Preliminary studies on respiratory activity in speech

Peter Ladefoged (Linguistics Department, UCLA, Los Angeles)
Gerald Loeb (Biomedical Engineering, USC, Los Angeles)

Introduction

This paper presents some data on two aspects of the use of the respiratory system in speech. The experiments to be reported use improved emg techniques to replicate some of the research conducted at Edinburgh University in the 1950s. The major papers from the earlier research have been summarized by Ladefoged (1967). The basic findings of this early research were twofold. Firstly, the mean power for speech is provided by actions of the respiratory muscles controlling the pressure of the air in the lungs. The stretching of the lungs after a maximal inspiration generates a force, the elastic recoil, which can produce a comparatively high pressure, the so-called relaxation pressure. In these circumstances the external intercostals and the diaphragm actively regulate subglottal pressure by preventing lung volume from decreasing too rapidly. As the volume of air in the lungs (and, correspondingly, the relaxation pressure) decreases, less external intercostal activity is needed. As lung volume decreases still further, and the subglottal pressure required for speech becomes greater than the force generated by the elastic recoil of the lungs, various expiratory muscles become active. The Edinburgh studies showed that the internal intercostals were active in most conversational speech, where lung volumes are similar to those in quiet respiration. At higher levels of lung volume the external intercostals played a significant role in checking the fall of the rib-cage, thus decreasing the pressure generated by the elastic recoil of the lungs. At lower levels of lung volume other muscles (such as rectus abdominis and the internal obliques) supplemented the action of the internal intercostals.

These results are summarized in figure 1 (redrawn from Ladefoged, Draper and Whitteridge 1959). The upper part of the figure consists of two records (1) the volume of air in the lungs relative to the mid-respiratory level, obtained by placing the speaker in a body plethysmograph; (2) the subglottal pressure as indicated by a balloon in the esophagus just below the vocal folds. Subsequent work showed that this balloon, high in the esophagus, was recording a pressure in between that of the pressure of the air in the trachea immediately below the vocal folds and the intra-pleural pressure. It is therefore only an indicator of the true subglottal pressure and is not given a numerical value in this figure. The speaker produced two normal respiratory cycles, and then took a deeper breath before counting from 1 to 32 at a normal conversational level. The lower part of figure 1 is a schematic representation of the activity of the respiratory muscles derived from many records of the activity during similar actions (two quiet respiratory cycles followed by counting as far as possible on a single breath).

The Edinburgh findings were supported by similar results reported by Newsom Davis et al.(1970), Campbell (1968) and Sears and Newsom Davis (1968). These early studies demonstrated what is now regarded as motor equivalence between the various muscles in the respiratory system in maintaining the mean subglottal pressure. They showed that when maintaining a subglottal pressure for a given voice level (soft, medium or loud), different muscles were used, depending on the volume of air in the lungs.

