Speculations on the control of speech

Peter Ladefoged

Speech, like most skilled movements, is best described as a goal directed activity. The speech centers in the brain send out instructions for the speech apparatus to achieve certain goals. But it is not at all clear how these goals are defined. One possibility has been outlined by proponents of Articulatory Phonology (Browman & Goldstein, 1986, 1992, Saltzman & Kelso 1987, int. al.), They suggest that the goals are a series of articulatory gestures, being careful to point out that we must distinguish between the higher level gestural goals and the lower level system that implements them. When you produce a word (e.g. *cat)* their notion is that the speech centers in the brain do not issue instructions for the use of particular muscles. Instead they issue instructions equivalent to "Make the following complex gestures: voiceless aspirated velar stop, follow this by a low front vowel, finally make a voiceless glottaled alveolar stop". The different parts of each of these instructions specify goals for different parts of the speech mechanism. "Voiceless aspirated" is an instruction for a certain laryngeal gesture, "velar stop" a lingual gesture, and so on.

We can never be sure what the goals of the speech centers in the brain are. We cannot find out by asking speakers what they are trying to control. If you ask an unsophisticated speaker of English what they are controlling when they say *cat* they will probably say that they don't know, or that they are trying to produce a 'c' followed by an 'a' and then by a 't'. Speakers of an African tone language do not know what they are doing when they say a word on a high tone. They do not know of their own knowledge as the lawyers say (i.e. without being prompted or instructed) the goals of the gestures they are making. They are simply intending to say a particular word that happens, in the view of linguists, to have a high tone. Sophisticated speakers might say that they are trying to produce a high pitch. But, as we will see, it is not clear what they mean by 'pitch' in these circumstances. Is it a sound or certain articulatory actions?

Determining the goals used by the speech centers of the brain is further complicated by the fact that we are apt to be biased by our linguistic analyses and think that the goals are the objects we use in our linguistic descriptions. It is an assumption that there are separate goals for vowels and consonants, just as one might assume that there are separate goals for the movements of the arms and legs when running to catch a ball. But running to catch a ball might be considered as a single action, just as producing a syllable or a word might be a coordinated whole.

A further problem is that speech may be goal directed, but the goals may not be achieved. Observation of what happens when we talk will not provide direct evidence of the targets involved. We can only deduce what they might be, and try to formulate a set of statements — a formal model — that will account for our observations. The hypothesis advanced here is that there are at least three different kinds of goals involved in talking. Sometimes what matters most is achieving a particular articulatory gesture, sometimes there are auditory targets, and sometimes it seems that the aim is to achieve certain aerodynamic conditions.

Articulatory targets

The notion of articulatory targets has been well described by articulatory phonologists, and needs little further illustration. As noted, most investigators assume that there are articulatory targets corresponding to each consonant. Thus Löfqvist and Gracco (2002) give an elegant account of the gestures involved in producing /t, d, k, g/ in contexts involving /i, a, u/, with the stops having a "virtual target" beyond the roof of the mouth. Similarly Recasens (2002) is able to explain the coarticulations involved in producing /p, n, l, s, ʃ/ in contexts involving /i, a, u/ in terms of the articulatory goals. There is a large amount of evidence to suggest that when we try to make a consonant we are, in general, aiming at an articulatory target

Auditory targets (1). Tone and intonation

The first example of a different goal oriented approach that will be considered in this paper is the notion of pitch control. What are speakers of English doing when they say a sentence with a particular intonation? Of course, from one point of view, they are simply trying to produce a particular meaning. We must always remember that speakers are not consciously trying to produce a particular pitch pattern. If they have been given some instruction in phonetics, or if they are trained singers, then they may know how the vocal folds determine the pitch of a sound. Given some analytical knowledge, they may be able to make deliberate laryngeal adjustments. But this is not what happens in everyday life, The speech centers in the brain act to produce utterances in ways that are beyond our conscious thought processes. You can tell people to flex their biceps, and those of us who have watched tough guy movies will know what to do. But only the speech experts know how to increase the tension of the vocal folds.

Accordingly, we should reword the questions we were asking and say instead: what are the speech centers in the brain doing to implement the desire to produce an utterance with that particular meaning? Put this way the answer almost certainly is that, with regard to what we linguists call tone and intonation, they are trying to produce a given pitch pattern. Extending the question, we can ask: how are they doing this? The primary control of the pitch of a speech sound is the tension of the vocal folds, and we might imagine that the speech centers achieve a pitch pattern by producing a particular sequence of vocal fold tensions. There is some evidence in favor of this notion, and some that goes against it.

The pitch of a sound is determined not only by the tension of the vocal folds, but also by aerodynamic factors such as the pressure drop across vocal folds and the rate of flow between them. It seems that the speech centers usually disregard the aerodynamic factors and control only the tension of the vocal folds when trying to produce a certain pitch. After a voiceless aspirated stop the vocal folds often begin vibrating at a higher rate because of the greater flow of air. Conversely there may be a drop in the rate of vibration during a voiced consonant when the flow is less, producing a lower pitch that may not be relevant (Silverman 1986).

These points are illustrated in figure 1. The speaker was asked to produce a series of similar sentences such as *Whatever pie you have in mind...* and *Whatever buy you have in mind...* saying each of them with the same intonation pattern. As the parts of these sentences in figure 1 show, there is a considerable difference between the beginning of *buy* (which is virtually level) and the beginning of *pie* (which descends rapidly). The high pitch at the beginning of *pie* is due to the high rate of airflow for the [p^h], which continues into the beginning of the vowel, producing a higher rate of vibration of the vocal folds. Even more noticeable is the drop in each phrase for the [**v**] in *whatever*. The speaker's intended pitch on *-ever* was presumably fairly level. The drop in pitch was simply because the airflow dropped when the [**v**] was produced.

