UCLA Working Papers in Phonetics

Number 94

December 1996

UCLA Working Papers in Phonetics

Number 94

December1996

Table of Contents

Gestural economy 1
Ian Maddieson

David Abercrombie and the changing field of phonetics 7
Peter Ladefoged.

The IPA and a theory of phonetic description 12
Peter Ladefoged

The IPA and the phonetics/phonology interface 20
Peter Ladefoged

Rate effects on French intonation: 26
Prosodic organization and phonetic realization
Cécile Fougeron and Sun-Ah Jun

The sounds of languages 52
Peter Ladefoged

Phonetics 66
Patricia A. Keating

Focus realization of Japanese English 110
and Korean English intonation
Motoko Ueyama and Sun-Ah Jun

Gestural economy

Ian Maddieson

ABSTRACT

This paper outlines a theory of gestural economy in language structure, with illustration partly drawn from recent studies of Ewe sounds using electro-magnetic articulography and video. It argues that languages tend to be economical in the number and nature of the distinct gestures used to construct their inventory of contrastive sounds. Evidence of such economy is more wide-spread than evidence for 'polarization' of contrast.

INTRODUCTION

Many linguists have observed that languages show a tendency to construct their inventory of contrastive sounds in a way that is at least partially symmetrical. For example, a language with the stops /p, t, k/ is far more likely to also have /b,d, g/ than to have /ɖ, ɟ, ɢ/. If sounds are regarded as composed of features, this tendency can be expressed in terms of maximum exploitation of compatible feature combinations. The number of features needed to form a given number of contrasts is thus economized.

This paper argues that a similar pattern can be seen in the articulatory organ-ization of the sounds of a language. That is, there is an analogous tendency to be economical in the number and nature of the distinct articulatory gestures used to construct an inventory of contrastive sounds, and it is this (rather than a more abstract featural analysis) that underlies the observed system symmetry. More-over, this tendency can be seen as an aspect of a more general principle that can be given the name 'Gestural Economy'. It is suggested that evidence of such economy is more widespread than evidence for opposing theories of 'polarization' of contrast.

There are three principal strands to the argument. First, there is the well-known evidence that languages as a whole favor certain articulatory positions and movements, which are by-and-large those that are more efficient and involve less extreme movements. Second, within a language a given articulatory gesture is often exploited for several distinct segments, for example, nasals and stops usually occur at the same places of articulation and complex segments are built up out of gestures used in simple ones. Third, articulations are not generally displaced from the 'economical' positions or otherwise modified when a language has an added contrast at a nearby place. That is, evidence for systematic use of polarization strategies is lacking.

Only a very brief review of the first point will be provided. The second point is supported by a demonstration that the labial and velar gestures in simple bilabial and velar stops are largely similar to those in labial-velar stops in Ewe. The third point will be supported by showing that one of the best-known hypothesized polarization effects is spurious: labio-dental fricatives in Ewe do not ordinarily involve use of an 'enhancing' elevation of the upper lip in these segments.

What is meant by a gesture?

Before proceeding to any further discussion, it may be useful to characterize what is meant by a gesture in the present context. This term is not intended to refer to a primitive element in the organization of phonology (as in Articulatory Phonology [1]), nor to an articulatory invariant. Here, it simply refers to a typical movement trajectory for a given articulatory subsystem in realizing a given phonetic contrast, bearing in mind the initial conditions for the start of the gesture, anticipation of the following context, and any competing demands of other simultaneously specified aspects of the current phonetic element of which the gesture is a component. It is thus a recasting of the traditional phonetic notion 'place of articulation' in dynamic terms and with the focus on the movements of active articulators as much as on the sites at which constrictions are formed.

INVENTORY STRUCTURE

Cross-linguistic studies of segment inventories show that languages tend to favor many of the same segments [2]. The stops /p, t, k, b, d, g/, the fricatives /f, s, ʃ/, and the nasals /m, n, ɲ, ŋ/ are more common than other segments of their respective classes. Moreover, languages show a strong tendency to have small 'families' of sounds that share common articulatory positions. This already emerges from the listing of common sounds above, where the sets /p, b, m/, /t, d, n/ and /k, g, ŋ/ share - in traditional phonetic terms - the same place of articulation.

These commonalities are part of the motivation for the proposal of gestural economy. The common articulatory gestures that are shared by such families of sounds are proposed as being in themselves efficient and economical, but further 'economy' is achieved by re-using the same gesture in a variety of segments and by resisting uneconomical modification in the interests of generating larger acoustic distinctions between competing sounds. The remainder of the paper will illustrate these two points using simple and complex stops as an example of re-use of common gestural patterns, and labio-dental fricatives as an example of the absence of modification.

SIMPLE AND COMPLEX STOPS

The similarity of the component gestures in labial-velar stops to those in simple bilabial and velar stops in Ewe has been discussed some detail in [3]. In that paper, evidence for an overall similarity of the gestures in doubly- and singly-articulated stops was illustrated with data from one speaker in an experiment using electromagnetic articulography [4]. Data from a second speaker is presented in Figures 1-3.

Figure 1 shows the time course of the vertical movement of the lower lip during the word /apaa/ *'job'*. In this figure and the next two, the movement data have been converted to standard scores so that they can be plotted on the same scale and with the same origin. Release of the consonant closure, determined from the acoustic record of the utterances, is at 300 ms. This point is used as the line-up point for aligning repetitions, and each of the figures represents the mean of ten repetitions.

Figure 2 shows the vertical movement of a point on the back of the tongue during the plain velar stop in the word /aka/ *'charcoal'*.

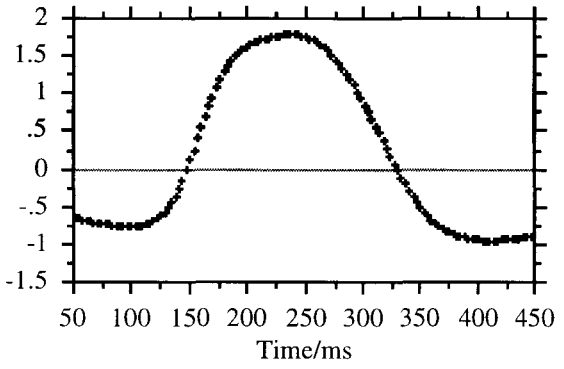


Figure 1. Normalized mean vertical movement of the lower lip in /apaa /

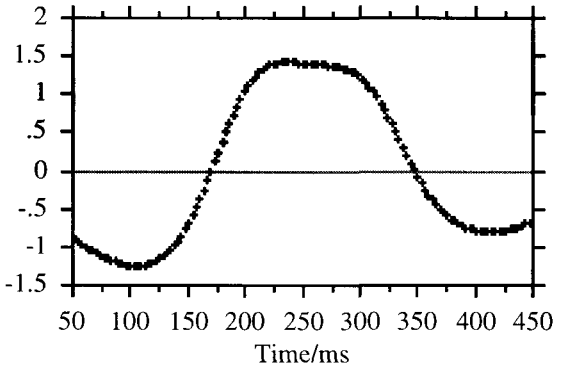


Figure 2. Normalized mean vertical movement of the tongue back in /aka/.

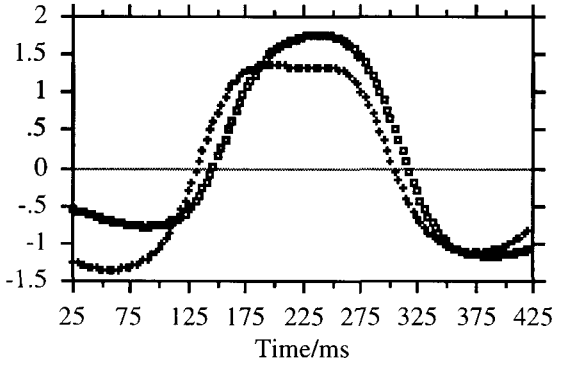


Figure 3. Normalized mean vertical movement of the lower lip and tongue back in /ak͡pa/.

The corresponding movements of both the lower lip and the tongue back during the word /ak͡pa/ *'too much'* are shown in Figure 3. The velar gesture, plotted with small crosses, leads the labial one by a few milliseconds, but both gestures are in all salient particulars like those in the simple stops /p/ and /k/. This, of course, need not be the case. The movements in the doubly-articulated stop might well have had substantially different time courses, or differed in their shape, or in their amplitude (amplitude cannot be read off the normalized plots shown here, but is very comparable in the unnormalized data.)

Some differences between the oral gestures in voiced and voiceless simple stops at the same place, and between the simple stops and the components in doubly-articulated stops were indeed observed and reported in [3], but with one exception these can be accounted for as contextual effects, due to the demands of other specified aspects of these segments. The exception concerns a backward movement of the tongue body in doubly-articulated stops that is absent in simple velars. Its explanation remains undetermined, but it could be an aerodynamically induced consequence of a double closure in the oral tract.

Apart from this detail, Ewe doubly-articulated labial-velar stops appear to be made in the simplest way possible – by combining the well-rehearsed movements that are used in simple labial and velar stops. We hypothesize that languages take maximal advantage of such opportunities for limiting the number of distinct gestures employed, as part of a general preference for gestural economy.

ABSENCE OF POLARIZATION

Ewe is well-known as one of the relatively small number of languages with a contrast between [ɸ, β] and [f, v]. It has been claimed [5] that Ewe speakers (and perhaps speakers of some other languages in the same area which also contrast [ɸ, β] with [f, v]) 'enhance' the bilabial/labio-dental contrast among fricatives by using an active raising gesture of the upper lip in the production of the labio-dentals. According to this view, the structure of the set of phonologically significant distinctions in the language has a direct influence on the production of a sound type – a labio-dental

fricative – that is among those that are the most highly favored in the world's languages [2].

We propose that labio-dental fricatives are favored because this is an optimal place for creating fricatives. It requires precise positioning of only one active articulator rather than two as for a bilabial, a relatively small movement compared to, say, a linguo-labial or interdental. Labio-dentals are also acoustically readily distinct from all fricatives produced further back – except perhaps [θ].

From a gestural economy perspective, these virtues would be expected to be retained, rather than disturbed because of a contrast with less economical sounds. The articulatory target might become a bit more precisely defined, constraining the variability in order to protect the contrast, but that is all.

The two Ewe speakers' productions of bilabial and labio-dental fricatives were also investigated using electromagnetic articulography. These speakers showed no upward movement of the upper lip for [f, v]. The upper lip in words such as /eve/ *'two'* remained in the same position as in words like /eke/ *'sand'* [6]. The upper lip lowers quite substantially for [ɸ, β], resulting in a visibly higher lip position for the labio-dentals than for the bilabials. However, this is not due to raising the upper lip in the labio-dentals.

In order to study this question in greater depth, 17 Ewe speakers were videotaped saying words contrastind bilabial and labio-dental fricatives, and words containing velar stops in the same vowel environments. In addition, a tape made earlier of another speaker was analyzed. Both frontal and lateral views of the lips were examined on a frame by frame basis.

All of these speakers come from the northern part of the Aŋlo dialect area, where the vowel /e/ is pronounced as a mid front vowel [ɐ̟] rather than as [ə]. Of the population of 20 speakers thus assembled, two show some clear raising of the upper lip in labio-dentals, and two others show some smaller adjustment of the upper lip position either forward or upward. More typical articulations are illustrated in Figures 4-6. These figures are digitized frames from the videotape of speaker 15 showing the culminating phase of the word-medial consonants. Figure 4 shows the lip position in /aɸa/ *'shout'*. The upper lip is lowered and drawn inwards to meet the lower lip. It entirely covers the upper teeth. In /afa/ *'half'* in Figure 5, the upper teeth are only partly covered by the upper lip, and the upper lip is in a position almost identical to that seen in the velar stop in /aka/ *'charcoal'* in Figure 6.

Figure 4. Position of the lips at the center of the consonant in /aɸa/ 'shout'.

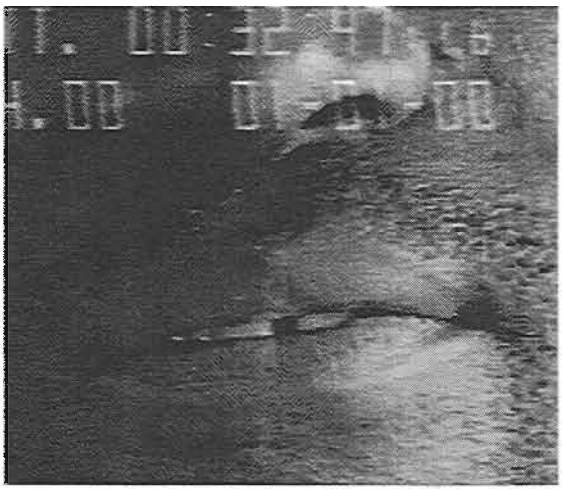


Figure 5. Position of the lips at the center of the consonant in /afa/ 'half'.

Figure 6. Position of the lips at the center of the consonant in /aka/ 'charcoal'.

Most of the Ewe speakers do not raise the upper lip to produce labio-dentals, but a few do. To compare if this is greater than the cross-speaker variability that one might find in another language without a bilabial/labio-dental contrast, 22 speakers of Sɛlɛ were videotaped. This language, spoken by a people who are neighbors of the Ewe, has only one labial fricative of any kind, [f]. Two of the speakers were discarded due to their lack of teeth. Of the remaining 20, two showed a clear raising of the upper lip during [f], three others showed some raising or fronting. Because a more extensive wordlist was taped with the Sɛlɛ speakers, it was also possible to note that the speakers who tended to raise the upper lip for [f] often had a rather similar gesture with certain other consonants, such as [s] and [ɲ].

These data suggest that the occurrence of a raising gesture for labio-dental fricatives is not in any way associated with the presence in the same language of a contrasting bilabial place of articulation for fricatives. Labio-dentals are typically produced in the same way – without an added upper lip gesture – regardless of inventory structure.

REFERENCES

[1] Browman, C. P. & L. Goldstein (1990), "Articulatory Phonology: an overview", *Phonetica,* 49: 155-80.

[2] Maddieson, I (1984), *Patterns of Sounds,* Cambridge: C. U. P.

[3] Maddieson, I. (1993), "Investigating Ewe articulations with electromagnetic articulography", *Forschungberichte des Instituts für Phonetik und Sprachliche Kommunikation, München,* 31: 181-214. (Also see *UCLA Working Papers in Phonetics* 85, 22-53.)

[4] Perkell, J. S., M. Cohen, M. Svirsky, M. Matthies, I. Garabieta & M. Jackson (1992), "Electromagnetic Mid-sagittal Articulometer (EMMA) systems for transducing speech articulatory movements", *Journal of the Acoustical Society of America*, 92: 3078-96.

[5] Ladefoged, P. (1993), *A Course in Phonetics* (3rd Edition), New York: Harcourt Brace Jovanovich.

[6] Ladefoged, P. & I. Maddieson, (1996), *Sounds of the World's Languages,* Oxford: Blackwells.

David Abercrombie and the changing field of phonetics

Peter Ladefoged
[To be published in *Journal of Phonetics*, 1997)

David Abercrombie (1909-1992) played an interesting role in the development of phonetics. Before his time the field was not considered to be a separate university subject in which one could get a degree. The notion of general phonetics as a discipline hardly existed until after World War II. Abercrombie helped define and shape the field. But by the time of his death two events had occurred that changed the role of phonetics: the Chomskyan revolution had made syntax rather than sound systems the major object of study in linguistics; and the needs of communication engineers had become more important than those of language teachers. Abercrombie's view of phonetics is now less central.

David Abercrombie was *the* British phonetician following Jones in the middle of the twentieth century. He re-defined the subject, creating general phonetics as a university discipline that had not previously existed. But by the time of his death on 4 July 1992, new technology and a world with different needs had led to yet another change in the nature of the field.

Abercrombie was born on 19 December 1909 into an idyllic Edwardian world. His father was the poet Lascelles Abercrombie, who came from a large, prosperous Manchester family. For most of his first five years David lived in Ryton, a tiny hamlet in Gloucestershire. The American poet Robert Frost shared the house for part of that time, and the well known war poet Rupert Brooke was a visitor. Abercrombie particularly liked Brooke because he treated him like an adult.

When the first World War came, the family moved to Liverpool, where Lascelles Abercrombie, who was unfit for military service, became an inspector of munitions. They stayed in Liverpool until 1922, as after the war Lascelles became a Lecturer in Poetry at Liverpool University. David was sent to school at Liverpool College, suffering from the bullies of the time. When his father was appointed Professor of English Literature at the University of Leeds he was transferred to Leeds grammar school. He went on to the university and received a third class honours BA in English from Leeds in 1930. That was the highest formal academic level he ever achieved, perhaps explaining why he, an interesting, innovative, meticulous scholar of the highest level in his own right, never set much store on academic credentials.

He originally intended to continue his studies by taking an MA in English Language. writing a thesis on i-mutation, but a meeting with the dominant phonetician of the time, Professor Daniel Jones of University College, London, led to a change of plans. Jones cast his spell and Abercrombie became a student in his department. In order to understand Abercrombie's later contribution to the field of phonetics, it is necessary to understand the status of phonetics in 1930. Daniel Jones was at the height of his power. Over the preceding quarter of a century he had built up a remarkable department, instilling in his junior colleagues great ability in the production and perception of a wide range of sounds. They were skilled teachers of the pronunciation of many of the world's major languages. But the emphasis was on the individual languages and their sounds, and not on the nature of spoken language as a whole.

Abercrombie's colleagues at University College included J. R. Firth, who went on to become Britain's first Professor of General Linguistics, Ida Ward, who became Professor of West African Languages, Hélène Coustenoble, a notable French scholar who helped Abercrombie acquire his impeccable French pronunciation, and Stephen Jones, who was in charge of the UC phonetics laboratory and showed him the value of instrumental records of speech. All these and others, such as Bronislaw Malinowski, the anthropologist at the London School of Economics (where Abercrombie taught English as a Second Language and French from 1934 to 1938) and C. K. Ogden, the philosopher and inventor of Basic English, had a great influence on him. But none of these scholars, including Daniel Jones, shared his vision of phonetics as a basic university subject.

During World War II Abercrombie developed his view of the nature of phonetics. In 1939 when war broke out he had been teaching for a year at the British Council Institute at Athens. Others on the staff there included the phoneticians Ian Catford and Elizabeth Uldall, and the

novelist Lawrence Durrell, whom Abercrombie recruited to teach Basic English. The German invasion of Greece in 1940 pushed the British first into Cyprus, and then into Egypt, where Abercrombie was employed in various government activities, as well as being a Lecturer at the University of Cairo. He met Mary Marble, an American journalist who was working for the U.S. Office of Strategic Services (the OSS, the forerunner of the CIA) in 1943. Through his own official duties he had access to the British dossier on her, which must have been to his satisfaction, as they were married in 1944. At the end of the war, worrying, as he told his new wife, that he had no market value, he returned to England. Fortunately he found that he was able to take up his former position at the London School of Economics.

In 1947 Abercrombie was appointed as a Lecturer in Phonetics in the Department of English (his and his father's old department) at the University of Leeds. This took him closer to the ambition he had been nurturing to teach phonetics as a general university subject, but it was still not what he had planned. In 1948, when Edinburgh University invited him to start a Phonetics Department, he jumped at the chance. The following year saw the first year-long course in Phonetics, attended by his wife and members of staff. The only regular student was his step-daughter, who was just beginning her first year at university.

From then on phonetics flourished at Edinburgh. Elizabeth Uldall had already joined the staff as a lecturer in 1949, and others soon followed. (The author of this paper was an Assistant Lecturer, later Lecturer, from 1953 to 1961.) By 1965, at its peak, the department had a staff of 12, including three who went on to become well known professors (Gillian Brown first at Essex, now at Cambridge, Klaus Kohler at Kiel, and John Laver who now holds a personal chair at Edinburgh), as well as Walter Lawrence, the retired designer of the first parametric speech synthesizer, PAT. In 1964 Abercrombie was appointed to a personal Chair in Phonetics; this became an established chair in 1967. He retired in 1980, but continued to be active in the field.

What was Abercrombie's view of phonetics as embodied in his Ordinary Course (the Edinburgh term for a year-long general introductory course in a subject)? It began, not surprisingly, with an account of the speech production mechanism. But it was a much more complete account of human phonetic capabilities than had been heretofore available for beginning students. Early on students were introduced to the possibilities of different airstream mechanisms, as Abercrombie was well aware of the interest stimulated by discussions of clicks and ejectives. At the same time as this account of the set of possible speech sounds was being developed, students were introduced to phonetic transcription, gradually becoming experts in transcribing their own and others' speech. Abercrombie was the first person to make clear that there were many factors underlying the distinction between a broad and a narrow transcription. He pointed out that one transcription could be narrower than another because it used more specific symbols, such as **ɹ** instead of **r**, or symbols with diacritics such as **d̪** instead of **d**. Alternatively it could be narrower in quite a different way, namely in that it used a greater number of symbols, distinguishing allophones such as English initial **tʰ** and final **ʔt**.

Abercrombie's Ordinary Course introduced students to the sounds of a wide range of languages and many different accents of English, with Scottish English and Scots dialects being given a prominent place. Abercrombie was far from an advocate of his own upper class English accent, RP, as the most important form of English pronunciation. His egalitarian views on accents of English were no doubt shaped by teaching in Scotland, having an American wife and step-children, and his own non-elitist politics.

The Ordinary Course also included lectures on instrumental phonetic techniques and acoustic phonetics, usually given by other members of his staff. Abercrombie saw the necessity for students to have some laboratory experience to round out their phonetic studies. He also stressed the importance of students becoming practically adept phoneticians, and not just experts in the theory of phonetic description. Ear training and performance exercises were an important part of the courses that he taught, often occupying 40% of the teaching time (two out of the five teaching hours a week).

Abercrombie never wrote a book corresponding to the full Ordinary Course. His introductory book, *Elements of General Phonetics* should perhaps have been titled 'Topics in General

Phonetics', as it leaves out much that he considered to be at the core of the subject. The book begins with one of his major contributions to linguistic thought, the clear distinction between a language and the medium for expressing that language. A language is a system of rules for organizing abstract lexical items into sentences. The medium, which can take several different forms, is the method for conveying messages in that system. The medium may be the physical sounds that phonetics describes, or the letters and devices used in written communications, or the bumps of Braille, or the waving of semaphore flags. Abercrombie points out that the medium, be it sounds or written letters or anything else, is an artifact created by humans. As such, as well as conveying linguistic information, it conveys something about whoever produced it. It does this by what Abercrombie calls indexical features. Thus speech provides an index of the group to which the speaker belongs, a mark of the personal characteristics of the individual, and information on the speaker's physical or mental states such as excitement or drunkenness. The medium also has aesthetic properties which come to the fore in poetry (a natural interest of Abercrombie's), advertising slogans, and songs.

The main part of the book following this introductory material is an account of the mechanisms involved in the production of speech. What Abercrombie has to say on this topic is now commonplace, but when it appeared it incorporated many points, such as the nature of stress and an account of the possible airstream mechanisms, which had previously been available only in technical publications. There are also good accounts of basic topics such as the structure of syllables, phoneme theory and assimilation. It does all this in the most clear and simple way possible. Abercrombie took immense pains with his writing. He made sure that each thought followed logically, and was clearly expressed. Irrelevant points were cut out and difficult expressions simplified. As he once said to me, 'It is often difficult to get each sentence exactly right, but it is worth spending hours trying to do so.'

Some of the limitations of *Elements of General Phonetics* are due to its incomplete coverage of the field, but others can be ascribed to Abercrombie's aesthetic susceptibilities. In the Foreword he wrote: 'I hope that I have been able to show that it is possible to present the subject, or at least its elements, without disfiguring the text with the somewhat repulsive diagrams of the vocal organs and the exotic phonetic symbols which, for the general reader, are apt to make it seem unattractive.' It is probably not feasible to do this. Not only are diagrams essential to show movements of the vocal organs that would need a thousand words to describe, but also phonetic symbols, some of them somewhat exotic, are at the heart of work in the field, and the text would have been more helpful in leading students on to further study if it had included more phonetic transcription.

The development of a theory stating clearly what is implied by different types of phonetic transcription is one of Abercrombie's major achievements. But unfortunately his views on proper publication prevented this work from receiving immediate attention. He was delighted when he heard from a friend of his on the editorial board that his book, *English Phonetic Texts* (1964) had been accepted for publication by Faber and Faber, notable publishers of poets such as T.S. Eliot. But as a result an important book never became widely available to the phonetic community. Faber and Faber published an edition of only 1,000 copies, and had no real interest in promoting a book so different from the stock in trade of their regular list. So the careful exegesis and exemplification of different types of transcription, a subject on which Abercrombie was probably the world's leading authority, has taken much longer to have its full impact.

In addition to his wide knowledge of the theory and practice of different styles of phonetic transcription, Abercrombie was the foremost authority on the history of phonetics. One could ask him about almost any technical term in phonetics and he could tell you when it was first used and what its original meaning was. His paper on Isaac Pitman (1937) and his communication to the Philological Society on 'Forgotten phoneticians' (1949) were early work in this area. Throughout his career he taught courses discussing the works of the nineteenth century phoneticians, Bell, Ellis, and Sweet. At the time of his death he was still working on his study of the English phonetician William Holder, whose *Elements of Speech* was published in 1669. (This study has now been put into publishable form by John Kelly, his student, now a Reader in Phonetics at the University of York,.)

The other main area of Abercrombie's research was the study of prosody and rhythm. Sometimes he was able to combine this with his historical interests, as in his paper on 'Steele, Monboddo and Garrick', in which he describes Garrick's performance of the soliloquy 'To be or not to be' in an 18th century production of *Hamlet*, based on an early publication by Joshua Steele. At other times his work in this area reflected his long standing interest in poetry, as in his 'A phonetician's view of verse structure' (1967), or his concern with the contribution of phonetics to the teaching of English as a foreign language, as in his paper on 'Syllable quantity and enclitics' (1964). He was not a believer in the strict isochronicity of stressed syllables in English as might be evidenced (but is not) in laboratory records; but he showed very nicely how 'silent stresses', which he wrote with a stress mark in parentheses ('), might occur to maintain the rhythm, as at the ends of the first, second and last lines in a limerick:

An 'elderly 'lady from 'Ryde (')
Ate 'too many 'apples and 'died. (')
The 'apples fer'mented
In'side the la'mented
Making 'cider in'side her in'side. (')

Abercrombie did not do much work in the phonetics laboratory himself, although he was always very encouraging of the endeavours of others, even being a subject in a number of experiments. He wrote a paper on palatography, and another on speech synthesis in parametric terms. These and other valuable contributions are included in his two collections of papers, *Studies in Phonetics and Linguistics* (1965) and *Fifty Years in Phonetics* (1991). The latter book includes an essay with the same title as the book, which is in itself an excellent appraisal of his work. A complete bibliography of his publications up to 1980 appears in Uldall (1981)

When Abercrombie retired in 1980 there was an abortive attempt to fill his chair. However, two of the three subject heads in the department decided that filling the phonetics chair was not a departmental priority. Although this was probably not in their minds, in doing so they were reflecting the fact that Abercrombie's definition of the field of phonetics was becoming less appropriate. The reasons for studying speech differ from generation to generation. In Abercrombie's heyday phonetics was important in a variety of ways. Abercrombie saw speech as the primary means of conveying linguistic information ('the medium of spoken language' as he would say), and also as a source of sociolinguistic information ('indexical behavior' in his terminology) and personal data ('idiosyncratic information'). He was also concerned with the relation between speech and poetry, speech and writing, and speech as a window into the mind. In his view, phonetics should be of interest to anyone with a natural curiosity about life.

Times have now changed, and although many of these aspects of speech are still of concern, our motives for studying them are somewhat different. Since the advent of Chomsky, who is clearly one of the most powerful thinkers of the second half of this century, it is language, not speech, that is the most fashionable object of study, and syntax, rather than phonology, is generally seen as the central core of language. The grammar of a language includes its phonology and how the sounds are related to phonetic substance, as well as its semantics and how utterances are related to observable meanings, but the study of language has at its heart the morphology of words and the syntax of sentences. Phonetics is thus now seen by linguists as on the periphery of general linguistics. Abercrombie's department, originally the Department of Phonetics, became first the Department of Phonetics and General Linguistics, and then through further amalgamation with the School of Applied Linguistics, simply the Department of Linguistics.

The kind of phonetics that Abercrombie taught is no longer the centrepiece of many university departments because it is no longer the centre of so much research activity. There is still much to be learned about sounds and sound systems, but, largely due to the organization of phonetic knowledge by Abercrombie and people like him, the bases of phonetics are quite clear. The same is not true of syntax, where ongoing research is continually leading to new ways of looking at the fundamental premises of the field. A textbook such as Abercrombie's *Elements of General Phonetics* (1967) can still stand as a valid account of much of the subject, something which is not true of any elementary textbook on syntax written nearly thirty years ago.

The study of other phonetic topics, such as the role of speech in conveying sociolinguistic and personal information, have not diminished in importance, but they have also changed in many ways. Nowadays we are less concerned with helping students acquire a particular accent, such as a native French pronunciation, and more concerned with straightforward description of different accents so that our speech recognition machines can handle them.

Nowadays there are fewer departments teaching anything like the Ordinary Course in Phonetics. It is interesting to consider what Abercrombie might have done, if he were once again a young person asked to start a Department of Phonetics. He would probably place the same emphasis on distinguishing between language and medium. He would also require phoneticians to be skilled performers in the tradition of Bell, Sweet and Jones, which he followed. He would have looked askance at events at a recent scientific meeting (the XIVth International Congress of Phonetic Sciences), where it transpired that several leading participants were unable to produce clicks and ejectives in words. We can speculate on what he would have thought about the new emphasis on Communication Engineering. He was always eager to keep up with the latest technical advances, acquiring in 1950 one of the first Kay sound spectrographs outside the United States for his department. When, in 1953, he heard about speech synthesized by Walter Lawrence on the Parametric Artificial Talker, PAT, he enthusiastically endorsed the idea of research in this field at Edinburgh. Nobody else in the Faculty of Arts at that time had government funding. Abercrombie led the way in securing for his department a contract for basic scientific research. In his later years he felt that the subject of phonetics in the form in which he had helped to establish it was somewhat threatened by the rapid technical advances and government research funds that led to the establishment of a very large Centre for Speech Technology Research at Edinburgh, which he viewed as swallowing up his former department. But if he had been a young person in charge, starting again, he might well have realized that that was the way of the future.

Abercrombie is recognized by many phoneticians as an enormously important figure in their lives. I always feel honoured when people referred to me as his pupil, although it was not until after I had finished writing my own textbook (Ladefoged 1975, 1993) that I fully realized how much I owed to him. As I wrote in the Preface 'My greatest debt is to David Abercrombie, from whom I first learned what I took to be the commonly accepted dogma of phonetics; only later did I discover that many of the ideas were his own contribution to the field.' (My personal debts of course go far beyond the merely intellectual knowledge he bestowed on me. I was very lucky to have him as a teacher and a friend.) But many of the same phoneticians who acknowledge him also recognize that his era has passed. He organized the subject for his time, and earned his own place in the history of phonetics. But now we are moving on.

Acknowledgments.

This paper is based on a memoriam for David Abercrombie to be published in the Proceedings of the British Academy. I am grateful to the Academy for allowing me to use the material in this way. I am indebted for much helpful information to Mary Abercrombie, Mary Brown, John Laver and Elizabeth Uldall. The views expressed are solely my own.

References

Abercrombie, D. (1964) *English phonetic texts.* London: Faber and Faber.

Abercrombie, D. (1965) *Studies in phonetics and linguistics.* London: Oxford University Press.

Abercrombie, D. (1967) *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

Abercrombie, D. (1991) *Fifty years in phonetics.* Edinburgh: Edinburgh University Press.

Ladefoged, P. (1975) *A course in phonetics.* Orlando: Harcourt Brace.

Uldall, E. T. (1981) Bibliography: The published works of David Abercrombie. In *Towards a history of phonetics.* (R. Asher & E. Henderson), pp. 283–288. Edinburgh: Edinburgh University Press.

[The following two papers are related, and will probably be re-written as one. They were prepared for a conference and a festchrift in Korea.]

The IPA and a theory of phonetic description

Peter Ladefoged

The International Phonetic Alphabet (the IPA) is a set of symbols that is intended to represent all the distinctive sounds of the world's languages in terms of well defined phonetic categories. This paper will consider how well it succeeds in this aim, and the extent to which it can be considered a theory of phonetics.

What is meant by 'the distinctive sounds of a language'? There are several assumptions implicit in this notion. The first might be that a language is composed of a set of discrete sounds; but, as every phonetician knows, this is obviously untrue. Acoustic data, such as the spectrogram of the Korean word **jaju** 'booing' in Figure 1, shows that there are no boundaries between the segments. There is no way to separate the consonants from the vowels. A dramatic change in the formants occurs at the point indicated by the vertical line. At this moment the front cavity, which is increasing in size as the body of the tongue moves back for the vowel **u**, becomes sufficiently large to accommodate the resonance associated with the second formant. From then on the downward movement is that of the second formant rather than the third, as it was at the time just to the left of the arrow. But this point cannot be said to be the boundary between **j** and **u**. The downward movement is what characterizes **j**, and most of this occurs later.

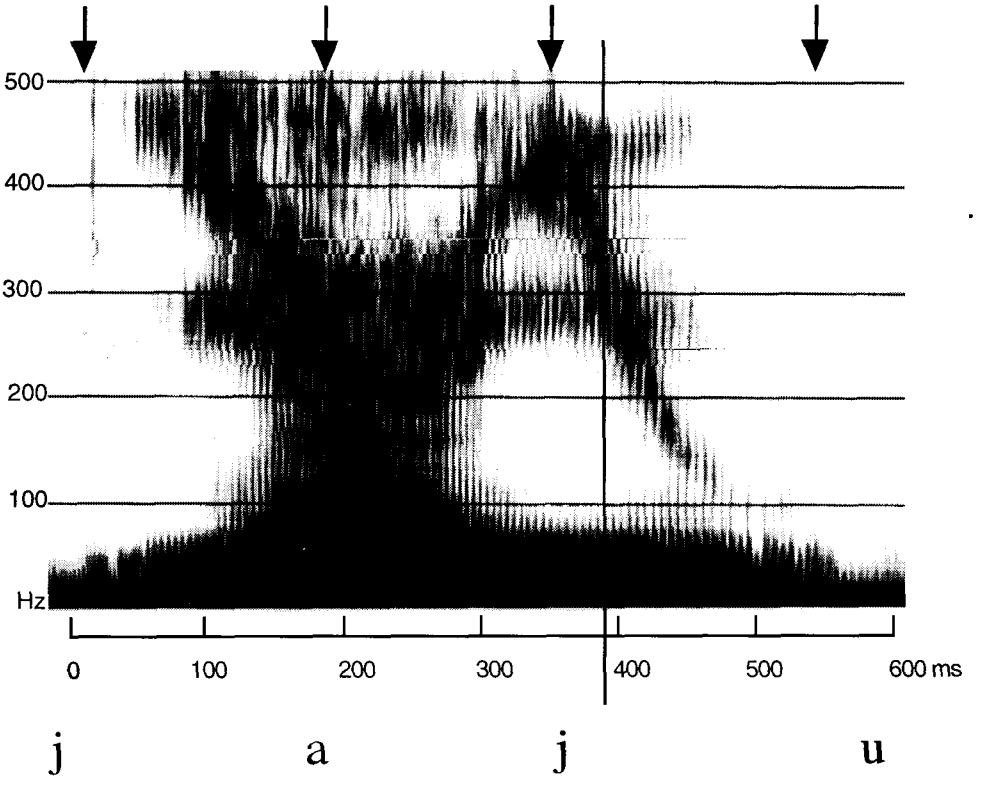


Figure 1. A spectrogram of the Korean word **jaju** 'booing'.

Although we cannot mark the boundaries between segments, Figure 1 does allow us to find their centers. At the points marked by the arrows above the figure there are either steady states or turning points in the formant trajectories. These points allow us to determine the minimum number of segments that there are. It might also seem (Nolan 1995) that we can tell how many segments

there are in a particular word because there are only a limited number of places in that word where changes can occur. Thus the English word 'big' can be said to have three segments because it can be changed at the beginning into 'fig' or 'pig', in the middle to 'bag' or 'bug'. or at the end to 'bin' or 'bit'. Similarly, 'brink' can be changed into 'blink', 'blank', 'plank', etc., showing that it has five. But this line of argument has problems. Changing the end of **plæŋk** 'plank' will necessarily involve a change of what we regard as two segments. There is no way of changing the **ŋ** without also changing the **k**. Once it is changed to, for example, **plænt** 'plant' (involving a change of **ŋk** to **nt**), it can be changed to **plænd** 'planned', changing just one of the last two segments and thus verifying the notion of five segments; but the separability of **ŋk** is not obvious. Nor, to take another example, is it clear that 'pig' and 'big' each have only three segments. It would be possible to regard 'pig' as having four segments, **phɪg**. Then the change from 'pig' to 'prig' would be regarded as changing **phɪg** to **pɹɪg**. Our reasons for considering 'pig' to have three segments and not four are phonological, not phonetic. We often cannot determine how many segments there are in a word by looking at phonetic data such as spectrograms. Segments are the products of a phonological analysis.

When considering the use of the IPA symbols, it does not really matter whether we regard the segments they represent as convenient descriptive fictions determined by our phonological analysis, or as real units that speakers manipulate. But in judging the success of the IPA as a way of describing the distinctive sounds of languages we still have to consider what we mean by distinctive. Do we mean simply contrasting within the language? Or are we concerned with whether a particular sound in one language is distinct from that in another? There is never any problem in finding a set of symbols that will represent all the distinct sounds, the phonemes, in a single language. It is possible to find distinct symbols or combinations of symbols for even a language such as !Xõó, a Khoisan language spoken in the Kalahari desert, which has 83 different ways of beginning a syllable with a click and 33? other consonants. But a more subtle problem arises when we try to say whether a sound in one language is distinct from that in another. Is the **p^{h}** in English **p^{h}ul** 'pool' the same as the **p^{h}** in Korean **p^{h}ul** 'grass'? Most observers would say that they are not the same; and if I used my English pronunciation when trying to say the Korean word, I would be regarded as having a foreign accent. But can we prove that they are distinct sounds?

It is not at all easy to show that there is a measurable difference between two sounds in different languages. For the case in point we would have to measure relevant properties of the two sounds in comparable words in English and Korean. We could assume that the differences are in the actions of the glottis and measure the Voice Onset Time (VOT). We could also quantify the phonation type in terms of the spectral slope or the relative amplitudes of the formants. We would have to record the two words as produced by at least 20 speakers of each language, so as to be sure that we were capturing properties of the language, and not just those of individual speakers. The Korean and English speakers should be matched for sex, age and size (height and weight), which would certainly be a challenging task. It is unlikely that the differences are due to physical differences between speakers of English and speakers of Korean, but this is always a possibility. Disner (1983; see also Ladefoged 1983) showed that there were systematic differences between the seven vowels of Italian and the seven vowels of Yoruba, and speculated that these differences could be due to the different facial characteristics of Italian and Yoruba speakers. At the end of our investigation of 20 matched speakers of Korean and English we might be able to show that the **p^{h}** in English **p^{h}ul** 'pool' is not the same as the **p^{h}** in Korean **p^{h}ul** 'grass'. But it is clear that this is not the kind of difference that the IPA was set up to try to capture.

Korean offers us another example of the difficulties the IPA has in meeting its goals. Having considered the word **p^{h}ul** 'grass', we can now ask what is the best IPA representation of the word **p*ul** 'horn'? As is well known, the Korean so-called fortis stops are unique among the world's languages; there is nothing like them anywhere else. I have here and elsewhere (Ladefoged 1971, Ladefoged 1993, Ladefoged and Maddieson 1996) used an asterisk to indicate this particular phonetic quality. This is not an officially sanctioned IPA symbol, but the IPA does not provide any way of marking this sound as distinctive from those of other languages. It can of course distinguish the three sets of Korean stops by transcribing them as, for instance, **p^{h}**, **p'**, **b**, etc. with the fortis stop being marked by a following apostrophe, But this symbol is officially reserved

for ejectives, sounds that occur in, for example, the American Indian language Lakhota. In this language the word **p'o** 'foggy' begins a voiceless bilabial ejective, a very different sound from that used in Korean **p*ul** 'horn', in which the glottalic airstream mechanism does not build up pressure behind the articulatory closure.

Thus far we have seen two ways in which the IPA is incomplete. It cannot (and, as we will see, should not) symbolize small, language specific, differences in parameters such as VOT; and it does not have distinct symbols for some really distinctive sounds. There are a number of other cases of this latter type that we will discuss in an attempt to find out why they have not been included. But first we should note that there are cases of distinctive sounds that occur in only a limited number of languages for which IPA symbols *have* been specially created. Why has this occurred?

The prime example of the latter category — a special symbol for a sound of a particular language — is the symbol **ɧ** for the so-called labial velar fricative in Swedish. This is an especially interesting case because it seems fairly clear that labial velar fricatives are impossible to produce; sounds with two sources for a fricative noise cannot occur (Ladefoged and Maddieson 1996:329). Nevertheless the IPA chart includes a symbol for labial velar fricatives. They were thought to occur in some Swedish dialects, but now it appears that the situation (somewhat simplified) is that some Swedish dialects have a velarized palato-alveolar fricative in their phonological inventory where others have a palatalized bilabial fricative and yet others have variants of these; but no dialect has a labial fricative that is simultaneously a velar fricative. The alphabet devised for the study of Swedish dialects in the nineteenth century provided a symbol that has been used for all the dialectal variants, each of which could be described using other IPA symbols and diacritics. This special Swedish symbol was incorporated in the version of the IPA produced in 1949 largely because of the influential Swedish phoneticians at that time.

Until recently (International Phonetic Association 1989) the IPA included another example of a special symbol for a sound that occurred in a particular language. The Japanese syllabic nasal **ɳ** was recognized as being different from all other sounds, despite the fact that it is typically a velar nasal **ŋ**. This sound occurs in words such as **ni.p.po.n** 'Japan', which can be divided into four mora, as indicated by the periods (the IPA symbol for a syllable division). Because the syllabic nasal has a special place in Japanese phonology (it does not occur before a vowel; it is always a single mora by itself), it came to have a special symbol. At the 1989 Kiel convention the International Phonetic Association decided that from a phonetic point of view there was nothing unusual about this sound (however unusual it might be phonologically), and withdrew recognition from the symbol **ɳ**.

Further examples of unique sounds that occur in a limited number of languages are not hard to find. Passy (1899) and Catford (1992) have described a bidental fricative in the Shapsug dialect of the North Caucasian language Adyghe, in which friction is caused by air passing between the clenched teeth. Everett (1982) described a sound in the Mura language Pirahã in which the tongue tip contacts the alveolar ridge and then moves forward so that it protrudes out of the mouth, pointing down towards the chin. Ladefoged and Everett (1996) described a sound in the Chapakuran languages Wari' and Oro Win in which an alveolar stop is released with a groove in the center of the tongue in such a way that the lips may be set vibrating. None of these sounds has been given a special symbol in the IPA.