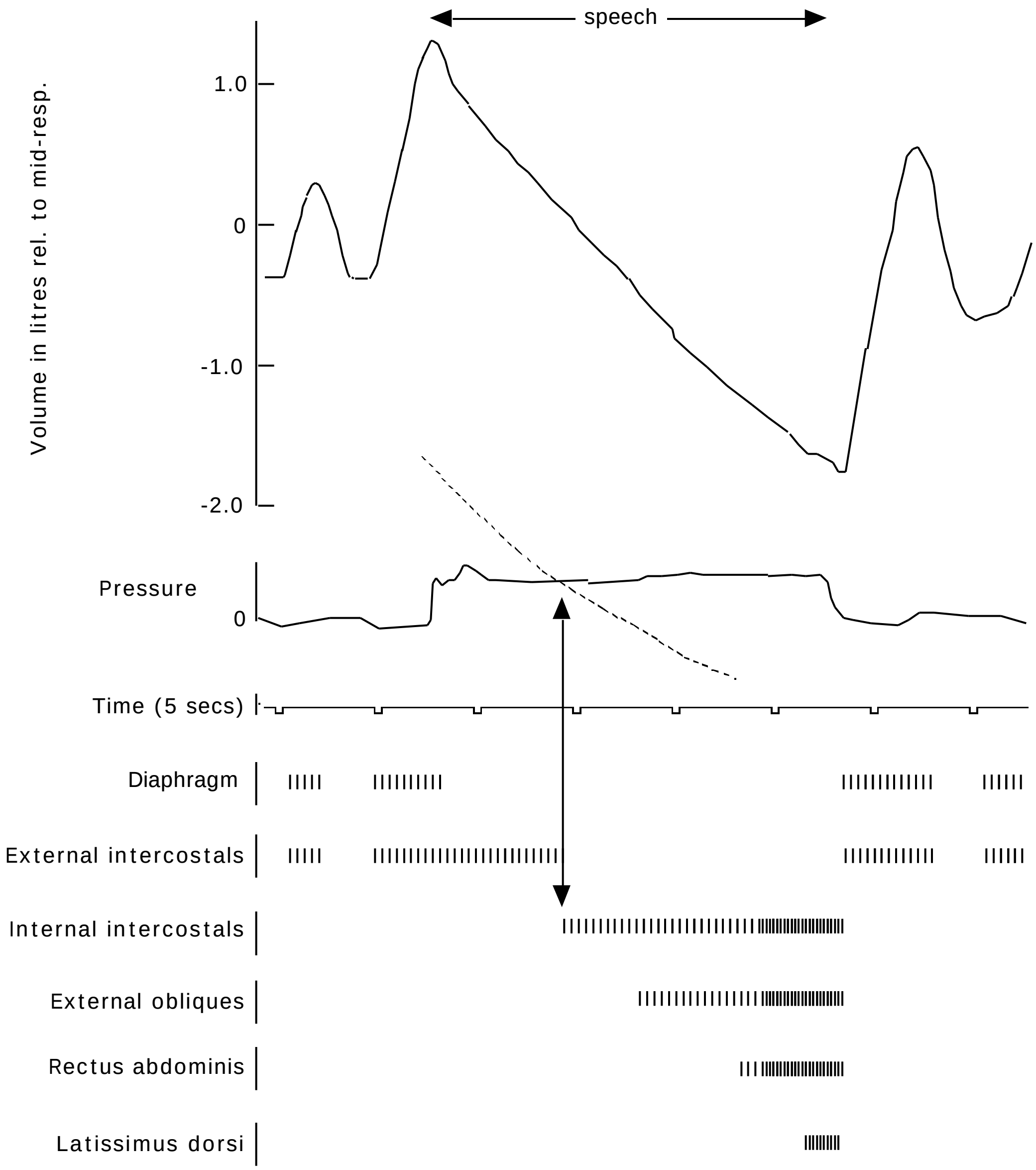

Figure 1. Upper part of the figure, the volume of air in the lungs and the esophageal (sub-glottal pressure) during respiration and counting from 1 – 32 at a conversational loudness. Lower part of the figure, a diagrammatic representation of the muscular activity that was observed to accompany speech during such pressure and volume changes. The dashed line indicates the relaxation pressure associated with the corresponding volume of air in the lungs.

The second major claim of the Edinburgh research was that stressed syllables are produced by increases of respiratory power above the mean, usually by additional activity of the internal intercostals. Although there was considerable qualitative data supporting this hypothesis, the early studies produced little quantitative data. At that time, before the use of computers allowed investigators to calculate the mean activity in several repetitions of a phrase, the most reliable way of quantifying the degree of muscular activity was to observe the rate of firing of a single motor unit. Finding a single unit when using comparatively large needle electrodes was difficult.

The Edinburgh group published only one record of this kind. The unmodified data is reproduced here in figure 2. On this occasion the rate of firing varied from around 5 to over 30 per second, with peaks in the rate occurring just before the stressed syllables. Burst of activity before stressed syllables were easy to see (though not to quantify) in many of the photographs of the oscilloscopes on which the data were recorded. The Edinburgh group was therefore reasonably certain that stressed syllables in conversational speech were typically accompanied by internal intercostals activity. However, the role of the respiratory musculature in relation to local peaks in subglottal pressure is still somewhat controversial (see, e.g., Ohala, 1990).