If the goal of the speech centers was to produce a given pitch pattern (and not just a particular vocal fold tension pattern), we might reasonably ask why there was no attempt to vary the tension of the vocal folds in order to compensate for aerodynamic factors. Experiments have shown that speakers are conscious of these small changes in pitch. When listening to someone else they can use the pitch changes to help them decide whether a word begins with a voiced or a voiceless consonant (Ohde 1984). But the data in figure 1 show that the speech production centers in the brain disregard aerodynamic factors, making no attempt to compensate for them. They arrange for a pattern of vocal fold tension that would produce the pitch pattern required for the sentence if there were no aerodynamic effects. So it might seem that the goals are organized in terms of the vocal fold tension.

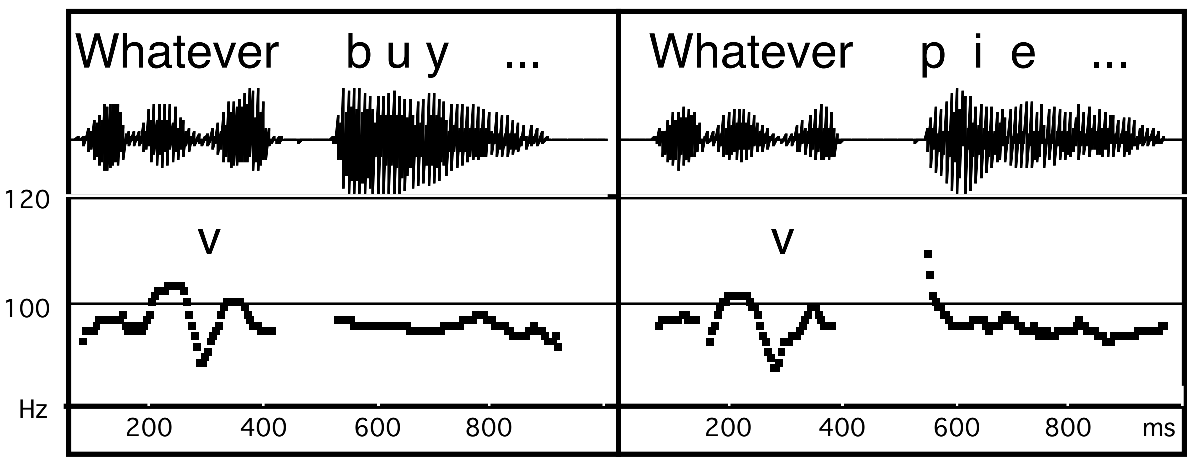

Figure 1. The waveform (upper panels) and fundamental frequency (lower panels) during parts of sentences beginning *Whatever buy* ... and *Whatever pie*. Note the drop in pitch during the voiced consonant [**v**], and the high pitch at the beginning of the word *pie*.

There are, however, two kinds of data suggesting that the speech centers are sometimes aiming directly for a pitch pattern, rather than a vocal fold tension pattern. The first comes from studies of pitch variation in vowels. When the tongue is pulled up to produce a high vowel such as [**i**], the hyoid bone is also pulled up. This in turn produces a pull on the thyroid cartilage; and when the thyroid moves upward the vocal folds are further stretched and a higher pitch is produced. A speaker producing a set of words on the same tone will have a pitch that is 2-5 Hz higher on words that contain [**i**] in comparison with other words (Peterson and Barney 1952, Ladd and Silverman 1984). Whalen & Levitt (1995) discuss 58 studies involving 31 languages in all of which the F0 of high vowels was higher than that of low vowels in similar phonetic contexts. Whalen, Gick, Kumada, & Honda (1998) point out that there is no fully agreed explanation of this finding, but they provide evidence that it is not a learned action but an automatic one. What it is important from our point of view is that tension of the vocal folds has been increased and there has been no compensatory action reducing it. The tension of the vocal folds has been disregarded in the control of pitch.

It might be possible to retain the notion that the production of pitch is organized in terms of the tension of the vocal folds. We could say that, in gestural terms, the speech centers are trying to produce a particular sequence of vocal fold gestures in which the target is not the tension of the vocal folds themselves, but just that part of the tension that is achieved by the laryngeal muscles. This possibility, however, is excluded by other data indicating that sometimes the goal may be the pitch and not the vocal fold tension, however defined.

Speakers do not all behave in the same way when it comes to making pitch variations. Ladefoged (1967:44-46) reported experiments in which 11 speakers produced a set of statements and questions such as *He's a pervert* and *He's a pervert?* An approximation to the subglottal pressure was obtained from a balloon in the esophagus attached to a pressure sensor. All 11 speakers had a rising intonation in the question form. Although actual values of the subglottal pressure could not be computed for these speakers, it was apparent that 4 of the 11 had a rise in subglottal pressure around the time in the sentence when the pitch increased. This rise in pressure alone may have been sufficient to cause the rise in pitch, and there may have been little or no increase in the tension of the vocal folds. Herman, Beckman and Honda (1996) also report a "strong correspondence" between the subglottal pressure and the final part of the intonation contour in questions as opposed to statements.

Further evidence that speakers may produce pitch changes by varying the subglottal pressure comes from an unreported experiment by Lin Mao-can and myself (Ladefoged lab notebook, 7 May 1983). The speaker recorded 16 disyllablic words, each representing a possible sequence of the four tones of Putonghua. At the same time his subglottal pressure was monitored by a balloon in the esophagus attached to a pressure sensor. There was a very high correlation between the pitch curves and the subglottal pressure curves. It was likely that, as in the case of the four subjects producing questions described above, the pitch variations in this speaker's Chinese words in citation form were produced by the variations in subglottal pressure.

These experiments show that meaningful pitch changes in speech can be produced in very different ways. Some speakers use primarily laryngeal muscles, others place greater emphasis on subglottal pressure, and some speakers use both systems in varying degrees. When there are different ways of achieving the same result it is best to regard the result as the goal, and regard the way of achieving it as a lower level set of possibilities. It is likely that for all speakers in most circumstances the goal of the speech centers controlling an utterance is simply to achieve a particular pitch pattern.