So what does it take to get a special symbol for a sound in the IPA? A cynical answer is that it takes a phonetician who speaks the language and is on the Council of the International Phonetic Association. The IPA is a collection of symbols that has to be approved by a vote of this body. But perhaps a better answer is to think of the IPA as a somewhat ill-formed theory of phonetics. A theory in this sense is something that tries to account for a body of data, in this case the range of sounds that languages use. The core of a phonetic theory is a set of categories that define possible sounds. Within this theory, the IPA is the way of defining the symbols that show the relations among the categories. Thus **p** is a shorthand way of designating the intersection of the categories voiceless, bilabial, and plosive; **m** is the intersection of the categories voiced, bilabial, and nasal; and so on.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993, corrected 1996)

CONSONANTS (PULMONIC)

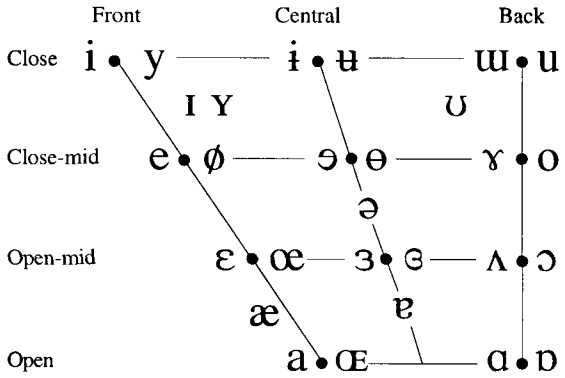
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	’	Examples:
ǀ	Dental	ɗ	Dental/alveolar	p’	Bilabial
ǃ	(Post)alveolar	ʄ	Palatal	t’	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	k’	Velar
ǁ	Alveolar lateral	ʛ	Uvular	s’	Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ʍ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
w	Voiced labial-velar approximant	ɺ	Alveolar lateral flap
ɥ	Voiced labial-palatal approximant	ʃ͡ʒ	Simultaneous ʃ and x
ʜ	Voiceless epiglottal fricative		
ʢ	Voiced epiglottal fricative		
ʡ	Epiglottal plosive		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

SUPRASEGMENTALS

ˈ	Primary stress	ˌfoʊnəˈtɪʃən	
ˌ	Secondary stress		
ː	Long	eː	
ˑ	Half-long	eˑ	
˘	Extra-short	ĕ	
|	Minor (foot) group		
‖	Major (intonation) group		
.	Syllable break	ɹi.ækt	
‿	Linking (absence of a break)		

DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. n̥

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰
◌ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼
◌̹	More rounded	ɔ̹	◌ʷ	Labialized	tʷ dʷ
◌̜	Less rounded	ɔ̜	◌ʲ	Palatalized	tʲ dʲ
◌̟	Advanced	u̟	◌ˠ	Velarized	tˠ dˠ
◌̠	Retracted	e̠	◌ˤ	Pharyngealized	tˤ dˤ
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝ (ɹ̝ = voiced alveolar fricative)
◌̩	Syllabic	n̩	◌̞	Lowered	e̞ (β̞ = voiced bilabial approximant)
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘
◌˞	Rhoticity	ɚ a˞	◌̙	Retracted Tongue Root	e̙
◌̋	Extra high	˧̋	◌̌	Rising	˨˩̌
◌́	High	˧́	◌̂	Falling	˧̂˥
◌̄	Mid	˧̄	◌᷄	High rising	˧᷄˥
◌̀	Low	˩̀	◌᷅	Low rising	˩᷅˧
◌̏	Extra low	˩̏	◌᷈	Rising-falling	˩᷈˩
↓	Downstep		↗	Global rise	
↑	Upstep		↘	Global fall	

TONES AND WORD ACCENTS

LEVEL: ˧̋ or ˧́ Extra high; ˧́ High; ˧̄ Mid; ˩̀ Low; ˩̏ Extra low; ↓ Downstep; ↑ Upstep

CONTOUR: ˨˩̌ or ˧̂˥ Rising; ˧̂˥ Falling; ˧᷄˥ High rising; ˩᷅˧ Low rising; ˩᷈˩ Rising-falling; ↗ Global rise; ↘ Global fall

The IPA chart shown in figure 2 is a one page account of a general phonetic theory. We can begin our discussion of this theory by noting that it is easy to show that nearly all the terms in the chart are needed in order to account for phonemic differences within languages. In this paper we will limit the discussion to consideration of possible places of articulation. Figure 3 shows languages that contrast the places of articulation in the consonant chart. There are contrasts between either voiceless stops or fricatives (or sometimes both) at nearly all the possible pairs of places of articulation. Thus the labiodental fricative **f** contrasts with the bilabial fricative **ɸ** in Ewe, a Niger-Kordofanian language. Malayalam, a Dravidian language, has six contrasting places of articulation for stops and nasals. Toda, another Dravidian language, has the same contrasts but also has additional contrasts among voiceless fricatives. The only blanks in the chart are almost certainly accidental in the sense that we have not yet found a language that exhibits the contrast, although there could well be one. Retroflex, palatal, uvular and pharyngeal sounds are all comparatively rare and it is not surprising that there are no known languages contrasting some pairs of these places of articulation. Nevertheless, the fullness of the matrix in Figure 3 demonstrates the necessity for all the places of articulation in the IPA consonant chart.

The next question is whether the IPA categories for place of articulation are not only necessary but also sufficient to characterize all the distinct sounds that occur in languages. Ladefoged and Maddieson (1996), in their survey of the sounds of the world's languages, find that they need to consider a greater number. In a chart similar to that in Figure 3 they list 17 places of articulation, 6 more than the 11 given in the IPA consonant chart. Part of the increase is due to the different aims pursued by Ladefoged and Maddieson. They are concerned with more than the phonemic distinctions that occur within languages. They also "take note of differences between languages." They want to describe "those segmental events that distinguish one language or accent from another and which are also sufficiently distinct to serve as potential conveyers of lexical contrasts for speakers of other languages." (Ladefoged and Maddieson 1996:3.) How distinct segments have to be in order to meet the criterion of being potentially able to form lexical contrasts is not clear; it is essentially a judgment call.

One of the extra possibilities listed by Ladefoged and Maddieson is epiglottal. There is no judgment needed here as epiglottal fricatives contrast with pharyngeal fricatives in at least one language, Agul (North Caucasian). The IPA allows this possibility by listing epiglottal fricatives among a collection of 'other symbols' showing sounds that could not easily be placed on the chart. (It is not clear why they could not have been placed on the chart; there could easily have been another column 'Epiglottal', which would have had three entries, one more than there now are for 'Pharyngeal'.)

The other five extra possibilities listed by Ladefoged and Maddieson can be handled by diacritics, small marks that can be added to a symbol to change its value. In a section below the main consonant chart, the IPA lists 31 diacritics. These include ways of indicating whether an articulation is apical or laminal. Ladefoged and Maddieson point out that it is not clear whether there are contrasts between apical and laminal articulations at the same place on the upper surface of the mouth, but it is plain that some languages typically have laminal dental articulations and others have apical dental articulations; similarly some languages have apical alveolar articulations and others have laminal alveolar articulations. It seems unlikely that there could be lexical contrasts of these kinds.

One of Ladefoged and Maddieson's other additions can be handled by the diacritic indicating that an articulation is made further forward. This diacritic can be used to mark the distinction between dental and interdental articulations. Californian English speakers — and many other speakers of Western General American dialects — typically pronounce the initial consonant of words such as 'thin, thanks' with an interdental articulation in which the tip of the tongue protrudes between the teeth. Speakers of most forms of British English have a dental articulation for these sounds, keeping the tip of the tongue behind the upper teeth. This difference is never used contrastively within a language, and it is not at all clear that it is ever likely to do so.

		BILABIAL	LABIODENTAL	DENTAL	ALVEOLAR	POSTALVEOLAR	RETROFLEX	PALATAL	VELAR	UVULAR	PHARYNGEAL
		p ɸ	f	t̪ θ	t s	t̠ ʃ	ʈ ʂ	c ç	k x	q χ	ʡ ʕ
LABIODENTAL	f	Ewe									
DENTAL	t̪ θ	Malayalam	Toda								
ALVEOLAR	t s	Malayalam	Toda	Malayalam							
POSTALVEOLAR	t̠ ʃ	Hindi	Toda	Toda	Toda						
RETROFLEX	ʈ ʂ	Malayalam	Toda	Malayalam	Malayalam	Toda					
PALATAL	c ç	Malayalam	Hungarian	Malayalam	Malayalam	Logba	Malayalam				
VELAR	k x	Malayalam	Gaelic	Malayalam	Malayalam	Hindi	Malayalam	Malayalam			
UVULAR	q χ	Quechua	German	Urdu	Quechua	Urdu		Jaqaru	Quechua		
PHARYNGEAL	ʡ ʕ	Agul	Arabic	Dahalo	Dahalo	Dahalo			Dahalo	Agul	
GLOTTAL	ʔ h	Hawaiian	Arabic	Dahalo	Dahalo	Dahalo	Kuvi	Margi	Hawaiian	Ubykh	Agul

Figure 3. A matrix showing languages that have contrasts between either voiceless stops or fricatives (or sometimes both) at the various places of articulation listed in the IPA consonant chart.

Another diacritic, the one indicating that an articulation is made further back, can be used to differentiate between the two kinds of retroflexion observed by Ladefoged and Maddieson, the post-alveolar articulation in Indo-Aryan languages such as Hindi, and the more retracted articulation involving the under side of the tongue used in Dravidian languages such as Malayalam. This is another contrast that is never used to distinguish words within a language, although in this case the articulations are sufficiently distinct to produce an easily audible difference that might well be contrastive in some language.

Linguo-labials, the remaining additional place of articulation noted by Ladefoged and Maddieson, occur in Tangoa and other languages spoken in Vanuatu. They are formed by an articulation using the blade of the tongue and the upper lip. Tangoa has stops, nasals and fricatives with this articulation, contrasting with both bilabial and alveolar articulations. The IPA includes a diacritic specifically for these sounds. It is a little hard to consider this diacritic as a mark that can be added to a symbol to change its value in the same way a other diacritics. The change in articulation from a dental or alveolar stop to a linguo-labial stop is comparable with the difference between a dental or alveolar stop and a palatal stop. One suspects that if linguo-labials had been as common as retroflex, palatal, or uvular sounds, they would have had their own unitary symbols.

There are reasons for and against including diacritics in a system of phonetic representation such as the IPA. A reason for including them is that they are convenient ways of showing classes of sounds. All sounds with a small circle beneath them are voiceless; all sounds with the diacritic ̡ are palatalized. Many people in the International Phonetic Association have argued that the IPA should go further in this direction, and recognize the diacritic ˇ, which has been widely used to mark (and thus group together) the palato-alveolar sounds ʃ, ʒ, t͡ʃ, d͡ʒ, writing them instead as š, ž, č, ǰ.

At least at first glance a reason for not using diacritics is that their use goes directly against the IPA principle of having a separate symbol for each distinctive sound. But as our knowledge of linguistic phonetic events increases, it becomes more and more apparent that this principle should be dropped. There are several hundred distinct sounds used in the world's languages, and there is no way in which this principle — one symbol one sound — can be maintained in an alphabet such as the IPA.

There is, however, another disadvantage to the use of diacritics that must be recognized. A diacritic can be added to any symbol, and we thus lose the notion of the IPA being all *and only* the symbols that are needed for specifying the distinctive sounds that occur in the languages of the world. Our phonetic theory is becoming too powerful in that diacritics can be used to specify impossible sounds such as laminal uvulars or voiceless voiced segments. To some extent it has always been the case that the IPA theory is too powerful, and efforts have been made to remedy the problem by noting on the chart that certain articulations are judged to be impossible. But the possibility of adding diacritics in an unrestricted way is alarming. The best solution would be for the International Phonetic Association to drop both the principle requiring it to aim for a separate symbol for each distinctive sound, and the principle requiring it to avoid diacritics. In return it should adopt a set of statements specifying the symbols that each diacritic can modify.

Achieving this aim will be difficult. The IPA is organized and approved by the Council of the International Phonetic Association, a group of 30 very diverse phoneticians with widely different beliefs about some aspects of the IPA. Some are linguists, some language teachers, some speech communication engineers, and some speech pathologists. They do not all share my desire to regard the IPA as embodying a theory of phonetic description confined to what goes on in languages. The Council is also very conscious of the fact that it is the guardian of a tool used by many people who are not professional phoneticians. All this leads to a justifiable reluctance to change anything. It is vital for the IPA to be a stable set of symbols whose core is not changed. Addition of a few extra symbols or diacritics for rare sounds will not affect the majority of users; but a change in the underlying principle on which the IPA is based may have far ranging effects. For my part I would like to see alternative possibilities allowed, such as recognizing š, ž, č, ǰ as well as ʃ, ʒ, t͡ʃ, d͡ʒ. But I doubt it will happen. Democracy always has its problems.

References

Catford, J. C. (1992). Caucasian phonetics and general phonetics. In C. Paris (Ed.), *Caucasologie et mythologie comparée Actes du Colloque international du C.N.R.S.* (pp. 193 - 216). Paris: Peeters.

Disner, S. F. (1983) *Vowel quality: The relation between universal and language-specific factors.* Doctoral dissertation, University of California, Los Angeles.

Everett, D. L. (1982). Phonetic rarities in Pirahèa. *Journal of the International Phonetic Association, 12*, 94-96.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics* ((Midway reprint 1981) ed.). Chicago: University of Chicago Press.

Ladefoged, P. (1982). 'Out of chaos comes order': Physical, biological, and structural patterns in phonetics. In M. P. R. Van den Broecke & A. Cohen (Eds.), *Proceedings of the 10th International Congress of Phonetic Sciences* (pp. 83-96). Dordrecht: Foris Publications.

Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). New York: Harcourt Brace.

Ladefoged, P. and Maddieson, I. (1996) *The sounds of the world's languages.* Oxford: Blackwells.

Nolan, F. (1995) The handbook of the IPA (draft version). *Journal of the International Phonetic Association.* 25.1.

Passy, P. (1899). *Les Sons du Français* (5th edition ed.). Paris: Association Phonétique Internationale.

The IPA and the phonetics-phonology interface

Peter Ladefoged

There are not many complex theories that can be summarized on one side of a single sheet of paper. General phonetics, the theory required for describing the sounds of languages, might be considered to be one of them. It is summarized in the chart of the International Phonetic Alphabet (henceforth the IPA) shown here in Figure 1 [in previous paper]. The chart, which actually consists of three separate charts and a set of tabulated material, shows the principal phonetic categories required in the description of speech. The symbols on the chart are shorthand forms of descriptions using these categories. Thus **p** designates the intersection of categories with values [voiceless], [bilabial], and [stop]. **i** denotes [high], [front], [unrounded], [vowel].

The organization of the chart is particularly interesting because, perhaps due to chance or perhaps because of the innate wisdom of the International Phonetic Association, it nicely highlights some of the problems in the interface between phonetics and phonology. Phonology provides an account of the abstract units that surface as the phonetically observable sounds of the language. The IPA symbols have a dual character that has sometimes been overlooked. They are equally suitable for representing the phonological contrasts in a language, and for showing the phonetic detail in particular utterances.

Phonologists typically divide the sounds of languages into vowels and consonants (and other possibilities that we shall consider later). These are also the two major subdivisions of the IPA chart. There is a further sub-chart of the IPA showing the symbols and categories required for sounds made with other airstream mechanisms, such as clicks, implosives, and ejectives. This sub-section separates out sounds that phonologists regard as more highly marked, in the sense that they are less likely to occur in the world's languages.

The final sections of the chart include a set of what are labeled as 'other symbols' and a table of diacritics that provides for more detailed specification. In addition a set of symbols for suprasegmentals are provided so that properties of whole words and sentences such as tone, stress and intonation can be symbolized.

To what extent does the IPA chart represent an adequate theory of phonetics? In one sense the answer is superbly, in that there is hardly a sound in the world's languages that cannot be represented. One might quibble that it does not capture the full essence of a few esoteric sounds, such as the bicussive dental fricative in Adgye described by Catford (1993), or the dental plosive induced bilabial trill in Wari' described by Ladefoged and Everett (1996). But these are minor concerns. In another sense however, the chart does not represent the best possible theory, in that it does not show some relations among sounds, nor does it make some distinctions that are relevant for a theory of linguistic phonetics. The theory with which we will compare the IPA chart is based on that outlined in the final chapter of Ladefoged and Maddieson (1996). It is adumbrated in Ladefoged (1973, 1992), and presented more fully in Ladefoged (1997). There is a vast amount of phonetic detail that should be considered in presenting this theory, but we will give only a bare outline here.

The basic notion of this expansion of traditional phonetic theory is that most sounds can be described in terms of modal values of certain phonetic parameters. The values are largely categories on the IPA chart such as [back] and [front] for vowels, and [plosive], [fricative] and [approximant] for consonants. The parameters are similar to familiar terms for groups of values, such as Place and Manner of articulation, but with some differences. Values of parameters are always given in square brackets, e.g. [high], [low], and names of parameters are always capitalized, e.g. Vowel Height. It is also an important part of the theory to recognize that there are

some sounds that cannot be described in terms of values of general phonetic parameters; they have to be listed separately, rather like the 'other symbols' on the chart.

A major difference between this theory and the traditional IPA approach is that phonetic parameters are defined as physical properties that can be measured. They are, with few exceptions, scalar variables, things that can be more or less. For example, Vowel Height is defined in terms of the frequency of the lowest resonance of the vocal tract (F1). Vowels can have any degree of vowel height; but some values, the modal values, are more likely to occur than others. Techniques for measuring values of some of the parameters are not well developed, and in this brief account of the theory we will be able to sketch only an outline of the relevant physical properties.

A list of the principal phonetic parameters and modal values corresponding to IPA places of articulation is given in Tables 1. The modal values include some terms not on the IPA consonant chart: [linguolabial], [interdental], [apical], [laminal], [sublaminal]. These values are discussed in Ladefoged and Maddieson (1996).

Table 1. Parameters and modal values corresponding to IPA places of articulation.

PARAMETER	MODAL VALUES
Labial	[bilabial], [labiodental]
Coronal	[linguolabial], [interdental], [dental], [alveolar], [post-alveolar]
Apicality	[laminal], [apical], [sublaminal]
Dorsal	[palatal], [velar]. [uvular]
Radical	[pharyngeal]. [epiglottal]
Laryngeal	[glottal]

The IPA chart can be taken as regarding the place of articulation as a single variable (or, alternatively, as a set of independent binary categories). If we want to consider this as a single parameter it could conceivably be measured as the distance from the glottis to the place of articulation. From a phonological point of view it is best to consider five separate variables, each measurable as the normalized distance between the point of articulation and some landmark within the vocal tract, such as the lips (for the parameter Labial), the tip of the incisor teeth (the Coronal parameter), the tip of the uvular (Dorsal) and the glottis (Radical and Laryngeal). In addition we need a parameter, Apicality, which is measured along the surface of the tongue. There are no established procedures for making any of these measurements. They might be made using palatographic or MRI techniques, or by an analysis by articulatory synthesis procedure, similar to that used by Ladefoged (1976) in his description of how to put one person's tongue in another person's mouth. In any case, values have to be normalized in the sense that they have to be scaled in respect to the size of the speaker's head.

The major reason for replacing the single variable, Place of Articulation with the five parameters, Labial, Coronal, Dorsal, Radical and Glottal, is that it allows us to state and delimit possible combinations of simultaneous articulations. Many different combinations of these five parameters occur, sometimes with the same degree of stricture, as in **k͡p**, and **t͡p**, and sometimes with one primary and one secondary articulation as in $\mathbf{k^w}$ and $\mathbf{k^j}$. If the two articulations have the same degree of stricture they must involve different parameters; there are no languages that use sounds such as **k͡q** or **t͡ʈ**. When there are two different degrees of articulation such as a stop and a semivowel, then both articulations can involve the same parameter, as in a labialized bilabial stop.

A second reason for this set of parameters is that it allows us to distinguish the most complex articulatory region. In one, and only one, of the regions two choices have to be made simultaneously. Coronal sounds must also have a value of the Apicality parameter. Of course in

many languages, including English and French, it does not matter whether a Coronal stop is made with an apical or laminal articulation; but from a phonetic point of view it must have one or the other articulation. In many languages, such as Malayalam and O'odham, the distinction between the two possibilities is an import part of a phonological contrast. Ladefoged (1997) provides a formal way of representing the relations between parameters that require joint specification as well as other kinds of dependencies among parameters.

The IPA chart sets out the manners of articulation for consonants as labels for the rows in the consonant chart. If we are thinking in terms of parameters and values these labels must be reorganized as shown in Table 2. There is only one parameter, Stricture, that has several modal values. This parameter is measurable as the degree of opening of the vocal tract. There are obvious modal values associated with complete closure, [stop], a small opening sufficient to cause turbulence, [fricative], and a more open vocal tract, [approximant]. In between these values are others that may be associated with weak stops and weak fricatives.

Two of the other parameters, Tap and Trill, have only a single modal value each. Possible ways of relating them to each other and to the Stricture parameter are considered by Ladefoged (1996), but it seems likely that these two parameters will always remain the least obviously scalar properties. For most linguistic purposes a sound is either a trill or a tap or neither.

The remaining parameters in Table 2, Nasal and Lateral, are more clearly scalar properties, each with two modal values. Variations in the amount of nasality, the extent of the velic aperture, are common and need no further comment. It is also true that a sound can be more or less Lateral, if laterality is defined as the proportion of air flowing over the sides as opposed to over the center of the tongue. Some versions of final **l** in English are partially lateral in the sense that the tongue is narrowed from side to side, but there is no complete central closure, so some of the airstream flows out laterally and some centrally.

Table 2. Parameters and modal values corresponding to IPA manners of articulation.

PARAMETER	MODAL VALUES
Stricture	[stop], [fricative], [approximant], [vowel]
Tap	[tap]
Trill	[trill]
Nasal	[nasal], [oral]
Lateral	[central], [lateral]

The terms used on the IPA chart for characterizing vowels and some approximants are readily expressible as modal values of parameters as shown in Table 3. The first two, Height and Backness, are well established as parameters. Although there is still some discussion concerning the correct physical measures (acoustic variables such as the frequencies of the first two or three formants seem most appropriate), all observers agree that there is a vowel space within which most of the vowels of the world's languages can be described. It is, however, not quite so clear how many modal values should be recognized. Some languages cannot be adequately described in terms of only three possibilities for each parameter, [high], [mid], [low], and [front], [central], [back]; but perhaps these languages, which happen to include well known languages such as English, German and Danish, should be regarded as atypical in their use of the vowel space, in that they require more than the modal possibilities of the principal vowel parameters.

Table 3. Parameters and modal values corresponding to terms used on the IPA chart for describing vowels and some approximants.

PARAMETER	MODAL VALUES
Height	[high], [mid], [low]
Backness	[front], [central], [back]
Rounding	[spread], [rounded]
Tongue Root	[advanced], [retracted]
Rhotic	[rhotacized]

Rounding is also a well established parameter. There are, however, other views. Ladefoged and Maddieson (1996) consider Rounding to be composed of two independent parameters, Protrusion, with possible values [protruded] and [retracted], and Compression, with possible values [compressed] and [separated]. They do this because they are seeking an all inclusive form of description that can be used to describe any speech sound. This is not the aim of the theory being developed in this paper, which stipulates that there are two groups of sounds, those that should and those that should not be described in terms of values of general phonetic parameters. Within this theory there is little justification for replacing the parameter Rounding with the two parameters Protrusion and Compression. There are very few languages that need the extra parameters. The fact that they are required for much studied languages such as Swedish is irrelevant to the issue of whether these languages should be regarded as oddities in the phonetic world. In this paper we will consider only a single parameter, Rounding, measurable in terms of a combination of lip opening and protrusion as suggested by Fant (1960).

It might be thought that the remaining two parameters listed in Table 3, Tongue Root and Rhotic, should also be excluded from the set of parameters that are generally applicable in descriptions of the sounds of the world's languages. However, in many different parts of the world (the Caucuses, East Africa, West Africa, South East Asia) there are unrelated languages that have vowels distinguished by [advanced] versus [retracted] Tongue Root. This feature is definitely required for describing the vowels of the world's languages. It could be measured directly in terms of the position of the root of the tongue; there are also other possibilities suggested by factor analyses of tongue shapes (Jackson 1988, Nix et al. 1996).

The parameter Rhotic applies to vowels in remarkably few languages, by chance including the two most widely spoken languages, English and Standard Chinese. These two languages (or, at least, General American English and Pekingese Chinese) are phonetically odd in having phonologically contrastive r-colored vowels. A more important reason for including Rhotic in the set of parameters is that it is needed for the description of approximants such as different forms of ɹ. Like the parameters Height and Backness, it can be measured in terms of an acoustic variable, in this case the frequency of the third formant.

The IPA chart recognizes only two states of the glottis in the consonant chart; sounds are either voiced or voiceless. There are, however, diacritics for aspiration, breathy voice and creaky voice. One way of arranging these possibilities in terms of parameters and values is shown in Table 4. The Glottal Stricture parameter can be considered to be quantified in terms of the distance between the arytenoid cartilages, being furthest apart in voiceless sounds, brought slightly together in sounds with breathy voice, more closely approximated in regularly voiced sounds, even closer together in creaky voice, and pressed tightly together in a glottal stop. More complex phonation types can occur, but this single parameter provides a sufficient description of the glottal states for most linguistic purposes. The timing of the occurrence of these states with respect to the articulatory gestures they accompany is specified by means of the Glottal Timing parameter, for which the traditional measure is VOT (Voice Onset Timing). This table also includes a parameter, Glottal Movement, that accounts for ejective and implosive sounds. It is probably the rate of glottal

movement upward or downward that has to be measured, but little work has been done on quantifying this parameter. It is, however, known that some languages have weak ejectives whereas others have more forcible ejectives; and some languages have weak implosives and some strong (Ladefoged and Maddieson 1996).

Table 4. Parameters and modal values corresponding to terms used on the IPA chart for describing laryngeal activity.

PARAMETER	MODAL VALUES
Glottal Stricture	[voiceless], [breathy voiced], [modal voice], [creaky voiced], [closed],
Glottal Timing	[aspirated], [unaspirated]
Glottal Movement	[raising], [lowering]

The remaining parameters account for further differences in sounds due to the airstream mechanisms. The IPA does not recognize fortis–lenis contrasts, noted in Table 5 as possible values of a Pulmonic parameter. It has been shown by Dart (1987) that Korean fortis and lenis stops differ in subglottal pressure, a measure of this parameter. This table also includes the Velaric parameter, associated with clicks, which can be quantified in terms of the degree of oral suction. Some sounds, such as the k͡p in Yoruba, have a very weak degree of oral suction and do not qualify as modal clicks. In Zulu clicks the oral suction may have a value of –250 cm H_2O (Thomas, p.c.). This is more than an order of magnitude greater than the positive pressure in a regular plosive.

Table 5. Parameters and values for other aspects of sounds due to variations in the airstream mechanism.

PARAMETER	MODAL VALUES
Pulmonic	[fortis], [lenis]
Velaric	[click]

The parameters we have been discussing neither can nor should encompass the complete range of sounds that occur in the world's languages. We have already mentioned some rare sounds, such as fricatives produced between clenched teeth and bilabial trills produced by specially shaped dental plosives. Swedish vowels with compressed rounding should also be considered as rare sounds. Ladefoged and Maddieson (1996) describe a large number of other unusual sounds that occur in relatively few languages. Languages can use any sound that can be produced by the vocal organs with sufficient ease to enable it to be integrated into the stream of speech. We can never know the full range of this larger set of sounds. Some unusual sounds may have occurred in languages that existed in the past; others may appear in future languages. We know that many possible speech sounds have not been observed, although they are no more complex than Khoisan clicks or Toda trills. For example, there are no known languages with pulmonic ingressive fricatives or pure tone whistles. Nearly everybody can make these sounds, and, with a little practice, can integrate them into the flow of speech. But we do not want to set up special parameters to deal with them just in case they might turn up in some language.

The theory of phonetics outlined here makes it clear that there are two sets of sounds. The first set can be described in terms of the parameters listed above, It contains the more common sounds that participate in a wide range of general linguistic processes. The other set consists of the rarer sounds that have been observed in only one or two languages. The central nature of some sounds as opposed to the peripheral nature of others is captured by the IPA consonant and vowel

charts that contain the core sounds, complemented by lists of other symbols and diacritics that are necessary for describing other sounds. The same notion is more formally expressed by a theory in which there a set of parameters with modal values that occur in language after language, leaving other sounds for more *ad hoc* description. At the moment the boundary between the two sets of sounds is not clear. A future step in the development of phonetic theory is to make the distinction between the two sets more explicit.

Acknowledgment

A number of the ideas in this paper (notably those concerned with multiple articulations and degrees of laterality) came from my colleague Ian Maddieson, who has been, as always, very helpful.

References

Catford, J. C. (1992). Caucasian phonetics and general phonetics. In C. Paris (Ed.), *Caucasologie et mythologie comparée Actes du Colloque international du C.N.R.S.* (pp. 193 - 216). Paris: Peeters.

Dart, S. (1987). An aerodynamic study of Korean stop consonants: Measurements and modeling. *Journal of the Acoustical Society of America,* 81(1), 138-147.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Jackson, M. T. T. (1988). Analysis of tongue positions: Language-specific and cross-linguistic models. *Journal of the Acoustical Society of America,* 84, 124-143.

Ladefoged, P. (1975). *A Course in Phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1976). How to put one person's tongue inside another person's mouth. *Journal of the Acoustical Society of America,* 60, S77.

Ladefoged, P. (1992). The many interfaces between phonetics and phonology. In W. U. Dressler, H. C. Luschitzky, O. E. Pfeiffer, & J. R. Rennison (Eds.), *Phonologica 1988* (pp. 165-179). Cambridge: Cambridge University Press.

Ladefoged, P. (1997). Linguistic phonetic features. In W. Hardcastle and J. Laver (Eds.), *A Handbook of the Phonetic Sciences.* Oxford: Blackwells.

Ladefoged, P. and Maddieson, I. (1996) *The Sounds of the World's Languages.* Oxford: Blackwells.

Nix, D. A., Papçun, G., Hogden, J., & Zlokarnik, I. (1996). Two cross-linguistic factors underlying tongue shapes for vowels. *Journal of the Acoustical Society of America,* 99(6), 3707-3717.

Rate effects on French intonation: Prosodic organization and phonetic realization

Cécile Fougeron and Sun-Ah Jun

ABSTRACT

This study shows that an acceleration of rate modifies French intonation in both its phonetic realization of the f_0 contour and the prosodic organization of a text. We found that the effect of rate is not constant throughout the text but varies depending on position of the speech material within the text. In the first part of the text, the acceleration of rate induced modifications both in the shape of f_0 contour and the prosodic organization, while in the second part of the text, modifications are found only in the prosodic organization. Furthermore, interspeaker differences are found in strategies to increase speech rate. Two of three speakers reduced their pitch range and pitch displacements by lowering their f_0 maxima more than their minima. They also reduced the number of phrases and often did not realize an initial high tone. Conversely, the third speaker did not change her pitch range; her pitch displacements were reduced by raising the f_0 minima while keeping the maxima constant, and she did not modify the prosodic structure of the text. These results are compared to rate-based variation in the kinematics of other articulators. A model of the articulation of intonation is proposed assuming that the articulator of intonation and other articulators are governed by the same control mechanism.

1. INTRODUCTION

The effect of speaking rate has mostly been observed on the segmental level, with modifications in the temporal and spatial characteristics of speech. In general, three kinematic variables have been observed to vary in order to increase speaking rate. Speakers may reduce the spatial magnitude of articulatory movements, resulting in target undershoot (e.g. Lindblom, 1963, 1964; Kent & Moll, 1972; Gay, 1981). They may adjust the speed of transitions between successive targets by increasing the velocity of their movements (e.g. Abbs, 1973; Kuehn and Moll, 1976). They may increase the overlap between successive articulatory gestures by modifying their phasing (e.g. Engstrand, 1988; Munhall & Löfqvist, 1992; Krakow, 1993). These three variables are not mutually exclusive and can interact with each other in order to shorten articulation time. In fact several studies have shown that speakers vary in how these three parameters are employed as speaking rate increases (e.g. Kuehn, 1973; Kuehn & Moll, 1976; Ostry & Munhall, 1985).

The effect of speaking rate in the suprasegmental domain has also been studied, but mostly focusing on variation in duration (e. g. for French, Bartkova, 1991; Keller & Zellner, 1995). However, the effect of rate on intonation has rarely been studied. Recently, Caspers and van Heuven examined in detail the effects of rate on pitch movements in Dutch (Caspers & van Heuven, 1991, 1993, 1995; Caspers, 1994). They looked at variation in the number and type of pitch events induced by an increase in speech rate. They found that the tonal configurations used were similar at normal and fast rate. They also found that the number of melodic boundaries was reduced at fast speech rate while the number of pitch accents remained constant. Reducing the number of boundaries means reducing the number of phrases. This reduction in the number of phrases at fast rate has also been discussed for French by Vaissière (1983) and for Korean by Jun (1993). In Jun's study of Korean, as the rate increased, the number of Accentual Phrases (a phrase smaller than an Intonational Phrase) decreased by approximately 24%, averaged across five speakers. Interestingly, it seems that not all phrase boundaries are equally affected at fast

rate: Caspers and van Heuven found in Dutch that Intonational Phrase boundaries were more likely to be deleted when the boundary was optional.

Variation in the phonetic realization of the fundamental frequency (f_0) contour has also been observed at fast rate. Kohler (1983) found that in German there was a general raising of the phonatory level, both in frequency and intensity, at fast rate. Also, the level of the f_0 peak (the highest point in a sentence) was maintained whereas the level of the f_0 valleys were raised at fast rate. In consequence, the pitch displacements between peaks and valleys were reduced. A different pattern was found in Dutch by Caspers and van Heuven: both the level of peaks and valleys was raised at fast rate. As for pitch displacements, they observed variation between speakers as they raised their peaks and valleys in different proportions. One speaker raised his peaks more than his valleys, resulting in an increase in pitch displacements, while the other speaker raised his valleys more than his peaks, resulting in a decrease in pitch displacements.

In sum, it seems consistent across studies that an acceleration of rate leads to the reduction of the number of phrases. However, the effect of rate on pitch displacements and on the level of peaks and valleys seems to vary.

In this study, we investigated the effect of an acceleration of speech rate on the phonetic realization as well as the phonological structure of French intonation. By *phonetic realization*, we mean the shape of the pitch contour: the level of peaks and valleys, pitch displacements, and pitch range (highest minus lowest f_0 value). By *phonological structure*, we mean the prosodic organization of an utterance which is cued by its prosodic phrasing, the type of tones used, and its tonal pattern. Both phonetic and phonological representations were examined because the intonation of an utterance is variation in fundamental frequency and, at the same time, is the reflection of an abstract prosodic organization.

In addition, we examined the effect of rate on the intonation pattern of a text instead of isolated sentences as used by Caspers and van Heuven or Kohler. We expect that because the discourse structure is richer in a text than in isolated sentences, the intonation pattern of a text will provide phrases of various sizes with a larger inventory of boundary tone types as well as more variation in pitch range than those of an isolated sentence (Hirschberg & Pierrehumbert, 1986). The effect of rate acceleration on intonation may not be constant throughout the text, but may vary depending on the discourse structure and the location within the text. Therefore, we divided the text into two parts according to the discourse structure, and compared the modifications made in these two parts. Since local variation in articulation rate has been observed within long stretches of speech (Brubaker, 1972; Miller, Grosjean & Lomanto, 1984), it is possible that rate effect on f_0 will not be constant throughout the text. Moreover, we hypothesize that a speaker may use a different strategy towards the end of the text where f_0 declination induces a lowering of the f_0 top-line ("plateau") and a reduction of pitch range (Vaissière 1983). We also hypothesize that important boundaries in the discourse structure will vary less than other boundaries, as found in Caspers and van Heuven (1995).

Next, since speakers vary in their strategies for increasing speech rate, we tested for speaker variations both in terms of phonetic realization and phonological organization of intonation by comparing the speech of three French speakers.

Finally, to understand the global mechanism involved in speech production at fast rate, we compared the adjustments made in the intonation domain with that in the segmental domain. Therefore, we tried to use a terminology similar to that used in articulatory studies. We interpret f_0 contour in terms of *pitch target* (level of f_0 peak and valley), *pitch displacements* (pitch rise and fall), and *pitch movement velocity* (slope of pitch rise and fall). In this approach, we assume

that a pitch contour consists of a sequence of pitch targets and each target has an articulatory goal, laryngeal and/or subglottal. That is, we hypothesize that there is an active articulation of intonation, and its adjustments to increase rate would be governed by the same control mechanism as for other articulations.

2. METHOD

2.1. Subjects and speech material

Three Parisian French speakers (two female and one male) in their twenties participated in the experiment. Subjects were asked to read the text "La bise et le soleil" ("The North Wind and the Sun"), given in Table I, at self-selected normal and fast rates. Speakers were told to read the story in a lively manner, but no special instruction was given regarding the phrasing of the text. The recordings were made in a sound booth, and the speakers were asked to read the text three times at normal rate followed by three times at fast rate. Each time a speaker misread some part of the text, she or he had to begin it again from the beginning. No speech "error" or hesitation breaks were therefore included in the recordings.

The text was chosen for its narrative aspect (a tale) as well as for its syntactically complex construction leading to a large variety of phrases. The full text was analyzed for one of the female speakers (1F, the first author). For the other two speakers (2F, 3M) the comparison was limited to the first half of the text, where most variation between fast and normal rate was found for Speaker 1F.

Table I: *Text "La bise et le soleil". The numbers in parentheses correspond to the boundary codes used in Figures 5 and 6. The break between the two parts and the codes in parentheses were not written in the version read by the subjects. Word by word translation into English is given below the French text.*

La Bise et le Soleil First part: La bise (1) et le soleil (2) se disputaient (3), chacun (4) assurant (5) qu'il était (6) le plus fort (7). Quand ils ont vu (8) un voyageur (9) qui s'avançait (10), enveloppé (11) dans son manteau (12), ils sont tombés (13) d'accord (14) que celui (15) qui arriverait (16) le premier (17) à le lui faire (18) ôter (19) serait (20) regardé (21) comme le plus fort (22). Second part: Alors (23), la bise (24) s'est mise à souffler (25) de toutes ses forces (26), mais plus elle soufflait (27), plus le voyageur (28) serrait son manteau (29) autour de lui (30). Finalement (31), elle renonça (32) à le lui faire ôter (33). Alors (34), le soleil (35) commença à briller (36) et au bout d'un moment (37) le voyageur, réchauffé (38), ôta son manteau (39). Ainsi (40), la bise (41) dut reconnaître (42) que le soleil (43) était le plus fort (44).

The North Wind and the Sun First part: The Wind (1) and the Sun (2) were arguing (3), each (4) claiming (5) that he was (6) the stronger (7). When they saw (8) a traveler (9) who appears (10), wrapped (11) in his coat (12), they made (13) an agreement (14) whoever (15) should succeed (16) first (17) in making him (18) take it off (19) would be (20) considered (21) the stronger (22). Second part: Then (23), the North Wind (24) began to blow (25) with all his might (26), but the more it blew (27), the more the traveler (28) wrapped his coat (29) around himself (30). Finally (31), it gave up (32) in making him take it off (33). Then (34), the Sun (35) began to shine (36) and after a moment (37) the traveler, all warmed up (38), took off his coat (39). So (40), the North Wind (41) had to acknowledge (42) that the Sun (43) was the stronger (44).

2.2. Rate characteristics

In order to assess the relative speaking rate within and across speakers, acoustic durations were measured for each syllable in the text. Table II illustrates the differences found between the two rates for the three speakers in terms of total duration of the text, number of pauses, total duration of the pauses, articulation rate (number of syllables per second excluding pause), speaking rate (including pause), and average syllable durations and their coefficient of variation. Although the actual articulation rate in the normal condition is comparable across speakers (5.2~5.9 syll/s), the acceleration of rate from the normal to fast conditions varies across speakers (1.3~2.5 syll/s). The articulation rate at normal rate observed in this study is comparable to that found in previous studies on conversational speech in French: 5.73 syllables/s in Malécot, Johnson & Kizziar (1972) and 5.29 syllables/s in Grosjean & Deschamps (1975).

Table II: *Rate characteristics per speakers. (n = normal rate, f = fast rate)*

	spk 1F				spk 2F		spk 3M	
	1st part		2nd part		1st part		1st part	
	n	f	n	f	n	f	n	f
total text (s)	16	10.8	21.3	14.7	15.4	11.1	16.5	9.3
# pauses	6	4	8	6	7	6	9	4
total pause (s)	3.8	1.1	4.3	1.8	3.5	1.3	4	3.3
articulation rate (syll/s)	5.6	7.1	5.2	6.9	5.9	7.2	5.5	8
acceleration (syll/s)	1.5		1.7		1.3		2.5	
speaking rate (syll/s)	4.3	6.4	4.1	6	4.6	6.3	4.1	7.4
syll. (ms)	177	140	192	144	168	139	182	125
(variation coefficient, %)	(0.4)	(0.4)	(0.4)	(0.3)	(0.4)	(0.3)	(0.4)	(0.4)

2.3. Measurements

In this study, the terms 'f_0' and 'pitch' are used interchangeably. Pitch tracks of the readings were extracted and analyzed using Entropic Research Laboratory's XWAVES+ speech analysis software. Several acoustic measurements were taken for each tonal pattern, and these measures are schematically illustrated in Figure 1. For each f_0 peak and valley, the pitch level at the peak (hereafter called *f_0 maximum*) and at f_0 rising onset and falling offset (both points hereafter called *f_0 minima*) were collected. The difference in Hz between the minima and the maxima gives the magnitude of the pitch movements, which we will call pitch *displacements* in frequency (Hz) between these pitch targets (equivalent to the "excursion size" of the rise and fall in Caspers's (1994) measurements). The *velocity* (rate of change) of the pitch movement was calculated by dividing the displacement (in Hz) by the *transition* time between minimum and maximum (in ms). For each speaker, *pitch range* was also calculated by taking the difference between the highest- and the lowest-f_0-value in the text. Attention must be directed to the fact that in our terms *pitch range* is not the *averaged* range of f_0 displacement (as in Caspers, 1994), but the *maximal* range of f_0 variations for a particular speaker in the data considered. All measurements were compared at fast and normal rate (averaged across three repetitions) in order to evaluate the modification in the shape of f_0 contour at fast rate.

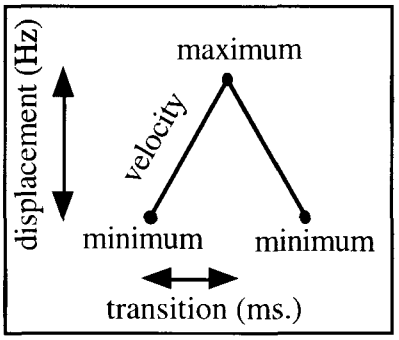


Figure 1: Schematic drawing of measurements taken from a f_0 contour.

To examine modifications in the prosodic organization, the prosodic structure of the text was qualitatively compared at fast and normal rate. The model of French intonation developed in Jun & Fougeron (1995) was used for the labeling of each pitch event. Following Pierrehumbert and others (Liberman, 1975; Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986), we assumed in this model that a tune is composed of a sequence of underlying tones and the intonational structure of a sentence is hierarchically organized. Two prosodic levels, and therefore two types of boundaries, are defined for French. The lowest tonally defined prosodic level is the *Accentual Phrase* (AP), which has an underlying tonal representation /L Hi L H*/. The Hi tone is realized at the initial stressed syllable ("accent secondaire") and the final H* tone is realized at the phrase final full syllable ("accent primaire"). Each L tone is realized at the syllable preceding the H-toned syllable. However, the initial Hi tone is often not realized and APs often surface with a tonal pattern [LLH]. This AP level roughly corresponds to the 'prosodic word' in French (Vaissière, 1992) or Intonation Group (Mertens, 1993), and is higher than the Tonal Unit of Hirst & Di Cristo (1984, in press). A prosodic level higher than an AP, and the highest level in this model, is the *Intonational Phrase* (IP). An IP has a big final lengthening with a boundary (%) tone and is optionally followed by a pause. Examination of the intonational patterns at fast rate in the present study leads us to propose an intermediate level between AP and IP, which we call the *Intermediate Phrase* (ip). This level is not tonally distinct from AP but differs from the AP by the degree of final lengthening and the height of the phrase final peak. Like AP, this level is not followed by a pause, but unlike AP, it has a medium final lengthening. Its final lengthening is smaller than that at the end of an IP, but the break is felt to be bigger than the one at the end of an AP. That is, 'ip' is considered as a small IP. Though this level is not phonologically distinct from the other two levels, we used this level for the purpose of describing the modification of intonation at fast rate.