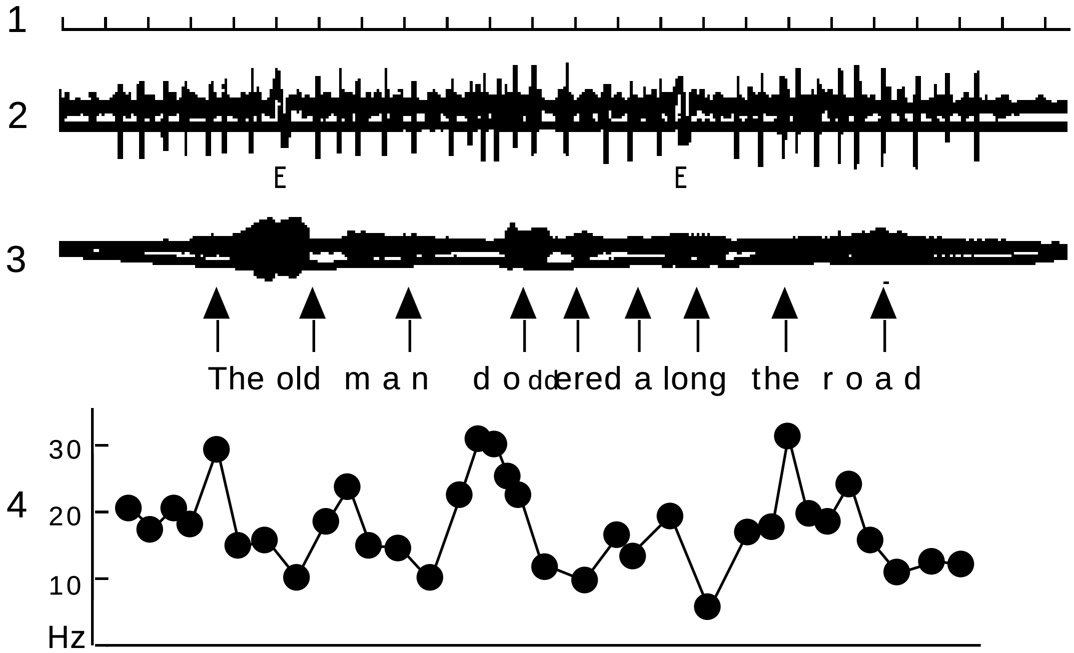

Figure 2. Internal intercostals activity during the phrase *The old man doddered along the road.* (1) Time marker, 100 ms intervals [redrawn]. (2) Internal intercostals action potentials (electrical activity associated with action of the heart indicated by E. The superimposed slightly descending thick line is the record of the lung volume). (3) Audio (with a superimposed faulty pressure record). (4) Instantaneous frequencies of the single motor unit recorded in (2) in impulses per second [redrawn].

The Edinburgh studies have been justifiably criticized on a variety of grounds by Hixon and Weismer (1995). They noted that the pressure records did not take account of the intra-pleural pressure. (This was first noted by Kunze (1964), resulting in corrections to the original reports in the summary by Ladefoged (1967).) They correctly claimed that there were “(a) errors in establishing a backdrop of mechanical information; (b) discrepancies between data and the statements about them.” They also, with less justification, stated that there were (c) “counterpredictive features between data and other knowledge about breathing” — if this is true it may be that the other knowledge is wrong — and “(d) inadequacies in acquiring, portraying and interpreting electromyographic data…” so that the results are “for the most part invalid”. The present paper shows that this is incorrect. The EMG data was largely correctly acquired, portrayed and interpreted, subject to the instrumental limitations of its time. Nevertheless Hixon and Weismer's review of this nearly 50 year old research did show the necessity of trying to replicate it, using modern methods to measure intramuscular EMG, airflow and pressure.

Experiment 1. Respiratory activity for maintaining mean subglottal pressure

The first set of new data concerns the way in which subglottal pressure is maintained during an utterance as lung volume varies. Three subjects participated in these experiments, one man and two women. Various instrumental problems were encountered in the experiments with the male subject, but overall results were consistent with those for the other two subjects, who provided a firmer basis for our conclusions

Custom fine platinum alloy electrodes were inserted into the internal intercostals, the external intercostals and the rectus abdominis (cf. Loeb and Gans, 1986). The exposed portion of the wire was 3 mm in length. The rectus abdominis electrodes were inserted about 5 cm from the midline. Several insertion sites were tried for the external and internal intercostals. The best recordings from the internal intercostals were obtained from electrodes inserted into the eighth intercostal space in the mid-axillary line, and the best recordings from the external intercostals were obtained from electrodes in the same space either in the same location or 9 cm posterior to the internal intercostal insertion point. Muscles were identified by their functions, assuming that it was the external intercostals that were active on inspiration, and that the internal intercostals showed no activity during quiet respiration, but were active in forced expiration. The subjects spoke into a mask (SciconRD) that recorded the oral airflow and also contained a microphone for recording the audio signal. Before speaking subjects breathed in and out normally for two respiratory cycles. They then repeated the syllable [pa] as many times as they found convenient on a single breath. The audio signal, the EMG data from three sites, and the rate of flow of air from the mouth were recorded onto a computer using SciconRD's system for multi-channel AC and DC recording.

A typical set of records is shown in Figure 3. The only processing that has been done is to provide a lung volume record by taking each sample point in the oral airflow record and replacing it by the running sum of all the samples up to that time. This gives an estimate of the relative lung volume at each moment. The external intercostal record is from a monopolar electrode in the eighth intercostal space at the mid-axillary line and a nearby surface electrode. The internal intercostal record is from a second monopolar electrode inserted by the same needle in the same intercostal space. Because the length of the inserted wires differed, this electrode was seated approximately 5 mm deeper. It was recorded differentially with respect to the same surface electrode. The rectus abdominis record is from bipolar wires inserted into that muscle 5 cm from the mid-line.

The first EMG record shows external intercostal activity during the two cycles of quiet respiration and during the inspiration before speaking. This activity continues during the first part of the utterance in which the rib cage has to be supported to prevent there being too high a subglottal pressure. As the utterance continues and the volume of air in the lungs decreases, the external intercostal activity diminishes and the internal intercostals activity (the second EMG record) increases. There is some internal intercostal activity right at the beginning of the utterance, occurring simultaneously with the external intercostal activity. Rectus abdominis activity, shown in the third EMG record, also occurred as soon as the utterance began, coming in more strongly after the internal intercostal activity has begun increasing.