Now let us return to the tendency for the pitch of vowels to be higher after aspirated stops as illustrated in figure 1. This tendency may be said to be due to the aerodynamic situation. The greater rate of flow after aspirated stops produces a faster rate of vibration of the vocal folds. The speech centers may be aiming to produce a certain pitch, but do not (or cannot) adjust the vocal fold tension so as to allow for the greater flow rate during a small part of the vowel. This exemplifies a situation in which the speech centers aim at a certain pitch goal, but do not achieve it.

As we have noted, however, listeners are aware of these variations in the rate of vibration of the vocal folds. They disregard them, perhaps because they cannot do otherwise, when producing a given tone and intonation, but nevertheless they can hear them and use them when judging vowel quality, and assessing the differences between voiced and voiceless consonants (Ohde 1984). Sometimes these pitch variations lead to changes in the goals for producing a sentence. In Korean, the increase in fundamental frequency after voiceless aspirated and fortis obstruents is higher than the increase due to the higher rate of flow through the glottis (Jun 1996). Speakers have noticed the pitch difference between these consonants and other sounds and now the speech centers use laryngeal muscles and aim for a higher pitch, so as to make these sounds more distinctive. The variations in fundamental frequency that are usually disregarded have become phonologically relevant, so that these consonants are now marked by phonologically high tones.

Auditory targets (2). Vowels

Next, more speculatively, let us consider the possibility that consonantal gestures involve primarily articulatory goals, but that vowels have mainly auditory/acoustic targets. Consonant gestures are things that you can feel, whereas vowel movements have to be heard to be appreciated. This is in accord with the finding that a statistically improbable number of languages tend to have five or seven vowels (Maddieson 1984), with one so-called low central vowel and the same number of so-called back vowels as front vowels. As Liljencrantz and Lindblom (1972) and others have shown, these findings are explicable in acoustic/auditory terms. The favored sets of vowels are those that have their formants well distributed in the acoustic vowel space. But, as we will see, they do not have their vocal tract shapes well distributed in terms of the possible set of tongue shapes.

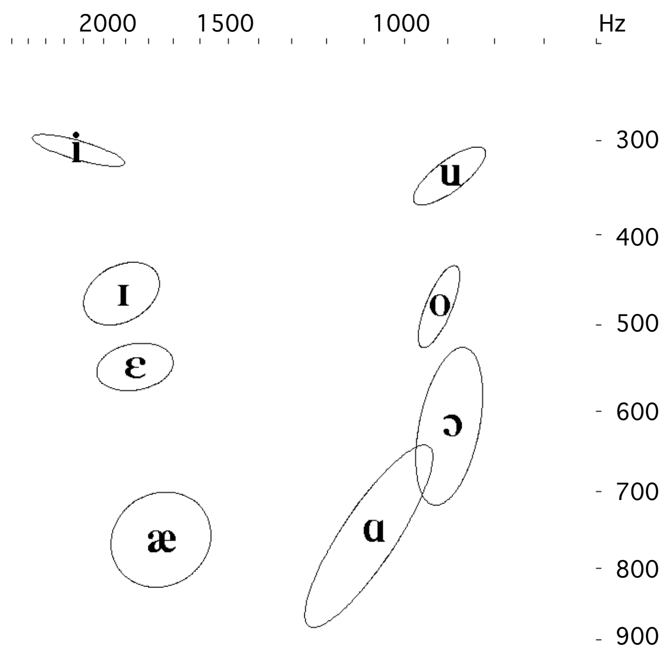

Figure 2. A Bark scale formant plot of the vowels in *heed, hid, head, had, hod, hod, hawed, hood, who'd* as spoken by 5 speakers of American English. The ellipses have their axes on the first two principal components and radii of one standard deviation around the mean of each vowel.

There are always individual differences between speakers, and it is not advisable to attempt to make linguistic generalizations on the basis of data from only one speaker. But there is a lack of data giving the mean formant frequencies and the corresponding tongue shapes for a set of speakers of a language with five or seven vowels. We can, however, make the point that vowels are best regarded as being organized in an auditory space rather than an articulatory one by first considering a set of English speakers producing different vowels. Figure 2 shows the mean frequencies of the first two formants of the front vowels in*, heed, hid, head, had,* and the back vowels in *hod, hawed, hoed, who'd* for five speakers of American English (Harshman, Ladefoged & Goldstein 1977). In the case of the vowels in *hid, head, had, hod* and *hawed* a steady state part of the second formant was selected. For the more diphthongal vowels in *heed, hoed,* and *who'd* a point shortly after the first consonant was selected. A cine x-ray film was made of the speakers as they were producing these vowels. The frame in the film corresponding to the time in the acoustic record used for measuring the formants was located and the vocal tract shape traced.

The formant chart in figure 2 shows the mean formant values for the five speakers. The ellipses have their axes on the first two principal components of the dispersion of the points for each vowel, and radii of one standard deviation around the means. If a low vowel [a] were substituted for the vowels [æ] and [ɑ], the pattern world be similar to that of a typical seven vowel language.