Following this model, f_0 points relevant to the prosodic organization (initial H tones, H and L phrasal and boundary tones) were labeled by hand for each repetition of the text. The transcription was done separately by each author. When compared later, the transcriptions showed good agreement. Both the phrasing of the text and the tonal realization of the APs were compared between normal and fast rates. Figure 2 gives an example of the types of qualitative reduction found in the prosodic organization. The upper graph shows the waveform, prosodic labeling, f_0-track, and orthographic transcription of one sentence produce by speaker 3M at normal rate. The lower graph shows the same sentence produced by the same speaker at fast rate. AP boundaries are marked by a parenthesis ( ), and IP boundary tones are marked with %. The letters A, B, and C refer to specific examples of reduction. In (A), the IP-boundary at the

end of "... le lui faire ôter" is reduced into an ip-boundary at fast rate (medium high pitch excursion, medium lengthening, no pause). This kind of reduction will be notated as [IP => ip] (in Table IV later). In this case, the sentence that consists of two IPs at normal rate is produced as one IP at fast rate. Similarly, in (B) the two Accentual Phrases at normal rate, "qui arriverait" and "le premier", are grouped in one AP at fast rate. Here, the reduction is done by a deletion of the Accentual Phrase final boundary (notated [AP => Ø] in Table IV). (C) shows an example of a change in the realization of the underlying tonal pattern of an Accentual Phrase. The initial Hi tone in the Accentual Phrase "serait regardé" is not realized at fast rate (notated [Hi => (Hi)] in Table IV).

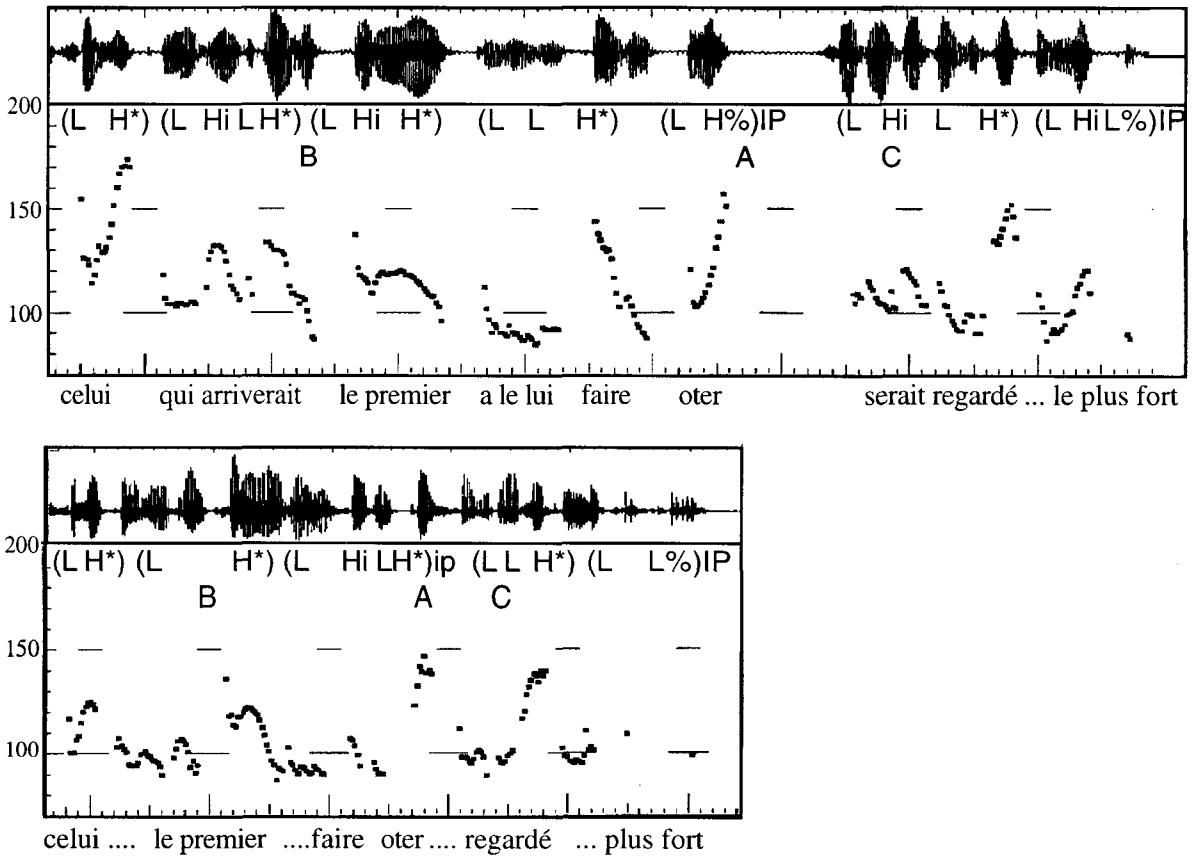


Figure 2: Example of "qualitative" reduction found in the prosodic organization. In the upper graph is given the waveform, prosodic labeling, f_0-track, and orthographic transcription of one sentence produced by speaker 3M at normal rate. The lower graph shows the same sentence produced by the same speaker at fast rate. The letters A, B, C refer to specific examples of reduction discussed in the text. For the prosodic labeling: L = low, Hi = initial High, H = pitch accent for the primary stressed syllable in Accentual Phrase (AP), ip = intermediate phrase, IP = intonational Phrase, ( ) = AP boundaries, % = IP boundary tone.*

3. RESULTS

3.1. Rate effect depending on the position in the text: first vs. second part.

In this section, we examine the effect of a rate increase on the production of the whole text for one subject (speaker 1F). The total duration of the text at normal rate was 37 seconds on average. In order to see whether the speaker applied a consistent strategy to increase speaking rate throughout the whole text, we compared two successive parts of the text at fast and normal rate. The text was divided into two parts at a semantic and narrative juncture. The first part consists of two long sentences and is the introduction of the tale, and the second part consists of 4 sentences and is the development and conclusion of the tale. As seen in Table II, the first part is shorter than the second part (16 vs. 21 s at normal rate). In general, at normal rate the first part is phrased in six Intonational Phrases, and the second part is phrased in thirteen Intonational Phrases, including five one-word IPs which are generally sentence adverbs or connectives. However, both parts generally include the same number of Accentual Phrases (22 APs) at normal rate.

3.1.1. Rate acceleration in the first vs. second part of the text:

Before comparing the change in pitch configuration induced by rate acceleration in the two parts, we examined any major differences in the rate of speech adopted in these two parts. As shown in Table II, the acceleration of 'articulation rate (syll/s)' is similar in both parts of the text, with an acceleration of 1.5 syll/s in the first part and 1.7 syll./s in the second part. The reduction of the number and duration of the pauses is also comparable in the two parts. Since a substantial variation in articulation rate can be averaged out when articulation rate is measured over large stretches of speech (Miller et al., 1984), we calculated the articulation rate for each "chunk" of continuous speech delimited by a pause (similar to "a run of pause-free speech" in Miller et al. (1984). This can allow us to test if the acceleration of rate is constant across smaller units of speech and if it varies depending on the position of the unit in the text. Since these chunks correspond to an Intonational Phrase at normal rate but not always at fast rate, we called them "C" to avoid confusion. The articulation rate of each C at normal and fast rate is shown in Figure 3.

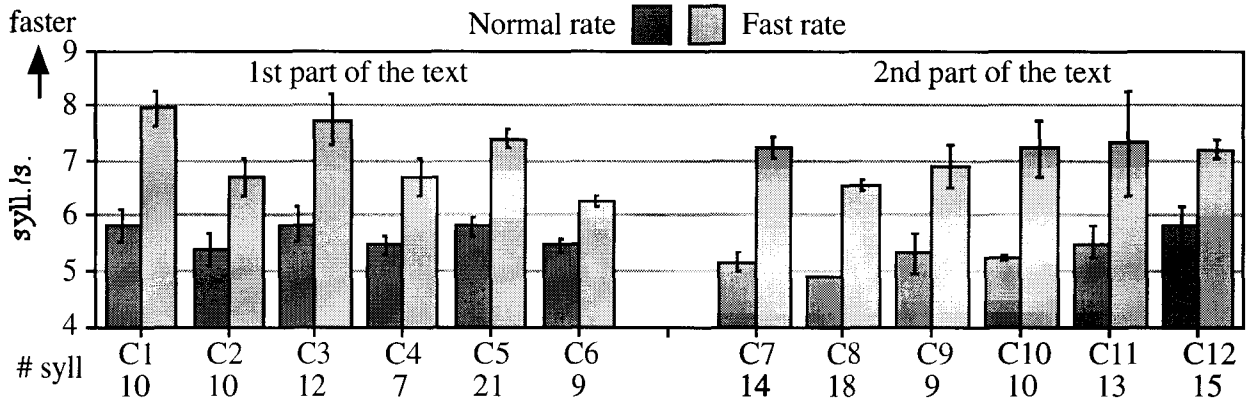


Figure 3: Articulation rate (syllables/s) for each "chunk" (C) of the whole text at normal (dark gray shaded bar) and fast rate (light gray shaded bar). (see text for the definition of C) Below each C, the number of syllables within each C is given. Speaker: 1F.

The number of syllables per unit C is given underneath each column, and the variation in articulation rate within a C and across the three repetitions is given by the standard deviation at the top of each bar. It is shown that this variation in articulation rate is either similar at both rates or greater at fast rate. Furthermore, when we compared the articulation rate between the successive Cs at both normal and fast rate, no systematic trend was found depending on the position within the text. These results do not support Brubaker (1972)'s observation of a progressive increase of rate from the first to the sixth and the last sentence in a paragraph. On the contrary, our results corroborate Miller et al. (1984)'s observation that during a given speech interval (in their case, the response in an interview), articulation rate does not increase or decrease gradually, but changes course a number of times (see their fig. 1). In our data, the acceleration of rate is comparable throughout the whole text.

3.1.2. Modification in the shape of f_0 contour:

Table III shows the reduction in the shape of f_0 contour at fast speech rate for all three speakers. Here we discuss only the data from speaker 1F; other two speakers' data will be discussed in Section 3.2. The differences in pitch range between fast and normal rate are presented in Figure 4 for the first and the second part of the text (see also the last three rows of Table III). In the first part, pitch range was considerably reduced at fast rate (32%) because the highest f_0 value at normal rate (360 Hz) was lowered at fast rate (300 Hz), and the lowest f_0 value was slightly raised from 140 Hz at normal rate to 150 Hz at fast rate. In contrast, there was very little reduction of pitch range in the second part of the text (6%), where both the highest and lowest f_0 values were only slightly lowered at fast rate.

Table III: *Reduction in the shape of f_0 contour at fast speech rate. Mean values in Hz at normal (n) and fast (f) rates. (n-f)/n refers to the percentage (%) of reduction at fast rate compared to normal rate. Velocity in Hz/ms, other measures in Hz.*

	spk 1F, first part			spk 1F, 2nd part			spk 2F			spk 3M		
	n	f	n-f/n	n	f	n-f/n	n	f	n-f/n	n	f	n-f/n
f_0 maxima	262	226	14%	223	212	5%	286	280	2%	175	145	17%
f_0 minima	192	179	7%	176	168	4%	209	218	-4%	108	104	4%
rising displ.	68	49	27%	48	44	8%	77	62	19%	70	42	40%
falling displ.	65	51	22%	47	45	5%	72	55	23%	70	45	36%
rising velocity	.36	.3	16%	.32	.29	10%	.47	.45	5%	.32	.25	22%
falling velocity	.38	.33	15%	.3	.26	11%	.4	.36	11%	.36	.29	19%
pitch range	213	145	32%	124	116	6%	204	205	0%	139	101	27%
highest f_0-peak	357	297		264	249		350	355		225	188	
lowest f_0-valley	144	152		140	133		146	150		86	87	

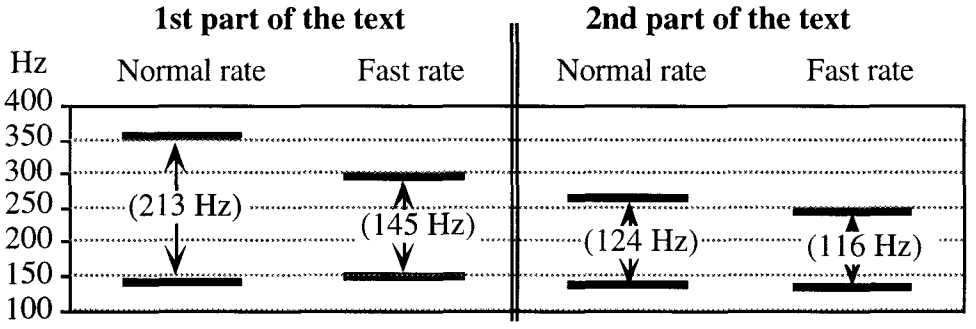


Figure 4: f_0 range (highest-lowest f_0 value) at normal and fast rate in the first and second part of the text for speaker 1F.

Regarding the level of the pitch targets, an acceleration of rate induced an overall lowering of both the f_0 minima and the maxima. Figure 5 illustrates the difference in pitch level for the f_0 maxima in the two parts of the text at normal (black bar) and fast rate (white bar). For each peak, the number given on the abscissa corresponds to the boundary-code written in parentheses after each prosodic phrase in the text given in Table I. In the first part of the text (upper graph), it appears that f_0 maxima are in general lower at fast rate than at normal rate, and this lowering is statistically significant ($t = 5.48$; $p < .0001$; $df = 47$).

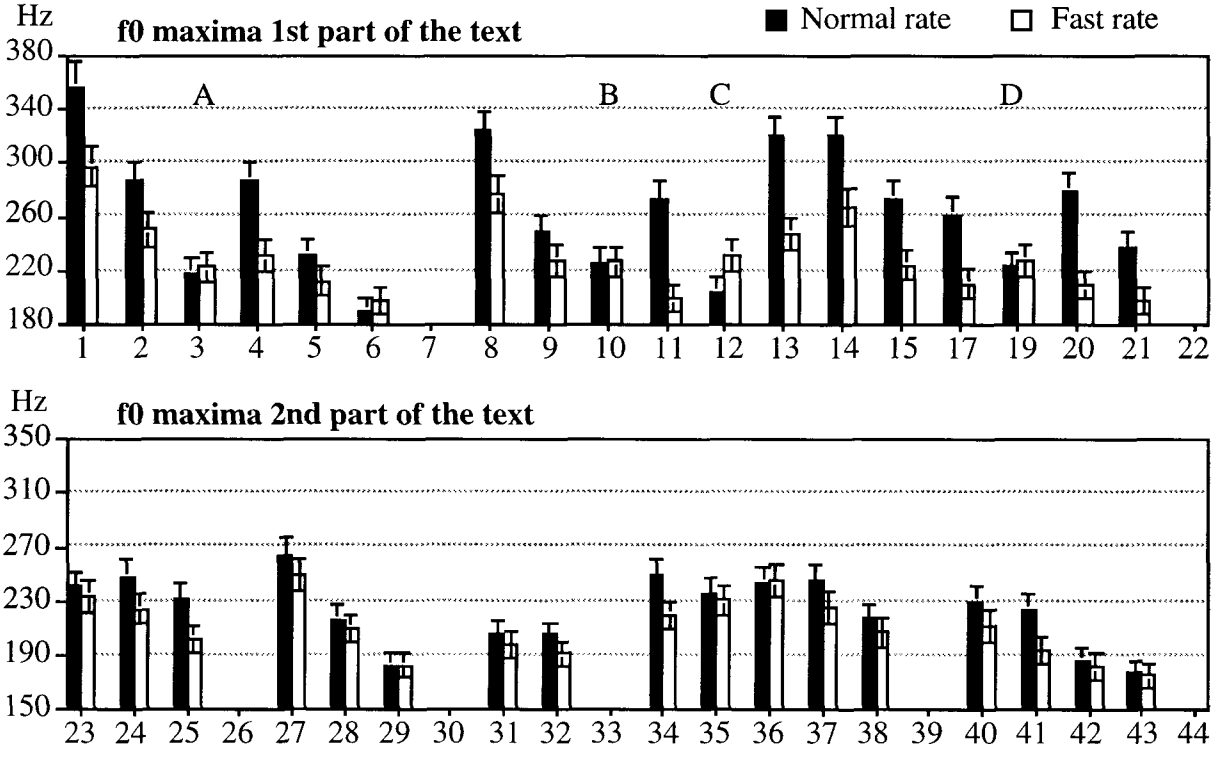


Figure 5: f_0 maxima in the first (upper graph) and second part (lower graph) of the text for speaker 1F. For each maximum, the number on the abscissa corresponds to the boundary code given in Table I. For each number, f_0 maxima are given at normal (black bar) and fast rate (white bar). Letters A, B, C, and D indicate the major prosodic boundaries that do not reduce at fast rate.

However, at certain boundaries, written in the figure as A to D, the height of the peaks at normal rate is either the same or higher at fast rate. These boundary points correspond to the major Intonational Phrase boundaries with a H boundary tone (H%) in the text. These are "... se disputaient" (A), "... qui s'avançait" (B), "... son manteau" (C), and "... faire ôter" (D). These major boundaries were not reduced at fast rate. In the second part of the text (lower graph), the reduction of the maxima is on average much smaller (5%) than that observed in the first part (14%). This lowering of the maxima in the second part is significant ($t = 4.14$; $p < .0001$; $df = 48$) though some of the boundaries (6 out of 17) are not lowered at fast rate. However, unlike the observation made for the first part, these boundaries did not correspond to major Intonational Phrase boundaries. The behavior of the minima, not illustrated here but given in Table III, follows the same tendency as that of the maxima. In both parts, the minima were lowered at fast rate, but the lowering was significant only in the first part of the text ($t = 3.43$; $p < .001$; $df = 50$), not in the second part of the text ($t = 1.84$, $p = 0.07$ (ns); $df = 60$).

Comparison of the effect of rate on the two kinds of f_0-targets (minima and maxima) shows that, in the first part of the text, f_0 maxima were lowered much more at fast rate (14%) than were f_0 minima (7%). As a consequence, there was a noticeable reduction of pitch displacements in the first part of the text (22-27%). In the second part, however, the lowering was similar for both the maxima (5%) and the minima (4%), resulting in a very small reduction of displacements (5-8%). Here f_0 minima values were calculated from either f_0-rising onset or f_0-falling offset. But when we examine f_0-rising displacements and f_0-falling displacements separately, the rising displacements are always reduced more than the falling displacements in the whole text.

For velocity, there seems to be a reduction at fast rate relative to normal rate when we look at the percent of reduction in Table III (15 ~ 16% for the first part and 10 ~ 11% for the second part). As for other measurements, velocity reduced more in the first part than in the second part, although the difference is very small. However, this change in velocity is hard to interpret since the difference in raw values between normal and fast rate is very small (less than 0.1 Hz per ms.). What is important here is that the velocity never increased at fast rate. That is, it is never the case that f_0 movement is faster at fast rate compared to normal rate.

3.1.3. Modification in the prosodic organization of the text:

Regarding the prosodic organization of the text, speaker 1F changed both the phrasing of the text and the realization of the underlying tonal pattern at fast rate. The results obtained in the two parts of the text are given in the first two columns of Table IV. Descriptive examples of the three types of organizational modification observed were given in Figure 2 and explained in Section 2.3. In the first and second part of the text, an acceleration of rate induced a reduction in the number of prosodic phrases. Almost half of the Intonational Phrase boundaries were reduced to lower level boundaries, ip or AP (50% in the first part, 42% in the second part), and 22% of the Accentual Phrase boundaries were deleted in both the first and the second part. Moreover, the underlying initial high tone (Hi) of the Accentual Phrase was very often not realized at fast rate (63% and 73% for the first and the second part, respectively).

Table IV: *Modification in the prosodic organization at fast speech rate. Frequency of occurrence per type of prosodic reduction and per speakers. (IP = Intonational Phrase, ip = intermediate phrase, AP = Accentual Phrase, Hi = initial high tone in an AP)*

	spk 1F		spk 2F	spk 3M
Phrasing	1st part	2nd part	1st part	1st part
IP => ip or AP	50 %	42 %	0 %	55 %
AP => Ø	22 %	22 %	0 %	28 %
Tonal realization				
Hi => (Hi) (/LHLH/ => [LLH])	63 %	73 %	4 %	81 %

In sum, depending on the position in the text, different patterns were adopted by this speaker under increased rate. In the first part of the text, an acceleration of rate induced some reduction in the shape of f_0 contour; mainly a lowering of both f_0 maxima and minima, a reduction in the displacement of the pitch movements, and a reduction of pitch range. The speaker's adaptation to the time constraint also resulted in a restructuring of the prosodic phrasing of the text with fewer prosodic groups and a non-realization of certain underlying tones. In the second part of the text, however, we observed a similar change in the prosodic organization but very little change in the phonetic realization of the f_0 contour.

3.2. Speaker variation.

Next we looked at inter-speaker differences in the way they modify their intonation at fast rate by considering two additional speakers. As shown in the previous section, more differences in f_0 contour between fast and normal rate were found in the first part of the text for speaker 1F. Therefore, we examined inter-speaker variation only for this first part of the text. As shown by the rate characteristics for each speaker given in Table II, the articulation rates of the three speakers are similar in the normal condition (5.5 to 5.9 syll./s). But the self-selected "fast rate" is different for the three speakers. The male subject (speaker 3M) has the biggest rate acceleration (2.5 syll./s increase), and the second female speaker (speaker 2F) has the smallest rate acceleration (1.3 syll./s increase). In this section we will present the modifications in both the shape of f_0 contour and the prosodic organization for these speakers.

3.2.1. Modification in the shape of f_0 contour:

The shape of f_0 contour for the male speaker (3M) showed a pattern of reduction very similar to the pattern presented above for speaker 1F in the first part of the text (see Table III). His pitch range was reduced by 27% at fast rate with a lowered highest-f_0-value (225 Hz to 188 Hz) and a stable lowest-f_0-value (86 to 87 Hz). When all pitch maxima and minima were compared, we found that this speaker also lowered significantly both the maxima ($t = 10.58$; $p < .0001$; $df = 44$) and the minima ($t = 2.29$; $p = .02$; $df = 48$). Like speaker 1F, this speaker achieved a reduction in f_0 displacements by lowering the maxima (17%) more than the minima (4%). This reduction in pitch displacements was quite considerable for both rising (40%) and falling (36%) movements.

In contrast, the pattern shown by speaker 2F was totally different. She modified neither the height of her highest-f_0-value (350-355 Hz), nor her lowest-f_0-value (146-150 Hz) at fast rate. Therefore, no reduction of pitch range was observed for this speaker. She also showed a

different pattern in the reduction of f_0 maxima and minima. Figure 6 illustrates the differences in pitch level for f_0 maxima (upper graph) and f_0 minima (lower graph) at fast (white bars) versus normal rate (black bars) for this speaker. f_0 maxima often have either the same height at both rates or are inconsistently raised or lowered at fast rate, and the difference was not significant ($t = 1.04$; $p = 0.3$; $df = 51$). On the contrary, almost all f_0 minima were slightly raised at fast rate compared to at normal rate. This raising of f_0 minima at fast rate was small (4%) but significant ($t = 4.57$; $p < .0001$; $df = 60$). As a consequence, this speaker also reduced her overall pitch displacements at fast rate, but with a totally different strategy: she raised her f_0 minima whereas the other two speakers lowered their f_0 maxima.

None of the speakers increased the velocity of pitch movements at fast rate. However, following the difference in displacement pattern mentioned earlier, we found that speakers 1F and 3M reduced the velocity of the rising movement more than that of the falling movement while, speaker 2F reduced more the velocity of the falling movement.

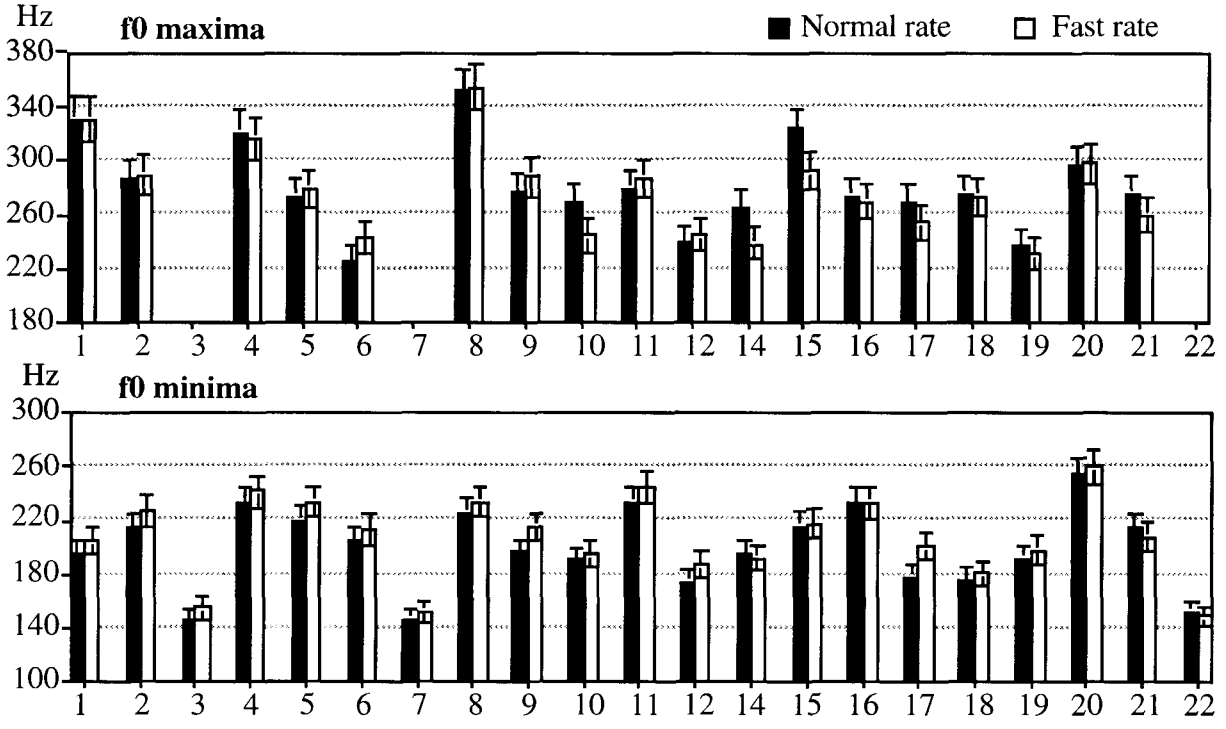


Figure 6: f_0 maxima (upper graph) and minima (lower graph) in the first part of the text for speaker 2F. The numbers on the abscissa correspond to the boundary codes given in Table I. For each number, f_0 values for normal and fast rate are given in the black and white bar, respectively.

3.2.2. Modification in the prosodic organization of the text:

Table IV (columns 1, 3, and 4) shows the modifications in the prosodic structure by means of intonational phrasing and the tonal pattern change for three speakers. Speaker 3M again showed a similar pattern to that of speaker 1F in the prosodic organization (see Section 3.1.3). When the rate is faster, speaker 3M reduced the number of phrases by reducing the strength of the Intonational Phrase boundaries into lower boundaries (55%) and by deleting some Accentual

Phrases boundaries (28%). Also, most of the initial high tones (Hi) in the Accentual Phrase were not realized for this subject (81%).

Speaker 2F differed from the other two speakers in that she did not modify her prosodic organization of the text. At fast rate, the phrasing and the realization of the underlying tonal pattern were exactly the same as in the normal rate. That is, the number of Accentual Phrase and Intonational Phrases was preserved at fast rate and most of the Accentual Phrase initial high tones observed in normal rate were also realized in fast rate.

In conclusion, our three speakers used different strategies in their production of intonation at fast speech rate. For three speakers, we found two patterns: two of the speakers changed both their f_0-contour and prosodic structure, whereas the other speaker changed only the f_0-contour, keeping the prosodic structure of the text the same . The modification of f_0-contour was also different between the two patterns: two speakers reduced the displacements of their pitch movements by lowering their f_0 maxima, whereas the other speaker obtained a reduction in pitch displacements by raising her f_0 minima, keeping the maxima constant.

4. DISCUSSION

4.1. Articulation of intonation at fast rate.

In this section we will review results found in both segmental and suprasegmental studies and compare them with our observations; we will try to explain the variation in the strategies used for increasing the rate in different parts of the text and between different speakers. Then, we will sketch a model of the articulation of intonation following the Task Dynamic framework (Saltzman & Munhall, 1989; Browman & Goldstein, 1990; Saltzman, 1995).

An acceleration of rate means a reduction in the time of articulation. As reviewed in the literature on 'segmental' articulations, a shortening of articulation-time can be achieved in three ways: successive gestures can be realized in a shorter time if the magnitude of their movement is reduced, and/or if the velocity of their movements is increased, and/or if the overlapping between gestures is increased. In this paper we have looked at modifications in intonation at fast rate assuming that a f_0-contour is the output of a laryngeal articulation. Here the articulatory gesture is not the spatial displacement of one articulator like the tongue, toward another, but it is a variation in the frequency of vibration of the vocal folds. We assume that the strategy used for increasing speech rate is the same for all articulations, laryngeal and supralaryngeal. As a consequence, we present the modifications observed in the shape of f_0-contour using a terminology similar to that used for describing other articulations: movement displacements correspond to the variation between high and low frequencies of vibration; velocity is the speed of the change between high and low frequencies; overlap is the phasing of the gestural activation for high and low frequencies of vibration. By adopting Jun & Fougeron's (1995) intonation model, we interpreted a tonal pattern as a succession of high and low pitch targets. For convenience we will call these pitch movements toward high and low targets, H-gesture and L-gesture, respectively.

4.1.1. Variation in displacements and/or velocity of pitch movements at fast rate.

Segmental studies have shown that velocities and displacements of articulatory movements can be affected by rate increase, and this result varies across studies, articulators and speakers. For example, Kuehn & Moll (1976) found that one out of five subjects always increased velocity, two usually increased velocity, and two reduced displacements with a concurrent decrease in

velocity. Regarding the effect of rate on intonation, an acceleration of rate has generally been found to be correlated with a *reduction in pitch displacements.* This has also been found in German (Kohler, 1983) and Dutch (Caspers, 1994). However, one of Caspers's subjects increased pitch displacements at fast rate. Regarding the velocity of pitch movements, Kohler (1983) notes that the fall-rise glides are generally leveled at fast rate, which seems to mean that the slopes of f_0 are less sharp. Caspers (1994) found that for the accent-lending rise in Dutch, both speakers shortened and steepened the rising pitch movement at fast rate, whereas for the accent lending fall, the shape of the f_0-contour was preserved at fast rate. However, in our study, it was never the case that the slope of f_0 movements were sharper at fast rate compared to normal rate. Therefore, it seems that the acceleration in the transitions between successive pitch targets at fast rate was achieved by a reduction in the displacements of f_0 movements, but not by an increase in the velocity of f_0 slopes. On the contrary, we have observed a *reduction in velocity* that is small in absolute values (always less than 0.1Hz/ms) but that is still important relative to the velocity of the slope at normal rate (5 to 22% increase depending on the speaker). This reduction in velocity may be the result of the reduction in displacements (as found by Kuehn & Moll, 1976). In sum, in our data an acceleration of rate is achieved by reducing pitch displacements rather than by increasing velocity of pitch movement.

4.1.2. Pitch target undershoot at fast rate: linear rescaling, overlap or deletion of pitch gestures.

A reduction of displacements can cause *target undershoot* (Lindblom, 1963, 1964). In Lindblom's terms, target undershoot is directly proportional to the duration of the lapse between successive motor commands, allowing or not allowing the articulators to reach their target before the next set of commands arrives. Laryngeal articulation can respond to the same time constraint by undershooting successive high and low pitch targets. Ohala and Ewan (1973) reported that, in singing, raising of f_0 takes longer than lowering of f_0. Following Lindblom's reasoning, if a command toward a minimum of f_0 arrives to the muscles before the completion of the preceding f_0-rising movement, an undershoot of f_0 maxima is likely to occur. And, if the rising movement is intrinsically longer than the falling movement, this rising movement has more chance to be cut off before the target is reached. This hypothesis of a duration-dependent undershoot of high target is supported by the speech behavior of speakers 1F and 3M. These speakers undershoot f_0 maxima by reducing their f_0-rising-displacements more than their f_0-lowering-displacements, although the difference is small (27 vs. 22 % for speaker 1F and 40 vs. 36 % for speaker 3M). Speaker 2F, who does not modify the height of f_0 maxima, does not show this tendency — the reduction in falling displacements (23%) is even slightly bigger than that of rising displacements (19%). However, this hypothesis does not hold if we consider that speaker 2F undershoots her f_0 minima rather than the maxima. For her, f_0 minima are higher at fast rate while f_0 maxima are kept constant. Kohler (1983) also found a raising of f_0 minima at fast rate for his three German speakers. In his study, only four pitch events are taken into account: f_0 maximum in the first word in the sentence, which is also the highest peak in the sentence, and three more points in the sentence which are all valleys. At fast speech rate, all points except the peak are raised, resulting in a reduction of pitch displacements. In sum, it seems that when the rate is faster, either high or low pitch targets can be undershot, and which target is undershot seems to be speaker-dependent.

As we mentioned above, target-undershoot is related to a reduction of articulatory displacements. This reduction of displacements can be driven by different mechanisms: a linear

rescaling or an overlap of articulatory gestures. In a *linear rescaling* hypothesis, the size of the articulatory gestures is scaled in proportion to change in duration. As a consequence, these gestures keep the same overall shape (same peak velocity) but their duration and displacements are reduced. In an *overlap* hypothesis, the reduction of displacements is a function of the degree of overlap between successive (not-rescaled) gestures. This hypothesis can also be called a *truncation* hypothesis. In our data the two H- and L-gestures share the same articulator (the vocal folds) so their overlap results in a blending of these two gestures. Truncation is one kind of blending in which the competing demands on one articulator made by two opposing gestures cause one to be cut off by the other (Harrington, Fletcher, & Roberts, 1995). If the truncation of a gesture is extensive it can result in the undershoot of the articulatory target. Based on our measurements we cannot distinguish which of these hypotheses, linear rescaling or truncation, is at work in our data.

Another possible strategy to increase rate is to simply *delete* some of the gestures because there is no time to articulate them. Munhall & Löfqvist (1992) observed that the two laryngeal gestures corresponding to two voiceless consonants in English can be realized at fast rate with a single smooth laryngeal gesture. They underline that two possible explanations can account for that phenomenon: "[...] the fastest utterances that exhibit only a single smooth glottal movement could still be composed underlyingly of two separate laryngeal gestures. [...] On the other hand, at some degree of overlap a reorganization may occur and a single laryngeal movement may be 'programmed' [...]" (p.122). Although several pieces of evidence favor the first hypothesis, in which the two overlapped underlying gestures surface in a single gesture, the authors underlined that it is difficult to distinguish between these two hypotheses. In our study we have shown that the underlying initial high tone (Hi) of an Accentual Phrase is often not realized at fast rate (at least for two speakers). It is possible to extend the question raised in Munhall and Löfqvist (1992) to our Accentual Phrase tonal pattern at fast rate: (1) Is it still composed of four underlying gestures /L H L H/, or (2) is it the case that the Accentual Phrase pattern is reorganized into two gestures /LH/ ? If the first case, what we have called until now the "non-realization" of the initial high tone would mean a complete overlap of this H-gesture by the adjacent L-gestures. If the second case, there would be a reorganization of the tonal pattern of the Accentual Phrase so that, when time is short, an AP is planned with only two gestures, L and H. We are inclined to favor the overlap hypothesis because of the fact that the non-realization of the initial high tone in French is not limited to a fast rate condition. The initial high tone is also not realized when the number of syllables within an AP is fewer than four syllables (Jun & Fougeron 1995). When the number of syllables is three, a trace of the initial high tone is sometimes seen with a small rise. This is parallel to Munhall & Löfqvist's (1992) traces of glottal opening gestures of two voiceless consonants showing the change from the case where there are two separate peaks, to the case where there is one big peak with a small shoulder, to the case where there is a single peak only. If a reorganization hypothesis is favored as opposed to an overlap hypothesis, we need to suppose that this reorganization takes place each time the temporal domain of realization of the f_0-contour is shortened (either due to rate acceleration or due to a small number of syllables). In this case, the assumption of an "invariant" underlying Accentual Phrase pattern would not be needed any more. However, as we have seen in our data, some initial high tones remain the same at fast rate, showing that the initial high is present underlyingly in those cases, and suggesting that it is present more generally at fast rate.

4.1.3. Adding an intonation tier to the task dynamics model of speech production.

The increasing body of data on the influence of prosody on articulatory gestures has suggested the need of introducing a prosodic module in a model of speech production (for example, see Saltzman, 1995; Byrd, Narayanan, Kaun, & Saltzman, 1996). Our results push forward this idea, on the basis of the similarities between the strategies used to increase speech rate in the segmental and suprasegmental domains. We propose that a prosodic tier should be added to a Task Dynamic model of speech production (Saltzman & Munhall, 1989; Browman & Goldstein, 1990; Saltzman, 1995) to model the production of intonation. This tier would be superimposed over the existing tiers since the domain of realization of the "intonation" gestures spans a constellation of gestures. Here, we will sketch some of the basic ideas of this proposal, although it is still in a preliminary stage.

In Browman & Goldstein's (1990, 1992) Articulatory Phonology model, the Glottis is one of the articulators considered along with the Velum, the Tongue Body, Tongue Tip, etc. This articulator has one tract variable: the glottal aperture. We propose to replace the Glottis articulator by a Vocal Fold articulator. For this Vocal Fold articulator, we have two tract variables: one is the "glottal aperture degree" and is the same as that in the original model (with the three modes given in Browman & Goldstein, 1990, Appendix B, p.373); the other tract variable is the "vocal fold vibration frequency". For this tract variable, we propose two values: "high" and "low" (however, it is possible that a "mid" value is necessary for languages with a lexical mid tone). Thus, intonation would be modeled by a H-gesture (or "high frequency vibration-gesture") and a L-gesture (or "low frequency vibration-gesture"). Using this model, we can represent the articulation of intonation and its modification observed at fast versus normal rate as schematically shown in Figure 7. Here, we represent the H- and L- gestures as triangles showing the movement toward the H and L targets. But we do not consider these gestures to be rising- or falling- gestures because we do not believe that f_0 rise or fall is the underlying unit of intonation. In our framework, a tune is a succession of the underlying H and L tones, and the rising or falling movements are the interpolation between H and L. Thus, rising is only the movement toward the underlying H target and falling is only the movement toward the underlying L target.

Figure 7a represents the basic pattern of an Accentual Phrase at normal rate which consists of a succession of a L-gesture, H-gesture, L-gesture, H-gesture without overlap. Figure 7b represents the undershoot of f_0 maxima and reduction in pitch displacements at fast rate as observed for speaker 1F and 3M. In this representation, the second L-gesture is phased earlier relative to the onset of the first H-gesture. As a consequence, the overlapped H-gesture is truncated. Its rising displacement is reduced and the high target is undershot. In this example, it is the initial H-gesture that is truncated but this can happen to any H-gesture. Similarly, although it is not shown here, a L-gesture can be truncated by adjacent H-gestures. _Figure 7c represents the "non-realization" of initial high tone in an Accentual Phrase. This case is a more dramatic version of the overlap presented in 7b. The phasing of the second L-gesture is so early relative to the first H-gesture that this H-gesture is totally overlapped. As a result, the H-gesture is hidden by the L-gesture.

By assuming H- and L-gestures we can explain the output patterns and modifications common to laryngeal and supralaryngeal articulators. Further research is needed to develop this model in order to understand the mechanism involved in the adjustments of both the Vocal Fold articulator and other articulators in speech.

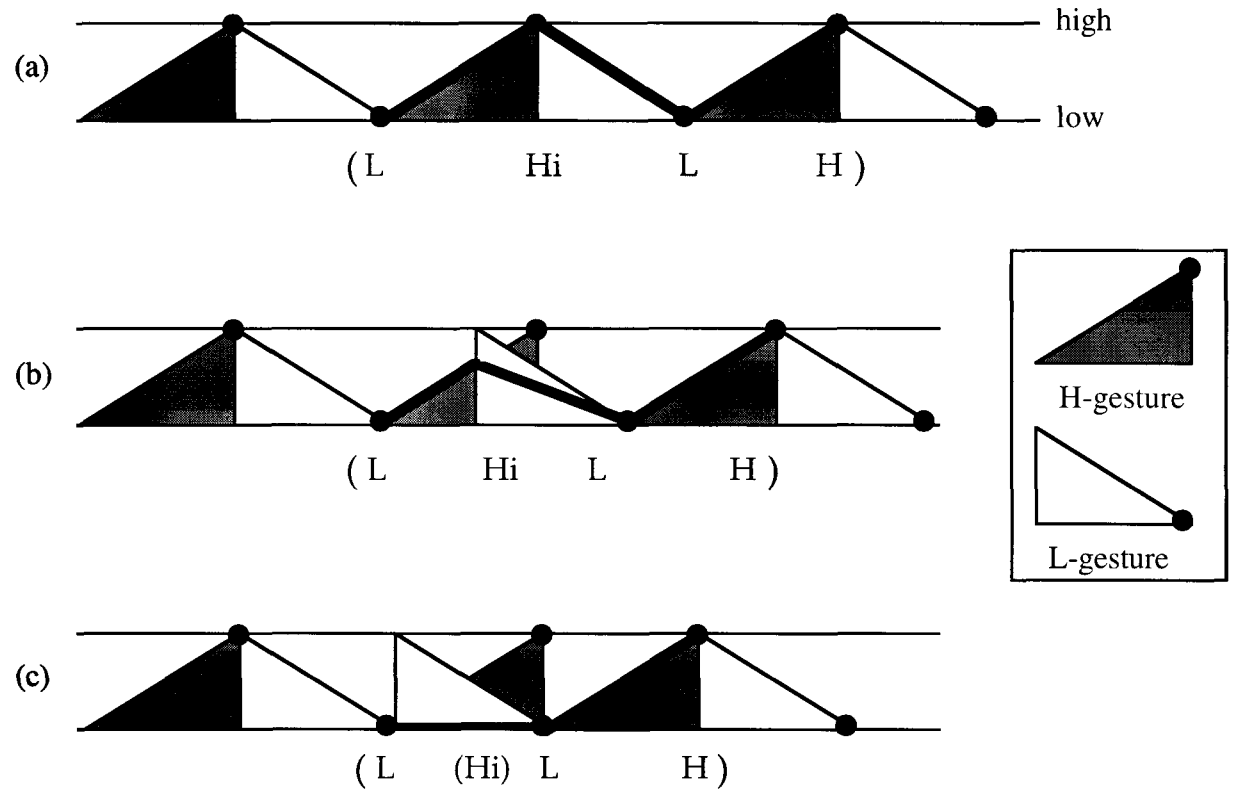


Figure 7: Schematic representations of the articulation of intonation model, with H-gesture (shaded triangle) and L-gesture (white triangle). Black dots indicate the pitch target for each gesture. The thick black line represents f_0 contour. In (a), four gestures (L, H, L, H) are successfully realized without overlap; in (b) H-gesture is partially overlapped by the following L-gesture; and in (c) H-gesture is completely overlapped with the following L-gesture, thus realized as [LLH]. Under each gestural representation is shown the tonal transcription of the Accentual Phrase (AP).

In sum, a consistent strategy to increase speech rate is a reduction in pitch displacements. However, this reduction is not systematically correlated with an undershoot of particular pitch targets. Rather, both high and low pitch targets can be undershot. More interestingly, target undershoot is not the only mechanism at work in rate acceleration. In our data, speaker 1F's and 3M's f_0 minima are on average lower at fast rate: this is an "overshoot" of low targets, rather than an "undershoot". Thus for these speakers there is a general lowering of the pitch targets and the reduction of pitch displacements is obtained only because f_0 maxima are lowered more than f_0 minima. Caspers (1994) found the opposite: a global raising of both f_0 maxima and minima at fast rate. Therefore, it seems that, at fast rate, H or L pitch targets can be undershot and the whole pitch level can be lowered or raised.

4.2. Variation in the compressibility of pitch movement at fast rate.

As observed in segmental studies, the strategies used to increase speech rate are variable. In our data, we have shown that the reduction in pitch range and pitch displacements varies depending on the speaker (2 patterns for 3 speakers) and depending on the position of the speech material in

the text (first vs. second part). Also, we found that the reduction of pitch targets varies depending on the linguistic strength of the boundary. In this section we will discuss the constraints on the compressibility of pitch movements to explain these differences.