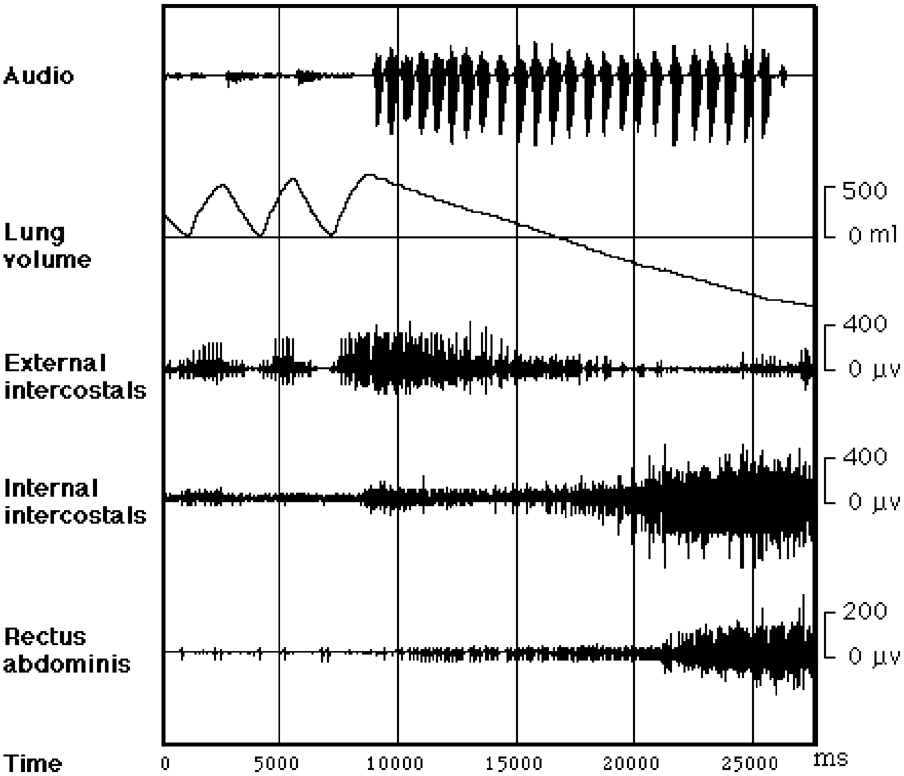

Figure 3. Records from speaker 3 producing two respiratory cycles followed by 25 repetitions of the syllable [pa] at a quiet conversational level.

Any EMG results must be examined carefully for evidence of cross talk. In practice there are two possibilities to be considered. Firstly, a record from a single electrode allegedly showing external intercostal activity may contain cross talk from internal intercostal activity, and vice versa. Secondly, records from two electrodes allegedly showing activity from two different muscles may in fact be recording at least some activity from the same muscle. The record in figure 4, which shows muscle activity during forced expiration by the second subject, illustrates the first problem. This record is from a single hooked wire inserted in the eighth intercostal space in the mid-axillary line and a surface electrode above it. There is a predominant single unit that fires on inspiration (and is therefore presumably in the external intercostals), and additional activity from multiple units during forced expiration, presumably from several fibers in the internal intercostals. Although it is possible to distinguish the activity from the different muscles by visual inspection, records of this kind have been disregarded in the analysis of our data.

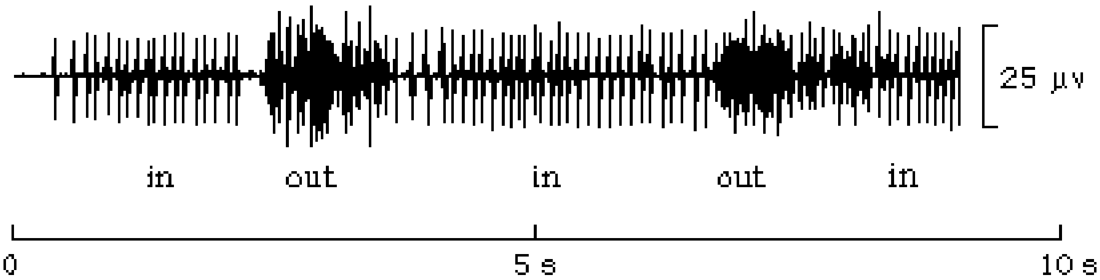

Figure 4. Activity from both internal and external intercostal muscles.

The second possible effect of cross-talk — records from two electrodes allegedly showing activity from two different muscles that may actually be from the same muscle — can be discounted by inspection of the data on an expanded time scale. For example, in the data shown in Figure 3, which shows a relatively small number of units recorded by each contact, it is possible to exclude cross talk by simple inspection of the expanded records. Figure 5 shows the external and internal intercostal activity during the last word uttered (after time 25000 ms), on a greatly expanded time scale and a slightly expanded signal scale. The presence of large unitary EMG signatures in one trace with little or no deflection at the same time in the other (and vice versa) is sufficient to rule out cross talk as an artifactual source of EMG signals when two muscles are active in an overlapping pattern.