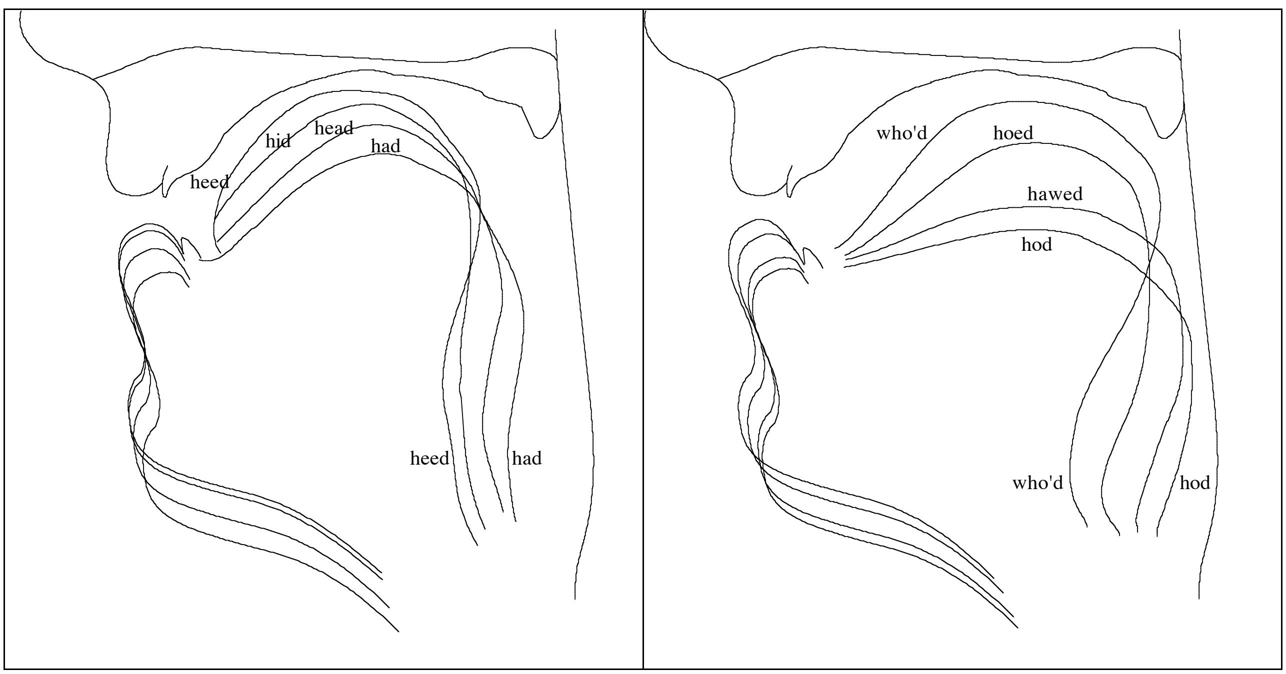

Figure 3. The mean tongue positions of 5 speakers of American English in the vowels in *heed, hid, head, had, hod, hood, hood*.

Figure 3 shows the mean tongue positions for these vowels. Calculating the mean tongue positions is more complex than calculating mean formant frequencies because the shape of the vocal tract differs from one individual to another. The procedure for obtaining a valid mean tongue shape has been described by Ladefoged, (1976) and Ladefoged & Lindau (1989). The tongue positions for each of the five speakers were traced from the x-rays made at the same time as the audio recording. The variations in tongue shapes characteristic of different vowels can be described in terms of two factors, the differences between individuals being ascribed to different personal weightings of these factors (Harshman, Ladefoged and Goldstein, 1977**)**. The weightings correspond to personal characteristics. An average speaker will have a weighting of 1.0 on both factors.

It is difficult to see how these vowels can be considered as being well organized in an articulatory space. The mean tongue positions in the first four vowels can be said to have similar changes in tongue height, but the other four vowels have very different tongue shapes, both from each other and from the first four vowels. The two sets of vowels also differ in that first four vowels are close together, but the other four vowels are further apart.

The data linking the articulations of five speakers with the corresponding formant frequencies can be used in another way. An articulatory/acoustic model (Ladefoged & Lindau 1989) can be used to reconstruct the tongue shapes appropriate for a language with five vowels. Figure 4 shows the first two formants of five vowels that are dispersed within the vowel chart in a manner similar to the vowels of Spanish. Figure 5 shows vocal tract shapes that will produce these sounds.

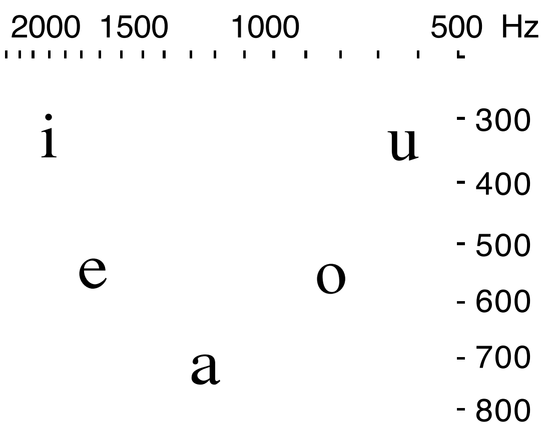

Figure 4. A bark scaled formant chart showing five vowels similar to those of Spanish

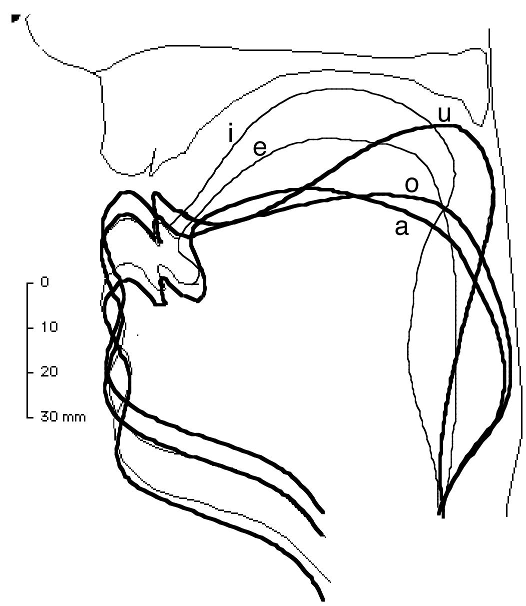

Figure 5. Vocal tact shapes that produce the formant frequencies shown in figure 4. The shapes for the vowels [a, o, u] are shown with heavier lines.