First, consider the variation in the production of the whole text by one speaker. We divided the text at a major narrative break in the story. Comparison of the modification in the shape of f_0 contour at fast and normal rate showed that the speaker reduced both her pitch displacements and pitch range in the first part, but not in the second part. This difference can not be explained by the fact that the rate is faster in one or the other part of the text: modifications in articulation rate and the number and length of the pauses are similar in the two parts of the text (Table II). Comparison of the acceleration of rate within smaller units of speech (Figure 3) also showed that the rate acceleration was not a function of the position in the text. However, the difference in pitch range between the two parts can explain the different strategies for increasing rate in the two parts. What we called pitch range in this study is the difference in Hz between the highest f_0 value and lowest f_0 value in the speech material considered. In consequence, this measure shows the range of f_0 variation. We observed that the pitch range at normal rate is a function of the position in the text: the pitch range is wider in the first part (213 Hz) than in the second part (124 Hz). This reduction of pitch range is partially due to the f_0 declination from early to late in the speech stream: f_0 top-line (plateau) lowers gradually while f_0 base line is quite stable (see Vaissière 1983 for a review, but see Sluijter and Terken 1993 for a constant difference between top-line and baseline in a paragraph in Dutch). When the rate increased, the lack of reduction of pitch range and of individual pitch displacements in the second part of the text may be caused by the fact that the pitch range is narrower or already compressed. It is not the case that pitch displacements are not reduced because the pitch range in the second part is small per se. If it were, we should see a difference in the pattern produced by the male speaker who has a smaller pitch range than that of the two female speakers, but the male speaker rather reduced his pitch displacements and pitch range at fast rate in a similar amount (27%) to that of one female speaker (32%). Therefore, the resistance to reduction in the second part of the text seems to depend on the fact that pitch displacements and pitch range in the second part of the text are already compressed and can not be further reduced at fast rate.

Similarly, variation on the degree of compressibility can explain the pattern of reduction of f_0 maxima at fast rate. In Figure 8, the amount of reduction of f_0 maxima at fast rate is plotted against the height of these maxima at normal rate. For the two speakers showing a lowering of f_0 maxima at fast rate (speaker 1F and 3M), there is a linear relationship between the reduction of the pitch targets and their f_0 height at normal rate: the higher the maximum is, the more reduced it is at fast rate (r^2=0.28 for speaker 1F in both parts, r^2=0.56 for speaker 3M). In other words, lower pitch maxima are less reduced than higher pitch maxima. We suppose that a high-pitch-target must have a minimal height to be still considered as a high-target (in production and/or perception), and if this height is already low at normal rate, it cannot be further lowered at fast rate.

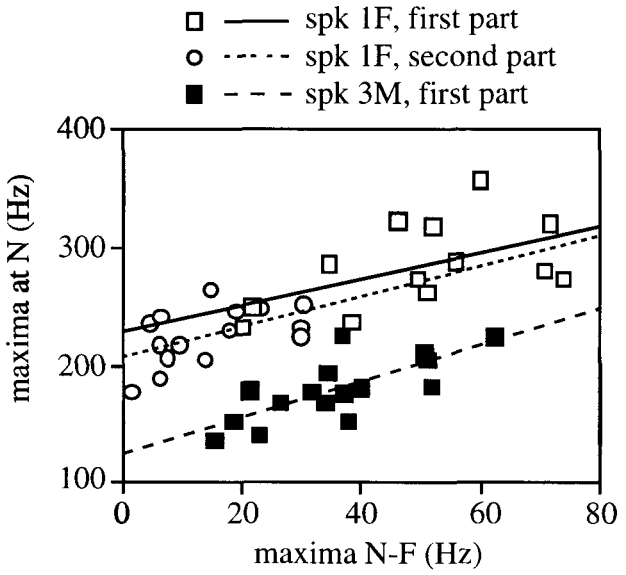


Figure 8: Difference between the height of f_0 maxima at fast and normal rate (maxima N-F in Hz) relative to the height of the maxima at normal rate (maxima at N in Hz).

Finally, consider the difference in the strategy used by different speakers to reduce pitch displacements at fast rate. We observed two strategies for three subjects in increasing rate: two subjects (1F and 3M) lowered f_0 maxima more than they lowered f_0 minima, one subject (2F) raised f_0 minima keeping f_0 maxima constant. Although only one of our speakers chose to undershoot pitch minima, this way of reducing pitch displacements seems to be more common since it was also found in Dutch for one of two speakers (Caspers, 1994) and German for all three speakers (Kohler, 1983). Thus, we will try to explain why speakers vary in their strategy to increase rate and why our two speakers chose to reduce their maxima rather than raising their minima.

First, the distribution of the pitch targets in the subject's pitch range can offer a plausible explanation for these two strategies. Here we will focus on the two female speakers since their pitch ranges are very similar. Figure 9 presents the distribution of the pitch targets (both minima and maxima) within the pitch range of speaker 1F (left panel) and speaker 2F (right panel). The vertical dashed line marks the mid-range in each distribution. We can observe that for speaker 2F the pitch targets are more equally distributed around the mid-range (58% of the pitch targets under, and 42% above the mid range) than for speaker 1F. For speaker 1F, the dispersion is more concentrated in the lower part of her range (74% of the pitch targets). We can assume that since speaker 1F uses the lower part of her pitch range more often, the movements toward the high targets located in the upper range may be more extreme, in that they require larger displacements than the displacements restricted to the lower region. Thus, when time is short, the targets at the extreme end of the upper range would be more likely to be reduced because they are more extreme. Also if the realization of these extreme targets requires a bigger effort because they are further away, it is possible that the speaker would prefer to reduce these more effortful gestures. A similar speaker variation has been observed by Kuehn (1976) for velic movements. Looking at the reduction in the displacements of velum movements at fast speech rate, he found that his two speakers varied in the strategy used to reduce movements. At fast compared to normal rate, one speaker produced less extreme high positions, keeping low velic

positions unaffected, whereas the other speaker produced less extreme low positions, keeping high velic positions unaffected. The difference in the choice of the target reduced at fast rate was accounted for by the fact that speakers may use different part of the range of velic movement available at their normal rate. The speaker who reduced the high position at fast rate can do so because at normal rate he raises the velum beyond the level necessary for a velopharyngeal closure. On the contrary, the other speaker raises the velum just high enough to achieve the closure at normal rate. Hence, high velum position could not be reduced at fast rate for this speaker without challenging the velopharyngeal closure.

Next, the variation in the choice of reducing low or high pitch targets can be the consequence of prosodic reorganization at fast rate. We observed that speaker 1F and 3M reduced about half of their IP boundaries into ip or AP boundaries. Since most boundaries were high tones, and the H% boundary tone of IP is in general higher than ip boundary tone or AP final high tone, this must have contributed to the reduction of high targets for these speakers. Along the same line, maxima may not have been reduced for speaker 2F because she did not reduce the strength of the prosodic boundary at fast rate. Since she keeps the maxima the same, she might have to reduce minima to reduce displacements at fast rate.

Finally, the interspeaker variation can be due to different styles of speech. For example, speaker 2F showed a particular pattern at fast rate: she preserved all prosodic groupings of normal rate; she did not undershoot high targets; she reduced low targets very little; she reduced pitch displacements but less than the other speakers; she reduced velocity but less than other speakers. Therefore, she seems to hyperarticulate her pitch movements by not reducing the movements and also by realizing the movements with faster velocity at both normal and fast rates.

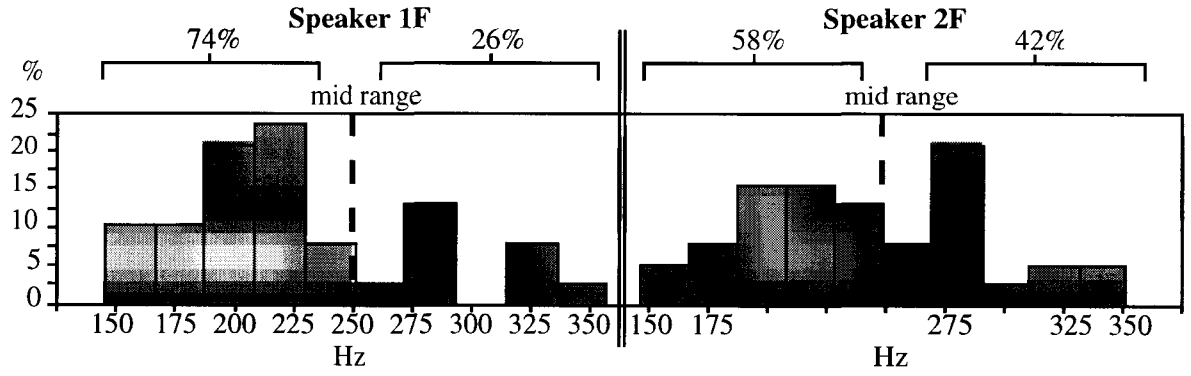


Figure 9: Distribution of the pitch targets for speakers 1F (left panel) and 2F (right panel) depending on their pitch range. The mid-value of their pitch range (mid-range) is shown by the vertical line in each graph. X-axis refers to frequency in Hz and Y-axis refers to the percentage of pitch targets occurring at these frequencies.

In sum, the production of intonation at fast rate is constrained by what appears to be a saturation effect. Small displacements, for the lower high-targets or for the second part of the text, are not further reduced at fast rate. It is possible that these displacements are small enough to be fully realized, even at fast rate. We expect that these small displacements would also be reduced if the speaking rate was further increased. The resistance to further reduction of pitch movement at fast rate could also be the result of a constraint on the minimal size that a pitch

movement must have (or the minimal height a high tone must have) in order to be perceived as a high tone. We also found that the compressibility of a pitch target depends on its linguistic function. In agreement with Caspers and van Heuven's observations in Dutch, we showed that at fast rate major Intonational Phrase boundaries are less likely to be reduced than less informative boundaries. Moreover, the choice of reducing high or low targets to achieve a reduction of pitch displacements is speaker dependent. We showed that the production at fast rate is constrained by the way speakers distribute their pitch targets in their pitch range and by their choice of reducing prosodic boundaries at fast rate.

4.3. Prosodic reorganization at fast rate.

In addition to the modification in the phonetic realization of intonation, we found that the prosodic structure of an utterance can be reorganized at fast rate at least for two out of three speakers. That is, the prosodic grouping (phrasing) and tonal pattern can be modified when the rate is increased. For example, IP becomes ip or AP, and an AP boundary can be deleted. As a result, a syllable carrying a particular phrasal (H*) or boundary (H%) tones at normal rate may be associated with a different tone at fast rate. An example of this is shown in Figure 2 with the letter 'B'. In this example, two APs, (qui arriverait) and (le premier), become one AP, (qui arriverait le premier), and the final vowel [E] in *arriverait* associated with H* at normal rate is toneless at fast rate. This means that the number of syllables in one AP is increased and tone to syllable association is reorganized. Also, since it has been shown that prosody conditions the phonetic realization of segments, we expect that segments would be realized differently depending on their new prosodic grouping. A segment that is IP final is lengthened and a segment that is IP initial is strengthened (Fougeron & Keating, 1996). If the segment is not in the same prosodic position at fast rate, it will loose these prosodically driven characteristics. Changing the prosodic grouping also means reorganization of the information structure: at fast rate more words can group into one prosodic unit, thus fewer prosodic units are observed. This reduction of the number of prosodic units has also been found in French (Vaissière, 1992), Korean (Jun, 1993) and in Dutch (Caspers, 1994). Even though the speech material is quite different in these studies, the reduction of Accentual Phrase boundaries in our study (22-28%) is similar to that in Korean (24%), and the reduction of Intonational Phrase boundary in our study (50-55%) is similar to that in Dutch (56%). However, while the deletion or reduction of boundaries are the results of the prosodic reorganization at fast rate, the absence of initial high tone (Hi) in an AP is not. As we proposed before (see Section 4.1.2 and 4.1.3), it is due instead to the phonetic realization of the tonal pattern of an AP showing more overlap between H- and L-gestures at fast rate.

So far, we have accessed the prosodic phrasing or organization only by its intonational cues. But phrasing is also marked by various degrees of final lengthening. In French, Intonational Phrase final syllables are longer than Accentual Phrase final syllables which in turn are longer than Accentual Phrase non-final syllables at normal rate (Hirst & Di Cristo, 1984, Pasdeloup, 1990, Jun & Fougeron, 1995). (For a similar final lengthening pattern following the hierarchical structure in English, see Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992). To see whether the durational cues of the prosodic organization are maintained at fast rate, we measured the durations of every syllable in the first part of the text for all speakers. Among these, we compared the syllables that have the same prosodic position at fast and normal rate. The result is shown in Figure 10.

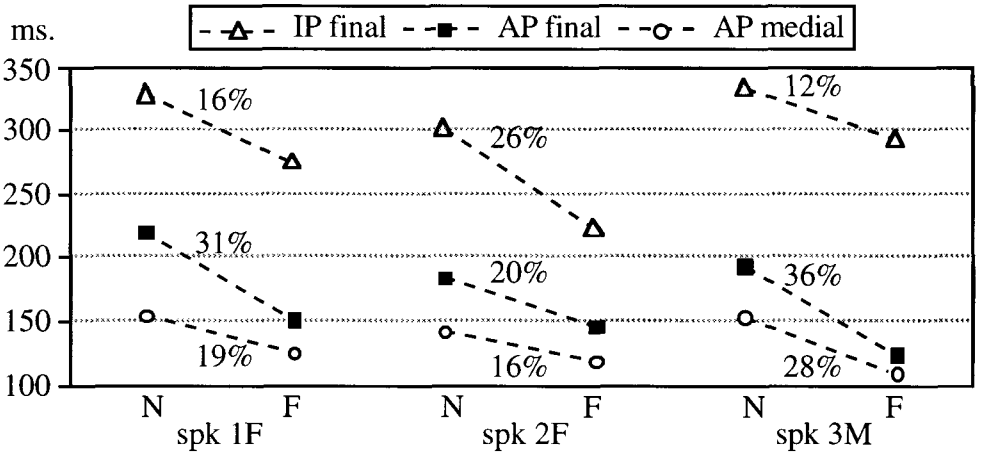


Figure 10: Duration of IP final, AP final and AP medial (also non initial high tone, Hi) syllable at fast and normal rate for the 3 speakers (the first part of the text only). For each position, the percent of reduction of the duration between normal and fast rate is indicated.

In this graph, the duration of Accentual Phrase final syllables, Intonational Phrase final syllables, and Accentual Phrase medial syllables in the first part of the text are plotted for normal and fast rate for each speaker. Accentual Phrase initial syllables that have an initial high tone (Hi) at normal rate are not included in the plot. The initial high-toned syllable has been claimed not to have a longer duration than the unaccented syllable in French (Pasdeloup, 1990). In our data, initial high-toned syllables are not always significantly different from lengthened Accentual Phrase final syllables. Therefore, we excluded the initial high-toned syllables from the comparisons. These durations are not standardized and therefore can be biased by intrinsic and contextual influences of segments. However, since in our data the same syllables are compared in normal and fast rate categories, their intrinsic/contextual properties will contribute the same way to each category. As shown in Figure 9, the hierarchical organization of syllabic duration depending on prosodic position is maintained at fast rate ($F(2, 688) = 292$, $p < .0001$). Intonational Phrase final syllables are still longer than Accentual Phrase final syllables, which in turn are longer than Accentual Phrase medial syllables.

However, the degree of reduction in the duration is not the same across the three prosodic positions: speaker 1F and 3M again pattern together by reducing the duration of AP final syllables the most (1F: 31%; 3M: 36%) compared to syllables in other prosodic conditions (IP final and AP medial reduction for 1F: 16-19% and 3M: 12-28%). Speaker 2F, on the other hand, showed a degree of reduction that is smaller from the highest to the lowest level in the prosodic hierarchy (IP final: 26%; AP final: 20%; AP medial: 16%), thus, the longer the duration at normal rate, the bigger the reduction at fast rate. We assume that this difference between speakers is due to their different strategies regarding the prosodic reorganization at fast rate. For speakers who reorganize the prosodic grouping (1F and 3M), IP boundaries maintained at fast rate are more likely to be the important ones. And as shown earlier the f_0 height of these boundaries did not change much at fast rate (see Figure 5). Thus, the duration of the syllables at this boundary would not change much either to preserve the strength of the boundary. However, since AP boundaries are less important in discourse structure (shown by their high deletion percentage at fast rate), the duration of this boundary is more free to reduce at fast rate (31-36%). As a consequence, AP final and AP medial syllables are less distinguished by duration at fast

rate, and for speaker 3M, this difference is not statistically significant. For this speaker it seems that there are only two prosodic levels at fast rate if we look at the duration. However, it is not a complete prosodic reorganization (reducing from 3 to 2 prosodic levels at fast rate) since the three prosodic levels are still marked by the tonal cue. This pattern shows that an AP boundary which can be cued by tone and duration in normal rate is cued only by tone at fast rate for this speaker. On the other hand, for speaker 2F, who did not change the prosodic grouping, the linguistic weight of the boundary (IP or AP) is not a factor of reduction in her pattern of duration. Rather the degree of reduction would be the function of the duration at normal rate: longer syllables are more reduced at fast rate. A similar pattern of reduction was found in the reduction of f_0 maxima for speaker 1F and 3M: the higher the f_0, the greater the reduction (see Figure 8).

In sum, we have shown that the prosodic grouping of an utterance can be reorganized at fast rate, but is speaker dependent. The pattern of prosodic reorganization is cued by the intonational pattern. At all prosodic levels, we observed a reduction in the number of prosodic groups at fast rate. Also, we showed that even with a compression in the temporal domain (rate acceleration), the durational cues for marking the prosodic structure of an utterance are generally maintained at fast rate.

5. CONCLUSION

In this paper, we have shown that an acceleration of rate affects intonation both in its phonetic realization and prosodic organization. Rate acceleration induces reduction of pitch range and pitch displacements between maxima and minima pitch targets, with small change in the velocity of pitch movements. The prosodic organization of intonation is modified: some phrase boundaries are reduced or deleted leading to fewer phrases. The realization of the underlying tones is also affected by speech rate. We also showed that these modifications of intonation at fast rate vary depending on the position in the text and the speakers. We believe that our study has shown the need to look at both sides of intonation — phonetic realization and phonological structure — when studying intonation.

In interpreting the result, we equated f_0 displacements, f_0 peaks and valleys, and f_0 slope with the kinematic parameters used in the studies of articulatory movements. We showed that at fast rate some dynamic properties of the f_0 contour are modified in a way comparable to the modifications found for other articulations. Therefore we conclude that the mechanism controlling speech rate is the same for both the articulation of intonation and for other articulations. We tried to model the observed modifications of intonation using L- and H- pitch gestures. We hope this will lead to further research on developing a model of speech production that includes prosody.

Acknowledgments

An earlier version of this paper was presented at the XIIIth meeting of the International Congress of Phonetic Sciences, Stockholm, Sweden, in 1995. We thank Pat Keating and Victoria Anderson for their comments and suggestions and J. Caspers and V. van Heuven for providing their articles. We also acknowledge the speakers for their participation in this experiment. The first author is supported by Allocation de recherche M.R.T. granted to the D.E.A. de Phonétique de Paris.

References

Abbs, J. H. (1973) The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations, *Journal of Speech and Hearing Research*, **16**, 175-200.

Bartkova, K. (1991) Speaking rate modelization in French: application to speech synthesis. In *Proceedings of the twelfth International congress of phonetic sciences*, pp. 482-485.

Beckman, M. & Pierrehumbert, J. (1986) Intonational structure in Japanese and English, *Phonology Yearbook,* **3**, 255-309.

Browman, C. P. & Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech, In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. (J. Kingston, & M. Beckman, editors) pp. 152-178. Cambridge: Cambridge University Press.

Browman, C. P. & L. Goldstein (1992) Articulatory Phonology: An overview, *Phonetica*, **49**, 155-180.

Brubaker, R. S. (1972) Rate and pause characteristics of oral reading, *Journal of Psycholinguistic Research*, **1**(2), 141-147.

Byrd. D., S. Narayanan, A. Kaun, & E. Saltzman (1996) Phrasal influences on articulatory detail. Paper presented at the Laboratory Phonology conference V. Evanston, IL.

Caspers, J. (1994). *Pitch movements under time pressure : effects of speech rate on the melodic marking of accents and boundaries in Dutch.* Doctoral diss. Holland Institute of Generative Linguistics, Univ. of Leiden.

Caspers, J. & van Heuven, V. J. (1991). Phonetic and linguistic aspects of pitch movements in fast speech in Dutch. In *Proceedings of the 12th international congress of phonetic sciences*, **5** (pp. 174-177).

Caspers, J. & van Heuven, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall, *Phonetica*, **50**, 161-171.

Caspers, J. & van Heuven, V. J. (1995). Effects of time pressure on the choice of accent-lending and boundary-marking pitch configuration in Dutch. In *Proceedings of Eurospeech.*, **2** (pp. 1001-1004), Madrid.

Engstrand, O. (1988) Articulatory correlates of stress and speaking rate in Swedish VCV utterances, Journal of Acoustical Society of America, **83**(5), 1863-1875.

Fougeron, C. & Keating, P. (1996) Articulatory strengthening in prosodic domain-initial position. In *UCLA Working Papers in Phonetics* , **92**, 61-87.

Fowler, C. (1977) *Timing control in speech production.* Bloomington, Indiana: Indiana University.

Gay, T. (1981). Mechanisms in the control of speech rate, *Phonetica*, **38**, 148-158.

Grosjean, F. & Deschamps, A. (1975) Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation, *Phonetica,* **31**, 144-184.

Harrington, J., Fletcher, J. & Roberts, C. (1995) Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data, *Journal of Phonetics,* **23**, 305-322.

Hirschberg, J. & Pierrehumbert, J. (1986) The intonational structuring of discourse. In Proceedings of the 24th annual meeting of the association for computational linguistics. (pp. 136-144). New York, NY..

Hirst, D. & Di Cristo, A. (1984) French intonation: a parametric approach, *Die Neueren Sprachen* , **83**(5), 554-569.

Hirst, D. & Di Cristo, A. (in press) *Intonation systems: A survey of twenty languages.* Cambridge University Press.

Jun, S.-A. (1993) *The Phonetics and Phonology of Korean Prosody.* Ph.D. diss. The Ohio State University.

Jun, S.-A. & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. In *Proceedings of XIIIth international congress of phonetic sciences.* Vol. **2** (pp. 722-725). Stockholm, Sweden.

Keller, E. & Zellner, B. (1995) A Statistical timing model for French. In *Proceedings of the XIIIth international congress of phonetic sciences*, Vol. **3** (pp. 302-305).

Kent, R. D. & Moll, K. L. (1972) Cinefluorographic analyses of selected lingual consonants, *Journal of Speech and Hearing Research,* **15,** 453-473.

Kent, R. D., Carney, P. J. & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control, *Journal of Speech and Hearing Research,* **17**, 470-488.

Krakow, R. (1993) Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate. In *Phonetics and Phonology: Nasals, nasalization, and the velum* (M. Huffman & R. Krakow, editors) Vol. **5**, pp. 87-116. Academic Press.

Kohler, K. (1983) f0 in speech timing. In *arbeitsberichte* (W.J. Barry & K. J. Kohler, editors) **20** (pp. 57-97). Institut für Phonetik, Universität Kiel.

Kuehn, D. P. (1976). A cineradiographic investigation of velar movement variables in two normals, *Cleft Palate Journal,* **13**, 88-103.

Kuehn, D. P. & Moll, K. (1976). A cinefluographic investigation of CV and VC articulatory velocities, *Journal of Phonetics,* **3**, 303-320.

Liberman, M. (1975) *The Intonational System of English*, Ph.D. dissertation. MIT. [Reproduced 1978 by the Indiana University Linguistics Club, Bloomington.]

Liberman, M. & Pierrehumbert, J. (1984). Intonational invariance under change in pitch range and length. In *Language Sound Structure* (M. Aronoff & R. Oehrle, editors) pp. 157-233. MIT Press.

Lindblom, B. (1963). Spectrographic study of vowel reduction, *Journal of the Acoustical Society of America,* **35** (11), 1773-1781.

Lindblom, B. (1964). Articulatory activity in vowels, *STL-QPSR* **2**, 1-5.

Malécot, A., Johnson, R. & Kizziar, P.-A. (1972) Syllabic rate and utterance length in French, *Phonetica,* **26**, 235-251.

Mertens, P. (1993) Intonational grouping, boundaries and syntactic structure in French. In *ESCA Workshop on Prosody, Lund WP.* **41** (pp. 155-159).

Miller, J., Grosjean, F. & Lomanto, C. (1984) Articulation rate and its variability in spontaneous speech: A reanalysis and some implications, *Phonetica,* **41**, 215-225.

Munhall, K. & Löfqvist, A. (1992) Gestural aggregation in speech: Laryngeal gestures, *Journal of Phonetics,* **20**, 111-126.

Ohala, J. & Ewan, W. G. (1973) Speed of pitch change, *Journal of the Acoustical Society of America*, **53**(1), p. 345 (Abstract)

Ostry, D. & Munhall, K. G. (1985) Control of rate and duration of speech movements, *Journal of the Acoustical Society of America*, **77**(2), 640-648.

Pasdeloup, V. (1990) *Modèle de règles rythmiques du français appliquées à la synthèse de la parole.* Doctorat Univ. Aix en Provence.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.

Saltzman, E. (1995) Intergestural timing in speech production: Data and modeling. In *Proceedings of the XIIIth international congress of phonetic sciences*, **2** (pp. 84-91).

Saltzman, E. & Munhall, K. G. (1989) A dynamic approach to gestural patterning in speech production, *Ecological Psychology,* **1**, 333-382.

Sluijter, A. M. & Terken, J. M. (1993) Beyond sentence prosody: Paragraph intonation in Dutch, *Phonetica,* **50**, 180-188.

Vaissière, J. (1983) Language-independent prosodic features. In *Prosody: Models and measurements* (A. Cutler & R. Ladd, editors), pp. 53-66. Springer-Verlag.

Vaissière, J. (1992) Rhythm, accentuation and final lengthening in French. In *Music, Language, Speech and Brain* (J. Sundberg, L. Nord, & R. Carlson, editors). Wenner-Gren, International Symposium Series, **59**, pp. 108-120. Stockholm

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. (1992) Segmental duration in the vicinity of prosodic phrase boundaries, *Journal of the Acoustical Society of America,* **91**, 1707-1717.

The sounds of languages*

Peter Ladefoged

Once upon a time the most important sounds were those of predators and prey, and possible sexual partners. As mammals evolved and signaling systems became more elaborate, new possibilities emerged. Now undoubtedly the most important sounds for humans are those of language. Spoken language, which always precedes written language, is our way of expressing awareness of what goes on around us. Because of this, language provides a storage system that represents our knowledge of the world. Our language and what we say and write in it bear the same relationship to our view of the world as a map does to the terrain it represents. Words and sentences are our way of forming maps that show what we think the world is like. Without language we cannot represent what we know.

As this series is named in honor of Darwin, it is appropriate to think about the origin of language. Unfortunately, nobody knows how vocal cries turned into language; but, then, Darwin did not know how life began. He was concerned not with the origin of life but with the origin of the various species he could observe. We will not consider the origin of language; but we will note the various sounds of languages, and discuss how they got to be the way they are. We will think of each language as a system of sounds subject to various forces.

The overall aim of a language is to represent as much information as possible with the least possible effort. Because there are so many ways of doing this, the sounds of speech are extraordinarily diverse. They are, however, constrained, first by what we can do with our tongues, lips and other vocal organs, and secondly by our limited ability to hear small differences in sounds. These and other constraints have resulted in all languages evolving along similar lines. No language has sounds that are too difficult to produce within the stream of speech (although, as we will see, some languages have sounds that would twist English speaking tongues in knots); and every language has sounds that are sufficiently different from one another to be readily distinguishable by native speakers (although, again, some distinctions may seem much too subtle for ears that are unfamiliar with them).

There are some additional factors that have shaped the development of languages, notably how our brains organize and remember sounds. If a language had only one or two vowels and a couple of consonants it could still have an infinite number of words; but many of the words would be very long, difficult to remember, and sound much alike. If words are to be kept short and distinct so that they can be easily distinguished and remembered, then the language must have a sufficient number of vowels and consonants to make many different syllables. But we don't want to have a large number of sounds that are all completely different from one another. It puts less strain on our cognitive capacities if the sounds of our languages can be organized in groups that are articulated in the same way. This is a principle that my colleague Ian Maddieson has called gestural economy. There is a pressure to select the sounds of a language so that they form a simple pattern within the vast set of possible speech sounds. Typically, if a language has one sound made by a gesture involving the two lips such as **p** as in *pie*, then it is likely to have others such as **b** and **m** as in *by* and *my* made with similar lip gestures; and if it has *pie, by, my,* and also a sound made with a gesture of the tongue tip such as **t** as in *tie*, then it is also likely to have other sounds made

* This paper is a draft of a chapter for a book based on the Darwin lectures on sound, intended for lay readers. British spelling is required, and no references are allowed.

with the tongue tip, such as **d** and **n** as in *die* and *nigh*. The sounds that evolve in a language form a pattern; and there is a strong pressure to fill gaps in the pattern.

Societies weight the importance of the various constraints — articulatory ease, auditory distinctiveness, and gestural patterning — in different ways, producing roughly 7,000 mutually unintelligible languages in the world. But despite the variations in sound that make each language distinct, there are common features that occur in all. For example, every language uses vowels (speech sounds that can be said on their own) and consonants (which generally can be sounded only with an accompanying vowel) to produce a variety of syllables; and all languages use the pitch of the voice in a meaningful way. In the following sections we will consider how languages use differences in pich, vowels and consonants to form different words.

Producing pitch changes in speech

When we listen to people talking we do not think of the sound as being that of a musical instrument. But the voice is an instrument that we all use to produce tunes when talking. Some people think that they cannot be producing tunes because they consider themselves to be tone deaf. But nobody is completely tone deaf, unless they are literally deaf in every respect. Everyone can hear and produce the tunes required in speech. We can make and distinguish statements and questions. We can use the pitch of the voice to make the subtle grammatical differences that are marked by punctuation in the written language such as "When danger threatens — your children call the police" as compared with "When danger threatens your children — call the police." Some different tunes on sentences are not even marked by the punctuation; the reader (or listener) has to get them from the context. For example, we can tell the difference between sentences such as "Jenny gave Peter *instructions* to follow" and "Jenny gave Peter instructions to *follow*". The first means that Peter was told what to do, and the second that he was told to come along after Jenny. The words are the same in the two sentences, but the meaning is different because of the differences in pitch and rhythm. Other pitch changes in speech can be used to convey other kinds of information. We can usually tell whether a speaker is angry or loving by listening to the tune without even hearing the words. Much of the emotional content of speech is carried by the pitch of the voice.

Being somewhat tone deaf does not mean that one cannot hear and produce different pitches accurately. It is just a matter of not being able to sing in tune. People for whom this is true probably did not have music in their background when they were young. Learning to sing is like learning a language, easier to pick up and do with perfection if you are a child, but more difficult when older. Singing differs from speaking in holding the pitch of the voice constant, usually for a syllable or two, and then jumping to the next note. In speech the pitch is always changing, never remaining the same, even within a single syllable. The pitch often goes down to mark the end of a sentence. It rises, at least in English, in questions that can be answered by yes or no. Statement such as "I'm going home" typically have a falling pitch; only yes or no questions such as "Are you going home?" usually rise at the end. However, it is almost impossible to generalize about the pitch changes that can occur in a language. The neutral way of asking the question "Where are you going?" is to say it with a falling pitch. But one can say "Where are you going?" with a rising pitch, and "Are you going home?" with a falling pitch, causing differences in emphasis and meaning. The tune of the voice in speech is one of the most difficult aspects of languages to describe, as it can be used to convey so many small nuances of meaning.

In English, and in most European languages, the meaning of a word remains the same irrespective of whether it is said on a rising pitch or a falling pitch. But this is not true in languages spoken in other parts of the world. In Standard Chinese the syllable **ma** has four different

meanings, depending on the pitch on which it is spoken; and in Cantonese, another language spoken in China, syllables can have up to six different meanings. Examples of words differing in pitch in each of these languages are given in Table 1. Differences of this type are called differences in tone. Standard Chinese is said to have four tones, and Cantonese has six tones on syllables of the type shown in Table 1. Although tonal differences are rare or non-existent in most languages spoken in Europe and India, well over half the languages spoken in the world use tones to differentiate the meanings of words.

Table 1. Examples of words differentiated by tone in Standard Chinese and in Cantonese.

STANDARD CHINESE **ma**				CANTONESE **si**			
媽	˥	high level	'mother'	詩	˥	high falling	'poem'
麻	˧˥	high rising	'hemp'	試	˧	mid level	'to try'
馬	˨˩˧	low falling rising	'horse'	事	˨	low level	'matter'
罵	˥˩	high falling	'scold'	時	˩	extra low	'time'
				使	˧˥	high rising	'to cause'
				市	˧	mid rising	'city'

The pitch of the voice depends mainly on the tension of the vocal folds, two small muscular flaps in the larynx. When the larynx is pulled forward the folds become longer and thinner so that they vibrate more quickly. The vibrations are the result of the air from the lungs sucking the folds together and then blowing them apart again. They behave in some ways like the lid of a boiling kettle. When the pressure of the steam below the lid becomes too great it is blown upward, releasing the pressure. When there is no pressure beneath it, it falls down; and then the pressure builds up again. In the case of the vocal folds there is a slight complication. The pressure of the air in the lungs will blow them apart, but, when the pressure is less, they do not simply collapse together. There is an additional mechanism drawing them towards one another by the air. The vocal folds are actively sucked together by the air passing between them. Air traveling at speed through a narrow gap drops in pressure (a fact that can be noticed by anyone in a vehicle when another vehicle traveling in the opposite direction passes it and the two vehicles are sucked towards one another). This suction causes the vocal folds to be sucked together faster than they were blown apart. The rapid, repetitive, coming together of the vocal folds is what produces the sound of the voice.

Vowels

The vibrations of the vocal folds are the source of energy in the production of vowels. Most languages use few more than the five vowels that can be represented by the letters **a**, **e**, **i**, **o**, **u**, as in the Spanish words in Table 2. Languages as diverse as Spanish, Hawaiian, Swahili and Japanese manage with just five vowel sounds. English has a comparatively large vowel inventory. Accents of English differ in their vowel qualities, but in some styles of southern British English there are 17 different vowels. We cannot find a single set of words differing in only the vowels as we almost did for the languages in Table 3, but we can demonstrate the possibilities that can occur by considering syllables beginning with **b** and ending with **d** or **t**, and those beginning with **h** and ending with **d**, as shown in Table 3. Some of the vowels in these words differ from others simply in a single quality, and others differ by being diphthongs — vowel sounds with changing

qualities in a single syllable. Many forms of English do not distinguish all these vowels — Californians, for example, have the same vowel in *hard, hod,* and *hawed* (making *father* and *bother* rhyme, both having the same vowel as in *author*), and many Scottish speakers do not distinguish the vowels in *hood* and *who'd* (or *look* and *Luke*). Other forms of English have a slightly larger number of vowels. (We are, of course, concerned with the sounds that can occur as vowels in English syllables, and not the various ways vowels can be spelled in the written language; English spelling is notoriously odd.) Other languages such as German and Swedish have an even greater number of vowels. The record for the greatest number of vowels (excluding those with special voice qualities and tones which we will discuss later) is probably held by Austrian German dialects that can produce different syllables by choosing among 26 different vowel sounds, 13 long and 13 short.

Table 2. Words illustrating the vowels **a, e, i, o, u** in Spanish, Hawaiian, Swahili and Japanese. (The qualities of the vowels are not exactly the same in each of these languages.)

SPANISH		HAWAIIAN		SWAHILI		JAPANESE	
masa	'dough'	**kaka'**	'duck'	**pata**	'hinge'	**ka**	'mosquito'
mesa	'table'	**keke'**	'surly'	**peta**	'bend'	**ke**	'hair'
misa	'mass'	**kiki'**	'rapid'	**pita**	'pass'	**ki**	'tree'
mosca	'fly'	**koko**	'blood'	**pota**	'twist	**ko**	'child'
musa	'muse'	**kuku'**	'thorn'	**puta**	'thrash'	**ku**	'suffering'

Table 3. English vowel sounds that can occur between **b** and **d** or **t,** and **h** and **d** in one southern British style of speech, in which 'r' after a vowel is not pronounced.

b__d	**b__t**	**h__d**
bead	beat	heed
bid	bit	hid
bayed	bait	hayed
bed	bet	head
bad	bat	had
bard	(Bart)	hard
bod(y)	bot(tom)	hod
bawd	bought	hoard
bode	boat	hoed
bud(hist)		hood
booed	boot	who'd
bide	bite	hide
bowed	bout	howd(y)
bud	but	Hudd
bird	Bert	herd
beard		
bared		hared

Vowel sounds can be produced on any pitch within the range of a speaker's voice. I can say the vowels in *heed, hid, head, had* on a low pitch, when the vocal folds are vibrating about 80 times a second (a low E), and then I can say them again with vocal folds vibrating 160 times a second (the E an octave above). The pitch of my voice will have changed, but the vowels will still have the same quality. Different vowels are like different instruments. One can play concert pitch A on a piano, a clarinet, or a violin. In each case the note will have the same fundamental frequency, but a different quality. Similarly, vowels will retain their individual qualities irrespective of the pitch produced by the vocal folds.

When we listen to a musical note, or a vowel, we can tell which instrument produced it, or which vowel it is, by the overtones —the higher frequencies—that occur. The reed of the clarinet or the vocal folds may be vibrating 120 times a second, but the sound that is produced at the mouth of the clarinet or the lips will contain characteristic groups of overtones at higher frequencies. For the vowel in *head* the most prominent overtones will be at about 550 and 1600 Hz, for the vowel in 'had' they will be around 750 and 1200 Hz, These overtones will occur although the vocal folds may be vibrating at any rate from about 80 to around 250 Hz for a male speaker.

To see how these higher frequencies arise, we can liken the air in the mouth and throat to the air in a bottle. When you blow across the top of a bottle the air inside it will be set in vibration. The note that the bottle produces will depend on the size and shape of the body of air that it contains. If the bottle is empty it will produce a low pitched note. Pouring water into it so that the body of air becomes smaller makes the pitch go up, as smaller bodies of air vibrate more quickly.

The air in the vocal tract (the tube bounded by the vocal organs) is set in vibration by the pulses of air from the vocal folds. Every time they open and close the air in the vocal tract above them will be set in vibration. Because the vocal tract has a complex shape, the air within it will vibrate in more than one way. Often we can consider the body of air behind the raised tongue to be vibrating in one way, and the body of air in front of it to be vibrating in another. In the vowel in 'head' the air behind the tongue will vibrate at about 550 Hz, and the air in front of it at about 1600 Hz.

The resonances of the vocal tract are called formants. Trying to hear the separate formants in a vowel is difficult. We are so used to a vowel being a single meaningful entity that it is difficult to consider it as a sound with separable bits. But it is possible to say vowels so that their component parts are more obvious. One possibility is to whisper a series of vowels. When whispering the vocal folds do not vibrate; they are simply drawn together so that they produce a random noise like that of the wind blowing around a corner. Because this noise is in the pitch range of one of the resonances of the vocal tract, we can hear that resonance more plainly. If you whisper you will not hear a note with a specific pitch; but you will be able to hear the changes in the vowel resonances. Try whispering *heed, hid, head, had, hod, hawed*; there will be general impression of a descending pitch.

Another way of making one of the resonances more obvious is to say a series of words on a very low pitch. Say *ah* on as low a pitch as you can, and then try to go even lower so that you produce a kind of creaky voice. Say the words *hard, hoard. hood, who'd* in this kind of voice. You may be able to hear not only the constant low buzzing sort of pitch associated with the vocal folds, but also a changing pitch in one of the overtones. When saying the words *hard, hoard. hood, who'd* this pitch goes down. If you say the words *heed, hid, head, had* with the same kind of creaky voice, you may be able to hear an ascending pitch.

The sound that you hear when whispering is mainly that of the vibrations of the air in the front of the mouth. The pitch changes associated with saying 'hard, hoard. hood, who'd' in a creaky voice are due to the vibrations of the air in the back of the vocal tract. This resonance is the lower in pitch of the two, and is called the first formant. The height of the bars in figure 1 shows the mean pitch of this formant in the vowels in *who'd, hood, hawed, hod, had, head, hid, heed.* The words are listed from left to right mainly in order of increasing frequency of the overtone that is heard when whispering (the second formant). The highest first formants will be when the

second formant is in the middle of its range. The lowest first formants will be when the second formant is very high (as you can hear it is when you whisper *heed*) or when it is very low (which does not occur in most dialects of English; there is no justification for placing the words *who'd, hood* on the left, other than the fact that they have low first formant frequencies). The solid curve in the figure shows the limits on the first and second formant frequencies that can occur, given an average size male vocal tract.

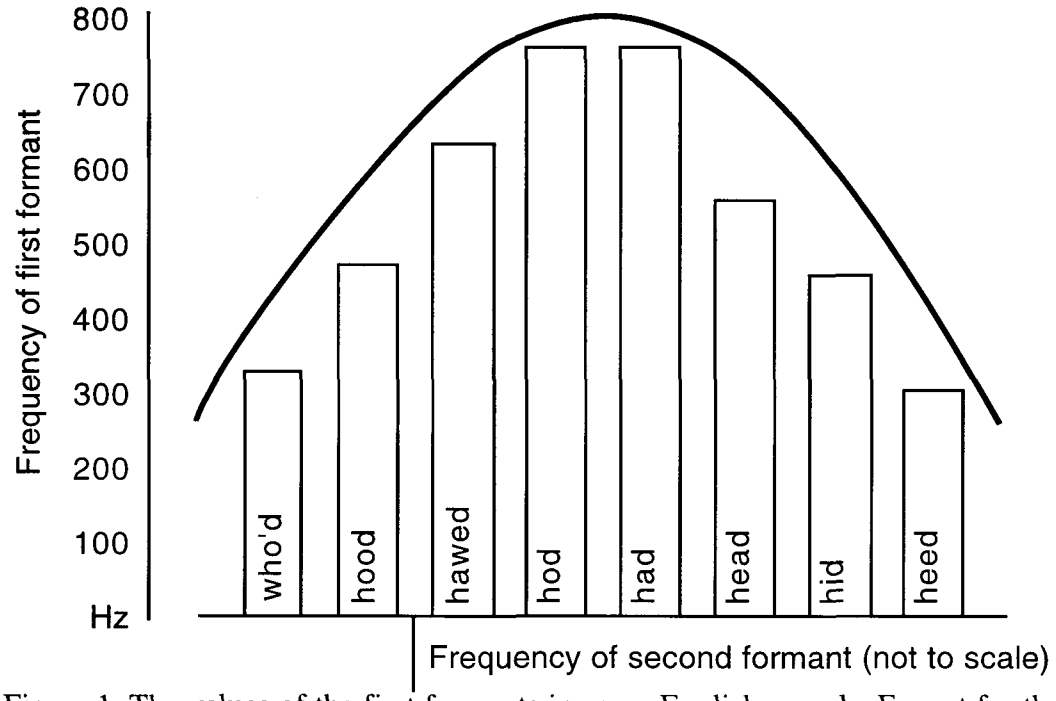


Figure 1. The values of the first formants in some English vowels. Except for the vowels in 'who'd, hood' the vowels are shown in order of increasing values of the second formant (not to scale). The solid curve marks the limits of the possible vowel space.

We can now see why so many languages have the five vowels **a**, **e**, **i**, **o**, **u**, pronounced as in the Spanish words 'masa, mesa, misa, mosca, musa'. Figure 2 shows the five Spanish vowels in relation to the boundaries of the formant space that can be produced by an average male speaker. These vowels are fairly evenly distributed near the perimeter of the vowel space. The vowels **i**, **a**, **u** are near the corners of this space, and thus as far apart from each other as possible. If we want a set of three vowels that will be as auditorily distinct as possible, these are the vowels to use —which is why so many languages have evolved so that they include these vowels. Most languages need more than three vowels in order to have a sufficient number of distinct syllables. If a language has five vowels, like Spanish and many other languages, they will be easy to distinguish if they are distributed in the possible vowel space as shown in figure 2. English uses a greater number of vowel qualities, making them distinct from one another by allowing them to vary in other ways such as length. The vowel in *heed* is different from that in *hid* not only by having a lower first formant, but also by being longer.