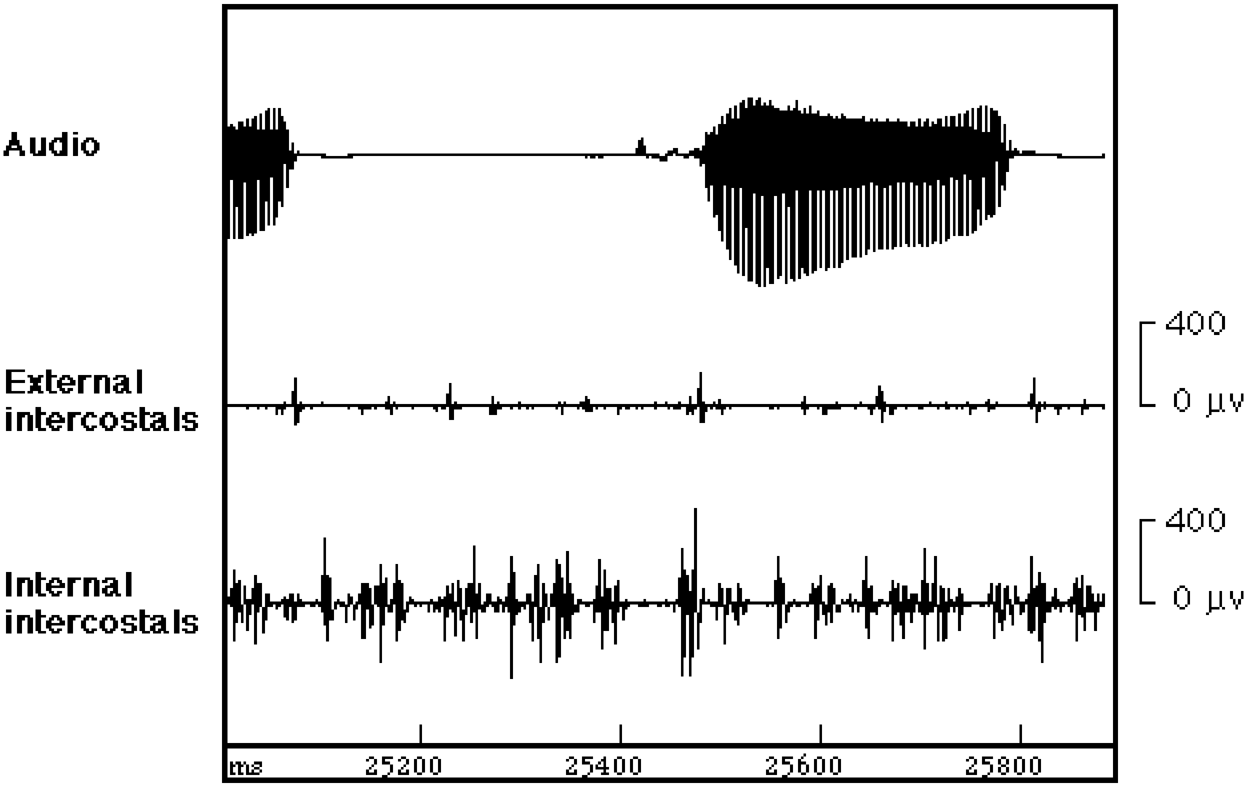

Figure 5. Part of Figure 3 (the last word) on an expanded time scale.

Given that there is no cross talk in the part of the data shown in Figure 5, we can be reasonably certain of our analysis of the whole utterance in Figure 3. A detailed visual inspection of this activity on an expanded time scale showed that the internal intercostal activity is not phase-locked to the external intercostal activity; there are no simultaneous spikes. It seems that both internal and external intercostal activity are occurring at the same time., and rectus abdominis activity also overlaps with external intercostal activity.

This simultaneous use of antagonistic muscles is a new finding that needs further investigation. The Edinburgh group expected a crossover between inspiratory activity and expiratory activity to occur at the point above the resting lung volume when the relaxation pressure was just sufficient to provide the power for vocal fold vibration. However, the pattern of muscular activity shown in Figure 3 is more complex. Both the rectus abdominis and the internal intercostals 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).

Each of the two female speakers produced records similar to this in each of 10 (speaker 2) or 7 (speaker 3) repetitions of this task. Instrumental difficulties in these studies precluded a quantitative analysis, but none of the observed data contradicted this pattern. These investigations of the maintenance of the mean respiratory power required for speech suggest that the earlier studies were on the right lines, but that speakers may achieve fine control by the simultaneous use of antagonistic muscles.

Experiment 2: Respiratory muscular activity associated with stressed syllables

The second set of new data concerns the notion that the respiratory muscles are actively used to produce local peaks in subglottal pressure. This finding has been widely accepted (e.g., by Hixon, 1987; but see also Hixon and Weismer, 1995). It has not, however, been demonstrated quantitatively to any great extent. As noted above, there was only one record in the early studies in which the activity of a single motor unit could be distinguished, and its rate of firing calculated. The second set of experiments provides a small step towards further quantifying the muscular activity, and illustrates a technique for measuring the respiratory activity associated with differences in stress. Similar techniques have been used in studies of articulatory actions at Haskins Laboratories, UCLA and elsewhere in the 1960’s and later.

Speakers 2 and 3 repeated the phrase *The old man doddered along the road* a number of times. (The same phrase that was used for calculating the rate of firing of the single motor unit in the early Edinburgh studies.) Speaker 3 did not produce the minimum of 10 valid repetitions deemed necessary to determine the average activity across utterances. Speaker 2 produced three utterances in which the phrase was repeated four times, and one in which it was repeated five times. One utterance in which the phrase was repeated four times on a single breath is shown in Figure 6. As can be seen, internal intercostal activity increased throughout this utterance.