Figures 4 and 5 show that the vowels of languages with five vowels (the most common system in the languages of the world) are well distributed in the auditory/acoustic space, but are less well organized in articulatory terms. The relations between articulations and the corresponding acoustic characteristics shown in these figure lend further credence to the notion that vowels are generally best considered to have auditory/acoustic goals. Accounts of universal tendencies among vowels are often more concerned with the sounds rather than the articulations. (There are exceptions to this notion, accounting for the ‘often’ in the previous sentence. Universal tendencies concerning lip rounding in vowels and in consonants may show that lip rounding is an articulatory goal, rather than an acoustic one, but this point will not be pursued further here.)

Johnson, Ladefoged & Lindau (1993) have also argued in favor of vowels having auditory targets. They point to the differences in the way in which individuals manipulate the tongue and jaw in order to produce the same vowel sounds, and propose an auditory theory of speech production. At least in the case of vowels, we produce what we want to hear. But it must be admitted that the evidence still allows for a particular vocal tract shape (as opposed to tongue and lip positions) to be the goal of the speech centers when producing a particular vowel. Individuals do not aim for one vocal tract shape at one moment and another shape that can produce the same formant frequencies at another. However, in the case of the next phonetic property that we will consider something like this occurs in that completely different movements are used to produce the same goal.

Aerodynamic target: stress

In some cases the goals of the speech centers cannot be considered to be auditory targets or articulatory gestures. Phrasal stress is a good example of a phonetic property that cannot be adequately described as an articulatory target or an acoustic one. Stress has been equated with a greater opening of the jaw (Fujimura 2000**),** and a greater jaw opening certainly occurs on many stressed syllables. But jaw opening is plainly not a defining characteristic. Some people hardly move the jaw at all, and anyone can produce properly stressed sentences while maintaining the jaw at a fixed height by holding an object between the teeth. These sentences sound completely natural, and are indistinguishable from those in which the jaw is free to move. Articulatory opening is equally not a defining characteristic. A sentence such as *We **keep** money* can be said with a strong emphasis on ***keep***. This emphasis can occur with an unusually close or with a more open articulation than usual for the vowel [i]. Stressed syllables can also be produced with different actions of the respiratory system, depending on the amount of air in the lungs, as will be illustrated later. These are all instances of motor equivalence, different systems being used to produce a goal that cannot be defined in articulatory terms.

Similarly, there do not appear to be any specific auditory/acoustic goals for stressed syllables. Intensity is sometimes regarded as the major characteristic of stress, but this is also not correct. Consider three sentences that are the same except for a contrastive stress that has been placed on a different word in each of them:). *I **see** three bees* (but I can't hear them), *I see **three** bees* (not a swarm of them), *I see three **bees*** (but no wasps). In these circumstances one might expect the stressed word to have a greater intensity than the other two words that have the same vowel. But, as the pitch and intensity records in figure 6 show, it is the pitch and the duration that indicates which word received the contrastive stress. In every case the stressed word has a high falling pitch and a greater length, but not a greater intensity. A dashed line has been drawn marking the intensity of the word *see*, showing that it is almost the same in all three phrases, irrespective of whether this word is stressed or not. In the first phrase, the word *bees* has the highest intensity although it is not stressed. In this particular set of phrases the pitch is the more important indicator of stress.

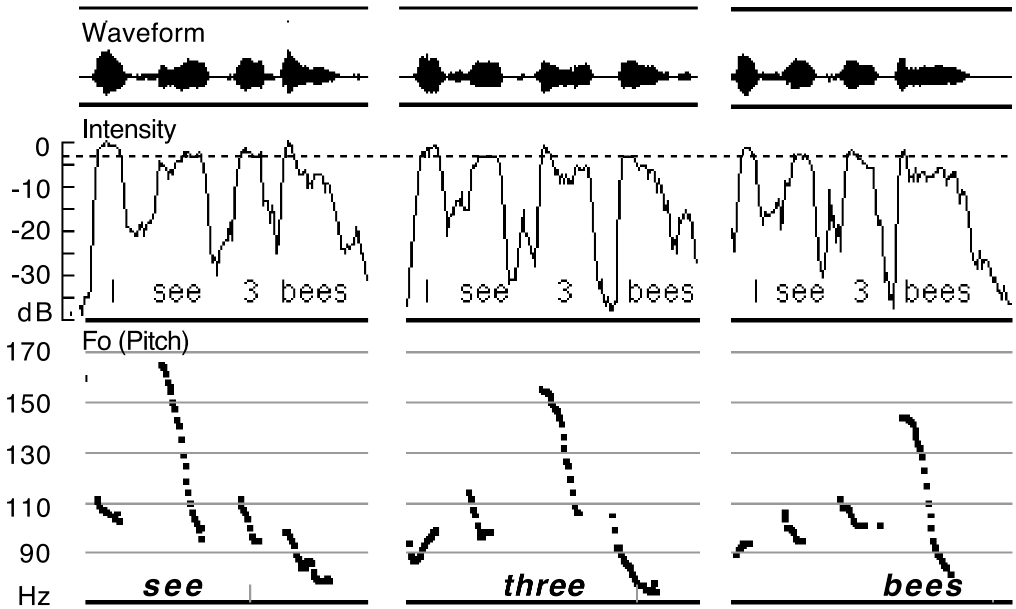

Figure 6. Waveform, intensity and pitch (fundamental frequency) records of *I see three **bees*** (but no wasps). *I see **three** bees* (not a swarm of them). *I **see** three bees* (but I can't hear them). The dashed line shows the mean intensity of the word *see*.

We should not, however, presume that a high falling pitch is always the most important correlate of stress. Figure 7 shows that it is possible to emphasize words without using an increase in pitch and a fall. This figure shows the waveform, pitch and intensity in the phrase *You **saw** what I meant.* The word *saw* has been emphasized as when replying to someone who had denied understanding an obvious statement. The pitch on *saw* is lower than that on *You* or *what I meant*. The intensity is higher and the vowel is longer than usual. These last two factors convey the information that this is the stressed syllable.