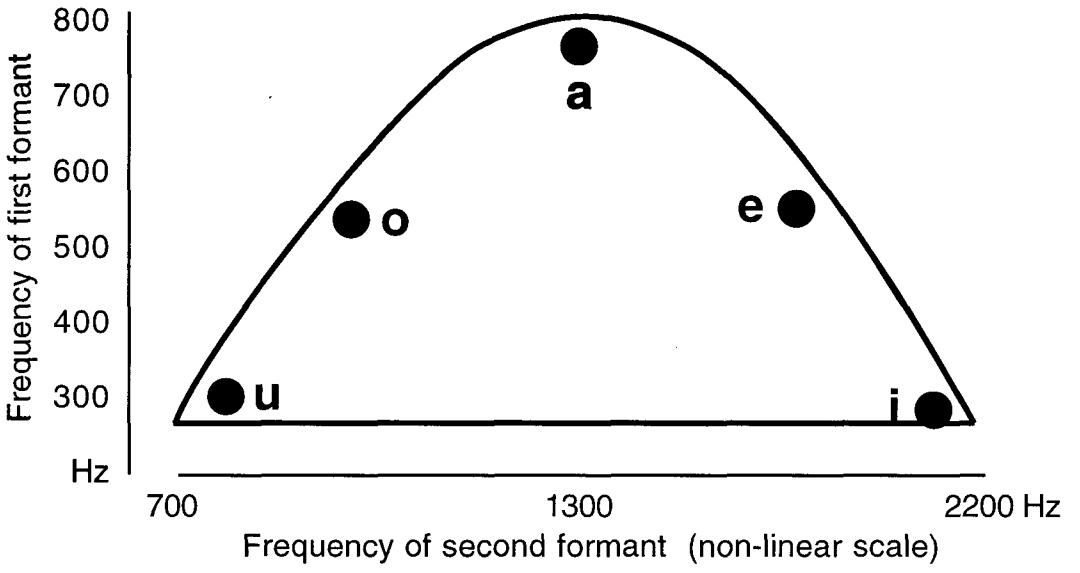


Figure 2. The possible vowel space, showing the five Spanish vowels.

So far we have been considering vowel systems in terms of only the first of the pressures acting on the sounds of languages — the need to make the sounds within a group as auditorily distinct as possible. But the vowel systems of the world's languages also show evidence of another constraint — the pressure for forming patterns. Given that the auditory space for possible vowels is somewhat triangular, the selection of the three most distant vowels **i**, **a**, **u** is obviously beneficial. It would be possible for languages to add just one vowel to these basic three, and, indeed, some languages do. But it turns out that far more languages have three, five, or seven vowels than have two, four, or six vowels. There are plenty of exceptions to this generalization. Many Italian dialects have six vowel systems. Standard Italian has seven vowels, including two types of **e** rather like the French *é* and *è*, and two types of **o**, but some dialects do not distinguish the two types of **e** in all circumstances, and some do not distinguish the two types of **o**. It seems likely that in years to come most forms of Italian will have symmetrical five or seven vowel system.

We have been describing vowels as if they were distinguished by only two formants, but actually the situation is more complicated. There is a third formant that is important for distinguishing some sounds, notably the r-colored vowel that occurs in American English pronunciations of words such as 'bird', and the French vowel that occurs in *tu* (you). There are also formants with even higher pitches that add to the overall vowel quality. We can see the more complete set of formants that occur by making a computer analysis of the sound waves in a set of words. The top part of Figure 3 shows the sound wave of the words 'bead, bid, bed, bad'; below it is a computer analysis showing the component frequencies in the form of a sound spectrogram. Time runs from left to right, as for the sound wave. The frequency scale (shown on the left) goes up to 4,000 Hz. The dark bands with white lines through them are the formants, with the degree of loudness (the intensity) of each formant being shown by the darkness of the band.

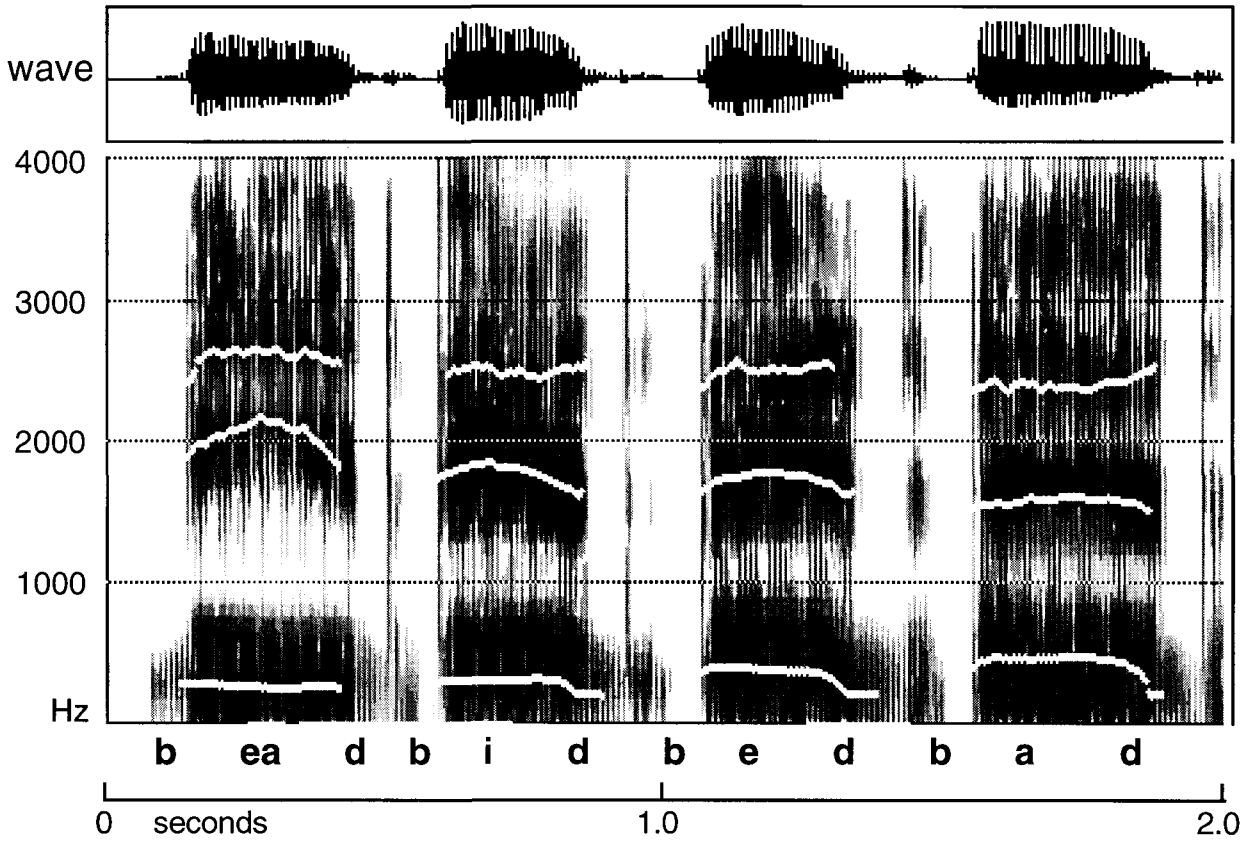


Figure 3. The upper part of the figure shows the sound waves produced when the author said the words "bead, bid, bead, bad." The lower part is a spectrogram of these sound waves in which the complex sound waves are split into their component frequencies (overtone pitches), the intensity (loudness) of each frequency being shown by the darkness. The three principal groups of overtones (the first three formants) are marked by white lines, labeled F1, F2 and F3.

In this spectrogram. the formants are far from straight lines. But if we consider just the centres of each of the syllables, you can see that the first formant frequency at that point gets steadily higher, just as you can hear it does when you say these words with a creaky voice. The second formant frequency in the middle of the vowel goes steadily down, as it does when you whisper them. The third formant also moves slightly down. If I had said just the vowels in these vowels, the formants would have formed more or less steady black bars (with the white lines I have drawn in the centre of them). The extensive movements, particularly of the second formant, are due to the consonants.

Consonants

Many consonants can be described in terms of the movements of the formant frequencies of the vowels. Consonants such as **b**, **d**, **g** are really just ways of beginning or ending vowels, as can be seen in Figure 3. When forming **b**, the lips are closed. As they open at the beginning of each of these words, the formants rapidly increase in frequency, as is particularly evident in the first three words. It is this movement that characterizes a **b** sound. At the end of each of these words the **d** is characterized by a downward movement of the first two formants, but an upward movement of the third formant, which is clearest in the third word.

Some consonants are produced without vibrations of the vocal folds. The spectrogram in Figure 4 shows the waveform and spectrogram of the English words *sin, shin, thin, fin*, each of which begins with a so-called voiceless consonant. In these consonants the sound is produced by air being forced through a narrow gap so that it becomes turbulent. The spectrogram shows that the hissing sound of **s** at the top left of the figure shows that it has random energy throughout a wide range of high frequencies. The **sh** sound has energy that is centred at a slightly lower frequency. If you make a long **s** you can hear that it sounds higher pitched than **sh**. The other two consonants, **f** and **th**, are less loud, and have less well defined frequency characteristics. The consonant, **n**, at the end of each of these words can be characterized in terms of the intensity (loudness) and frequency (pitch) of three formants, much in the same way as vowels. This sound is made by blocking the air from coming out of the mouth and allowing it to come out through the nose. There are two other nasal consonants in English, **m** and **ng**, as at the ends of the words *ram* and *rang*.

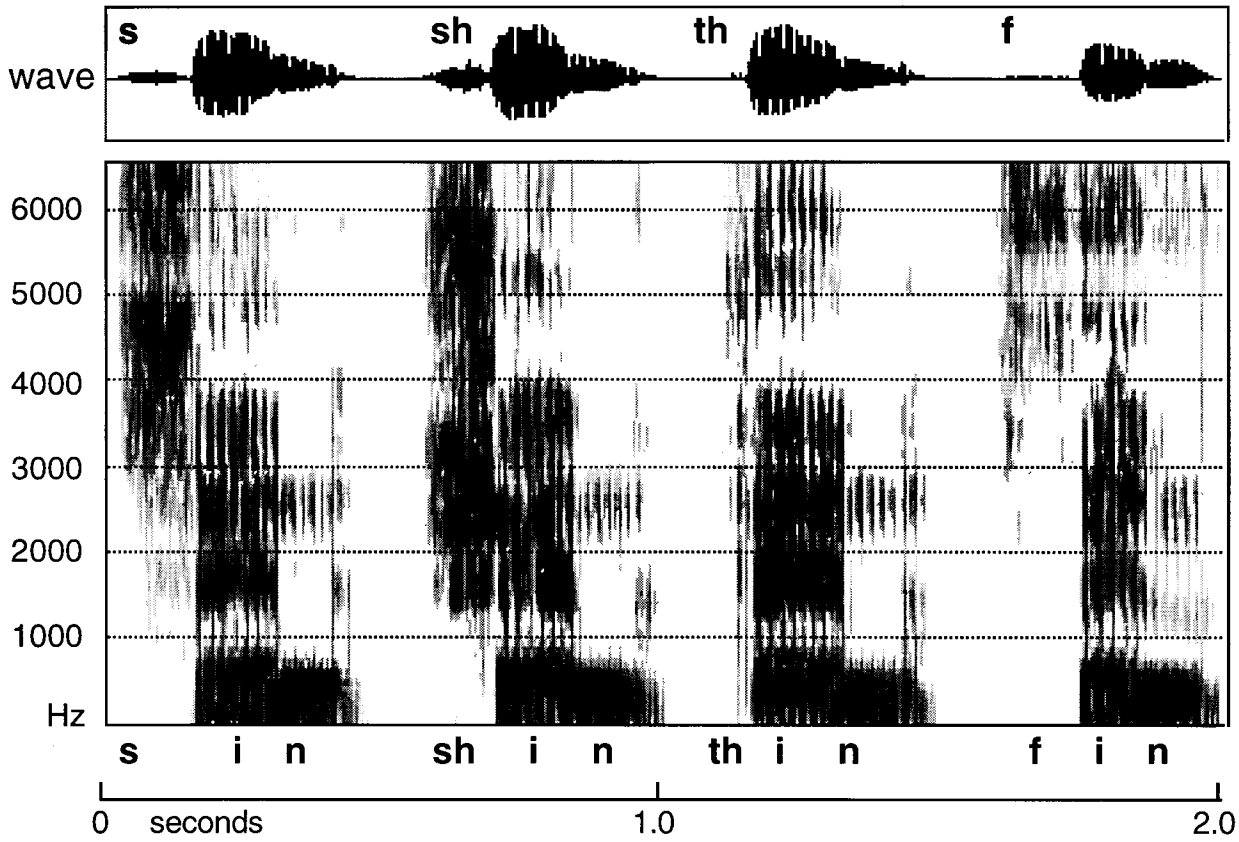


Figure 4. The upper part of the figure shows the sound waves produced when the author said the words *sin, shin, thin, fin* The lower part is a spectrogram of these sound waves. Note that the frequency scale extends higher in this figure than in the previous figure.

Acoustic parameters

We cannot consider the acoustic structure of all the consonants of English in this chapter, but they can in fact be fairly well described in terms of a limited number of parameters. Table 4 summarizes the eight most important acoustic parameters of speech. We can characterize nearly all speech sounds in terms of the values of these parameters.

Table 4. The major acoustic parameters of speech and their auditory correlates.

ACOUSTIC PARAMETER	AUDITORY CORRELATE
Frequency of first formant	Pitch of first group of overtones
Frequency of second formant	Pitch of second group of overtones
Frequency of third formant	Pitch of third group of overtones
Amplitude of first formant	Loudness of first group of overtones
Amplitude of second formant	Loudness of second group of overtones
Amplitude of third formant	Loudness of third group of overtones
Centre frequency of the semi-random noise	Pitch of the voiceless components
Intensity of the semi-random noise	Loudness of the voiceless components

We can demonstrate the validity of these parameters by considering speech that has been synthesized by using variations in these, and only these, possibilities. The upper part of Figure 5 shows a spectrogram of the sentence *A bird in the hand is worth two in the bush* as produced by a speech synthesizer, originally by the British communications engineer John Holmes over 30 years ago. The lower part of the figure show the same sentence produced by me. These two sentences sound very much alike. There are differences in that, for example, I have a very short version of the first word, *A*, and no vowel at all in the word *in*. But the similarities far outweigh the differences.

Some sounds of the world's languages

As we have seen, many languages have five vowels fairly evenly distributed in the possible vowel space. There are also some well-favored consonants whose first merit is that they are as different as possible from vowels. Vowels are produced with very little obstruction of the vocal tract, and with vibrating vocal folds. In the best consonants the vocal tract is completely obstructed and the vocal folds are not vibrating. They are the sounds (or, more accurately, the silences) known as voiceless stops, the most common of them being **p**, **t**, **k**. About 98% of the world's languages have sounds that are something like these three sounds (although not necessarily exactly as the English versions), and the remaining 2% have sounds similar to two of the three.

The sounds **p**, **t**, **k** are very distinct from vowels, but they are less well distinguished from each other. If you say each of these sounds by itself, without a vowel after it, you will be able to hear slight differences in the pitch and loudness of the burst of noise that occurs (the last two parameters listed in Table 4). The differences arise from each of these stops being made at a different place within the mouth. There are also small differences in the movements of the formants in the adjacent vowels, similar to those we discussed for **b**, **d**, **g**. It is possible to make consonants in which the airstream is blocked at several different places in the mouth. Some languages, such as Malayalam, a Dravidian language spoken in India, have voiceless stops made by stopping the air at six different places within the mouth. But the necessity for auditory distinctiveness without having too great an articulatory complexity forces restrictions on the consonant space. Most languages find that it becomes too crowded when there are stops made at more than three or four places in the mouth. Our third constraint on the ways languages develop, the pressure for gestural patterning, is also very evident in the formation of consonant systems. Malayalam, for instance, has not only six voiceless stops, but also six voiced stops and six nasal consonants, all made with very similar articulatory gestures.

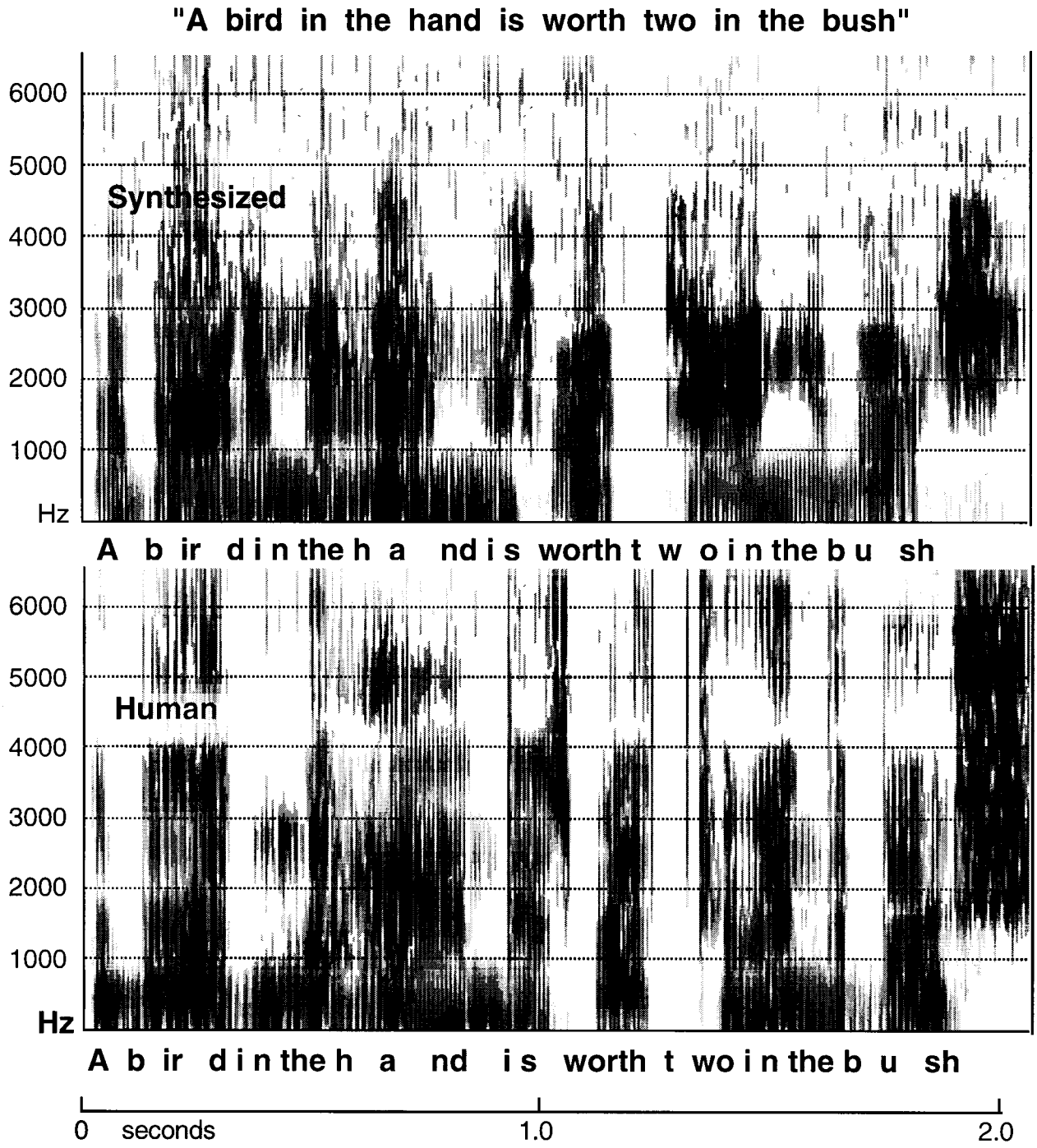


Figure 5. A spectrogram of the sentence *A bird in the hand is worth two in the bush* as produced by a speech synthesizer (upper part of the figure), and the same sentence produced by the author (part of the figure).

Sounds in which there is a there is a turbulent airstream, such as the English **s**, **sh**, **f**, **th**, are less common than stop consonants. Some languages, such as Hawaiian, have no sounds of this type. But others, such as Polish and Standard Chinese, have additional possibilities. The best sounds of this type in evolutionary terms — those that produce the loudest and most distinctive

sounds for the least effort — are the sibilants like English **s** and **sh**. Polish has not only two sounds rather like these English sounds, but also a third possibility, made with the tongue a little further back in the mouth. Table 5 shows words illustrating the six sibilants of Polish, three voiced and three voiceless. These words are given in the Polish orthography, and also in a phonetic transcription using the symbols of the International Phonetic Alphabet (the IPA). The IPA is an internationally agreed set of symbols that can be used for transcribing all the contrastive sounds that occur in the world's languages.

Table 5. Words illustrating the Polish sibilant sounds between two vowels. The columns on the left illustrate voiceless sibilants, those on the right illustrate voiced sibilants.

ORTHOGRAPHY	IPA	ENGLISH	ORTHOGRAPHY	IPA	ENGLISH
kosa	**kosa**	scythe	koza	**koza**	goat
kasza	**kaʂa**	kasha, groats	gaza	**gaʐa**	gauze
Basia	**baɕa**	Barbara (dim.)	bazia	**baʑa**	catkin

The Polish sibilants shown in Table 5 cannot be adequately described in terms of the parameters in Table 4, which listed only the centre frequency and the intensity (the pitch and the loudness) of the noise as characterizing voiceless sounds. To describe these sounds accurately we need to include more complex properties of the spectra of the noise. This becomes even more apparent when we try to describe sounds in some of the less well known languages such as Toda, a language spoken in the Nilgiri Hills in the south of India by a few hundred people. Toda has four voiceless sibilants — one more than Polish — all of which can be used to distinguish words. The four sibilants that can occur at the ends of words in Toda are shown in Table 6. Toda has never been written in the roman alphabet, so the words are given just in IPA symbols.

Table 6. Words illustrating the sibilant sounds in final position in Toda, a Dravidian language spoken in the south of India.

IPA	ENGLISH
koʂ̟	money
pɔs	milk
pɔʃ	language
pɔʂ	name of a clan

There are hundreds more consonants in the world's languages than can be described in this chapter. We will conclude by commenting on a few of the more unusual sounds, and noting why they might be less common. So far we have considered sounds in which the vocal folds are either vibrating (as in voiced sounds) or not (as in voiceless sounds). There are, however, intermediate possibilities. The few hundred speakers of Mpi, a language spoken in Northern Thailand, produce one set of vowels in which the vocal folds are vibrating somewhat laxly so that the voice is slightly breathy, and another set in which there is more tension in the back of the throat, so that the voice has a harsher, slightly creaky, quality. As Mpi also distinguishes words by tones, on similar lines to Cantonese which we discussed earlier, alphabetic segments such as **si** may have 12 different meanings as shown in Table 7. We noted earlier that most of the world's languages use differences in pitch to distinguish words. The subtler adjustments of the vocal folds required for breathy voice and creaky voice need greater articulatory precision than that required for regular voiced sounds. They are also hard for listener's to distinguish, particularly as a breathy or creaky voice quality may be simply the person's normal way of speaking. As a result these voice qualities are used by a comparatively small number of languages. But they often characterize individual

voices. Most of us can think of individuals who have something like a Marilyn Monroe breathy voice quality or a Louis Armstrong creaky voice quality.

The vocal folds can be used to produce not only differences in pitch and voice quality, but also sounds of a very different type. If they are held tightly together and then moved upwards, the air above them will be pushed out of the mouth. but if the lips are closed, or the tongue is making a closure against the roof of the mouth, the air in the mouth will be compressed. When the closure is released there will be a popping sound as the compressed air rushes out. Sounds of this type, which are called ejectives, occur in about 10% of the world's languages, including many American Indian languages. Examples from Quechua, the native language of the peoples who now live in the Inca regions of Peru and Bolivia, are given in Table 7 in an anglicized orthography. An apostrophe after a letter indicates that the pressure for the sound was produced by the tightly closed vocal folds being raised upward to form an ejective. Although these sounds are auditorily very clear and distinct, the articulatory mechanism is more complex than simply using outgoing air from the lungs to produce pressure. The increased complexity accounts for ejectives being comparatively rare. Note, however, that once a language has some ejectives, the principle of gestural patterning often results in there being aspirated and ejective sounds made with the same articulatory gestures as those used in other sounds in the language.

Table 7. Some Quechua words in a partially anglicized orthography. The apostrophe indicates that the preceding sound is an ejective, produced with air being pushed out by the closed glottis. The raised ^{h} indicates aspiration — an extra puff of air after the sound. The **ch** sound is similar to *ch* in English (but not exactly the same). The **q** corresponds to a sound made with a closure at the back of the mouth, near the uvular.

PLAIN		ASPIRATED		EJECTIVE	
chaka	'bridge'	**ch^{h}aka**	'large ant'	**ch'aka**	'hoarse'
kuyui	'to move'	**k^{h}uyui**	'to whistle'	**k'uyui**	'to twist
qalyu	'tongue'	**q^{h}alyu**	'shawl'	**q'alyu**	'tomato sauce

Clicks

Probably the most striking unusual sounds found in the world's languages are the clicks that occur in some of the languages spoken in Africa — and nowhere else (as part of a regular language). It seems likely that clicking sounds first developed among the ancestors of the people we now call the Bushmen. From there they spread to other tribes such as the Nama of Namibia. A few hundred years ago the Zulus, Xhosa and other Bantu tribes swept southward from central Africa and conquered the Nama and the Bushmen, taking them as wives and servants. They also took their click sounds into their languages.

Click sounds are used by many speakers of English, but not as part of the regular language. The sound that novelists write as *tsk, tsk* is used to express disapproval. Linguists call this a dental click, as it is made with the blade of the tongue touching the upper front teeth. Some people use a clucking sound as a sign of approval or encouragement. This is a lateral click, as air comes in at the side of the mouth. Clicks can also be made with the tip of the tongue curled up so that it touches the roof of the mouth behind the upper front teeth (a so-called alveolar click), and with the body of the tongue raised up against the hard palate in the centre of the mouth (a palatal click). Note that in all clicks the air comes *into* the mouth, as opposed to going out of the mouth as in almost all other sounds. This ingressive airstream is the major characteristic of a click.

Table 8 illustrates the 20 clicks of Nama, arranged in four columns and five rows, labeled with the technical terms that linguists use for describing these sounds. We cannot here give detailed explanations of all these terms; many of them are fairly self-evident from what has been said earlier. The symbols are those of the International Phonetic Alphabet. The distinctions between the columns (the different types of clicks) are fairly easy to hear, but some of the distinctions between rows (the different click accompaniments, such as aspiration) are fairly subtle. As always in a table with completely filled in rows and columns as here, we can see the pressure of gestural patterning being exerted on the sound system of a language.

Table 8. Words illustrating contrasting clicks in Nama. All these words have a high tone.

	DENTAL	ALVEOLAR	PALATAL	LATERAL
VOICELESS UNASPIRATED	**k|oa** 'put into'	**k!oas** 'hollow'	**k‡ais** 'calling'	**k‖aros** 'writing'
VOICELESS ASPIRATED	**k|^{h}o** 'play music'	**k!^{h}oas** 'belt'	**k‡^{h}aris** 'small one'	**k‖^{h}aos** 'strike'
VOICELESS NASAL	**ŋ|^{h}o** 'push into'	**ŋ!^{h}as** 'narrating'	**ŋ‡^{h}ais** 'baboon's arse'	**ŋ‖^{h}aos** 'special cooking place'
VOICED NASAL	**ŋ|o** 'measure'	**ŋ!oras** 'pluck maize'	**ŋ‡ais** 'turtledove'	**ŋ‖aes** 'pointing'
GLOTTAL CLOSURE	**k|ʔoa** 'sound'	**k!ʔoas** 'meeting'	**k‡ʔais** 'gold'	**k‖ʔaos** 'reject a present'

Clicks provide many puzzles for those interested in the development of language systems. Because they involve a sucking gesture, they seem difficult to integrate into the stream of speech. But this may not be true, as they have often been borrowed from one language into another. They were probably borrowed from the Bushman languages into languages like Nama. We know for certain that clicks are borrowed sounds in Zulu and other Bantu languages, as these languages did not have any clicks a few hundred years ago. But clicks are now very much part of their regular sound systems, appearing in print with the letter *c* for the dental click (as in the name of the warrior chief *Cetewayo*), the letter *x* for the lateral click (as in the name of the language *Xhosa*), and the letter *q* for a click in which the tongue tip curls up as it makes contact with the roof of the mouth. As a teacher of phonetics for many years, I have found it quite easy to get people to say words with clicks in them. Students find it much easier than, for example, learning to produce a trilled *r* sound, or to produce some of the sequences of consonants that occur in Polish. Clicks are also auditorily very distinct from other sounds, and, although the different accompaniments (the rows in Table 8) are hard to distinguish, there seems no obvious reason why most languages should not have at least two or three different types of clicks. Perhaps languages will evolve in this way. The Bushmen got there ahead of us and may have the most evolved language; but may be in two or three thousand years time most languages will have a few clicks among their consonants. And maybe, to end this story, when the languages of the world have fully evolved, we will all live happily ever after.

Phonetics

Patricia A. Keating

This is a draft of the phonetics chapter for a forthcoming introductory linguistics textbook edited by V. Fromkin. The chapters of the book are being written by different members of the UCLA Linguistics Department. In the published book, chapters will not be attributed to individual authors, but rather to the group as a whole. Two other notes about the organization of the textbook: first, the phonetics-phonology section is not at the beginning of the book but instead follows the morphology, syntax, and semantics sections, so that many basic linguistic ideas will be already known to the student; and second, all example sentences and words throughout the book are supposed to relate to Shakespearean themes and characters.

Table of Contents

A. Introduction
B. Tools for phonetic description
B1. Speech utterances: segments and suprasegmentals
B1.1. Segments
B1.2. Suprasegmentals
B2. Phonetic alphabets
B2.1. Problems with orthography
B2.2. The IPA and other phonetic alphabets
B3. Basic IPA symbols for American English
B4. Articulatory definitions of symbols
B4.1. The IPA chart
B4.2. Articulators
B4.3. Consonants
B4.3.1. Columns: Place of articulation
B4.3.2. Rows: Manner of articulation
B4.3.3. Voicing; other consonants
B4.4. Vowels
B4.4.1. The vowel chart
B4.4.2. Diphthongs
B4.5. Different pronunciations mean different symbols
C. Optional: Other IPA symbols
C1. Consonants and vowels
C2. Diacritics
D. Conclusion

A. Introduction

Linguistic structure is conveyed to a listener by speech. (Although linguistic structure can be conveyed to a viewer by sign or by writing, in this chapter we will consider only oral communication.) **Phonetics** is the study of the physical aspects of speech events, including: **speech production** (how speech is produced by the speaker, an instance of skilled motor performance), **speech acoustics** (the properties of the airwaves that transmit speech from speaker to listener), and **speech perception** (how speech is perceived by the listener). Phonetics is part of linguistics, but it is part of other disciplines as well, such as speech and hearing science, psychology, and engineering. **Linguistic phonetics** is a term sometimes used to describe the aspects of speech articulation, acoustics, and perception that are of interest to linguists. It includes the study of the speech sounds of a range of languages, generalizations about sounds that hold across languages, and the study of the relation of phonetics to other areas of linguistics.

In the next section we provide some concepts and terms that let us describe the sounds of speech.

B. Tools for phonetic description

B.1. Speech utterances: segments and suprasegmentals

B1.1. Segments

We will describe a speech utterance as first being composed of a sequence of individual **speech sounds**. Speech sounds, also called **segments** or **phones**, are sounds used in languages. As such they exclude various noises humans can make that are not used in languages, including sounds made with the hands, or sounds made by inhaling. Speech sounds are generally divided into two types, consonants (abbreviated C) and vowels (abbreviated V). **Consonants** are sounds in which a significant constriction is made somewhere in the vocal tract--a narrowing that interferes with the flow of air out of the mouth--so that there is at least some reduction in the energy of the sound. **Vowels** are sounds in which no such constriction is made; the air flows out of the mouth relatively freely and the sound is loud and strong. Consider the word "Macbeth", which has the shape CVCCVC. Can you feel the constriction for each of the four consonants? (They are: at the lips, at the back of the mouth, at the lips again, on the teeth.) All languages have consonants and vowels. This is a point on which students often confuse writing and speech. Even languages which do not use alphabetic writing have consonant and vowel sounds. However, languages do differ in *which* consonants and *which* vowels they have.

The idea that English utterances can be divided into a succession of segments, one consonant or vowel after another, can seem completely obvious to literate speakers of the language --most people feel they hear speech this way. They also think that different instances of a given consonant or vowel are the same-- that every "b" is alike, for example. But these are psychological idealizations from the physical speech signal, in two ways. First, if we look at actual speech, it is not so obvious that there are

separate sounds in succession, because there are not always sharp boundaries between them. Figure 1 shows a **spectrogram** of the sentence "Tell me where is fancy bred". A spectrogram is a frequency by time display in which the stronger frequency components are highlighted. Abrupt changes in these components give clear visual and auditory boundaries between speech segments. The word "fancy" shows fairly abrupt boundaries between its segments, but the other words do not do so in every case. The end of "tell" and the beginning of "me" is hard to discern, as is the end of "where" and the beginning of "is", as well as all the segments inside "where". Most intervals of speech contain information about two or even more speech segments because adjacent sounds overlap. Much of the signal shows changing transitions between segments, where information from both is present. This is particularly true of "where" and "bred", because of the r sounds. That is, while the ordering or sequencing of sounds is usually clearly supported by the speech signal, the feeling that we could make a clean slice between each pair of segments is an idealization from the signal.

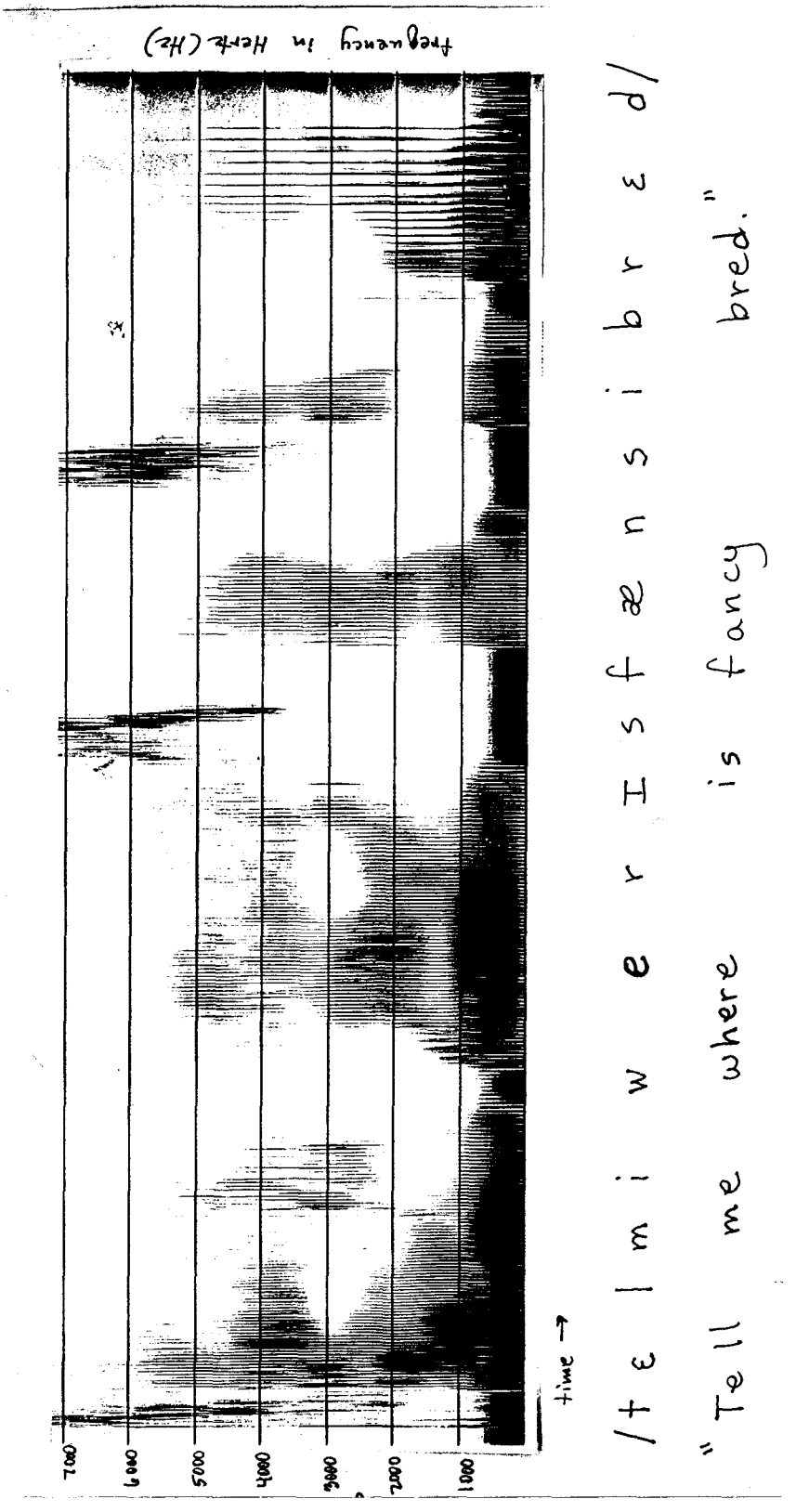
FIGURE 1 (spectrogram with orthographic transcription) on next page

SEGMENTATION AND LEARNING TO READ AN ALPHABET

Alphabets such as the one used for English are based on segmentation into individual sounds. It makes sense, then, that it should be easier for someone who already has psychological representations of words as composed of segments to learn alphabetic reading and writing. Reading experts talk about "phonological (or phonemic) awareness" as a prerequisite to reading--when a child can focus on individual sounds in a word, that child is better able to learn the relations between sounds and letters (or "phonics"). Otherwise, the child has to learn whole-word patterns of written and spoken words. Some examples of phonological awareness include the ability to say if two words rhyme or if they begin with the same sound, or (a harder task) to say how many sounds a word contains, or what sound comes after another sound in a word. (You will notice that these skills are relevant in the exercises below.) At the same time, someone who does not already have segmental representations of words is likely to acquire them as a result of learning to read. It seems that phonological awareness and reading ability feed each other as the child learns to read, and success in one predicts success in the other. (Reference: Gough, Ehri, and Treiman 1992.)

Second, if we look at speech signals, we see that it is not true that all "b"s, or all of any other sound, are the same. It is not the case that a language uses a small number of speech sounds over and over, always exactly the same each time. There are many small differences between different instances of what we think of as the "same" sound, not only across languages and across speakers of the same language, but also within the speech of any one speaker. In Figure 1, the two r sounds in "where" and "bred", and the vowels of "tell" and "bred", for example, look different. These kinds

Figure 1. Wideband spectrogram, orthography and broad transcription of "Tell me where is fancy bred."



of differences are completely normal and we see similar things across languages. But adults do not hear these small differences, or do so only with training, so that our normal perception is different from what we can measure.

EXERCISE 1: For each set of three words, which one begins with a different speech sound? Consider only the *first* sound in each word.

Example: **scale - state - shall** -- "shall" begins with a different sound

1. **countenance - king - cheer**
2. **sister - she - cease**
3. **equal - eyes - even**
4. **again - opponent - all**
5. **throne - thy - these**
6. **character - chaste - coldly**
7. **heart - where - who**
8. **jelly - giving - gentlemen**
9. **admiration - against - appears**
10. **every - each - else**

EXERCISE 2: How many speech sounds does each of the following English words contain? For each sound, say whether it is a consonant (C) or a vowel (V).

Example: **still** = 4 sounds, CCVC (there is only one "l" *sound* even though there are two letters)

1. **yet**
2. **seems**
3. **boot**
4. **have**
5. **privy**
6. **walks**
7. **dumb**
8. **theme**
9. **health**
10. **grizzly**

B.1.2. Suprasegmentals

In addition to the individual consonant and vowel qualities (segments, or segmentals), there are larger-scale properties of utterances which are usually referred to as **suprasegmentals**, or **prosody**. "Suprasegmental" means "above the segment", i.e. suprasegmental properties are most apparent over a string of segments. These properties include variations in loudness, duration and pitch, as well as variation in the degree of energy or effort put into the articulation of each sound. These generally function to make some element more prominent than others. At the lowest level, vowels are generally more prominent than consonants. A **syllable** is a string of segments bound together by the fact that one of the segments (usually a vowel) is more prominent than the others. Roughly, each vowel is the head of a syllable, and adjacent consonants (if any) belong to the syllable along with the vowel. For example, the word *cat* has one syllable, *adult* has two, and *oasis* has three. If there is no vowel, a consonant may be prominent, as in the second syllable of words like *little, children, button*, and *paper* in many dialects. Although these words are spelled with vowel letters in the second syllable ("e" or "o"), these letters are usually not pronounced.

Syllables in turn can vary in prominence. Variations in loudness, duration, and effort together produce differences in **stress**: one syllable appears stronger than others in the same word. In some languages such stress differences can distinguish one word from another. For example, in English, the noun "(a) convict" is stressed on the first syllable, while the verb "(to) convict" is stressed on the second syllable. Here the stress is indicated by underlining. Here are some words, all with three syllables, but with stress on either the first, second, or third syllable:

(1)

first syllable stress	second syllable stress	third syllable stress
vanquisher	discretion	incorrect
funeral	advantage	overwhelm
impotent	remembrance	entertain

Variations in duration, and to some extent loudness, also produce differences in rhythm. Languages sound different one from another in part because of their characteristic rhythms. English rhythm allowed Shakespeare to use a pattern called iambic: *weak-strong-weak-strong-weak-strong* etc.. Variations in the pitch of the voice give rise to an overall melody for an utterance. In some languages, the pitch of the voice is specified for each vowel in each word; a **tone** is such a pitch specification. In a tone language, a sequence of consonants and vowels will have different meanings depending on the pitch(es) of the voice used to speak that sequence. English is not a tone language, but most languages of Africa and Southeast Asia, and some native American languages, are tone languages. For example, in Kana, a language spoken in Nigeria, "be" (where "e" is pronounced [e], which will be defined later) with a Low tone means "to fence", "be" with a Mid tone means "home or compound", and "be" with a High tone means "fight". Finally, in all languages, tone or not, pitch of the voice is also used to convey things about whole utterances, and this is called

intonation. In English, some kinds of questions are characterized by a rising melody at the end: "Is Hamlet upset?", and "Still here, Laertes?", both usually with a rising melody on the last word, as compared to "What did Polonius say?" and "I shall obey you, madam.", both with a falling melody near the end (e.g. at the end of "Polonius" and "obey"). More generally, however, the amount and kind of variation in the overall prosody of an utterance conveys information about the speaker's attitude towards the utterance. Consider, for example, "To be, or not to be; that is the question:". It can have a flat or an animated intonational melody; some parts can be stretched out or speeded up. Some segments can be more forcefully articulated. All of these suprasegmentals help render an actor's interpretation of this utterance.

EXERCISE 3: For each set of three words, one of them has the main stress on a different syllable than the other two. Mark that word. If you are not sure where the main stress in a word is, look it up in a dictionary.

Example: **question - scholar - tonight** -- mark "tonight"

1. **expressed - surprised - triumph**
2. **luxury - malicious - ministers**
3. **porcupine - secrecy - illusion**
4. **possess - answer - gracious**
5. **extravagant - revolution - disposition**

EXERCISE 4: Underline the vowel in the stressed syllable in each word. (If two letters spell the vowel sound, underline both of them.)

Example: **foolishly** = f<u>oo</u>lishly

1. **extravagant**
2. **intermission**
3. **encounter**
4. **hospitality**
5. **unworthy**
6. **reputation**
7. **unwillingly**
8. **childishness**
9. **philosophy**
10. **messengers**

B2. Phonetic alphabets

In what follows, we will first consider how to represent what we will call the "basic" sounds of a language. By basic sounds we mean the minimum number of sounds needed to represent each word in a language differently from all other words, in a way that corresponds to what native speakers think are the same sounds in different words. That means ignoring differences between two sounds that do not distinguish different words and which native speakers are unaware of. Only after we have established some symbols for transcribing these basic sounds, will we go on to consider how finer details, and suprasegmentals, can be transcribed.

B.2.1. Problems with orthography

Phonetic alphabets are sets of symbols used for representing the speech sounds that occur in utterances. The fundamental principle of most phonetic alphabets is that *each symbol should represent only one sound, and each sound should be represented by only one symbol.* Depending on what we mean by "one sound", most standard orthographies of the world's languages violate this principle in one or both directions. English is bound to violate it, since it uses only 26 letters to represent some 38 basic sounds (which will be presented in section B3 and again in the next chapter). So some letters do double-duty in English spelling, like "y" and all of the vowel letters; and combinations of letters are used for single sounds, like "sh", the first sound in "ship". Violating the principle in both directions at once, the combination "th" uses two letters to spell two distinct single sounds, the first sounds in "thin" and "this".