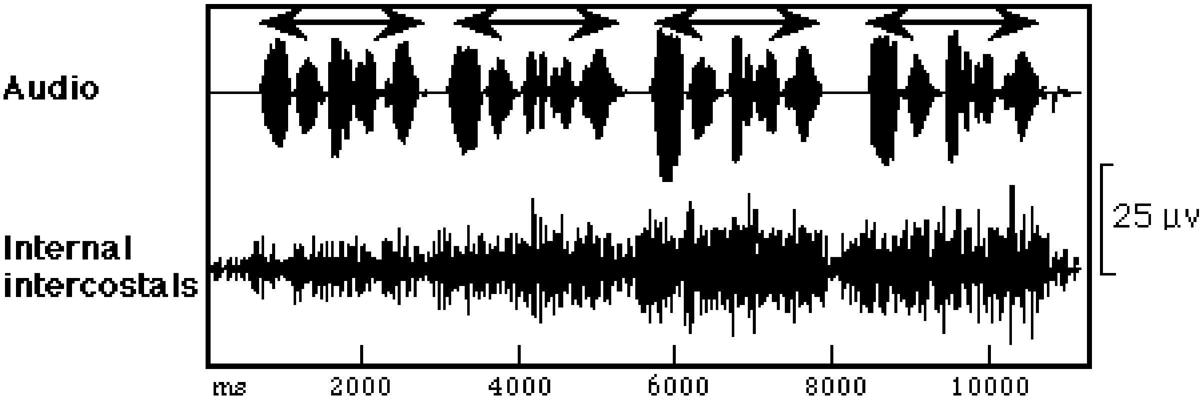

Figure 6. Internal intercostal activity during four repetitions of the phrase *The old man doddered along the road* by speaker 2. The boundaries of the four utterances are shown by the linked pairs of arrows at the top.

In a single record it is difficult to see increases in activity associated with particular syllables, because the large number of action potentials recorded masks the overall temporal pattern of firing. This problem may account for the fact that Adams (1979) found no evidence of localized bursts of internal intercostal activity corresponding to stressed syllables when she recorded EMG activity from the internal intercostals. Her data may reflect the activity of a very large number of motor units firing almost simultaneously, so that the action potentials are difficult to observe in the photographic records that she was making. But just because localized bursts of activity are not observable in records of the form shown in Figure 6, it does not follow that they do not occur.

In order to determine whether increases in internal intercostal activity occur during particular parts of a phrase, EMG activity must be averaged across several utterances. In previous studies, activity has typically been averaged over 20 or more repetitions of a particular phrase (Hirose, 1971; Kewley-Port, 1973). In this study, however, only 17 utterances were available. EMG activity in the internal intercostal muscle was sampled at 3,000 Hz. The absolute value of each of these samples was determined, and then each of these values was replaced by the mean of itself and the preceding 59 samples. As there were small variations in the length of each phrase, the activity was time matched at three points in each utterance. For each utterance the moving average procedure was started 300 ms before the vowel of *The* and calculated for 3,000 points. It was then restarted 300 ms before the release of the first [d] in *doddered*, and calculated for a further 3,000 points. Finally, it was restarted 300 ms before the release of the last consonant, the [d] in *road,* and calculated for 3,000 more points. We intend to use a more sophisticated time warping procedure in future experiments. The mean of the absolute values of each of the 9,000 points in each of 17 utterances was determined. Figure 7 compares the activity of the internal intercostals calculated in this way with the audio recording of one of the repetitions of the phrase. The speaker varied the way in which she said this phrase, but there were usually stresses on *old*, *man*, the first syllable of *doddered* and *road*.