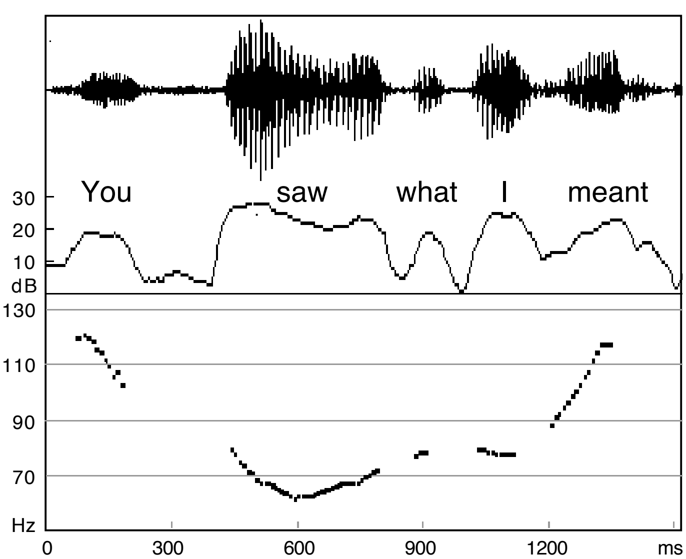

Figure 7. The waveform, pitch and intensity in the phrase *You **saw** what I meant.*

If the speech centers in the brain are not trying to produce a given articulation or a particular intensity or pitch pattern when producing a stressed syllable, what is the appropriate goal? The answer is that word stress always uses a greater amount of respiratory energy. The goal is probably an increased pressure below the vocal folds. There is evidence (Ladefoged 1967, Ladefoged and Loeb 2002) that stressed syllables use greater respiratory energy. But this energy is not always produced in the same way. Figure 8 (adapted from Ladefoged and Loeb 2002) shows the most common way. The mean activity of the internal intercostals, the principal respiratory muscles for pushing air out of the lungs during speech, have been averaged over 17 repetitions of the phrase *The old man doddered along the road*. The audio recording of one of the repetitions of the phrase is also shown. The speaker was slightly inconsistent in the way in which she said this phrase. There was always a stress on *man*, on the first syllable of *doddered* and on *road*, and on some occasions there was also a stress on *old*. The onsets of the stressed syllables are marked by open arrows.

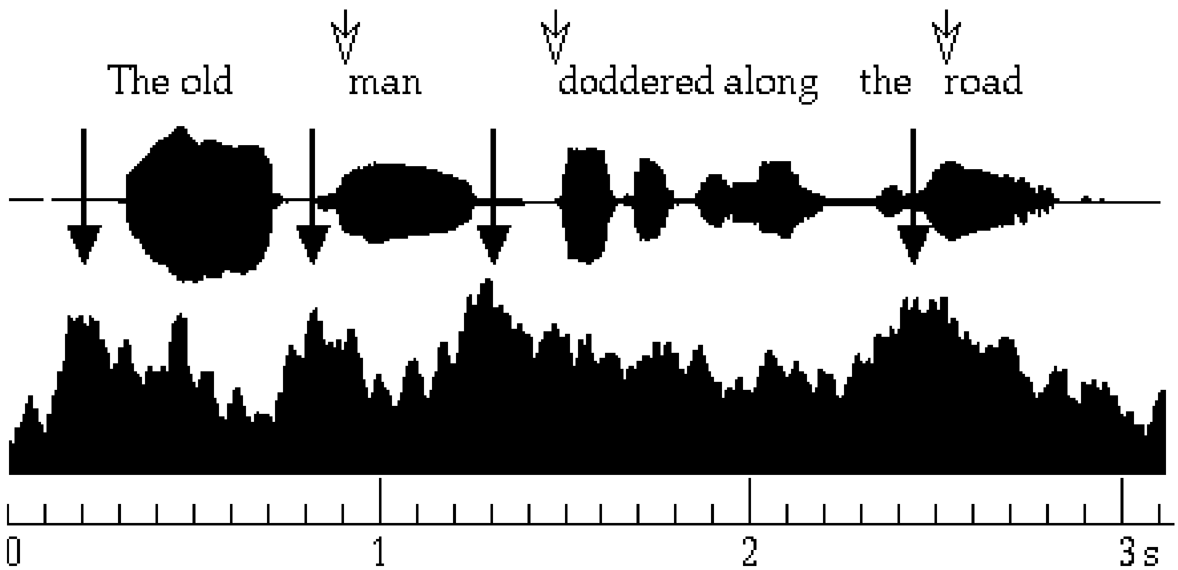

Figure 8. A single audio record of *The old man doddered along the road* above the mean EMG activity of the internal intercostals during 17 repetitions of this phrase. The open arrows mark the onsets of the stressed syllables and the solid arrows mark four of the major peaks in the average activity. (For details of the procedure see Ladefoged and Loeb 2002)

There are a number of peaks in the internal intercostal activity. Solid arrows mark four peaks with high amplitudes that are maintained for over 100 ms. The first peak occurs before the beginning of the utterance, and presumably reflects the generation of the respiratory power necessary to start vocal fold vibration. The second peak occurs before the word *man*. The third peak has the largest summed amplitude and occurs shortly before the first vowel of *doddered*, the syllable that carries the first primary stress in the utterance. At the end of the phrase, shortly before the stressed syllable *road*, there is another peak. There is also a sharp peak, not marked by an arrow in the figure, near the beginning of *old* in which the activity is confined within a 60 ms interval. This peak possibly reflects the occasional inconsistency in the stressing of the phrase. There are no peaks in the EMG activity for the lengthy stretch of the utterance containing the unstressed syllables *..ered along*.