A common kind of violation occurs when the orthography was standardized some time ago, and pronunciations of some words have changed in the meantime. For example, there are many pairs or sets of words in English that are spelled differently but pronounced the same, such as "rite"/"right"/"write"/"wright". These words used to be pronounced differently, in accord with their spelling differences. However, as the language changed they came to be pronounced the same, so that now we have four ways to spell the same sound sequence. For a standard orthography to keep one symbol corresponding to one sound, spellings would have to be updated as the sounds of the language changed. Also, violations occur because different speakers of a language have different pronunciations. No standard orthography can keep one symbol equal to one sound for all speakers of a language; by definition its goal is to make the language readable and writable by all its speakers. For example, there are many word pairs/sets in English that are pronounced the same by some speakers, but distinguished by other speakers, as in (2). For these words, the spelling violates the one-symbol/ one-sound principle for some speakers but not for others. Read these and decide whether you pronounce them the same or not.

(2) Example words sometimes pronounced the same:

1.	witch	which	
2.	horse	hoarse	
3.	morning	mourning	
4.	sot	sought	
5.	bawdy	body	
6.	father	farther	
7.	Mary	merry	marry
8.	poor	pour	pore

More fundamentally, all orthographies violate the principle to the extent that orthographies tend to represent only the basic sounds of the language, not the variants that they may show in particular combinations or positions. The spelling of a word, by itself, does not really tell you how to pronounce that word. You need the native speaker's knowledge of the language to do that. (This is a property that spellings share with the pronunciations of words indicated in basic sounds in dictionaries.) For example, many English speakers pronounce sequences of "tr" (as in "train"), "tw" (as in "twin"), and "t-y" (as in "got you") as if they contain a sound sequence like "chr", "chw", or "ch", a detail of pronunciation that could not be guessed from the standard spelling. (Though it could be guessed from informal spellings like "gotcha".)

In sum, orthographic transcriptions of words do not unambiguously represent every aspect of pronunciation that we might want to represent. Phonetic alphabets are distinct from the orthographic system used for any language. In the best cases, phonetic alphabets are capable of representing different pronunciations for a single word -- how a word is usually pronounced, or how some particular speaker pronounced it on some particular occasion. A phonetic transcription can indicate enough about a pronunciation that someone who does not know the language being transcribed can nonetheless read the transcription and pronounce it fairly accurately.

EXERCISE 5. Each of the following English words contains two instances of the letter "s" or the letter "c" . The letter "s" can spell the sound of either "s" or "z". The letter "c" can spell the sound of either "s" or "k". For each word, decide whether the two letters are spelling the same sound, or two different sounds. .

Example: **season** the two "s"s are pronounced differently, as "s" then "z"

1. **Francisco**
2. **pastors**
3. **resolves**
4. **sometimes**
5. **surprised**

6. **disposition**
7. **wisdoms**
8. **secrecy**
9. **spirits**
10. **sensible**

EXERCISE 6. Each of the following words contains a silent letter. If this letter were removed from the spelling, the spelling would still represent how the word is pronounced. Pronounce each word, decide which letter is not sounded, and circle it.

Example: **answer**: "w" is silent ("anser" would still be a possible spelling for this word)

1. **guard**
2. **designed**
3. **black**
4. **witch**
5. **wrung**

B.2.2. The IPA and other phonetic alphabets

Over time and in different countries, many phonetic alphabets have been devised. The one with the most widespread acceptance is the alphabet of the International Phonetic Association (or IPA). This alphabet is called the International Phonetic Alphabet (also abbreviated IPA). Unlike most other phonetic alphabets, the IPA is an attempt to provide a symbol for every sound of every language. Another difference between the IPA and other alphabets is that the IPA comes from an organization whose members and Council discuss and vote on changes to it. Thus in 1989 an international convention met in Kiel, Germany to update the IPA, resulting in many recent revisions.

The advantage of the IPA is that because it is widely-studied and used, transcriptions using it can be interpreted by many readers. Therefore we use the IPA throughout this text. Nonetheless, it must be stressed that other systems are well-suited for other, more-limited purposes. Most linguists who work on particular languages have devised their own systems for those languages, systems which may or may not correspond closely to the IPA. American linguists in particular generally use a few phonetic symbols which are not part of the IPA, so you are likely to see these in other textbooks. This chapter will present the IPA, but your teacher may choose to depart from it. (Reference on phonetic symbols used in a variety of traditions, including pre-

1989 IPA: G.K. Pullum and W.A. Ladusaw (1986).)

Finally, it is important to note that even the IPA symbol set can be used in various ways for a given language, in two senses. First, two linguists may disagree about a sound quality they are trying to represent, so that each of them would chose a different IPA symbol for that sound. Second, the IPA explicitly allows the substitution of simpler symbols for more complex ones in a particular language if no confusion would result. We will take advantage of this provision with two symbols needed for English: simple [a] instead of [ɑ], and simple [r] instead of [ɹ].

B3. Basic IPA symbols for American English

(3) gives a minimum set of symbols sufficient for distinctively representing the consonants and vowels of many speakers of American English. Strictly speaking, #34 [ə] is not needed for this purpose, but we include it because it is customary and convenient to do so, since many speakers feel that it is a basic sound of English. On the other hand, unlike other texts we do not include a vowel [ɔ], since this vowel is being rapidly lost across the US. This and other vowels found in other dialects of English may be discussed by your instructor.

(3) Basic sounds of English using a minimal symbol set. Unless otherwise indicated by underlining, sound occurs at both beginning and end of word. All of these words appear in Shakespeare's plays.

	phonetic symbol	word illustrating it
1	p	pope
2	b	barber
3	m	mum
4	f	fife
5	v	vive
6	t	taunt
7	d	deed
8	n	nun
9	r	rare
10	θ	thousandth
11	ð	this, breathe
12	s	source
13	z	zanies
14	ʃ	shush
15	ʒ	measure
16	l	lull
17	tʃ	church
18	dʒ	judge
19	j	yoke

20	k	cook
21	g	gag
22	ŋ	singing
23	w	**we**
24	h	**he**
25	i	easy
26	ɪ	imitate
27	e	able
28	ɛ	edge
29	æ	battle, attack
30	a	father
31	o	road
32	ʊ	book, should
33	u	food
34	ə	aroma
35	ʌ	but
36	aɪ	**ride**
37	aʊ	house
38	oɪ	**boy**

Many of these symbols and their values are familiar from English orthography, but many are not. Note especially that #19, 25, 30, and 33 do NOT represent their most common English values, although they do in some words borrowed into English or in proper names.

With just these symbols, some notion of careful pronunciations of many words can be adequately conveyed. However, as noted above, English words can differ in stress. Therefore, to be able to distinguish more pairs of English words -- to be able to give words unique transcriptions -- we need to transcribe stress as well as consonants and vowels. Although different degrees of stress may be distinguished in some transcriptions, we will note only the most strongly stressed vowel in a word. This is called the **main stress** or **primary stress**. Following IPA usage, we will mark stress by a raised vertical tick before the stressed vowel (or before any preceding consonants): ['kanvɪkt], the noun form of "convict", but common American practice is to instead use an accent mark on the vowel: [ḱánvɪkt]. Stress will be shown in transcriptions like those above only where needed to distinguish two words, which is never the case in (3).

A transcription which uses only the minimal set of basic symbols can be said to be a **broad transcription**, or a **phonemic transcription**. (For our purposes we will use these terms interchangeably.) The pronunciations given in dictionaries are broad transcriptions. (Some dictionaries use IPA, some do not.) Broad transcriptions are often enclosed in slant brackets, e.g. /a/. When a transcription goes beyond this to indicate details of pronunciation, it can be said to be a **narrow (or narrower)**

transcription. The difference between a broad and a narrow transcription is one of degree: the more detail included, the narrower the transcription. Therefore, any word can be transcribed in more than one way. Narrow transcriptions are usually enclosed in square brackets, e.g. [a]. However, broader transcriptions are also sometimes enclosed in square brackets, to make clear that a transcription is not maximally broad or phonemic. (See also the next chapter on this point.) Note that a broad transcription is likely to be valid for more speakers on more occasions than a narrow transcription will be, since the broad transcription gives relatively less information about an exact pronunciation.

In a broad transcription, the set of allowed symbols is strictly limited -- e.g., those in (3). Words are transcribed by combining these, and only these, symbols. Sometimes, though, certain sound combinations sound different from any allowed symbol combination, so that students (and professional linguists!) may be unsure which symbols to use. For example, vowels before [r] can be quite variable in quality, and speakers may disagree over which basic symbols should be used. Is the vowel in "air" more like the vowel in "able" or in "edge"? Probably for most speakers it is somewhere in between. There are two approaches to such cases. In a broad transcription, some standardized transcription may be adopted, perhaps an arbitrary-seeming one. Thus in this chapter we will use the following vowel+/r/ combinations in broad transcription: /ir, er, ar, or, ur, aɪr, aʊr/.

The second approach is to expand the inventory of symbols from the minimal one, that is, to give a transcription that is somewhat narrower. This can be done both by providing additional symbols, and by supplementing symbols with **diacritics**, marks added to symbols to modify their values. We will provide a few options that go beyond the basic symbols, because there are some sound qualities that most native speakers prefer to have separate symbols for. (1) /t/ and /d/ are often pronounced in a quick and weak way, and sound the same; they can be transcribed [ɾ]: "city" ['sɪɾi], "ready" ['rɛɾi], "sanity" ['sænɪɾi]. (2) Note item #35, "bird", transcribed with no vowel symbol but instead with [r] with a small line under it. This line indicates that the /r/ sound has no accompanying vowels. As noted before, such consonants are said to be syllabic. In English, /r/ may be syllabic when either stressed or stressless; we therefore list this [r̩] as a basic sound of the language. For most speakers /m,n,l/ may also be syllabic, but only when stressless. These three syllabic consonants are therefore generally not included among the basic sounds of the language; instead, such words are broadly transcribed using a [ə] plus the consonant. The syllabic symbols [m̩ n̩ l̩] are used only in narrower transcriptions. In sum, we have ""bird" [bɝd] and "butter" [bʌɾr̩] in broad transcription, but "bottom" [barəm] in broad and ['barm̩] in narrow. (3) The consonants /p/, /t/, and /k/ are sometimes pronounced with an extra, h-like puff of air coming through the vocal cords and out of the mouth (called **aspiration**). They are then transcribed as [p^h], [t^h], [k^h]. (4) The sound at the beginning of a cough or at the beginning and middle of the phrase "uh-oh" is called "glottal stop", transcribed [ʔ]: ['ʔʌʔo]. This same sound can be made along with the

consonants /p/ /t/ or /k/, in which case they are said to be **glottalized** (or preglottalized) and are transcribed [ʔp], [ʔt], and [ʔk]. (5) The consonants /p t k b d g m n ŋ/ can all be pronounced without a sound from the mouth opening at the end of the consonant, in which case they are said to be **unreleased** and transcribed [pˀ] etc.. /p t k/ may be glottalized ([ʔpˀ]), unreleased ([pˀ]) or both ([ʔpˀ]). (6) Stressed vowels may sound like diphthongs. For example, the vowel /e/ may be transcribed as [eɪ] in narrow transcription. (7) Vowels can be marked with a raised tilde ˜ when they are pronounced with air coming out of the nose ("nasalization") next to consonants [m], [n], [ŋ].

(4) gives broad and narrower transcriptions of all the words in Table 1, plus the additional examples given above. In the broad transcriptions, stress is marked only when it is needed to distinguish this word from some other existing word, but in the narrower transcriptions it is marked whenever the word has two or more syllables. It is inevitable that some readers will pronounce some of these words differently from what is given here; the narrow transcriptions in particular are arbitrarily chosen, and purposely have been made to differ across examples. (Compare for example the various final /k/s in "yoke", "cook", "book", and "talk".) For some words, the broad and narrow transcriptions given here are the same; there are no special characteristics of any of the sounds to be noted in the narrow transcription. Some of the aspects of the narrow transcriptions shown here will be presented further in Section C2.

Providing a narrow transcription of an utterance is an advanced skill that goes well beyond a general introductory course. The particular details included here are intended to give a taste of the kinds of variation that can be observed. In addition, these details will enter into later discussions of English and other languages.

(4) Transcriptions of sample English words. Broad transcriptions use the symbol set in (3). Narrow transcriptions show some uses of IPA diacritics, including some not yet discussed. Brackets omitted after first item.

word (orthography)	broad transcription	narrower transcription
pope	/pop/	[pʰoʔpˀ]
barber	barbɻ	ˈbarbɻ
mum	mʌm	mʌ̃mˀ
fife	faɪf	fʌɪf
vive	viv	viv
taunt	tant	tʰãntˀ
deed	did	didˀ
nun	nʌn	nʌ̃nˀ
rare	rer	rer
thousandth	θaʊzɪntθ	ˈθaʊzɪ̃ntθ
this	ðɪs	ðɪs

breathe	brið	brið
source	sors	sors
zanies	zeniz	ˈzeɪniz
shush	ʃʌʃ	ʃʌʃ
measure	mɛʒɻ	ˈmɛ̃ʒɻ
lull	lʌl	ɬʌɬ
church	tʃɻtʃ	tʃɻtʃ
judge	dʒʌdʒ	dʒʌdʒ
yoke	jok	jokʰ
cook	kʊk	kʰʊk
gag	gæg	gæg˺
singing	sɪŋɪŋ	ˈsĩŋĩŋ
we	wi	wi
he	hi	hi
easy	izi	ˈizi
imitate	ɪmɪtet	ˈĩmĩtʰeʔt˺
able	ebl	ˈebl̩
edge	ɛdʒ	ɛdʒ
attack	ətæk	əˈtʰæk
battle	bætəl	ˈbæɾl̩
father	faðɻ	ˈfaðɻ
talk	tak	tʰaʔk
road	rod	rod˺
book	bʊk	bʊk˺
should	ʃʊd	ʃʊd
food	fud	fud˺
aroma	əromə	əˈromə
but	bʌt	bʌʔt˺
ride	raɪd	raɪd˺
house	haʊs	haʊs
boy	boɪ	boɪ
convict	kanvɪkt	ˈkʰãnvɪʔk˺t
air	er	er
or	or	or
city	sɪti	ˈsɪɾi
ready	rɛdi	ˈrɛɾi
sanity	sænɪti	ˈsænɪɾi
bird	bɻd	bɻd˺
butter	bʌtɻ	ˈbʌtɻ
bottom	baɾəm	ˈbaɾm̩

EXERCISE 6: For each word, a choice of broad transcriptions is given. Indicate which one is consistent with pronunciation of the word and with the set of IPA symbols for broad transcription used in this chapter and listed in (3).

Example: **"cat" (a) /cat/ (b) /kat/ (c) /kæt/ (d) /cæt/** Answer is (c).

1. "see"	(a) /see/	(b) /si/	(c) /cee/	(d) /sy/
2. "Fuji"	(a) /fuji/	(b) /fuge/	(c) /fudʒi/	(d) /fudʒe/
3. "class"	(a) /class/	(b) /klass/	(c) /clæs/	(d) /klæs/
4. "you"	(a) /you/	(b) /ju/	(c) /jou/	(d) /yu/
5. "spa"	(a) /spa/	(b) /spæ/	(c) /spo/	(d) /ʃpa/
6. "she"	(a) /she/	(b) /ʃe/	(c) /shiʔ	(d) /ʃi/
7. "sir"	(a) /sir/	(b) /sr̩/	(c) /ʃir/	(d) /ser/

EXERCISE 7: Give the regular English orthography for the following words, which are given in a fairly broad transcription but with a few extra symbols. Even if the pronunciation given here is not the same as yours, you should be able to figure it out. If you are not sure of the orthography, look the word up in a dictionary.

Example: **[mʌtʃ]** is "much"

1. **[naɪt]**
2. **['mjuzɪk]**
3. **['bælkəni]**
4. **[gost]**
5. **['mr̩si]**
6. **['merɪdʒ]**
7. **['feriz]**
8. **['berli]**
9. **['tʃrædʒəri]**
10. **['kʌntrimɪn]**

EXERCISE 8: Give broad transcriptions of the following words, as best you can for your own pronunciation. Or, look the words up in a dictionary and give that pronunciation in IPA symbols.

1. xerox

2. utopia

3. direct

4. photo

5. triumph

Even using the limited number of symbols of a broad transcription, many differences between speakers can be indicated. Indeed, because the limited symbol set has been chosen to cover just those differences that distinguish words, these differences between speakers will be the ones that listeners hear most readily. Consider some ways in which one of the authors of this text's pronunciations are possibly different from yours even in a broad transcription, as shown in (5). Try comparing the different pronunciations by saying both of them aloud.

(5) Two pronunciations (in broader transcription) of some English words.

word	author's pronunciation	more common pronunciation
1. forest	farɪst	forɪst
2. poor	por	pur
3. merry	mɛri	meri
4. marry	mæri	meri
cf. Mary	meri	meri
5. parade	pəreɪd	p̣reɪd
6. particular	pətɪkjəl̩r	p̣rtɪkjəl̩r

EXERCISE 9. Compare the pronunciations in (5) with those in a dictionary. First, copy out the pronunciation(s) as given in the dictionary. Then, if necessary, convert these into IPA by referring to the "pronunciation key" at the beginning of the dictionary. Compare your IPA-pronunciations to the ones in the table to see whether the author's pronunciations are recognized by the dictionary.

We have illustrated the pronunciations of individual words in isolation because that makes it easy to focus on individual sounds. The pronunciations of words in fluent, connected speech, however, can be very different from their pronunciations in

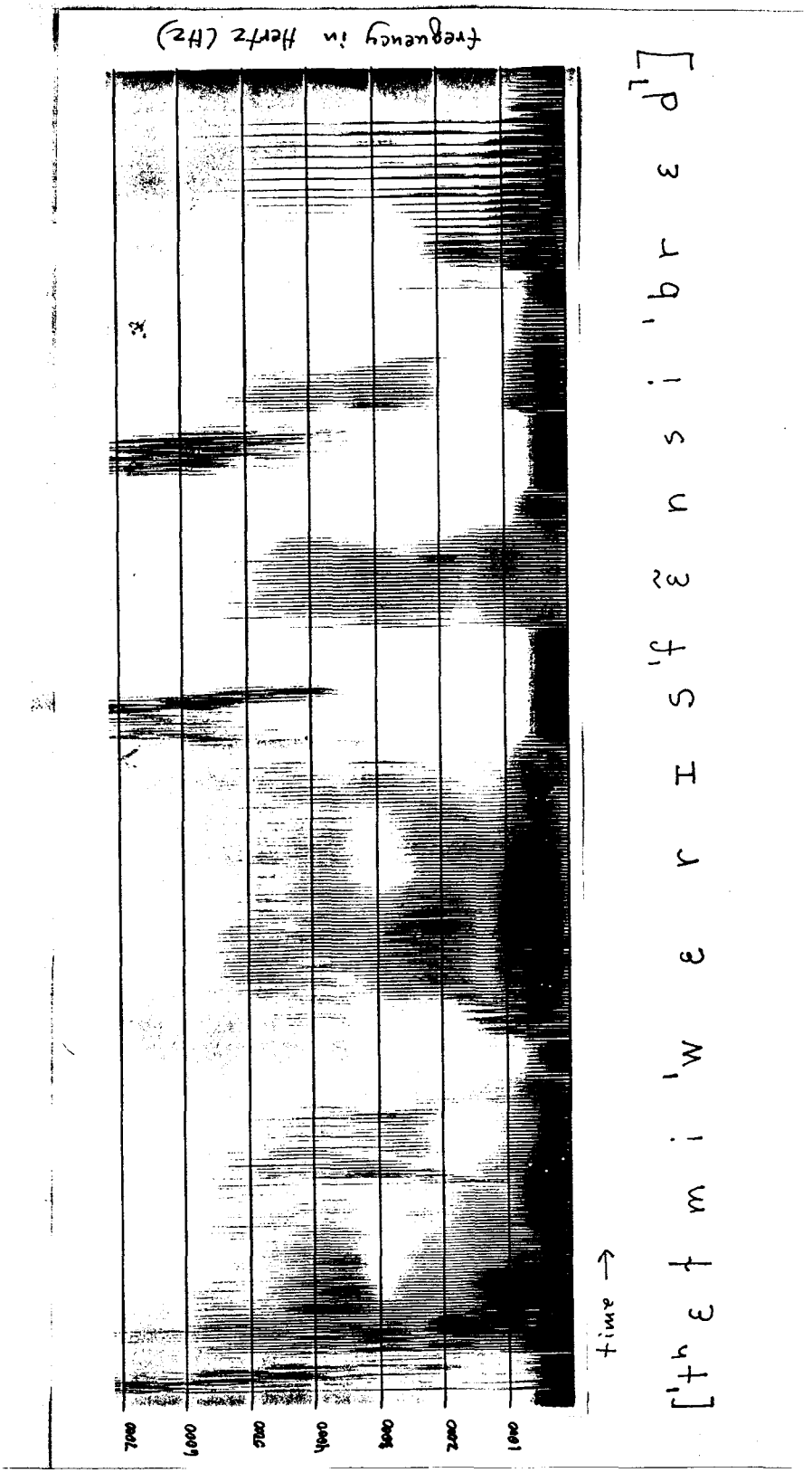
isolation. At UCLA we have studied the pronunciation of some words when they occur in recordings of spontaneous telephone conversations. Compared to speech which is read, spontaneous speech shows more reduced or deleted vowels and consonants.

Today, narrow transcriptions are usually done with a combination of listening and looking. A computer display lets a listener select some portion of a word and listen to it repeatedly, even slowing it down, while seeing its acoustic characteristics. Figure 2 shows a transcriber working at this kind of workstation. The phonetic transcription is entered into the computer so that it becomes attached to the speech utterance. Figure 3 shows the same spectrogram as in Figure 1, now with its associated narrow transcription. The /t/ in "tell" is aspirated, the vowel in "where" is /ɛ/, the /w/ is voiceless, the /z/ in "is" is partly voiceless, the vowel in "fancy" is nasalized [ɛ̃], and the final /d/ in "bred" is unreleased.

FIGURE 2 (a happy transcriber) not shown in this draft

FIGURE 3 (Spectrogram) on next page

Figure 3. Same spectrogram as in Figure 1, now with narrow transcription.



B4. Articulatory definitions of symbols
B4.1. The IPA chart

It would be helpful to be able to describe specifically what sounds these transcriptions are meant to convey, instead of just giving examples of words containing a sound and hoping you recognize some particular sound quality from them. In other words, we need a vocabulary for talking about sounds. The terms we use make reference to *where and how the sound is produced in the speaker's mouth.* Where in the mouth a sound is made is called the **place of articulation**, and how a sound is made is called the **manner of articulation**. The different components of sound production give us descriptions of sounds and also definitions of the phonetic symbols. The terms that we use are given as labels on the **IPA chart,** which organizes symbols with respect to articulation.

FIGURE 4 (1996 IPA chart) on next page

FIGURE 4 reproduces the entire most recent (1996) IPA chart. It consists of a consonant chart, a vowel chart, and lists of other symbols. Locate for yourself on this chart the symbols listed above. Some of our symbols are combinations of two different symbols from the chart. For example, [tʃ] is a combination of [t] and [ʃ]. You can see that we have used very few of the available symbols! Some of these other symbols will be discussed in section C below.

You can also see that the symbols are organized into a descriptive framework: the individual charts contain descriptive labels as well as just the symbols. For example, the symbol [b] is in a box or cell labeled "Bilabial" and "Plosive", and the symbol [w] is followed by the phrase "Voiced labial-velar approximant". The sounds represented by the symbols are defined by how the sounds are articulated. One wants an independent source of information about what the symbols mean, and articulatory definitions attempt to provide this. That is, for each sound a set of properties is given that more or less tell you what that sound is and how it is different from other sounds. Any transcription can be read in terms of these definitions, even if you do not know the language being transcribed.

Articulatory definitions are not the only possible kind; acoustic or auditory-perceptual definitions, derived from looking at spectrograms, could be used. When the IPA was set up over a century ago, there really was no choice, as speech articulation was much better understood than either speech acoustics or speech perception. Nowadays you will see descriptions of sounds in these other domains as well. For example, vowels are thought to be as well, or better, described in terms of their auditory qualities. (Indeed, IPA vowel descriptions are based in part on a tradition of equal auditory spacings between a subset of the vowel sounds.) Even one traditional phonetic term, "sibilant" (sometimes "strident"), refers to the particular loud, high-pitched, noisy sound of certain fricatives, more than it refers to a particular articulation. But we will limit our discussion here to articulation, in accord with the IPA descriptive terms on the chart.

Figure 4

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993, corrected 1996)

CONSONANTS (PULMONIC)

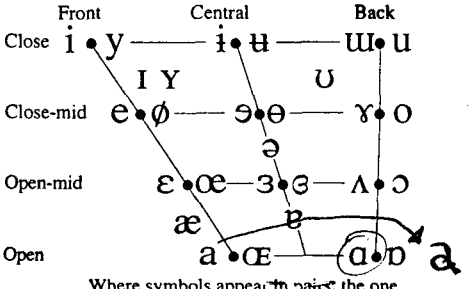
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	B			r					R		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ	ɻ		j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	'	Examples:
|	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
!	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

VOWELS



OTHER SYMBOLS

ʍ Voiceless labial-velar fricative
w Voiced labial-velar approximant
ɥ Voiced labial-palatal approximant
ʜ Voiceless epiglottal fricative
ʢ Voiced epiglottal fricative
ʡ Epiglottal plosive
ɕ ʑ Alveolo-palatal fricatives
ɺ Alveolar lateral flap
ʩ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

k͡p t͡s

SUPRASEGMENTALS

ˈ Primary stress
ˌ Secondary stress ˌfounəˈtɪʃən
ː Long eː
ˑ Half-long eˑ
◡ Extra-short e̯
| Minor (foot) group
‖ Major (intonation) group
. Syllable break ɹi.ækt
‿ Linking (absence of a break)

TONES AND WORD ACCENTS

LEVEL		CONTOUR	
é͡ or ˥	Extra high	ě or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˧˥	High rising
è ˨	Low	e᷅ ˩˧	Low rising
ȅ ˩	Extra low	e᷈ ˥˧˥	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
◌ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼	◌̻	Laminal	t̻ d̻
◌̹	More rounded	ɔ̹	◌ʷ	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̜	Less rounded	ɔ̜	◌ʲ	Palatalized	tʲ dʲ	◌ⁿ	Nasal release	dⁿ
◌̟	Advanced	u̟	◌ˠ	Velarized	tˠ dˠ	◌ˡ	Lateral release	dˡ
◌̠	Retracted	e̠	◌ˤ	Pharyngealized	tˤ dˤ	◌̚	No audible release	d̚
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ			
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝	(ɹ̝ = voiced alveolar fricative)		
◌̩	Syllabic	n̩	◌̞	Lowered	e̞	(β̞ = voiced bilabial approximant)		
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘			
◌˞	Rhoticity	ɚ a˞	◌̙	Retracted Tongue Root	e̙			

PHONETIC SYMBOLS AND COMPUTERS

With the advent of thousands of fonts for computer word-processing, free or inexpensive phonetic fonts have become widely available. Some fonts include a few "phonetic" characters, while others offer the whole symbol set approved by the IPA in 1989 in Kiel, Germany (the "Kiel" IPA). The WWW home page of the IPA is a good starting point for finding phonetic fonts. With these fonts, every symbol can be assigned to a key or key combination on the keyboard. For example, in one commercially-available IPA font, the character [ɔ] ("open o") is typed as Shift-o. However, the Kiel symbol set goes beyond just the font characters. It also includes a numerical code for every symbol, so that converting from one font to another is easy and unambiguous. For example, character [ɔ] is number 306. The Kiel conventions also include a system for machine-readable (ASCII-only) equivalents of every symbol. For example, character [ɔ] is ASCII "O" (upper-case o). This form of the IPA is useful for people doing computer transcription of long speech samples. The "ARPABET" is a different ASCII system developed just for English and used by the American speech technology (recognition and synthesis) community.

B4.2. Articulators

The IPA articulatory definitions are provided in part by arranging some of the consonant and vowel symbols into individual charts which are three dimensional in content but flattened out to 2 dimensions on paper. One dimension is in the horizontal arrangment, another is in the vertical arrangment, and the third is in the order of symbols that are written in pairs. The main consonant and vowel charts are both 3-dimensional in this way, but they use different sets of dimensions. To understand these labels, we need to first consider the speech production mechanism. Figure 5a shows what one of the authors' vocal tract looks like along the midline of the head during the consonant [s], in a scan taken by Magnetic Resonance Imaging, and Figure 5b is a schematic based on this figure. The vocal tract is the parts of the body used in producing sounds: the larynx, the pharynx, the oral cavity, and the nasal cavities. The pharynx and oral cavity together are sometimes called the oral tract. There are several vocal organs that can move independently (active articulators), and several anatomical structures that these moving articulators may move towards (passive articulators, also called places of articulation). The labels on the figure point out several different articulators and structures. They are listed in (6); additional structures not seen in the figure are listed in (7).

FIGURES 5a and 5b (articulators) on next 2 pages

Figure 5a. An author's vocal tract, along the midline, while saying /s/.

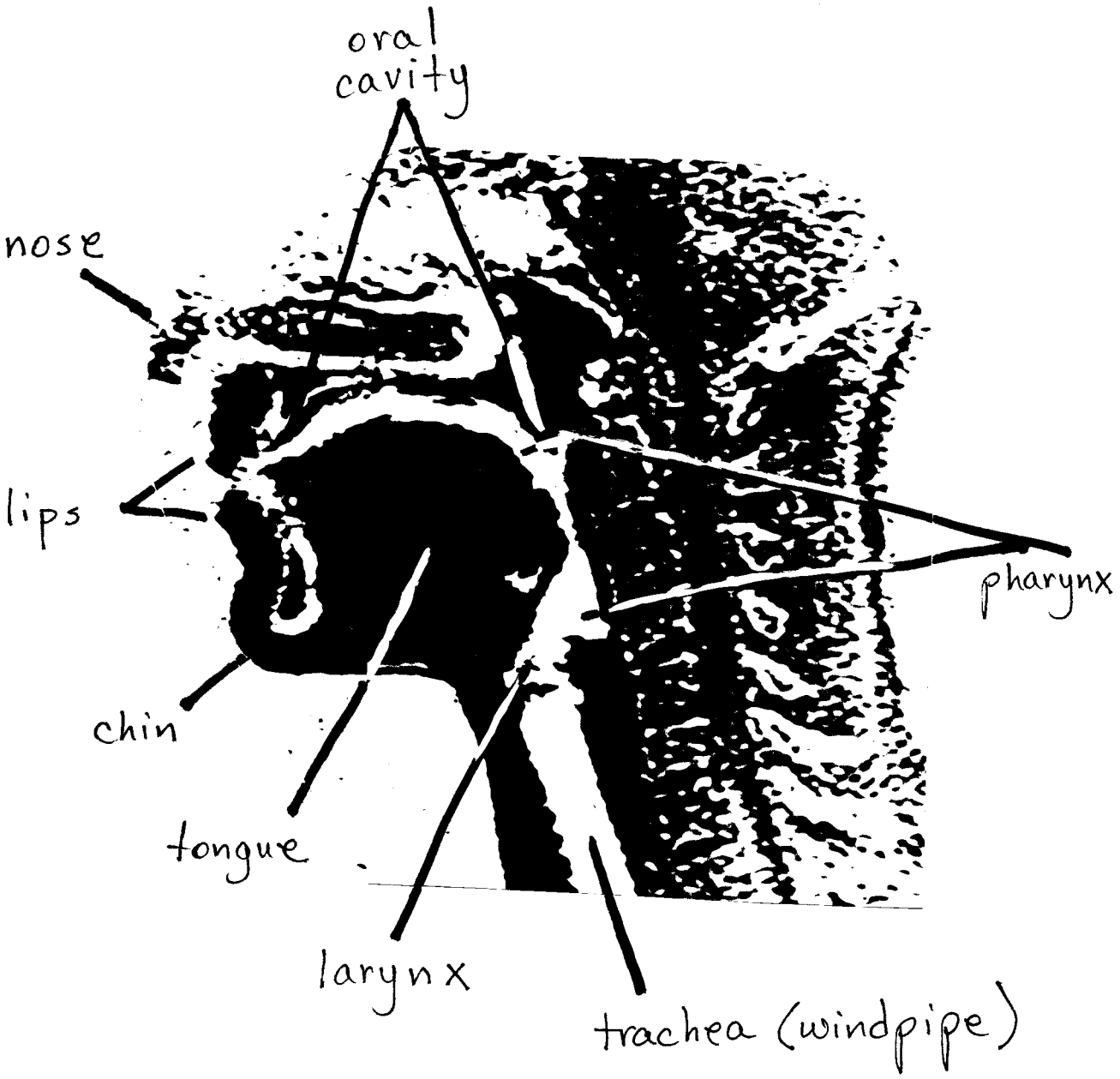
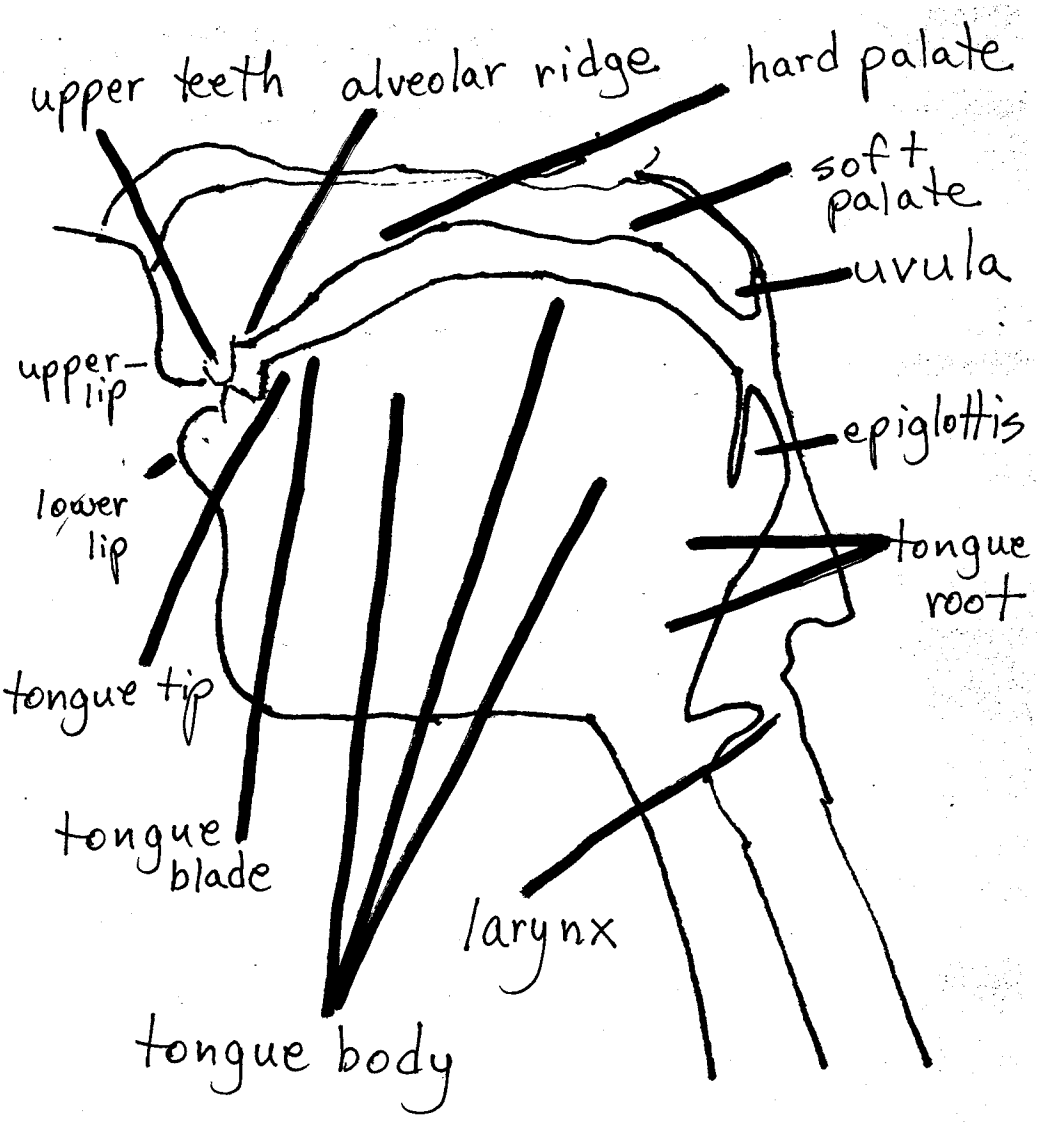


Figure 5b. Vocal tract traced from 5a, additional structures labeled.



For speech to be heard, sound energy must pass out of the vocal tract through the movement of air molecules. Therefore you can think of the immediate goal of speaking as setting air molecules into motion. Speech sounds are usually described in terms of acoustic energy sources, and filters that modify the sources. For many speech sounds, the source is voicing, which is the regular passage of puffs of air through the larynx as the vibrating vocal cords open and close. Those puffs of air are basically the same for all voiced sounds, but they are then filtered as they travel through the air spaces of the pharynx, oral cavity, and/or nasal passages. The various consonants and vowels, with their different configurations of the active articulators creating different patterns of constrictions, each have their own characteristic filter function.

(6) Articulator definitions

ARTICULATOR	DEFINITION
oral cavity	the top, curved-horizontal airspace of the vocal tract
upper lip	as an active articulator, moves independently to approach the lower lip; as a passive articulator, is approached by the lower lip (or tongue tip)
lower lip	as an active articulator, moves independently, and is also moved by the jaw, to approach the upper lip or the upper teeth; as a passive articulator, is approached by the upper lip
upper teeth	passive only; can be approached by the lower lip or by the tongue blade/tip
alveolar ridge	passive only; can be approached by the tongue tip or blade. A short flat stretch just above and behind the upper teeth. At its back edge, the palate turns more sharply upward (creating a corner). Not everyone has a ridge; the author's is more prominent off the midline, so it is not obvious in this figure. If you have no ridge per se, then the term can be taken to refer to the first half-cm or so behind the upper teeth.
tongue tip	an active articulator, moved by the tongue blade, of which it is the very end, or frontmost part (also called "apex")
tongue blade	an active articulator, moves independently of the rest of the tongue, or is moved by the rest of the tongue. About the first 3 cm of the tongue, behind which there is a point where the tongue can bend and flex. The tongue tip is the very end of the blade.
tongue body	active; the main part of the tongue; can be divided into a front part and a back part
tongue root	active; the bottom part of the tongue, which forms the front wall of the pharynx
epiglottis	active; the leaf-like appendage to the tongue in the pharynx
pharynx	the back, vertical airspace of the vocal tract between the uvula and the larynx; usually thought of as the passive articulator approached by the body or root of the tongue, but its walls can also function actively, squeezing the pharynx space
hard palate	the hard, bony, surface of the roof of the mouth
soft palate	the soft, non-bony part of the roof of the mouth behind the hard palate,

	also called the velum, which as an active articulator moves up and down at the top of the pharynx to either block off or open up the air passage to the nose, and as a passive articulator is approached by the tongue body
uvula	the hanging back tip of the soft palate; as an active articulator it can vibrate in the upper pharynx; as a passive articulator it is approached by the tongue body
larynx	the cartilage box at the bottom of the pharynx (and at the top of the trachea) housing the vocal cords; the box can move up and down

(7) Some other parts of the body used in producing speech

lungs	the usual source of airflow for speech (via the trachea)
vocal cords	two bands or folds strung front to back inside the larynx
glottis	the airspace between the vocal cords
sides of tongue	the figure, taken at the midline of the tongue, shows only its center; the two sides of the tongue can curl and roll independently of the center
the nervous system	controls all motor movements

EXERCISE 10: Give the term corresponding to the definition given.

Example: **the soft part of the roof of the mouth** = soft palate

1. **the end of tongue blade**
2. **airspace between uvula and larynx**
3. **bottom part of tongue forming front wall of pharynx**
4. **cartilage box at bottom of pharynx holding the vocal cords**
5. **active or passive articulator; moved by the jaw**

B.4.3. Consonants

With this background information in hand, let us turn back to the IPA chart, beginning with the main consonant chart at the top. The main consonant chart contains a subset of the consonants, those made with air flowing out from the lungs (**pulmonic egressive**). Recall that consonants are sounds produced with a significant constriction in the vocal tract. This constriction means that the flow of air through the oral cavity and out of the mouth is affected: either the airstream becomes noisy, or less air is able to flow out than would without the constriction.

B4.3.1. Columns: Place of articulation

The columns give information about the active and/or passive articulators of a sound: what articulator approaches what along the midline of the vocal tract. Some columns define both: bilabial, labiodental, glottal. Most columns just define the passive articulator, with the assumption that the normal active articulator is the part of the tongue closest to that passive area. (For some uvular and pharyngeal sounds, the "passive", i.e. destination, articulator may also be active.) The retroflex column primarily defines the action of the active articulator, the blade of the tongue raised up, with the assumption that it goes somewhere into the postalveolar region. The combination of kinds of information given by the columns is often loosely called "place of articulation". These column headings are listed in (8). Note that there is a single column for dental, alveolar, and postalveolar together. The base symbol shown in all the rows except the fricative row is for the alveolar place; the dental place is indicated by modifying that base symbol with the dental diacritic [̪] under it, and the postalveolar place is indicated by modifying that base symbol with the "Retracted" diacritic [̠] under it. There are also special (non-IPA) terms available for the active articulators alone: labial means one or both lips are active, and thus includes bilabial and labiodental; coronal means the tongue blade/tip is active, and thus includes dentals, alveolars, postalveolars, retroflexes, and sometimes palatals; dorsal means the tongue body is active, and thus includes velars, uvulars, and sometimes palatals and pharyngeals; radical means the tongue root is active, and thus includes pharyngeals and also epiglottals (not on the main consonant chart).

(8) IPA consonant column labels (places of articulation)

bilabial	the 2 lips, each both active and passive
labiodental	active lower lip to passive upper teeth
dental	active tongue tip/blade to passive upper teeth
alveolar	active tongue tip/blade to passive front part of alveolar ridge
postalveolar	active tongue blade to passive behind alveolar (When the passive place is specifically the back part or corner of the alveolar ridge, this is also called "palatoalveolar" or "alveopalatal")
retroflex	active tongue tip raised or curled to passive postalveolar (difference between postalveolar and retroflex: blade vs. tip)
palatal	tongue blade/body to hard palate behind entire alveolar ridge
velar	active body of tongue to passive soft palate (sometimes back of hard)
uvular	active body of tongue to passive uvula, OR active uvula
pharyngeal	active body/root of tongue to passive pharynx
glottal	the two vocal cords, each both active and passive

Some other places of articulation are not given columns on the chart, symbols for consonants with those places instead being listed under "Other Symbols". Two of these places are labial-velar (sometimes also called labiovelar) and labial-palatal, which each combine two separate columns of the main chart by having two primary

articulations. Labial-velars combine bilabial and velar articulations. The voiceless fricative [ʍ] is still found in some American dialects in words like "which" and "what" (but not "who", [hu]). The voiced approximant [w] is a very common sound in the world's languages. The labial-palatal combination is not used in any basic sounds of English. The epiglottal place, also not used in English, refers to the epiglottis as an active articulator in the lower part of the pharynx. The alveolo-palatal place lies between alveolar and palatal; the active articulator is the blade of the tongue. Note that this is different from "palatoalveolar" or "alveopalatal" (see above).

VISUAL KNOWLEDGE OF SPEECH

In this chapter, as in other textbooks, we discuss people's perception and knowledge of speech signals as if the only relevant sensory modality were hearing. Yet this is not correct. Listeners also have extensive experience *seeing* speech being spoken. They use this knowledge when they are listening, and may feel at a disadvantage when forced to listen without the visual information. That is, normal listeners generally use both kinds of information when they are available. Hearing-impaired listeners rely correspondingly more on the visual information. **Speech reading** (or "lip reading") is the name given to the use of visual information for speech processing by hearing-impaired "listeners". Speech reading can be explicitly taught, and your library may have teaching materials on this skill from early in this century. (Further reference: B. Dodd and R. Campbell, 1987.)