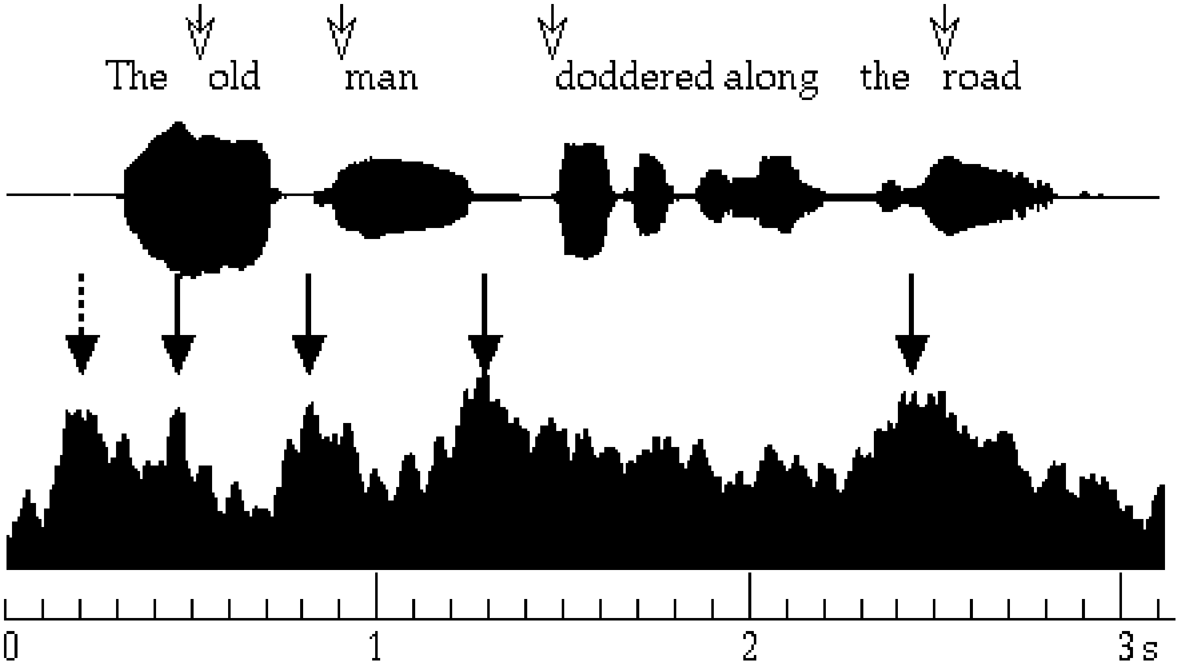

Figure 7. Above: a single audio record of *The old man doddered along the road*. Below: the mean of the absolute values of each of the 9,000 sampled points of the internal intercostal EMG in each of the 17 utterances of this phrase. Open arrows mark the onsets of the stressed syllables, and solid arrows mark four of the major peaks of internal intercostal activity. The dashed arrow at the beginning shows the increase in activity associated with the onset of speech.

Figure 7 clearly shows peaks in the internal intercostal activity, marked by the solid arrows, before each of the stressed syllables. In addition there is a peak at the beginning of the utterance (marked by a dashed arrow) that probably reflects the generation of the respiratory power necessary to start vocal fold vibration. The peak with the largest summed amplitude occurs shortly before the first vowel of 'doddered', the syllable that carries the primary stress in the utterance. At the end of the phrase, shortly before the stressed syllable 'road', there is another peak. There are no peaks in the EMG activity for the lengthy stretch of the utterance containing the four unstressed syllables '(dodd)ered along the'.

We cannot expect exact time locking of events between the average EMG signal over 17 phrases and a particular token of this phrase. But these results show that there are variations in the EMG activity during a phrase, and that the hypothesis that such variations are associated with the degree of emphasis required for particular syllables is not unreasonable.

Further confirmation of the role of the respiratory muscles in controlling stress come from the utterance shown in figure 3. Each of the 25 repetitions of the syllable [pa] can be considered as a stressed monosyllable. The external intercostals remain active for the first part of the utterance, including a number of repetitions of this syllable. Before each of these syllables 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 3, with its compressed time scale. Figure 8 shows just the EMG and the audio for the central part of figure 3. The lessening of the external intercostal activity just before each stressed syllable is apparent.

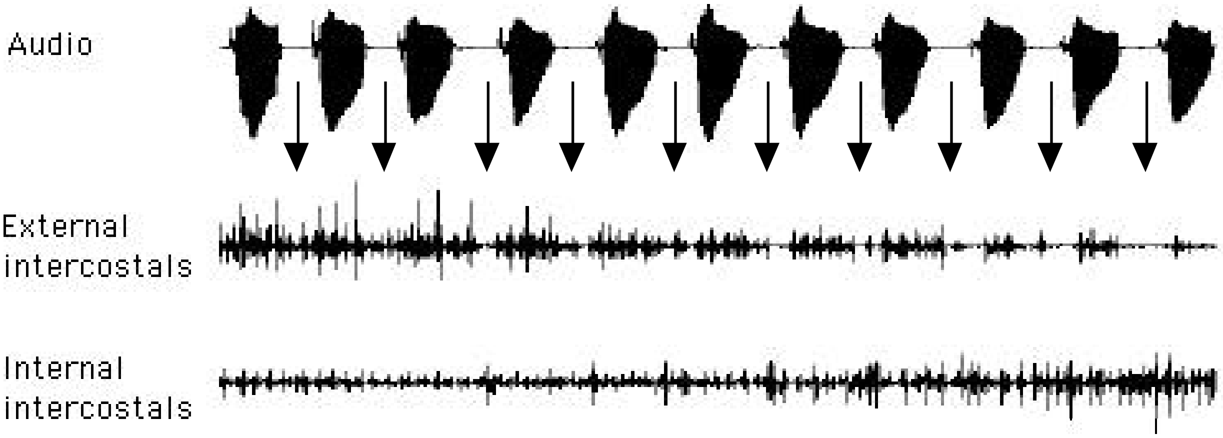

Figure 8. An enlargement of the part of Figure 3 from time 11,500 to 19,000 ms. The arrows mark onsets of stressed syllables and decreases of external intercostal activity.