The bursts of internal intercostals activity reduce the lung volume and produce a higher subglottal pressure. But these controlled muscular actions cannot be regarded as gestures that produce stresses. The increase in pressure is the goal of the activity, and it is not always produced by respiratory muscles pushing air out of the lungs. Consider what happens after a speaker takes in a breath. After an inspiration the lungs are like inflated balloons containing air under pressure. In these circumstances the problem is not how to push air out of the lungs, but how to stop it going out at too high a pressure. Without muscular action to hold the rib cage up, the pressure beneath the vocal folds would be too great.

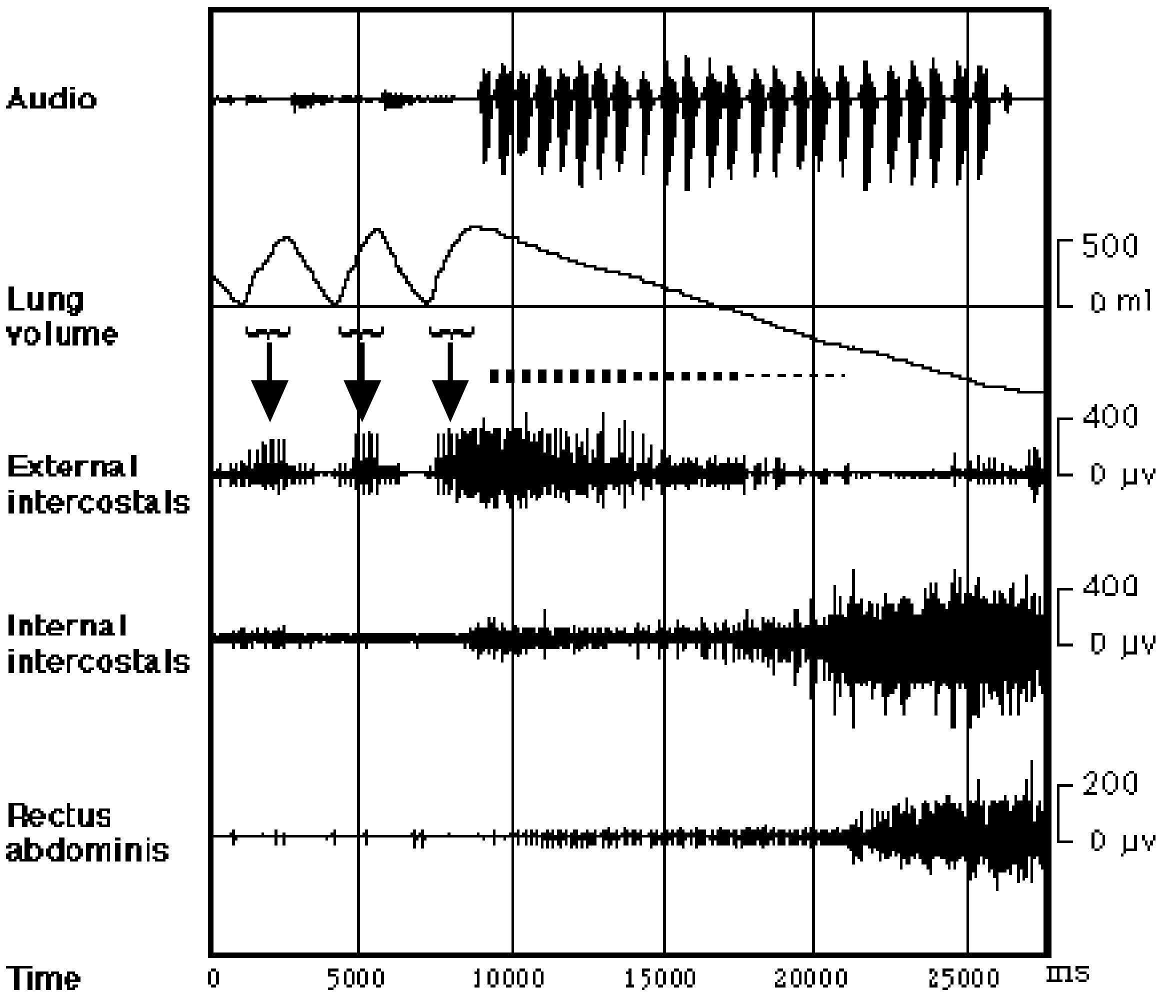

Figure 9. Audio, lung volume and EMG records from the external intercostals, internal intercostals and rectus abdominis while producing two respiratory cycles followed by alighter deeper inspiration and 25 repetitions of the syllable [pa] at a quiet conversational level (adapted from Ladefoged and Loeb, 2002).

Figure 9 shows what happens when a speaker takes two normal breaths and then starts talking after a third, slightly larger, inspiration. The external intercostals, the muscles that pull the rib cage up, are used to increase the volume of air in the lungs. They are active in all three inspirations shown in figure 9 at the times indicated by the arrows. The external intercostals remain active for the first part of the utterance where there is a diminishing dotted line. Thespeech in this figure consists of repetitions of the single stressed syllable [ma]. For each of these syllables in the first part of the utterance the external intercostals relax slightly so as to allow the pressure of the air below the vocal folds to increase. As the utterance proceeds the internal intercostals take over, producing contractions of the rib cage to increase the subglottal pressure. The relaxations of the external intercostals are difficult to see in figure 9, with its compressed time scale. Figure 10 shows just the EMG and the audio for the central part of figure 9. The lessening of the external intercostal activity just before each stressed syllable is apparent.

