic and articulatory study of vowel reduction in four languages'. *International Review of Applied Linguistics*, 7, 295-325.
- Fant, G. (1960), *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Flege, J. E. (1989), 'Differences in inventory size affect the location but not the precision of tongue positioning of vowel production'. *Language and Speech*, 32, 123-147.
- Flege, J. E. and Eefting, W. (1986), 'Linguistic and developmental effects on the production and perception of stop consonants'. *Phonetica*, 43, 155-171.
- Gay, T., Boé, L.-J., and Perrier, P. (1992), 'Acoustic and perceptual effects of changes in vocal tract constrictions for vowels'. *Journal of the Acoustical Society of America*, 92, 1301-1309.
- Grieser, D. and Kuhl, P. K. (1989), 'Categorization of speech by infants: Support for speech-sound prototypes'. *Developmental Psychology*, 25, 577-588.
- Hillenbrand, J. and Gayvert, R. T. (1993), 'Identification of steady-state vowels synthesized from the Peterson and Barney measurements'. *Journal of the Acoustical Society of America*, 94, 668-675.
- Jones, Daniel (1956), *An Outline of English Phonetics* (8th edn). Cambridge: W. Heffer & Sons Ltd.
- Jongman, A., Fourakis, M., and Sereno, J. A. (1989), 'The acoustic vowel space of Modern Greek and German'. *Language and Speech*, 32, 221-248.
- Juszyk, P. W. (1993), 'From general to language-specific capacities: The WRAPSA model of how speech perception develops'. *Journal of Phonetics*, 21, 3-28.
- Keating, P. A. and Huffman, M. K. (1984), 'Vowel variation in Japanese'. *Phonetica*, 41, 191-207.
- Ladefoged, P. (1993), *A Course in Phonetics* (3rd edn). Orlando: Harcourt Brace College Publishers.
- Laver, J. (1980), *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
- Liljencrants, J. L. and Lindblom, B. (1972), 'Numerical simulation of vowel quality systems: The role of perceptual contrast'. *Language*, 48, 839-862.
- Lindau, M. (1979), 'The feature expanded'. *Journal of Phonetics*, 7, 163-176.
- Lindblom, B. E. (1986), 'Phonetic universals in vowel systems', in J. J.



- Ohala and J. Jaeger, *Experimental Phonology*. New York: Academic Press, pp. 13–44.
- Lindblom, B. E. F. and Studdert-Kennedy, M. (1967), 'On the role of formant transitions in vowel recognition'. *Journal of the Acoustical Society of America*, 42, 830–843.
- Lindblom, B. E. F. and Sundberg, J. E. F. (1971). 'Acoustical consequences of lip, tongue, jaw, and larynx movement'. *Journal of the Acoustical Society of America*, 50, 1166–1179.
- Lisker, L. and Abramson, A. S. (1970), 'The voicing dimension: Some experiments in comparative phonetics', in B. Hála, M. Romportl, and P. Janota, *Proceedings of the 6th International Congress of Phonetic Sciences, Prague 1967*. Prague: Academia Publishing House, Czechoslovak Academy of Sciences, pp. 563–567.
- MacKay, I. R. A. (1987), *Phonetics: The Science of Speech Production* (2nd edn). Boston: Allyn and Bacon.
- Manuel, S. Y. (1990), 'The role of contrast in limiting vowel-to-vowel coarticulation in different languages'. *Journal of the Acoustical Society of America*, 88, 1286–1298.
- Manuel, S. Y. and Krakow, R. A. (1984), 'Universal and language particular aspects of vowel-to-vowel coarticulation'. *Haskins Laboratories Status Reports on Speech Research*, 77/78, 69–78.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975), 'An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English'. *Perception and Psychophysics*, 18, 331–340.
- Nearey, T. M. (1989), 'Static, dynamic and relational properties in vowel perception'. *Journal of the Acoustical Society of America*, 85, 2088–2113.
- Nittrouer, S. (1992), 'Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries'. *Journal of Phonetics*, 20, 351–382.
- Nittrouer, S. and Studdert-Kennedy, M. (1987), 'The role of coarticulatory effects in the perception of fricatives by children and adults'. *Journal of Speech and Hearing Research*, 30, 319–329.
- Ohde, R. N. (1994), 'The development of the perception of cues to the [m]-[n] distinction in CV syllables'. *Journal of the Acoustical Society of America*, 96, 675–686.
- Ohde, R. N. and Haley, K. L. (1997), 'Stop-consonant and vowel perception in 3- and 4-year-old children'. *Journal of the Acoustical Society of America*, 102, 3711–3722.
- Okalidou, A. and Koenig, L. L. (1999), 'Patterns of vowel-to-vowel coarticulation in Greek and English', in *Collected Papers from the Joint Meeting "Berlin 99": 137th Regular Meeting of the Acoustical Society of America and 2nd Convention of the EAA: Forum Acusticum*. Oldenburg: Deutsche Gesellschaft für Akustik e.V.
- Perkell, J. S., Svirsky, M. A., Matthies, M. L., and Jordan, M. I. (1991),

- 'Trading relations between tongue body raising and lip rounding in production of the vowel /u/', in O. Engstrand and C. Kylander, *Papers from the Symposium on Current Phonetic Research Paradigms: Implications for Speech Motor Control* [Volume 14 of PERILUS: Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm].
- Peterson, G. E. and Barney, H. L. (1952), 'Control methods used in a study of vowels'. *Journal of the Acoustical Society of America*, 24, 175-184.
- Polka, L. (1991), 'Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions'. *Journal of the Acoustical Society of America*, 89, 2961-2977.
- Sagey, E. (1986), 'The representation of features and relations in non-linear phonology'. Doctoral dissertation, Massachusetts Institute of Technology.
- Schmidt, A. M. (1996), 'Cross-language identification of consonants. Part 1. Korean perception of English'. *Journal of the Acoustical Society of America*, 99, 3201-3211.
- Sekiyama, K. and Tohkura, Y. (1991). 'McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility'. *Journal of the Acoustical Society of America*, 90, 1797-1805.
- Stevens, K. N. and House, A. S. (1955), 'Development of a quantitative description of vowel articulation'. *Journal of the Acoustical Society of America*, 27, 484-493.
- Stevens, K. N. and House, A. S. (1961), 'An acoustical theory of vowel production and some of its implications'. *Journal of Speech and Hearing Research*, 4, 75-92.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983), 'Dynamic specification of coarticulated vowels'. *Journal of the Acoustical Society of America*, 74, 695-705.
- Swadesh, M. (1934), 'The phonemic principle'. *Language*, 10, 117-129. [Reprinted in Joos, M. (1957), *Readings in Linguistics I*. Chicago: University of Chicago Press.]
- Terbeek, D. (1977), 'A cross-language multidimensional scaling study of vowel perception'. *UCLA Working Papers in Phonetics*, 37.
- Werker, J. F. (1989), 'Becoming a native listener'. *American Scientist*, 77, 54-59.
- Yngve, V. H. (1996), *From Grammar to Science: New Foundations for General Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Zlatin, M. A. and Koenigsknecht, R. A. (1975), 'Development of the voicing contrast: Perception of stop consonants'. *Journal of Speech and Hearing Research*, 18, 541-553.