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. Sometimes they are involved in pulling the rib cage down, and sometimes in simply allowing it to collapse more rapidly. Depending on the circumstances, stressed syllables may, or may not, involve other articulatory actions such as having a greater jaw opening, or a greater intensity, or a higher pitch. They may or may not have bursts of internal intercostals activity. But it appears they always use greater respiratory energy.

Conclusion

Our data provide an initial glimpse at important processes in the production of speech. On the basis of this data it seems that the pattern of muscular activity reported in the early Edinburgh studies is largely correct. The mean power for speech is provided by actions of the respiratory muscles that vary in accordance with the volume of air in the lungs. The external intercostals check the outgoing airflow when speaking with high lung volumes, and the internal intercostals, rectus abdominis and other muscles become more active as lung volume decreases. These results are consistent with the general hypothesis that the central nervous system is able to control subglottal pressure by using various combinations of passive elastic and active muscle forces, thereby achieving motor equivalence and acoustic invariance over a wide range of lung volumes as found in the early Edinburgh studies. The second aspect of the earlier studies is also supported by these experiments. For these speakers, stressed syllables are produced by increases of respiratory power above the mean, mainly by additional activity of the internal intercostals. The major refinement that should be incorporated in the general 1950s picture is that speakers may achieve fine control by the simultaneous use of antagonistic muscles.

We are well aware of the deficiencies in these studies. We have not attempted to assess the activity of all the intercostal muscles; muscles in different intercostal spaces may well behave differently. Nor have we examined variations in posture; our speakers were always seated, and there would undoubtedly have been systematic differences if they had been lying down or standing up. We have not recorded the movements of the rib-cage nor of the abdominal cavity that accompanied these muscular activities; we do not know the extent of the corresponding volume changes. But none of these problems lessen the value of what we have found. We consider these results worth publishing in that they indicate that the Edinburgh studies provided correct insights into the control of the power for speech, subject only to their lack of recognition of the important use of antagonistic muscles to achieve finer control.

References

Adams, C. (1979). English speech rhythm and the foreign learner. The Hague: Mouton.

Campbell, E. (1968). The respiratory muscles. In A. Bouhuys (Eds.), Sound production in man (pp. 135-140). New York: Annals of the New York Academy of Science.

Draper, M. H.,Ladefoged, P., & Whitteridge, D. (1959). Respiratory muscles in speech. Journal of Speech and Hearing Research, 2, 16-27.

Hirose, H. (1971). Electromyography of articulatory muscles: Current instrumentation and technique. Haskins Labs Status Reports, SR 25/26, 73-86.

Hixon, T.J. (1987). Respiratory function in speech. In T.J. Hixon (Ed.), Respiratory function in speech and song (pp. 1-54). Boston: College-Hill.

Hixon, T. J., & Weismer, G. (1995). Perspectives on the Edinburgh study of speech breathing. Journal of Speech and Hearing Research, 38, 42 - 60.

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

Kunze, L. H. (1964). Evaluation of methods of estimating sub-glottal air pressure. Journal of Speech and Hearing Research, 7, 151 - 164.

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

Loeb, G. E. (1990). The functional organization of muscles, motor units, and tasks. In M. D. B. & L. M. Mendell (Eds.), The Segmental Motor System (pp. 23 - 35). New York: Oxford University Press.

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

Loeb, G.E. and Gans, C. (1986). Electromyography for Experimentalists. Chicago: University of Chicago Press.

Netsell, R., Lotz, W. K., DuChane, A. S., & Barlow, S. M. (1991). Vocal tract aerodynamics during syllable productions: Normative data and theoretical implications. Journal of Voice, 5, 1 - 9.

Newsom Davis, J., & Sears, T. A. (1970). The proprioceptive reflex control of the intercostal muscles during their voluntary activation. Journal of Physiology, 209, 711-738.

Ohala, J. (1990). Respiratory activity in speech. In W. J. &. A. M. Hardcastle (Eds.), Speech production and speech modeling (pp. 23-53). The Netherlands: Kluwer.

Sears, T., & Newsom Davis, J. (1968). The control of respiratory muscles during voluntary breathing. In A. Bouhuys (Eds.), Sound production in man (pp. 183-190). New York: Annals of the New York Academy of Science.

Sharp, J. T., & Hyatt, R. E. (1986). Mechanical and electrical properties of respiraory muscles. In T. M. &. J. Mead (Eds.), Handbbok of physiology (pp. 385 - 414). Bethesda: American Physiological Society.

Shipp, T.,K.., I.,H.K., S., & P., M. (1988). Subglottal air pressure in spastic dysphonia speech. Folia Phoniatrica, 40(3), 105 – 110

Smitheran, J., and Hixon, T.J. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. J. Speech Hear. Disord., 46, 138-146.

Watson, P. J. (1999). What about the rib cage? ASHA Special Interest Division 3 Newsletter, 9.1, 3-4.

Weismer, G. (1985). Speech breathing: Contemporary views and findings. In R. Daniloff (Eds.), Speech science (pp. 47 - 72). San Diego: College-Hill Press.