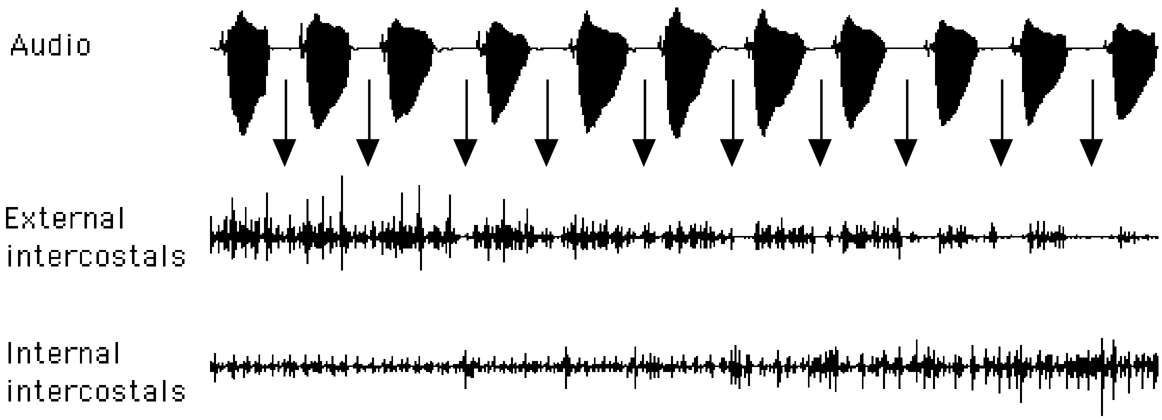

Figure 10. An enlargement of the part of Figure 5 from time 11,500 to 19,000 ms. The arrows mark onsets of stressed syllables and decreases of external intercostal activity.

There are other aspects of figures 9 and 10 that are worth noting when considering the control of speech. As soon as speech starts, there is activity of both the internal intercostals and the rectus abdominis, a muscle that decreases the size of the abdominal cavity, thus pushing the diaphragm upwards so that the lung volume is reduced. These expiratory muscles are active while the external intercostals are still preventing the rib cage from descending. This pattern of activity may represent co-contraction to stabilize the musculoskeletal systems against stochastic fluctuations and external perturbations by taking advantage of the intrinsic mechanical properties of the active muscle (Loeb, Brown and Cheng, in press; Hogan, 1984). It is evident that the respiratory muscles are operating as a complex system to control the pressure of the air in the lungs, and hence the respiratory power used for speech. There is no way in which these actions can be considered as always being the same muscular gesture. Sometimes they involve pulling the rib cage down, and sometimes simply allowing it to collapse more rapidly. Stressed syllables may, or may not, have a greater jaw opening; they may or may not have a greater intensity or a higher pitch. They may or may not have bursts of internal intercostals activity. But they always use greater respiratory energy.

Conclusion

This paper has considered some of the ways in the speech centers in the brain may be controlling speech acts. The major proposal is that we should not consider all aspects of speech to be controlled in the same way. For the most part, consonants are the result of trying to achieve gestural goals such as those specified by articulatory phonology. Tone and intonation are set as auditory goals consisting of pitch patterns. The targets for vowels are points or trajectories in an auditory/acoustic space. The targets for stressed syllables are increased subglottal pressure, achieved through actions of the respiratory system. We need to think sometimes in terms of an auditory theory of speech production, sometimes in terms of aerodynamic constraints, and at yet other times a gestural theory is more appropriate. We will not get much further in our investigation of the real nature of speech until we have discovered more about how speech is organized. We need to know what we are trying to do when we talk.

References

Browman, C. P. & Goldstein, L. (1986). Towards an articulatory phonology. Phonology Yearbook, 3,19-252.

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

Fujimura, O. (2000). The C/D model and prosodic control of articulatory behavior. Phonetica 57. 2.

Harshman, R. A., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. Journal of the Acoustical Society of America, 62, 693-707.

Herman, R, Beckman, M. & Honda, K. (1996) Subglottal pressure and final lowering in English. International Conference on Spoken Language Processing. 145-148.

Hogan, N. (1984). Adaptive control of mechanical impedance by co-activation of antagonist muscles. IEEE Transactions on Automatic Control, 29, 681-690.

Johnson, K.,Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. Journal of the Acoustical Society of America, 94, 701-714.

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

Ladd, D. R., & Silverman, K. E. A. (1984). Intrinsic pitch of vowels in connected speech. Phonetica. 41,31-40.

Ladefoged, P. (1967). Three areas of experimental phonetics. London: Oxford University Press.

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. & Lindau, M. (1989). Modeling articulatory-acoustic relations. Journal of Phonetics, 17, 99-106.

Ladefoged, P, & Loeb. G. (2002). Preliminary experiments on respiratory activity in speech. http,//www.jladefoged.com/respiratorystudies.pdf.

Liljencrantz, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems, The role of perceptual contrast. Language, 48, 839-862.

Lindblom, B. E., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. N. Li (eds.), Language, Speech and Mind, Studies in Honor of Victoria A. Fromkin (pp. 62-80). London and New York: Routledge.

Loeb, G.E., Brown, I.E. & Cheng, E. (in press). A hierarchical foundation for models of sensorimotor control. Experimental Brain Research.

Löfqvist and Gracco (2002). Journal of the Acoustical Society of America, 111, 2811-2827.

Maddieson, I. (1984). Patterns of sounds. Cambridge: Cambridge University Press.

Maddieson, Ian. (1997). Phonetic Universals. In W. J. Hardcastle and J. Laver (eds.), The Handbook of Phonetic Sciences, 619-639. Oxford: Blackwell.

Martinet, A. (1955). Economie des changements phonétiques (2nd ed.). Berne: Francke.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. Journal of the Acoustical Society of America, 75, 224-230.

Peterson, G. E., & H.L. Barney, H.L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24, 175-184.

Recasens, D, (2002). An EMA study of VCV coarticulatory direction. Journal of the Acoustical Society of America, 111, 2828-2841.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions, A task dynamic approach. Psychological Review, 94, 84-106.

Saussure, F. (1916). Cours de Linguistique Generale. New York: Philosophical Library.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. Journal of Phonetics, 23, 349-366.

Whalen, D. H.,Gick, B.,Kumada, M., & Honda, K. (1998). Cricothyroid activity in high and low vowels, exploring the automaticity of intrinsic F0. Journal of Phonetics, 27, 125-142.