As you learn in this chapter about how different sounds are articulated, think about which aspects of articulation are likely to be easily recoverable from a visual signal. Some consonant articulations are quite directly observable. For example, in bilabial stops like [b,p,m] the two lips come together. In contrast, most tongue articulations cannot be seen. It turns out that English listeners, in deciding what consonant they have heard, take into account the fact that they expect to see strong visual lip cues for bilabial consonants. The **McGurk effect** is the name given to the phenomenon in which listeners' percepts are determined by visual as well as auditory information. In now-classic studies, McGurk and MacDonald showed listeners videos of speakers saying certain syllables, but played carefully-synchronized recordings of other syllables. The listeners often reported hearing syllables that were different from both of these, yet they were generally unaware that the visual and auditory signals did not match. When the listeners were played audio [ma] and visual [ta], they reported hearing [na]; when they were played audio [ba] and visual [ga], they reported hearing [da]. That is, listeners seem to require visual confirmation to hear a bilabial consonant, and if they don't get that confirmation, they will hear the closest articulation. Interestingly, the effect has subsequently been shown to be larger in some languages compared to others, a finding which remains to be explained.

B4.3.2. Rows: manners of articulation

The rows give information about how close the active articulator comes to the

passive one, that is, how open the oral air passage is. Five such manners of articulation can be distinguished and are listed in (9).

(9) Constriction degrees for consonants

stop	touch and hold-to-seal (no flow of air out of the mouth)
trill	vibrate in airstream
tap/flap*	touch but don't hold (includes quick touch and fast sliding)
fricative	leave a gap and make noise in it
approximant	leave a big gap with almost free flow of air

[*We will not distinguish taps and flaps here, but simply note that the two terms are sometimes used interchangeably, with "flap" perhaps the more common term in the U.S.]

Most of these terms appear in the chart. The term "stop" does not; instead, the first two rows of the chart are two kinds of stops, plosive and nasal. In both of these, there is a complete seal in the oral tract. The two rows distinguish between stops in which the air, once released, flows out of the oral cavity ("oral" flow, this word not appearing on the chart because it is the typical case) or out of the nasal cavity ("nasal" flow). The plosives are oral stops and the nasals are nasal stops. All other rows of the chart can also be taken to be oral. Other rows of the chart distinguish between sounds in which the airflow is along the center of the oral cavity ("central" passage, this word not appearing on the chart because it is the typical case) or along one or both sides of the oral cavity ("lateral" passage). The row labels are given in (10). Affricates do not have a row on the chart; they combine a stop plus a fricative at the same place of articulation.

(10) IPA consonant row labels (manners of articulation)

plosive	a pulmonic-egressive, oral stop
nasal	a pulmonic-egressive stop, but not a plosive, because not oral
fricative	implies central; noise generated in air gap
lateral fricative	fricative whose air gap is on side(s)
approximant	implies central;
lateral approximant	approximant whose air passage is on side(s)

You can see that the consonant grid more or less defines an articulatory space within the vocal tract. The columns divide up the places of articulation moving along from the front (the mouth opening) to the back and then down to the bottom of the vocal tract, while the rows characterize the air passage from least to most open. To the extent that this is so, the boxes in the grid cover regions of the available space rather than exact values. For example, "velar" covers the entire soft palate and even the back of the hard palate. But in any one production of a velar consonant, only some of this region is the passive articulator. Exactly which part varies across productions. In this sense, many subtly-different sounds all count as "velars".

Another classification that is often made regarding these manners of articulation (though not made by the IPA) is that plosives and the two kinds of fricatives (and therefore also affricates) are called **obstruents,** while nasals, trills, taps/flaps, and the two kinds of approximants are called **sonorants**. The obstruents are the sounds in which the airflow is noisy (the air meets an obstruction) while the sonorants are the sounds in which the airflow is smooth. The sonorants are often further divided into glides (semi-vowels, or vowel-like central approximants) and liquids (r and l sounds).

Note that the symbol we are using for the American English r-sound is [r]. This symbol is used on the chart for an alveolar trill, but the American English sound is a post-alveolar or retroflex approximant. The IPA system allows substitutions of simpler symbols for unusual ones for a given language – as we are doing with [r] – if no confusion will result. To help avoid confusion, we have also indicated the place of [r] as we are using it on the chart.

B4.3.3. Voicing; other consonants

Inside many cells (or boxes) are pairs of consonants with the same row-and--column definition. These pairs differ in **voicing**, that is, in the activity of the vocal cords. Generally, if the vocal cords vibrate for all or part of the sound, it is said to be voiced and appears as the right member of the pair; if the vocal cords do not vibrate at all, the sound is said to be voiceless, and appears as the left member of the pair. Notice that all the unpaired sounds (the nasals, trills, tap/flaps, and approximants) are all placed to the right in their cells to show that they are all voiced.

The basic English consonants that are not on this main consonant chart are the affricates and the approximant [w]. For [w], it is because it is a combination of two articulations, bilabial and velar. Sometimes on phonetic consonant charts you will see [w] in the bilabial column, sometimes in the velar column, sometimes in both, sometimes in a special column labeled "labialvelar" or "labiovelar". The IPA now does none of these, instead listing it with other symbols that would otherwise require a new row or column on the main consonant chart. The affricates [tʃ] and [dʒ] are combinations of sounds that do appear on the chart. An affricate is a stop followed by a fricative made at a same or similar place which functions as a single sound. The two affricates of English are postalveolar (commonly called palatoalveolar), voiceless [tʃ] and voiced [dʒ].

To summarize IPA articulatory definitions, then, each sound is described primarily in terms of its oral articulation: which active articulator gets how close to which passive articulator. For example, in [p,b,m], the two lips act as both active and passive articulators, and they touch-and-hold to make a seal. Thus these sounds are all bilabial stops. But in addition, other articulations happen at the same time. The vocal cords are either vibrating ([b,m]) or not ([p]); the velum is either raised ([p,b]), in which case the stop is oral, or it is lowered ([m]), in which case the stop is nasal. When a stop is oral and pulmonic, as with [p,b], it is a plosive. Each symbol, then, is an abbreviation for a combination of properties. To "name" a consonant sound, you

list these properties, by convention in the order voiced/voiceless, then place, then manner. Thus [b] is a voiced bilabial plosive; [p] is a voiceless bilabial plosive; and [m] is a voiced bilabial nasal. It is worth noting that these terms are usually used in a kind of shorthand way that presupposes what is typical. Thus plosives and nasals are all stops, but since stops are usually pulmonic and oral the word "stop" is often used by itself to mean plosive; since nasals are usually stops and voiced the word "nasal" is often used by itself to mean voiced nasal stop; since laterals are usually approximants the word "lateral" is often used by itself to mean lateral approximant. For example, then, [m] may be referred to as a bilabial nasal rather than a voiced bilabial nasal, and [l] as an alveolar lateral rather than a voiced alveolar lateral approximant. However, it is never wrong to use all three or four terms, even if some seem redundant.

Sounds can be grouped together according to which properties they have in common, and sounds which share several properties are generally more similar than sounds which share few properties. This will be discussed in the second phonology chapter.

EXERCISE 11: Give the term corresponding to the definition given.

Example: **both lips** = bilabial

1. **tongue blade to ridge above upper teeth**
2. **tongue body to soft palate**
3. **make noise in a gap**
4. **plosives and fricatives as a group**
5. **vocal cord vibration**

EXERCISE 12: Provide the IPA symbol whose definition is given. Only IPA terms are used.

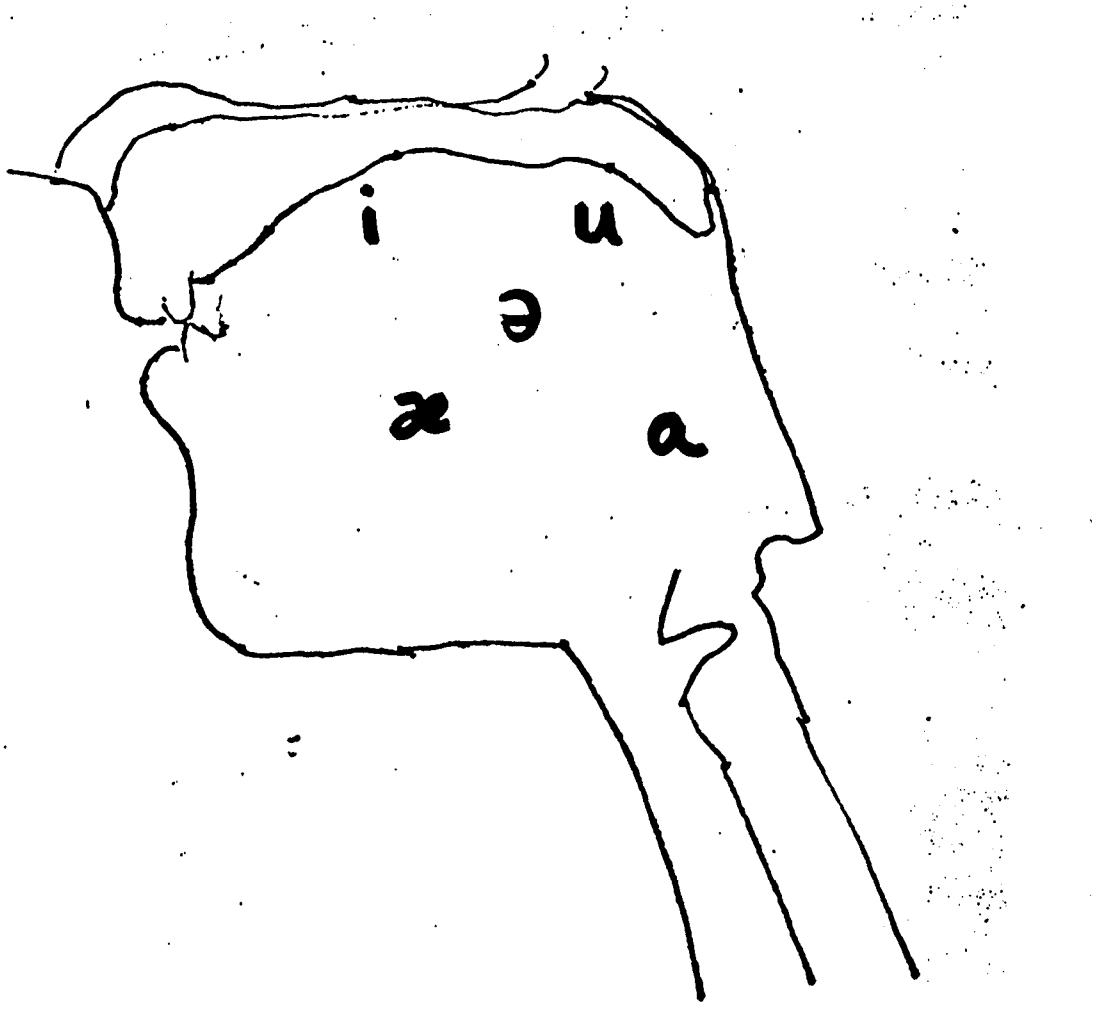
Example: **voiced alveolar plosive** = [d]

1. **voiceless velar plosive**
2. **bilabial nasal**
3. **voiced labiodental fricative**
4. **alveolar lateral approximant**
5. **glottal plosive**
6. **alveolar tap or flap**
7. **voiced postalveolar fricative**
8. **velar nasal**
9. **voiceless glottal fricative**
10. **voiced dental fricative**

B.4.4. Vowels

B4.4.1. The vowel chart

FIGURE 6. Vowel chart superimposed on schematic of vocal tract (from Figure 5b)



Next consider the main vowel chart. Like the consonant chart, the vowel chart encodes three dimensions, but the vowel dimensions are different from the consonant ones. The primary active articulator of vowel sounds is always some part of the tongue (almost always the tongue body), the passive articulator is some part of the midline of the outer surface of the vocal tract, and "how close" the active articulator comes to the passive is "not very". The vowel chart represents a kind of grid of the vocal tract in which the top of the chart is the roof of the mouth, and the left side of the chart is the front of the oral cavity, viewed on the speaker's left side, just as for the main

consonant chart. This is shown in Figure 6. The tongue moves around in this space for different vowel sounds, and the IPA definitions describe its location. That is, the chart's first two dimensions are the height of the tongue (relative to the roof of the mouth) and the backness of the tongue (relative to the front teeth), both usually referring to the highest point on the tongue's surface. Note that although for many vowels the sides of the tongue touch the sides of the palate, the descriptions focus on the midline of the tongue, because that is where the air flows.

The vowel chart is not as clearly divided into rows and columns as the consonant chart is. There are four height categories labeled "close", "close-mid", "open-mid", and "open". These are often also called "high", "higher mid", "lower mid", and "low", respectively. There are also three other rows between these that don't appear with labels. The row between "close" and "close-mid" is sometimes called "lower high" or "high lax". The row between "close-mid" and "open-mid" is usually called simply "mid". The row between "open-mid" and "open" is sometimes called "higher low".

The third dimension of the chart is the position of the lips. Rounded vowels usually have the lips constricted (pulled in close together) and protruded (pushed out from the face) so that when viewed from the front they make a circle, and when viewed from the side they project out. (Sometimes either the constriction or the protrusion is relatively weak.) For these vowels, then, there is a Labial articulation. Unrounded vowels do not have this feature; the lips are either in a neutral posture or they are spread out to press against the face. For the high vowels in particular, these different lip positions can make a large difference in the sound produced with a given tongue position. English uses unrounded front and central vowels; the rounding of the back vowels of English is somewhat variable across dialects. Nonetheless we use the symbols for rounded back vowels in broad transcription. (11) lists all the descriptive terms of the vowel chart.

(11) IPA vowel row and column labels

close	compared to other vowels, overall height of tongue is greatest; tongue is closest to roof of mouth (Also: "high")
open	compared to other vowels, overall height of tongue is least; mouth is most open (Also: "low")
close-mid, open-mid	intermediate positions (Also: mid/uppermid/lowermid)
front	compared to other vowels, tongue is overall forward
central	intermediate position
back	compared to other vowels, tongue is overall back (nearer pharynx)
rounded	lips are constricted inward and protruded forward

Locate the vowel symbols we have used for English on the IPA vowel chart. We are substituting simple [a] for the low back vowel of English, often more precisely

given as [ɑ]. (In fact American speakers vary so much that it is not obvious which symbol is more precise, especially given that the IPA offers no symbol for a low central vowel.) You will see that most of the English vowels are taken to be unrounded; only /u ʊ o/ are rounded. Also, except for the central /ə/ and open /a æ/, all of the vowels of English fall into front - back pairs (i-u, ɪ-ʊ, e-o, ɛ-ʌ). (For many speakers, however, the vowel we are representing with [ʌ] is central, i.e. IPA [ɜ].) To "name" a vowel sound, give these terms in the order row, column, rounding -- for example, [i] is a high front unrounded vowel.

The vowel chart can also be viewed not as a row-by-column grid but as more of a continuous representation of the two-dimensional vocal tract space, as seen in Figure 6 above. When it is used this way, vowel symbols are placed on the chart in a location that is meant to indicate the tongue's position in a speaker's mouth (or the corresponding auditory impression produced by the vowel for a listener). The dots next to the symbols on the charts are used as a reference grid of vowel qualities. So, the [i] in English may be placed somewhat off from the [i] on the chart, to indicate that the English vowel is not as extreme as the reference vowel. When the chart is used in this way, rounding distinctions do not influence the symbol locations.

B4.4.2. Diphthongs

Some vowels are represented as sequences of vowel symbols because the tongue and/or lips move from one position to another. Such vowels are called **diphthongs**. We can distinguish "large" diphthongal movements, which cross a large part of the vowel space, from "small" diphthongal movements, which involve only adjacent areas of the chart. The vowel transcriptions provided for English in this chapter represent only large movements: /aɪ, aʊ, oɪ/. Other books give small diphthongs for /e/ and /o/, and sometimes /i/ and /u/ as well. In general, the large diphthongs are diphthongs for most American speakers, while the small diphthongs are diphthongs for fewer speakers, including fewer California speakers. Our transcriptions are thus a compromise among the different possible pronunciations of English vowels as diphthongs vs. monophthongs, as with all our broad transcriptions.

EXERCISE 13: Give the term corresponding to the definition given.

Example: **made with the tongue overall forward** = front

1. **made with the lips pulled together and forward**
2. **vowel composed of a sequence of two vowel sounds**
3. **made with the tongue maximally low**
4. **made with the tongue maximally high**
5. **made with the tongue in an intermediate position in the front-back dimension**

EXERCISE 14: Provide the IPA symbol whose definition is given.

Example: **mid central unrounded vowel** = [ə]

1. **high (= close) front unrounded vowel**
2. **lower high front unrounded vowel**
3. **higher mid (= close-mid) back rounded vowel**
4. **low (= open) back unrounded vowel**
5. **high back rounded vowel**
6. **higher low front unrounded vowel**
7. **high front rounded vowel**
8. **lower mid front unrounded vowel**
9. **high central rounded vowel**
10. **lower high back rounded vowel**

B4.5. Different pronunciations mean different symbols

The IPA charts make a big distinction between consonants and vowels, giving them separate charts and descriptive terminology. This puts the glides in an odd position, since they are like both consonants and vowels. Their position on the consonant chart (as central approximants) obscures their similarity to vowels. Glides are sometimes also called semi-vowels because they are vowel-like. The glide [j] is only a bit more constricted than the otherwise-similar vowel [i], and the glide [w] is only a bit more constricted than the otherwise-similar vowel [u]. It may be hard to say in any given case whether a particular sound is more like a vowel or a consonant. Similarly, we use a syllabic consonant symbol, [ɻ̩], for a glide-like sound that the IPA also provides vowel symbols for ([ɚ] or [ɝ], r-colored vowels). But generally otherwise the definitions of the symbols are fixed by the chart; and if you are sure what a given sound is then you pretty much know what symbol to use for it in a narrow transcription. Where there are uncertainties or disagreements, it is about the nature of the sounds themselves. If you pronounce a sound differently from how the symbol is defined, then strictly speaking that is the wrong symbol for your pronunciation. For example, if you pronounce [s] as a dental fricative rather than an alveolar, then you will be troubled by the fact that it is in the "alveolar" column, and you will wonder how a dental should be represented. (The answer is [s̪].) Similarly, if you pronounce [r] as a retroflex rather than a post-alveolar approximant. But we can still agree to use the symbols for the most common pronunciations in a broad transcription, knowing that the IPA does provide the resources for making these distinctions in a narrower transcription.

ARTICULATORY INFORMATION FROM THE PHONETICS LABORATORY

How do we know that these articulatory definitions are accurate, for at least some speakers? Although much useful information about articulation has been acquired through careful introspection, there is also a tradition of laboratory data acquisition from the 19th century to the most current technologies. Information about place of articulation can be had from ultrasound, Magnetic Resonance Imaging (MRI) and X-rays, palatography, or magnetometry. In dynamic electropalatography a set of contact electrodes on the hard palate records where the tongue touches. In magnetometry an electromagnetic field is established around the head of a speaker and the location of one or more receiver coils is tracked. Information about manner of articulation can be had not only from these techniques but also from records of airflow and pressure. Information about voicing can be had from electroglottography, which records contacts between the vocal cords.

C. OPTIONAL SECTION: Other IPA symbols

C1. Consonants and vowels

The IPA provides many more symbols than we have used so far. This is in part because other languages use sounds that English does not use. The IPA is meant to provide a symbol for every basic sound of every language. It is beyond the scope of an introductory course to present and discuss all of the symbols of the IPA, but if you understand the conceptual framework you should be able to cope with many of these symbols when you encounter them. In the consonant chart, there are many consonants which English does not use, but which are combinations of phonetic properties which English does use. For example, English has bilabial stops (oral [p,b] and nasal [m]) and labiodental fricatives ([f,v]), but not the reverse combinations, bilabial fricatives ([ɸ, β]) and labiodental stops (nasal [ɱ]) -- yet you should be able to figure out what these sounds must be, just from their definitions. Similarly, if you know how to make a particular front unrounded vowel such as [i], you should be able to understand the idea of the corresponding front rounded vowel [y] (keep your tongue in the same position, but round your lips as if for a back rounded vowel); and if you know how to make a back rounded vowel such as [u], you should be able to understand the idea of the corresponding back unrounded vowel [ɯ] (unround your lips, or even smile, while keeping your tongue in the same position). Other sounds are a little harder because they are more different from the basic sounds of English, but some experimentation might yield new phonetic skill. For example, since you can make a palatal approximant (or glide) [j], as in "yonder", you can try raising the center of your tongue until you make a voiced palatal fricative [ʝ], and then keep raising the tongue until it makes a complete seal for a voiced palatal stop [ɟ]. Nasalize it and you have [ɲ], a sound of Spanish (where it is spelled "ñ"), French (where it is spelled "gn"), and many other languages.

On the other hand, the charts include some dimensions which are not used in any basic sound of English. All of the consonants of English, and most of the sounds on the chart, are pulmonic egressive, meaning that the air comes out of the lungs. Note the separate chart for non-pulmonic consonants: clicks, voiced implosives, and (voiceless) ejectives. These are all sound types in which the airflow is established in some other way. For example, in implosives, there is a downward movement of the larynx during the stop closure. This downward movement gives the voicing during the stop a special strong quality, and at the same time it expands the size of the air cavity in the vocal tract. When the oral stop (e.g. at the lips in the bilabial [ɓ]) is released, air can also flow into the vocal tract rather than out of it, making the release strong also.

C2. Diacritics

The lower right part of the IPA chart contains a chart of diacritics, of which a few were presented in section B3 above. We noted before that the cells of the charts cover ranges of articulations. Diacritics serve to narrow down those ranges, and transcriptions using them are often called "narrow". A narrow transcription is used to represent small differences between speakers or languages, to show how a basic sound's exact value changes depending on the surrounding sounds, and to show differences between speech that is more or less careful, etc. However, in many cases diacritics are needed even for the basic sounds of a language. For example, a language that has nasalized as well as oral vowels will necessarily use the nasalization diacritic. Many of the diacritics in the chart raise subtle definitional issues that go beyond the scope of an introductory text like this one. (12) lists some diacritics that can be used for English and which appear in this chapter or the phonology chapters.

(12) Some IPA Diacritics

aspirated	◌ʰ	noise in the glottis, especially at the end of a consonant
syllabic	◌̩	a consonant without a vowel
rhoticity	◌˞	r-coloring (seen especially in vowel symbols for ɹ̩, [ɚ] and [ɝ])
dental	◌̪	upper teeth are passive articulator (used to modify basic symbols in Alveolar column)
nasalized	◌̃	air flows through nose as well as mouth
unreleased	◌̚	noisy release of consonant hold is not heard
voiceless	◌̥	partial or no vocal cord vibration in an otherwise voiced sound
velarized	◌ˠ	tongue backing during some other primary articulation

Finally, above the diacritics chart are some additional diacritics specifically for

suprasegmentals. The first is for primary, or main, stress, and has been used already in this chapter. The second is for secondary, or weaker, stress, which we will not cover. The next three convey lengthening and shortening of a consonant or vowel relative to its typical duration. For example, the next chapter includes a detailed discussion of how English vowels are shortened before voiceless consonants, transcribed with ˘ over the vowel. On the other hand, in some languages consonants or vowels can be roughly doubled in length, transcribed with : after the symbol. The next four diacritics represent breaks or connections between segments; the period to mark syllable divisions will be used in later chapters.

The next two columns contain alternative ways of marking tones. Thus the Kana word for "to fence", with a Low tone, can be transcribed [bè] or [be˩]; the Kana word for "home or compound", with a Mid tone, can be transcribed [bē] or [be˧]; and the Kana word for "fight", with a High tone, can be transcribed [bé] or [be˥]. We have not considered the tonal phenomena covered by the other tonal diacritics.

Intonational rises and falls of the pitch of the voice can be indicated by rising (↗) and falling (↘) arrows. For example, in "Is Hamlet upset?", a fall on "Hamlet" and a rise on "upset" can be transcribed as in (13), while in "What did Polonius say?", a rise followed by a final fall can be transcribed as in (14).

↘ ↗
(13) Is Hamlet upset?

↗ ↘
(14) What did Polonius say?

The next two chapters will use a variety of symbols and diacritics from the chart; you can refer back to the chart then to see what kinds of sounds they represent. Phonetics textbooks that give more detailed treatments of a range of sounds from the world's languages include Ladefoged (1992), Rogers (1991), Catford (1988), and Smalley (1989), and at a more advanced level, Laver (1994) and Ladefoged and Maddieson (1996).

EXERCISE 15: Give a symbol with a diacritic according to the description provided.

Example: **aspirated voiceless bilabial stop at the beginning of "pat"** = [p^h]

1. **syllabic alveolar nasal at the end of "sweeten"**
2. **voiced dental stop in "breadth"**
3. **nasalized high front vowel in "lean"**
4. **unreleased final voiced velar stop in "hag"**
5. **partially voiceless alveolar fricative in "buzz"**

EXERCISE 16: Add in arrows to indicate the intonational rises or falls described.

Example: **"Virtue? A fig!" with a rise on the first word and a fall on the last**

↗ ↘
"Virtue? A fig!"

1. **"It cannot be." with a fall on "be"**

2. **"Put money in thy purse." with a fall on "purse"**

3. **"Thou art sure of me." with a rise and fall on "sure"**

4. **"Do you hear, Roderigo?" with rises on "hear" and "Roderigo"**

5. **"How, is this true?" with a fall on "how" and a rise on "true"**

D. Conclusion

Just as native speakers of a language have unconscious knowledge about other aspects of linguistic structure, they have phonetic knowledge. They use the speech signal to recover many kinds of linguistic and non-linguistic information from an individual utterance: the words, the phrasing of those words (and therefore some aspects of syntactic structure), the speaker's attitude towards the utterance, personal and group characteristics of the speaker (such as regional accent, or mood). They know how to produce a native-sounding utterance, and they can usually judge such things as whether a given speech sample is a possible utterance in their language and whether it was produced by a computer or a human voice. They can also tell whether it was produced by a native speaker of the language, and by a speaker of a similar dialect of the language, that is, they can detect what are colloquially called "accents". More generally, people can judge how similar or dissimilar two sounds are. People can also automatically relate a sound that they hear to the articulation of that sound, in that they can imitate sounds they hear.

An interesting question is how much these kinds of knowledge depend on knowing the native language. Another way to pose this question is as follows: by knowing the phonetics of one language (say, English), how much do you know for free about the phonetics of another language (say, Kikuyu)? If you think about what it's like to begin the study of a new language, you can see that the answer is, "not much". There is not much information we could recover from a Kikuyu speech utterance. We certainly could not tell if the speaker had an accent -- probably we would not even recognize American-accented Kikuyu as such. And the abilities that seem more general, such as judging how similar two sounds are, or imitating, have become colored by our knowledge of a native language. Our knowledge of English will

probably *interfere with* our phonetic abilities in Kikuyu.

SOUND CORRESPONDENCES ACROSS LANGUAGES

It seems that when people confront the sound system of a new language, they make correspondences between the sounds of the new language and their native sounds. As much as they can, they equate new sounds with known sounds, as long as the sounds are somewhat similar. Suppose the native language is English, which has a basic sound [ʃ], while the new language has retroflex [ʂ] or alveolo-palatal [ɕ], either of which is similar to [ʃ] but not quite the same. English speakers will think of this as the "funny [ʃ] in that language". But then, having identified the new sound as corresponding to a known sound, even though imperfectly, English speakers will tend to use their [ʃ] in speaking that language, and that means they will have an English accent. Worse, the new language might have two of these -- "two different [ʃ]s" -- and since both are seen as corresponding to the native [ʃ], they get pronounced the same. On the other hand, sometimes the new sound is so different from any native sound that no correspondence can be made. Suppose an English speaker encounters click sounds. These are patently so unlike any native sound that the speaker realizes they just have to be learned. The speaker may not make them quite right, and so will have an accent, but it will not necessarily be an accent from the native language.

End-of-chapter exercises

EXERCISE 17: Give the regular English orthography for the following words, which are given in a broad transcription. The pronunciations given may not be like yours, but the words should be identifiable nonetheless.

1. bʊk
2. onli
3. pepɻ
4. aʊt
5. rimaɪnd (or rəmaɪnd)
6. stap
7. hɛd
8. θɪŋk

EXERCISE 18: Give broad transcriptions for the following pairs of English words. The focus here is on what makes the two words in each pair different. Use a dictionary if you like, but use IPA symbols. (Since each student may transcribe his or her own pronunciations, there can be no single correct answer here.)

1. spot - Scot
2. weary - worry
3. cue - few
4. lose - loose
5. man - men
6. woman - women
7. attend - Athens
8. size - seize
9. show - shoe
10. put - putt

EXERCISE 19: Give the regular English orthography for the following words, which are given in a narrow transcription. Again, these pronunciations may not be like yours.

1. [ˈpʰlɛ̃nti]

2. [bʊkˀ]

3. [tʃrækʰ]

4. [ˈmoɾr̩]

5. [sɛ̃nts] (some speakers will have more than one possible answer for this!)

EXERCISE 20: Give a broad transcription for the following words, which are given in a narrow transcription.

1. [ˈwʌ̃nd̩rf̩l̩]

2. [ˈʔæp̩l̩]

3. [pʰʊʔt]

4. [me̯ɪ̯]

5. [wãtˀ] (hint: what would cause the vowel to be nasalized?)

EXERCISE 21: Compare the sounds in each set below. In each set, all but one are in a single row or column on the IPA chart. Give the name of that row or column, and circle the sound which does not belong.

1. m n r ŋ

2. p t k v d g

3. p t s l n

4. f v s z h k

5. i e u ɛ æ

EXERCISE 22: Same as above, but here all but one sound in each set belong to a class of sounds that goes beyond a single row or column of the chart--classes such as *labial, coronal, dorsal, obstruent, sonorant, approximant, stop, fricative, voiced, voiceless, rounded, unrounded*--or within a single row or column, such as *voiced stops, labial stops*. Give the name of the class of sounds, and circle the sound which does not belong.

1. β v ð z ʒ h

2. i e æ u ʌ

3. m l r j w

4. m n k l r j

5. θ f ð s z ʃ ʒ

EXERCISE 23: Here are the basic sounds of Burera, an Australian language. Compare this set with the basic sounds of English to answer the questions that follow.

/p t c k m n ɲŋ ɾ l r (as in English) j w i ɛ a ɔ u/

a. Which sounds of Burera are not basic in English (as in (3) in the chapter)? You do not need to define these, just indicate them.

b. What labial sound(s) of English are not in Burera?

c. What coronal sound(s) of English are not in Burera?

d. What dorsal sound(s) of English are not in Burera?

e. What glottal sound(s) of English are not in Burera?

References

Adams, Marilyn J. (1990). Beginning to read: thinking and learning about print. Cambridge MA: MIT Press.

Catford, John C. (1988). A Practical Introduction to Phonetics. Oxford: Clarendon Press.

Dodd. B. and R. Campbell (editors) (1987). Hearing by Eye: The Psychology of Lip Reading, Hillsdale NJ: Lawrence Erlbaum Assoc.

Gough, Ehri, and Treiman (editors) (1992). Reading Acquisition, Hillsdale NJ: Lawrence Erlbaum Assoc.

Ladefoged, Peter (1993). A Course in Phonetics, third edition. Fort Worth: Harcourt Brace Jovanovich College Publishers

Ladefoged, Peter and Ian Maddieson (1996). Sounds of the World's Languages. Oxford: Blackwell Publishers

Laver, John (1994). Principles of Phonetics. Cambridge UK: Cambridge U. Press.

Pullum, Geoffrey K. and W.A. Ladusaw (1986). Phonetic Symbol Guide. Chicago: University of Chicago Press.

Rogers, Henry (1991). Theoretical and Practical Phonetics. Toronto: Copp Clark Pittman Ltd.

Smalley, William A. (1989). Manual of Articulatory Phonetics, revised edition. Lanham, MD: University Press of America. First edition 1961.

Focus Realization of Japanese English and Korean English Intonation[1]

Motoko Ueyama & Sun-Ah Jun

1. Introduction

In earlier work on intonation, not much attention has been paid to the interaction of the intonation of first (L1) and second language (L2). At the same time, most previous works on second language acquisition have concentrated on the segmental level (for reviews, see Leather and James 1991; Flege 1987, 1995). Furthermore, only a few studies on the acquisition of L2 intonation have been done based on instrumental evidence (Gårding 1981 for Greek French and Swedish French; Todaka 1990 for Japanese English; Argyres 1996 for Greek English), and these studies did not consider how the phonology and phonetics of L2 intonation interact with those of L1 intonation, but only concentrated on the phonetic description of tonal shapes. Todaka (1990) compared Japanese speakers' English intonation with native English intonation adopting Pierrehumbert's (1980) model of English intonation. But he did not compare the intonation contours in terms of the phonological components of intonation. Rather, he described the intonation contour as a holistic pattern. For example, he showed that while pitch for English speakers rose gradually after focus, and the high pitch was sustained in an interrogative sentence, Japanese speakers lowered pitch after the focus and sustained low pitch in the same sentence. Schematic figures are shown in Figure 1.

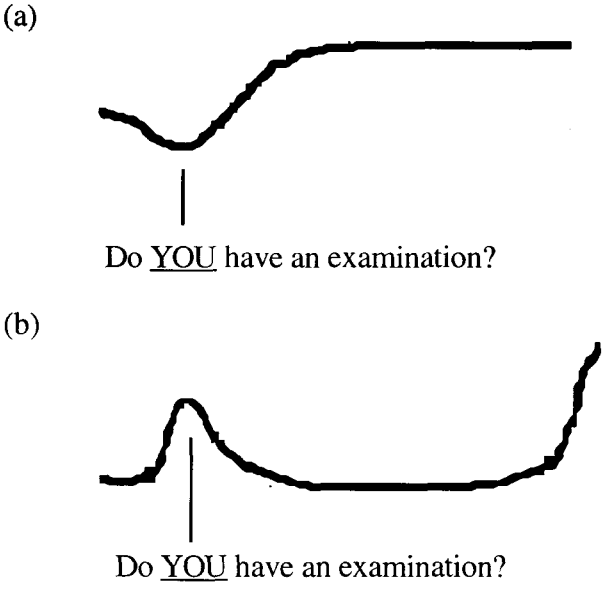


Figure 1. Schematics of intonation contour of English interrogative produced (a) by English native speakers and (b) by Japanese speakers.

In this paper, we adopted the phonological model of English intonation proposed by Pierrehumbert and her colleagues (e.g. Pierrehumbert 1980, Beckman & Pierrehumbert 1986, Pierrehumbert & Beckman 1988) to compare the phonological characteristics of English intonation produced by native speakers of English with those of English intonation produced by Japanese and

[1]This paper will be published in the proceedings of the 7th Japanese and Korean Linguistics Conference.

Korean speakers. In this model, "continuous" intonation contours are analyzed as sequences of H and L tones. These underlying tones are categorized as one of three types; pitch accents, phrasal tones, and boundary tones. The *pitch accent* is associated with the stressed syllable of the phrase, and by this association, the stressed syllable of a certain word gets pitch prominence. The *boundary tone* marks the end of a phrase and it is associated with an Intonational phrase. The *phrasal tone* covers the space between the last pitch accent and the boundary tone. In English, there are six types of pitch accents (H*, L*, H+L*, H*+L, L+H*, L*+H), two types of phrasal tones (L-, H-), and two types of boundary tones (L%, H%). In Pierrehumbert's model, these three types of tones (pitch accent, phrasal tone, and boundary tone) are hierarchically organized so that one Intonational phrase can have more than one intermediate phrase, and one intermediate phrase can have more than one pitch accent but should have at least one pitch accent. When there is more than one pitch accent in one intermediate phrase, the last pitch accent is the most prominent and labeled as the nuclear pitch accent. In English, focused words receive nuclear pitch accents. That is, if a word is narrowly focused, the stressed syllable of the word will have pitch accent and it will not be followed by any pitch accent, but only by a phrasal tone and a boundary tone. If the sentence is an interrogative, the stressed syllable of the focused word will have a low tone (L*) followed by a high phrasal tone (H-) and a high boundary tone (H%). This tonal contour is described as L* H- H%. If the sentence is a declarative, the stressed syllable of the focused word will have a high tone (H*) followed by a low phrasal tone (L-) and a low boundary tone (L%), i.e. H* L- L%. That is, the words after the focused word will be completely deaccented and any phrase boundary will be canceled, thus either a L-plateau or a H-plateau will arise. Schematic contours of interrogative and declarative sentences with sentence initial focus are shown in Figure 2. In the interrogative utterance, the final H% is upstepped after phrasal H- tone, showing a higher f0 value than phrasal H tone.

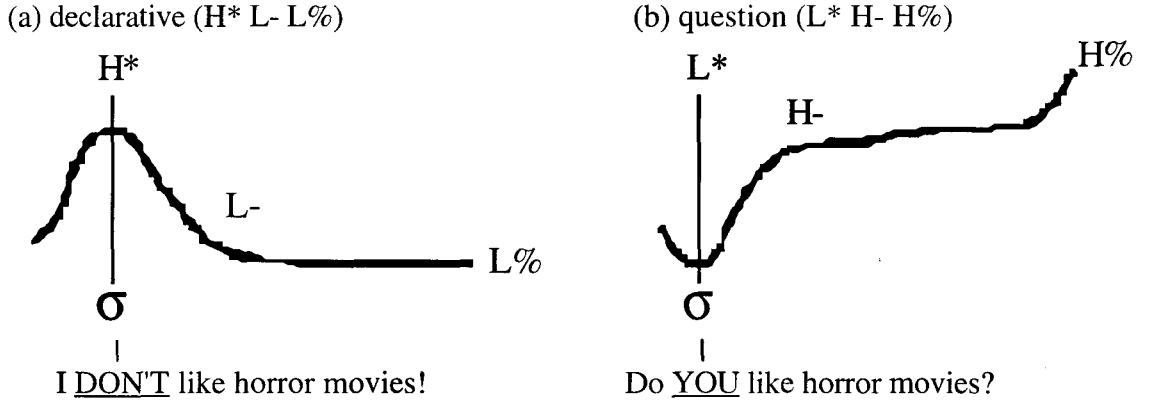


Figure 2. Schematic contours of interrogative and declarative in English

This model has been applied to Korean by Jun (1993) and to Japanese by Beckman & Pierrehumbert (1986) and Pierrehumbert & Beckman (1988). Thus we can easily identify the features of L1 (Korean or Japanese) intonation that are transferred into L2 intonation (English). This will allow us to analyze how the L2 intonation system interacts with the L1 intonation system.

The first goal of this paper is to examine the realization of English focus produced by Tokyo Japanese speakers and Seoul Korean speakers (henceforth Japanese English and Korean English, respectively) at different proficiency levels. The second goal is to analyze the phonology (underlying tonal sequence) and phonetics (actual realization) of L2 intonation produced by learners of different proficiency level, using Pierrehumbert's intonation model. We will try to answer to the following questions: (1). How does the L1 intonation system (Japanese vs. Korean) affect L2 (English) intonation patterns?; (2). Which factors characterize different proficiency levels?

To analyze L2 intonation produced by Korean and Japanese learners, we need to know how focus is realized in the two languages. Although the prosodic systems of Tokyo Japanese and Seoul Korean are typologically different in that Japanese has a lexical pitch accent with a H*L tonal pattern (Pierrehumbert & Beckman 1988) while Korean has a postlexically determined accentual phrase with a LHLH or HHLH tonal pattern (Jun 1993, 1996), the realization of focus is very similar, both phonetically and phonologically. Schematics of focus realization of interrogative and declarative in Korean and Japanese are shown in Figure 3. To help comparison, the focus patterns of English are shown again.

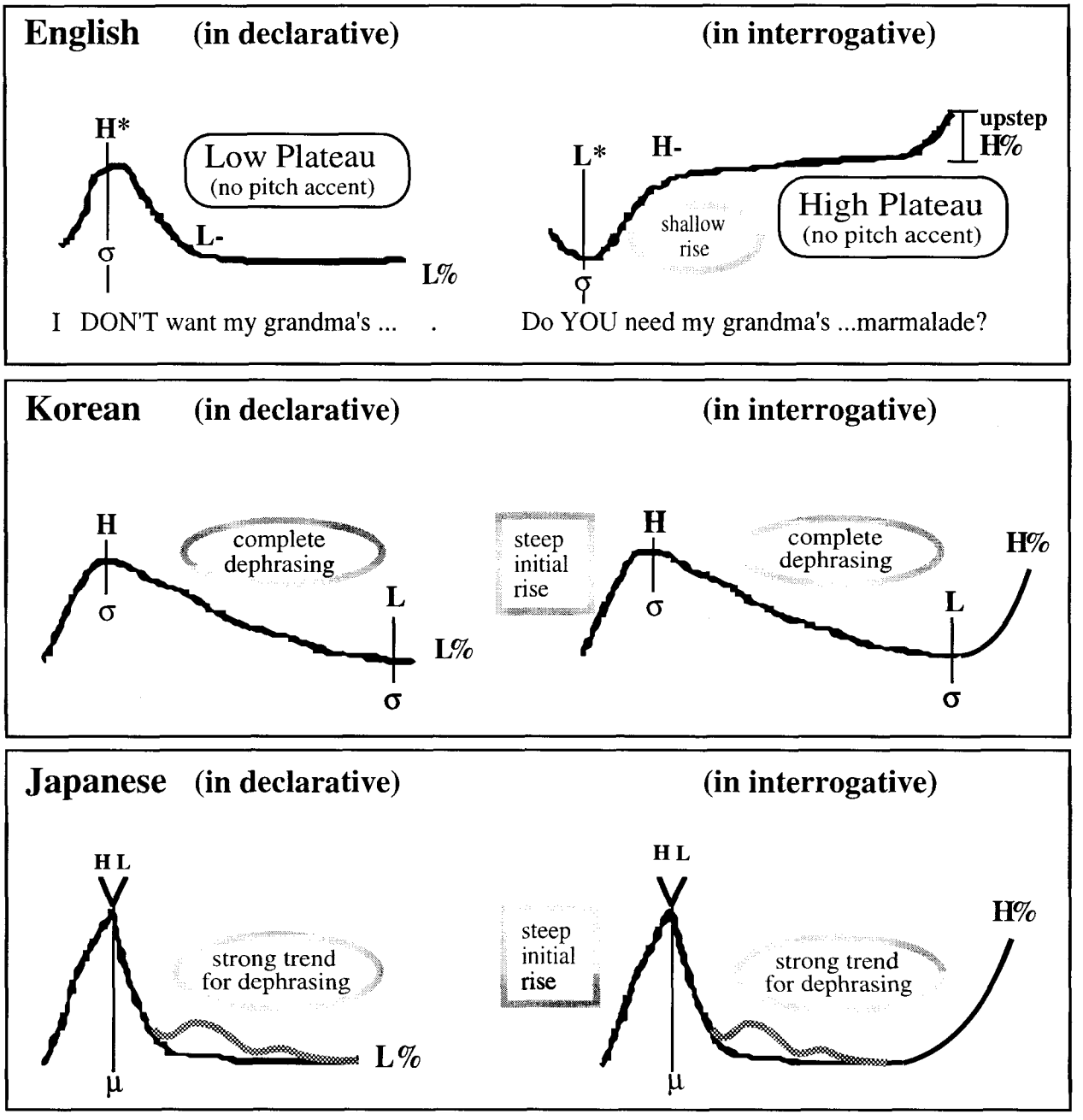


Figure 3. Schematics of focus realization in declarative and interrogative in English, Korean and Japanese.

In all three languages, focus is realized with higher pitch range and larger amplitude, but the tonal contour of the focused phrase differs across languages. In English, focus in interrogatives is realized as a low tone (L*) followed by a high plateau (H- H%) while focus in declaratives is realized in as high tone (H*) followed by a low plateau (L- and L%). In both cases, any pitch accent after focus is deaccented. In Korean, focus is always cued by a phrase initial H tone followed by a L tone regardless of the sentence type. The only difference between interrogatives and declaratives is in the boundary tone. The boundary tone for an interrogative utterance is high (i.e., H% or LH%, see Jun & Oh 1996) while the boundary tone for a declarative utterance is low (i.e., L%, HL%, or LHL%). As shown in the figure, there is complete dephrasing after focus in Korean. That is, there is no phrase boundary after the focused word until the boundary tone. In this respect ('no-pitch-accent' after focus) Korean is similar to English.

In Japanese, the tonal pattern of the focused phrase is similar to that of Korean, in that there is no different tone type for focus depending on the sentence types (interrogative or declarative). The two types differ only with respect to the boundary tones: H% for an interrogative and L% for a declarative focused phrase. But, unlike in Korean, dephrasing after focus is not always complete in Japanese. Pierrehumbert & Beckman (1988, Figure 4.7) show that there are cases where dephrasing does not occur, and a H tone on the word following the focused word is realized. But they also note that after focus dephrasing is common (p.105). When an accented word is focused, all following words within the phrase are dephrased, and the focused phrase shows a low plateau after the pitch accent on the focused word (H*L). Maekawa (1994) shows that though post-focus dephrasing is realized as a low plateau contour, the slope of the low plateau is sharper (negatively) when the focused word is followed by an accented word than when it is followed by an unaccented word. In any case, however, no high tone is realized after focus in both cases.

Another difference between Japanese and Korean focus realization is the degree of H tone sustaining. In Korean, the focused H tone is realized on the second syllable of the accentual phrase. When the phrase initial syllable has an underlying H tone, high tone maintains over two syllables. On the other hand, in Japanese, when the focused word is unaccented and the following word is accented, the focused phrase has a high tone covering a few moras from the initial mora of the phrase up to the accented mora of the second word. Thus, compared to Korean, a high tone can be sustained longer in Japanese than in Korean. Since the phrasal tone of English interrogative is a high tone covering many syllables after the focused word, up to the phrase final syllable, it is interesting to see how English high plateau is realized differently by Japanese learners and Korean learners.

So far, we have discussed differences in the intonational phonology of each language. But the realization of underlying tone sequences also differs across these languages. For example, the slope of initial f0 rise in the interrogatives differ between English and Korean/Japanese. In English, after the focused word L*, f0 reaches its peak around the *end of the following word.* On the other hand, in Korean and Japanese, f0 (after the phrase-initial L tone in the focused phrase) reaches its peak around the *second syllable/mora of the phrase*, thus creating a sharper rising slope than that in English. Korean and Japanese further differ from English in that a focused word initiates a new left-headed phrase in both Japanese and Korean, while it delimits a right-headed phrase final boundary in English (Venditti et al 1996). We will however not discuss this point further in this paper since this paper is concerned with the phrasing differences *after* the focused word.

Many studies on the segmental aspect of L2 speech learning show that the phonology and phonetics of L1 interacts with L2 speech production and the degree of interaction differs depending on the degree of proficiency in L2 (e.g., Weinreich 1953; Flege & Davidian 1987; Flege 1995). Therefore, we hypothesize that the phonetics and phonology of intonation of L1 will interfere with the acquisition of L2 intonation, and the degree of interference will differ depending on the degree of proficiency in L2. Under these assumptions, we will test the following hypotheses. Each hypothesis will be evaluated based on the learner's level of proficiency in L2.

1. Since Korean and Japanese have dephrasing after narrow focus, English dephrasing after focus should be easy to learn. Furthermore, dephrasing should be particularly easy for Korean learners, since it is obligatory in Korean but it is only a strong tendency in Japanese.

2. Since both Korean and Japanese have no high-plateau pattern comparable to the one found in English interrogatives while both languages have a pattern similar to low-plateau in English declaratives, the high-plateau pattern should be more difficult to learn than the low-plateau pattern.

3. Since Japanese speakers sustain high pitch across word boundary (e.g. between phrase-initial H and accent H*L on the second word) while Korean speakers do not, high-plateau should be easier to learn for Japanese learners than for Korean learners.

4. Since in interrogatives both Korean and Japanese raise pitch up to the peak with a sharper slope than English does, the initial rise of interrogative focus should be sharper in non-native speech than in native speech.

2. Experiment

2.1. Subjects

Two native speakers of American English participated as the control group, and Japanese learners of English and four Korean learners of English formed the experimental groups. All the speakers were females in their 20s or early 30s. To find a developmental path in L2 intonation acquisition, we compared different proficiency levels within each experimental group. The three Japanese learners were categorized as advanced, respectively, intermediate and beginning level, while four Korean learners were categorized as two intermediate and two beginning level speakers of L2 English. The following table shows the description of each learner with respect to the number of years of residence in the United States and their age of arrival in the States.

Table 1. Description of each learner with respect to years of residence in the States and their age of arrival.

Speakers	Years of Residence	Age of arrival
Japanese Advanced	5 years	16
Japanese Intermed.	5 years	20
Japanese Beginning	0 year	0
Korean Intermed. 1	6 years	18
Korean Intermed. 2	5 years	26
Korean Beginning 1	2 months	22
Korean Beginning 2	1.5 months	29

2.2 Corpus

Test sentences were designed to check whether focus realization differs between declaratives and interrogatives and whether the location of focus influences its realization in either type of sentences. We constructed two data sets; in Set 1 the length of a noun phrase varied from three words to six words while the location of focus was kept constant at the beginning of a sentence. In Set 2 the location of focus varied while the sentence length was kept constant. The sentences of Set 1 are reported in (2); those of set 2 are reported in (3). The words in bold are focused. To

trigger focus in each sentence, a monologue was designed for each interrogative sentence in Set 1, and a dialogue was designed for each declarative sentence in Set 1. (1) shows an example monologue and dialogue for set 1. For set 2, focus was triggered by putting a phrase in parenthesis right before each sentence. For example, the phrase "(not an apple, but)" was given for the sentence with focus on *potato*.

(1) Monologue for H-plateau (interrogative):
A1: None of my friends needs my grandma's (0~3 Ns) marmalade.
A2: Do **YOU** need my grandma's (0~3 Ns) marmalade?
Dialogue for L-plateau (declarative):
A: Do you want my grandma's (0~3 Ns) marmalade?
B: No. I **DON'T** want your grandma's (0~3 Ns) marmalade.

(2) A. Interrogative
1. Do **you** want my grandma's marmalade?
2. Do **you** want my grandma's orange marmalade?
3. Do **you** want my grandma's Mandarin-orange marmalade?
4. Do **you** want my grandma's homemade Mandarin-orange marmalade?

B. Declarative
1. I don't want your grandma's marmalade.
2. I **don't** want my grandma's orange marmalade.
3. I **don't** want my grandma's Mandarin-orange marmalade.
4. I **don't** want my grandma's homemade Mandarin-orange marmalade.

(3) A. Interrogative
1. Is this yellow **potato** a source of vitamins?
2. Is this yellow potato a **source** of vitamins?
3. Is this yellow potato a source of **vitamins**?

B. Declarative
1. This yellow **potato** is a source of vitamins.
2. This yellow potato is a **source** of vitamins.
3. This yellow potato is a source of **vitamins**.

2.3 Procedure

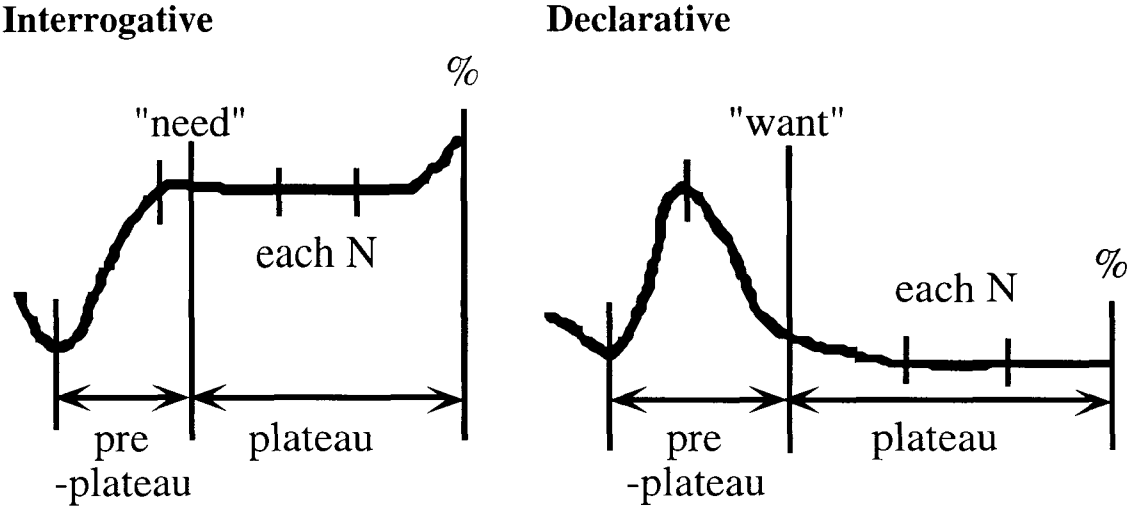
For Set 1 data, the dialogue/monologue sequences in each set were randomized and foil sentences were inserted pseudo-randomly so that each sequence of target sentences was separated by a foil sentence. Similarly, sentences in the Set 2 data were randomized and mixed with foil data. The entire list was repeated six times by each subject. Subjects' utterances were recorded in a sound-proofed room. The speech data were digitized and the pitch contours (fundamental frequency (=f0) tracks) were analyzed using Entropic's XWAVES+ software.

2.4 Measurements

For the phonological description of L2 intonation from a phonological perspective, the type of pitch accents and phrase boundaries occurring in each utterance was labeled using Pierrehumbert's (1980) model of English intonation. In addition, the number of pitch accents and phrase boundaries after the pitch accent and before the boundary tone was counted. For the phonetic analysis of L2 intonation, f0 and absolute time point at several points in each utterance were collected using Xwaves+ software. Each utterance was divided into two parts, *pre-plateau* and *plateau*, with reference to the center of the vowel in the verb, "need" or "want". These verbs were taken as reference points since they coincide with the beginning of plateau following the H or L

nuclear pitch accent of the focused word. Within the pre-plateau region, the lowest f0 and the highest f0 were measured together with their corresponding time values. Within the plateau region, the f0 and time value of the highest f0 point of each noun as well as the f0 and time value of the utterance-final point were measured. Figure 4 shows schematics of measurement points in the interrogative and declarative sentences.

Figure 4. Schematics of measurement points in the interrogative and declarative sentences.



3. Results and Discussion

3.1 Phonology of L2 English Intonation

Since dephrasing is one of the main characteristics of Japanese and Korean focus intonation, we hypothesized that dephrasing would be easy for both learners. In addition, since dephrasing is obligatory in Korean but only a strong tendency in Japanese, we expected that dephrasing would be easier for Koreans. However, the results shown in Table 2 suggest that this is not true. In this table, the number of stars (*) indicates the number of pitch accents in each phrase and 'phr' refers to the number of phrase boundaries between the focused word and the boundary tone. Since there is no pitch accent after focus (**you** or **don't**) in English intonation, we expect that there would be only one star and no phrase boundary in each sentence. Since each sentence type was repeated six times, there should be a total of six stars and zero phrase boundaries. This is true for the native English speaker (data for only one speaker is shown). In Table 2 on the next page, table cells showing the native pattern, i.e. 6 stars and 0 'phr', are shaded. Data for the Korean beginners and J-Adv's 3 noun data are not available.

For non-native speakers, four factors influence the degree of dephrasing. First, the degree of dephrasing differs depending on the proficiency level of L2 learners: the more fluent the speaker is, the greater is the degree of dephrasing. J-Adv is very close to the native English speaker, but the Japanese intermediate speaker shows a higher number of pitch accents (*) and phrase boundaries than the advanced speaker. J-Beg has the highest number of stars and phrase boundaries.

Second, the degree of dephrasing differs depending on the length of the controlled noun phrase (NP). For all the learners, it is easier to dephrase when the number of nouns in the NP is smaller. As the length of the NP increases, the number of pitch accents and phrase boundaries increases. That is, the degree of dephrasing correlates with proficiency level and with the length of NP. This suggests that L1 dephrasing is not positively transferred to L2. Rather, there may be an independent constraint on L2 speech learning. Less advanced L2 learners may be able to parse

fewer words within the same prosodic phrase. The size of the maximal phrase will increase as they become more fluent in L2.

Third, dephrasing was better for Japanese than for Korean speakers when we compared learners with the same level of proficiency. For example, Japanese and Korean intermediate learners show a similar pattern of dephrasing for H-plateau. However, the Japanese learner shows better dephrasing than the Korean learner for L-plateau. This suggests that Japanese downstep is being positively transferred to L2 production and the lack of downstep in Korean is negatively transferred to L2 production. In Japanese, the pitch accent of an accented word following another accented word is downstepped relative to the preceding pitch accent, i.e. it is realized with a lower f0 peak relative to that of the preceding pitch accent. Dephrasing would be easier for Japanese speakers with the help of L1 downstep. Alternatively, the speakers might have produced a certain noun with a pitch accent, but due to its downstepped lower pitch level, we did not perceive it as a pitch accent. In Korean, on the other hand, phrases are marked by a phrase-final H tone without downstep. Thus it may be harder for Korean speaker to reduce pitch level, and at the same time, it is easier for us to perceive the phrase boundary.

Finally, Table 2 shows that for all speakers are better at H-plateau dephrasing than at L-plateau dephrasing. This could have two possible explanations. One is that the fluctuation in the higher pitch range may be more difficult than the one in the lower pitch range. The other is that H-plateau is easier to learn than L-plateau because of Equivalent Classification hypothesis (Flege 1987, 1995). According to this hypothesis, a new pattern of L2 is easier to learn than a similar but not identical pattern. For example, since French /y/ is a new sound to English speakers while French /i/ is a similar but identical to English /i/, French /y/ is usually easier to learn by English speakers than French /i/. Thus, since H-plateau is a new pattern for both Japanese and Korean learners, while L-plateau is similar to L1 (Korean and Japanese) pattern, H-plateau may be easier to learn than L-plateau.

Table 2. Number of pitch accents (*) and phrase boundaries (phr) from focused word to the end of the utterance for Set 1. Each cell has six tokens. J-Adv = Japanese advanced speaker, J-Int = Japanese intermediate speaker, J-Beg = Japanese beginning speaker, and K-Int = Korean intermediate speaker. Table cells showing the native pattern, i.e. 6 stars and 0 'phr', are shaded.

High-plateau

	English		J-Adv		J-Int		J-Beg		K-Int	
NP	*	phr	*	phr	*	phr	*	phr	*	phr
3 N	6	0	6	0	6	0	16	6	6	0
4 N	6	0	6	0	10	2	19	6	8	2
5 N	6	0	6	0	17	4	18	6	17	2
6 N	6	0	6	1?	22	5	19	6	20	2

Low-plateau

	English		J-Adv		J-Int		J-Beg		K-Int	
NP	*	phr	*	phr	*	phr	*	phr	*	phr
3 N	6	0	----	----	10	2	16	6	17	2
4 N	6	0	6	0	15	3	20	9	20	3
5 N	6	0	6	0	19	6	18	6	25	6
6 N	6	0	10	3	25	7	23	7	34	5

When the phrase after focus is shorter than three words as in our Set 2 data, the difference between proficiency levels and that between H-plateau and L-plateau did not clearly emerge. Table 3 shows the number of pitch accents and phrase boundaries for Set 2 data. In the 'initial'

condition, the focused word is "potato", followed by two other nouns (*source & vitamins*); in the 'medial' condition, the focused word is "source", followed by one noun (*vitamins*); and in the 'final' condition, the focused word is the final word of a sentence (*vitamins*). As shown in this table, all the learners are better at dephrasing as the length of the phrase after focus decreases from 'initial' to 'final'. This tendency is very similar across H-plateau and L-plateau data. For this corpus, data for the Japanese learners are not available.

Table 3. Eng-1 and Eng-2 = native English speakers, K-Int 1 and K-Int 2 = Korean intermediate speakers, K-Beg 1 and K-Beg2 = Korean beginning speakers. Table cells showing the native pattern, i.e. 6 stars and 0 'phr', are shaded.

H-plateau

	Eng 1		Eng 2		K-Int 1		K-Int 2		K-Beg 1		K-Beg 2	
focus	*	phr	*	phr	*	phr	*	phr	*	phr	*	phr
initial	6	0	6	0	18	6	13	6	17	6	18	6
medial	6	0	6	0	12	0	7	0	12	0	12	0
final	6	0	6	0	6	0	6	0	6	0	6	0

L-plateau

	Eng 1		Eng 2		K-Int 1		K-Int 2		K-Beg 1		K-Beg 2	
focus	*	phr	*	phr	*	phr	*	phr	*	phr	*	phr
initial	6	0	6	0	17	0	16	6	18	6	18	6
medial	6	0	6	0	6	0	7	0	12	4	7	1
final	6	0	6	1	6	0	6	0	6	0	6	0

3.2 Phonetics of L2 English Intonation

So far, we have shown that the phonology of H-plateau was easier to learn by both Japanese and Korean learners of English. The next question is whether H-plateau is also phonetically easier to learn than L-plateau. That is, is sustaining high pitch easier than sustaining low pitch? The results show that L-plateau is easier to produce than H-plateau at least for advanced and intermediate speakers. Figure 5 shows the progression of f0 values starting from the verb and through each of the following nouns both in H-plateau and L-plateau utterances, produced by a native English speaker (EZ). Both the H-plateau and L-plateau of the English native speaker are very flat, and the H-plateau is even slightly rising as the length of NP increases.

Figure 6 shows f0 values for the H-plateau and L-plateau for the Japanese Advanced speaker, the Intermediate speaker and the Beginning speaker. The H-plateau of the Japanese Advanced speaker is not as flat as in the native speaker's H-plateau, but falls down after the verb. Her L-plateau, on the other hand, is similar to the native speaker's L-plateau except for one peak at the second noun, *homemade*, in the six-noun-NP sequence. This suggests that her H-plateau is phonetically "worse" than her L-plateau. This is interesting since her H-plateau is phonologically almost perfect, and indeed is phonologically better than her L-plateau (see Table 1). This suggests that the phonology and phonetics of L2 intonation are not necessarily learned at the same pace. For this speaker, the phonology of intonation is closer to native speakers than her phonetics. The H-plateau of the Intermediate speaker is also phonetically worse than her L-plateau, while her phonological L-plateau is worse than her H-plateau. Both the H-plateau and the L-plateau of the Beginning speaker show a wide range of pitch movement away from the plateau target, suggesting that she is not differentiating between H-plateau and L-plateau.

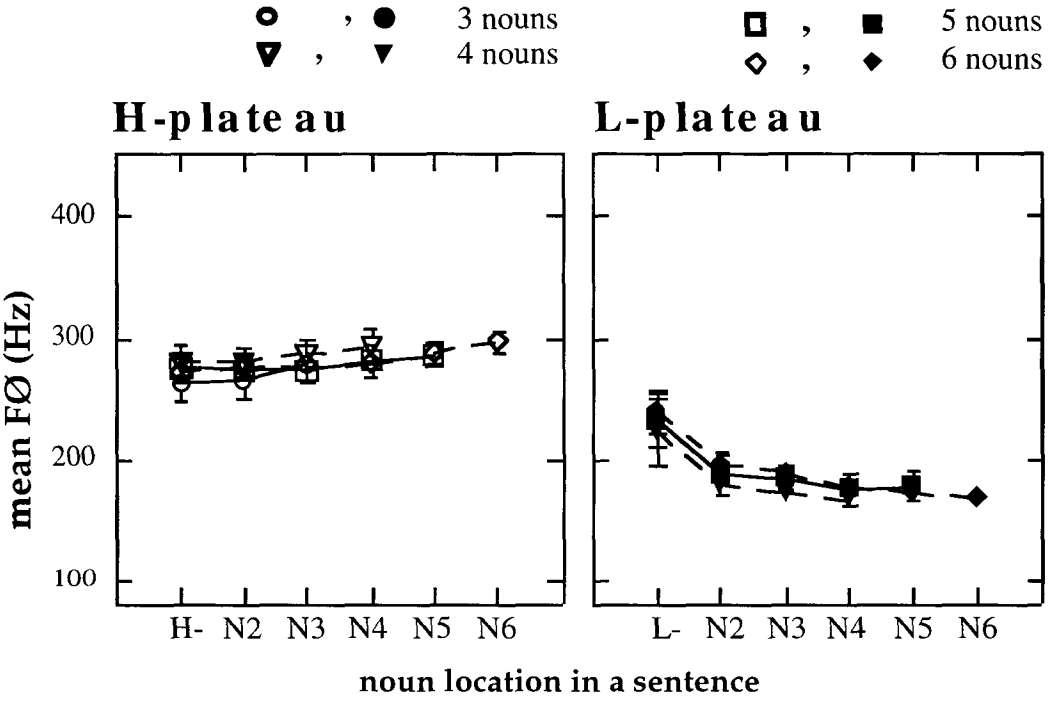


Figure 5. F0 values from the verb and through each of the following nouns both in H-plateau (left graph) and L-plateau (right graph) produced by Native English speaker 1.

A similar tendency was found in the phonetic data of Set 2, as shown in Figure 7. Fig 7a shows H- and L-plateau of the Korean Intermediate learners and Fig. 7b shows that of the Korean Beginning learners. Again, when the length of the phrase following focus is shortest, i.e. in the 'final' condition, all non-native speakers produced H-plateau and L-plateau very similar to those of the native English speakers. When focus was followed by two words, both intermediate and beginning level speakers produced the interrogatives as a L-plateau patterns except for the boundary tone. When the focused word was followed by one word, both group of speakers produced L-plateau closer to native pattern than the respective H-plateau.

We mentioned earlier that high tones are sustained longer in Japanese than in Korean within the same focused phrase. Thus we expect that sustaining a high pitch in producing English interrogative focused utterances should be easier for Japanese speakers. The data show that there is no significant difference between the learner groups. As illustrated in Figure 8, both Korean and Japanese intermediate learners show a similar degree of pitch drifting, with a better high pitch for short NPs (i.e. 3 Nouns). The general difficulty to sustain f0 in the high pitch range may be due to physical constraint against holding high pitch for a long time period.

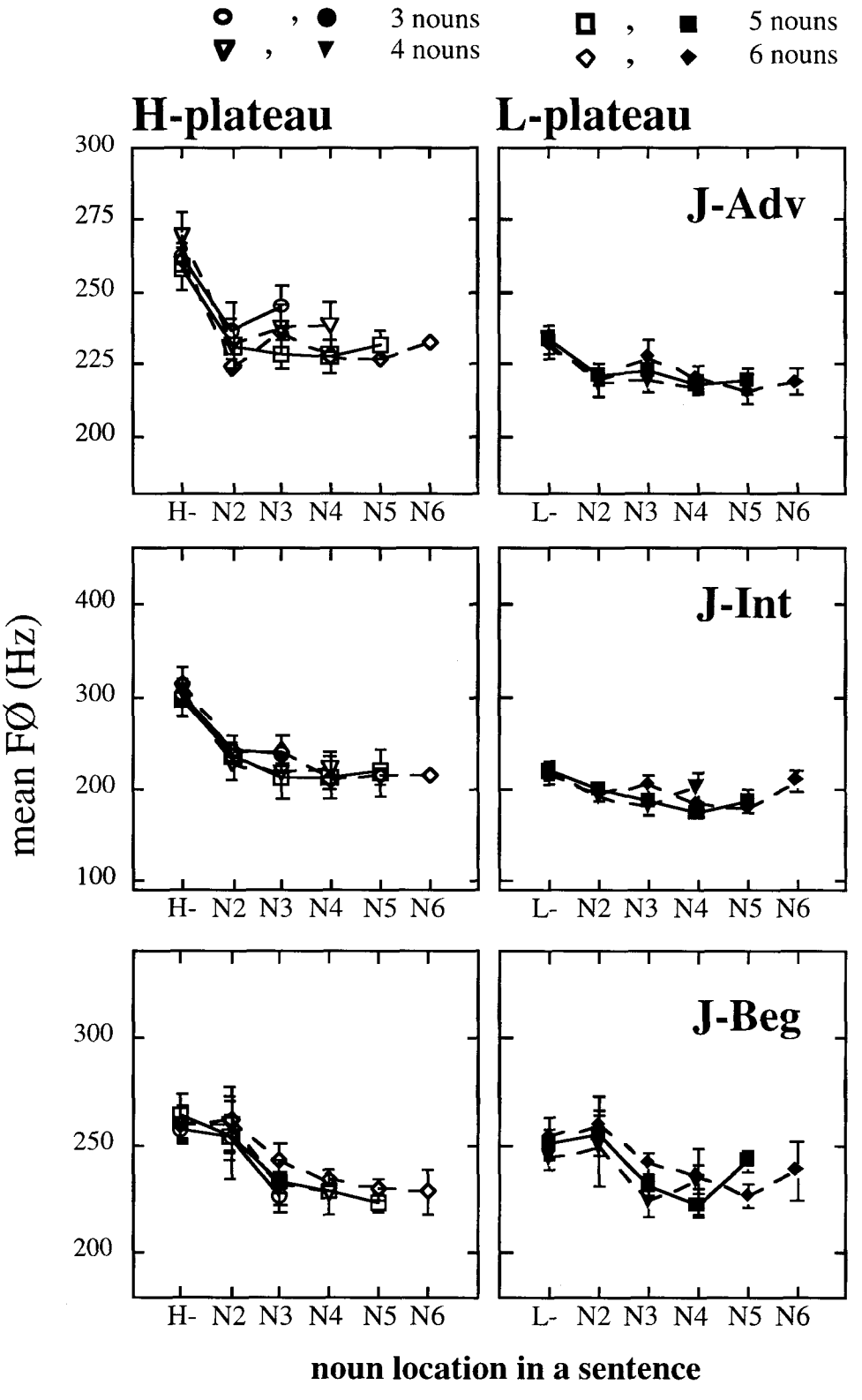


Figure 6. F0 values from the verb and through each noun in H-plateau and L-plateau for three Japanese speakers, advanced, intermediate and beginning level.

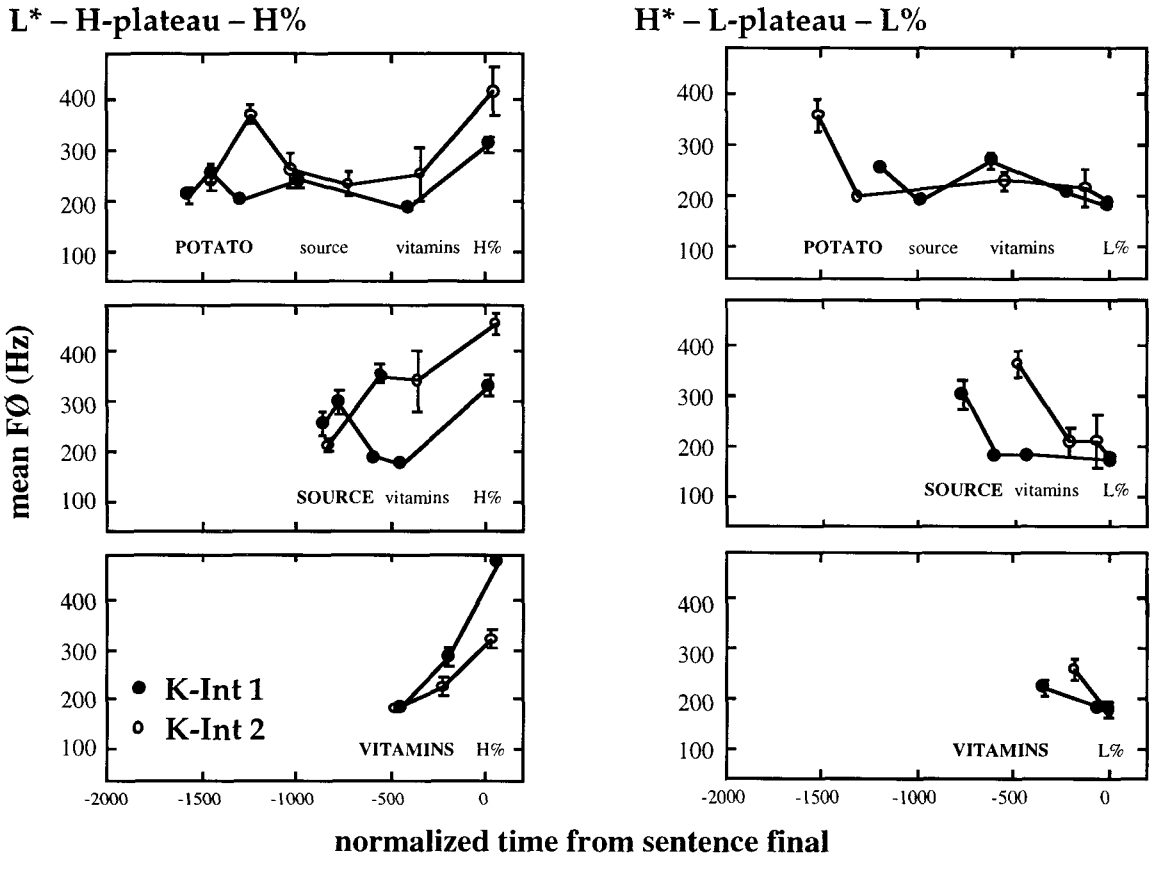


Figure 7a. F0 value of the focused word, the following noun(s) and the boundary tone (also the valley before peak for H-plateau) for two Korean intermediate learners.

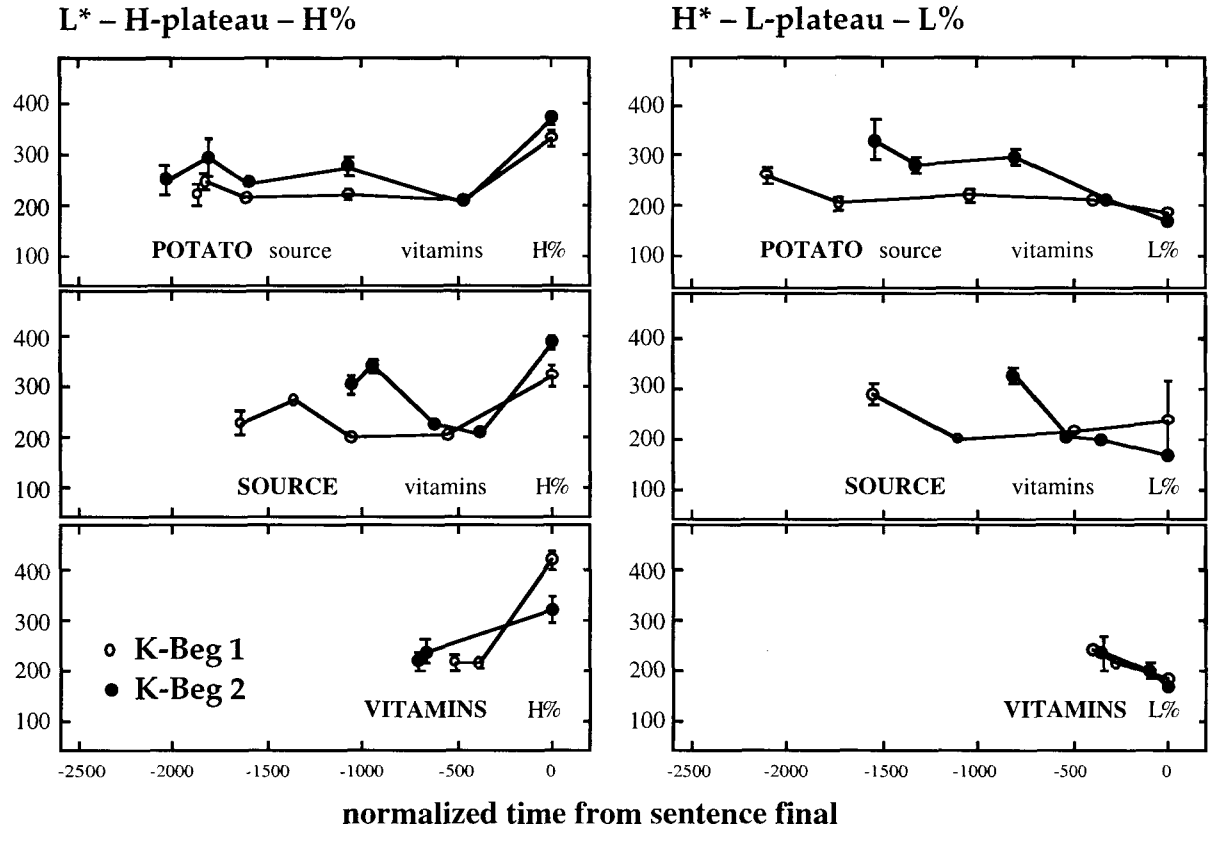


Figure 7b. F0 value of the focused word, the following noun(s) and the boundary tone (also the valley before peak for H-plateau) for two Korean beginning learners.

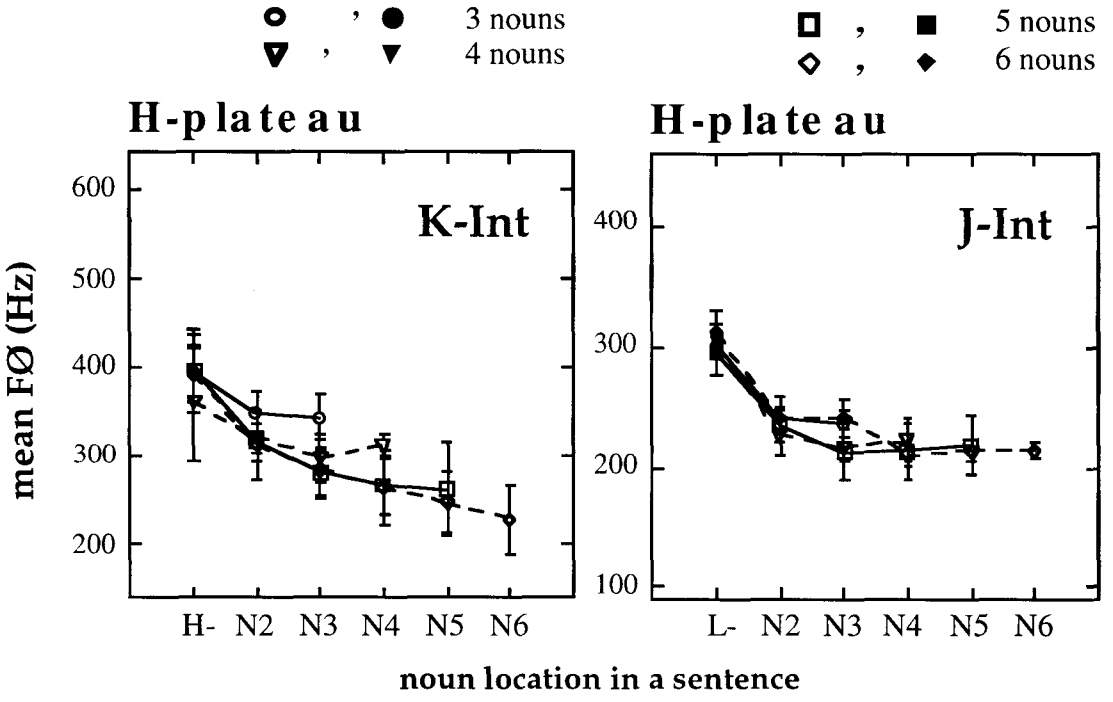


Figure 8. F0 value of H-plateau produced by Korean intermediate speaker 2 and Japanese intermediate speaker.

Finally, we hypothesized that the slope of initial rise in interrogative sentences should be sharper for non-native speakers than for native English speakers. As shown in Figure 9, this hypothesis is tenable. The figure displays the slope of the initial rise for each speaker. Short bars symbolize very shallow slopes and tall bars symbolize very sharp slopes. Results of ANOVA tests show that there is a significant main effect of the group factor on the slope of the initial rise ($F(6,158) = 125.782$; $p < .0001$). Scheffe's S posthoc test showed no significant difference among the bars that we represent with the same colors. Bars of different color are significantly different at 0.01 level. That is, the slope of the two English native speakers is significantly shallower than that of the non-native speakers. Among the non-native speakers, the slope is significantly shallower for the advanced and intermediate speakers than for the beginning speaker. The slope of the Japanese Advance Speaker (JA) was not significantly different from that of the Korean intermediate speaker 1 (KI1) and from that of the Japanese intermediate speaeker (JI). The slope of the Japanese intermediate speaker was significantly shallower than that of the Korean intermediate speaeker (KI2). These results suggest that the sharp slope in L1 pitch rise is negatively transferred, and that the degree of negative transfer decreases as the proficiency level increases.

To sum up, we found that we can characterize different proficiency levels in the production of English focus intonation in terms of the four different factors. A schematic representation of the four factors is presented in Figure 10. First, the slope of the initial rise in H-plateau is shallower for the native speakers, and it gets steeper as the proficiency level decreases. Second, the slope of H-plateau is slightly rising for native speakers, and flat or slightly falling for advanced speakers, but strongly falling for beginning level speakers. Third, the frequency of H-plateaus, that is, how often speakers produce high plateau, decreases as the proficiency level decreases: native speakers always produced H-plateaus; advanced or intermediate speakers were successfully producing H-plateaus about 50 percent of the times while beginning level speakers rarely produced H-plateaus. Finally, the number of pitch accents after focus increases as the proficiency level decreases. That is, compared to native speakers, dephrasing is more perfect in the advanced speaker than in the less advanced speakers.

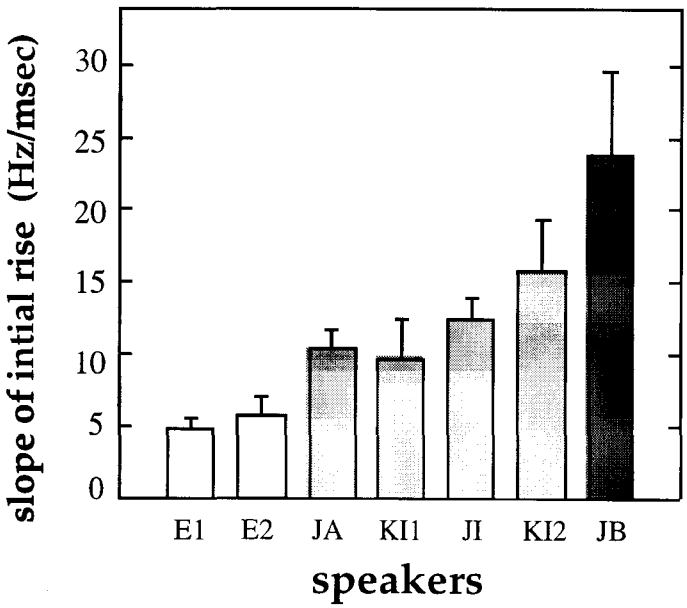


Figure 9. Slope of initial rise in interrogative (H-plateau) by native English speakers and all non-native speakers.

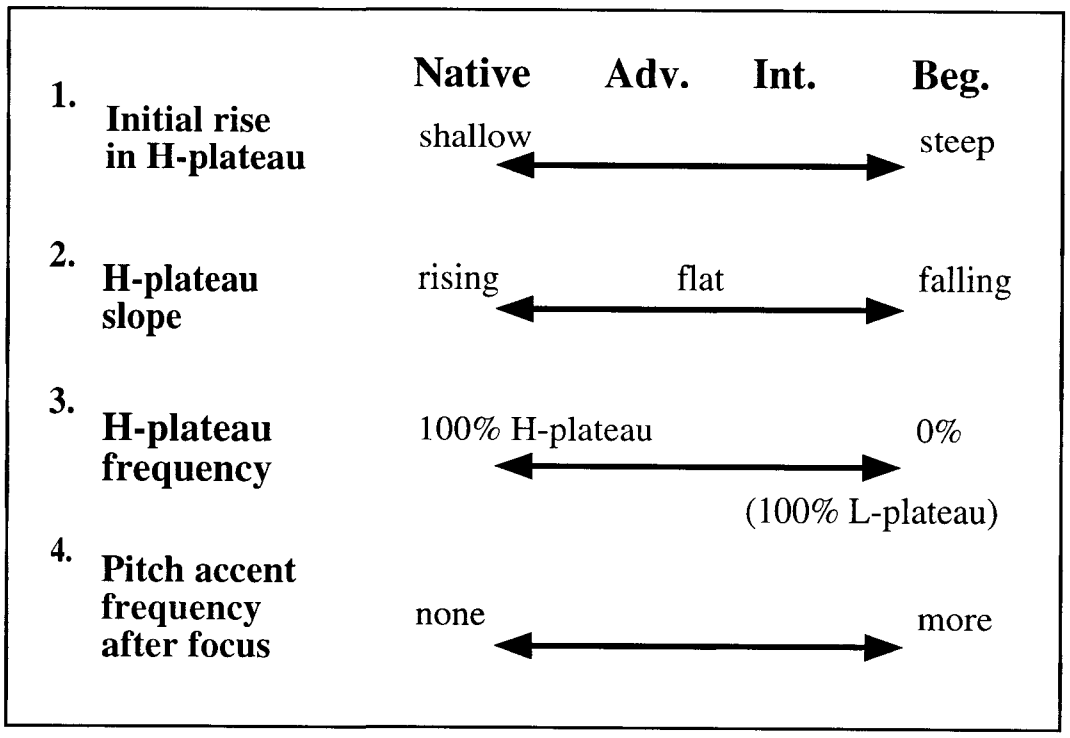


Figure 10. Schematic representation of factors characterizing different proficiency levels.

4. Conclusion

In conclusion, we have shown that L1 intonation system affects L2 (English) intonation pattern, irrespective of the L1 source. However, not all L1 features directly shape L2 intonation. Rather, they interact with universal constraints on speech production, such as the tendency to avoid dramatic fluctuations in the high pitch range, and with constraints on L2 speech learning, such as the tendency to reduce phrase sizes in the beginning stages of L2 learning. We also proposed four factors characterizing the level of L2 proficiency, based on the speaker's production of English focus intonation.

Acknowledgments

We would like to thank the members of the Phonetics Seminar group at UCLA, and especially Marco Baroni, Cécile Fougeron, Sean Fulop, Matt Gordon, Chai-Shune Hsu, and Patricia Keating, for their suggestions and comments. We also would like to thank I. Park for his help with programming, and the speakers who participated in the experiments.

References

Argyres, Z. (1996). *The Cross-cultural Pragmatics of Intonation: the Case of Greek-English.* MA thesis, University of California, Los Angeles.

Beckman, M. E. & J. B. Pierrehumbert. (1986). "Intonational structure in Japanese and English," *Phonology Yearbook* . 3: 255-309.

Flege, E. (1987) "The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification," *Journal of Phonetics* 15:47-65

Flege, E. (1995). "Second-language speech learning: Theory, Findings, and Problems", in W. Strange (ed.), *Speech Perception and Linguistics Experience: Theoretical and Methodological Issues in Cross-language Speech Research.* Timonium, MD: York Press.

Flege, E. & R.D. Davidian (1984) "Transfer and developmental processes in adult foreign language speech production," *Applied Psycholinguistics* 5: 323-347.

Gårding, E. (1981). Contrastive prosody: A model and its application. *Studia Linguistica* 35: 146-165.

Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody.* Ph.D. dissertation, The Ohio State University.

Jun, S.-A. (1996) "Influence of microprosody on macroprosody: a case of phrase initial strengthening," *UCLA Working Papers in Phonetics* 92: 97-116

Jun, S-.A. & M. Oh (1996) "A Prosodic analysis of three types of wh-phrases in Korean" *Language and Speech* 39(1):37-61.

Maekawa, K. (1994) "Is there 'dephrasing' of the accentual phrase in Japanese?", *Ohio State University Working Papers in Linguisticscs: Papers from the Linguistic Laboratory* 44: 146-165.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation.* Ph.D. dissertation, MIT. [published in 1987 by IULC, Bloomington: Indiana University Linguistics Club.]

Pierrehumbert, J. B. & M. Beckman (1988) *Japanese Tone Structure.* Cambridge, MA: MIT Press.

Todaka, Y. (1990). *An Error Analysis of Japanese Students' Intonation and Its Prosodic Analysis.* MA thesis, University of California, Los Angeles.

Venditti, J., S-A Jun, and M. Beckman (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In J. Morgan and K. Demuth (eds.) *Signal to Syntax.* Lawrence Erlbaum Assoc., Inc.

Weinreich, U. (1953) *Languages in Contact, Findings and Problems.* The Hague: Mouton.