Can the Energy Costs of Speech Movements be Measured? A Preliminary Feasibility Study

*Björn Lindblom **Seung-Jae Moon

Abstract

The main question addressed in this research was whether an adaptation of a standard exercise physiology procedure would be sensitive enough to record excess oxygen uptake associated with speech activity. Oxygen consumption was recorded for a single subject during 7-minute rest periods and an automatic speech task, also 7-minutes long and performed at three different vocal efforts. The data show measurable and systematic speech-induced modifications of breathing and oxygen uptake patterns. The subject was found to use less power for normal than for soft and loud speech. This result is similar to findings reported by experimental biologists on the energetics of locomotion. However, more comprehensive feasibility studies need to be undertaken on a larger population before solid and detailed conclusions about speech energy costs are possible. However, it appears clear that, for experimental tasks like the present one, i.e., variations in vocal effort, standard exercise physiology methods may indeed offer a viable approach to recording excess oxygen uptake associated with speech movements.

The status of 'articulatory ease' in phonetics and phonology

'*Articulatory ease*' is a traditional notion in linguistics that has often been invoked to explain both synchronic/diachronic aspects of sound systems and on-line phonetic patterns [1, 2, 3]. Despite the fact that it seems to have a strong common sense appeal its scientific status remains controversial. Ladefoged's position [4] is (i) that articulatory ease is language-dependent, (ii) that it cannot be measured and (iii) that therefore appeals to it are unscientific. In a paper on assimilation - usually seen as an articulatory process - Ohala [5] rejects articulatory factors in favor of a perceptual account. He argues that articulatory ease is likely to play a marginal role in shaping sound patterns and that invoking it makes explanations teleological.

As warnings against uncritical use of articulatory ease such statements are well taken, but, in the broader context of experimental biology, they appear overly pessimistic. This field presents a large literature on the energetics of locomotion in various species. Quantitative data are available on how humans and dogs walk and run, birds and bumblebees fly and how fish swim. A standard way of presenting results is to plot the amount of energy that the subject expends against traveling speed. The energy used is inferred from measurements of oxygen consumption made for subjects under steady-state conditions and therefore in an aerobic mode of oxygen uptake [6].

A parenthetical comment on the use of oxygen in these studies might be useful at this point. It is well known that all muscle contractions cost metabolic energy. This energy comes from oxidization of foodstuffs (whether carbohydrates, proteins or fats). Metabolic biochemistry is too complex to permit a brief summary (but see e.g., McArdle, Katch & Katch [7] for an overview). However, very conveniently for the experimental biologist, the bottom line of the input-output relationship is rather simple. It turns out that the input is

* Björn Lindblom is at the Department of Linguistics, University of Texas at Austin. He is also with the Speech Physiology and Perception Group at the Department of Linguistics, Stockholm University, Sweden.
** Seung-Jae Moon is currently at the Department of Foreign Languages and Linguistics as an exchange professor.He is permanently at the English Department, Ajou University, Suwon, Korea.

Manuscript Received:July 24, 2000

oxygen and food and that the output is free energy, water and carbon dioxide:

O_2 (+ food) → energy + H_2O + CO_2.

The above paragraph gives some idea of why a subject's O_2 consumption is used to provide an estimate of movement energy costs. We should note that such estimates have the advantage of being physiological. They take *all types of muscular activity* into account, both isotonic and isometric contractions. That implies that, in principle, they should give a much more accurate picture of energy costs than attempts to calculate the purely physical (bio-mechanical) work associated with a given task [8].

A typical example of this research is the study by Hoyt & Taylor [9] who measured energy consumption for horses walking, trotting and galloping. The subjects were observed as they moved freely and at speeds controlled by a treadmill. The energy used expressed per unit distance traveled and plotted against traveling speed formed U-shaped curves with distinct minima. Significantly, these minima were found to occur at speeds that subjects spontaneously adopted when moving freely and unconstrained by the speed of the treadmill. Such findings rest solidly on a large body of physiological studies [7] and have been reported for a number of species. Experimental biologists interpret them to suggest that locomotion is shaped by a criterion of 'minimum energy expenditure'.

Why should speech movements be different?

Are speech movements and whole body movements similarly organized? Since energy costs for speech are likely to be small in comparison with those of locomotion, it might be argued that they play no major role in shaping phonetic movements. It is true that, until speech energy costs can be reliably measured, we have no basis for settling that issue satisfactorily. However, evolution's tendency towards parsimony would make us expect the same rules to apply for small as for big movements.

Among phoneticians, it is widely believed that both speech and sound patterns have many characteristics that are most readily accounted for in terms of production constraints. Conceivably, we will ultimately be able to show that many of them derive from a minimum energy expenditure condition. For instance, in running speech, prosodic modulations and speaking styles produce both strong elaborated and weak reduced forms [10]. In our opinion, this segmental dynamics is an obvious candidate for an analysis based on energetics. Similarly, looking typologically at phonological systems, we observe a clear preference for low-cost motor patterns [11]. Hence, in line with that type of traditional thinking, it does not seem totally preposterous to hypothesize that minimization of energy expenditure plays a causal role in:

- the absence of vegetative movements and mouth sounds;
- determining the feature composition of phonetic segments (e.g., why are /i/ and /u/ universally 'close' vowels?);
- constraining the universal organization of syllabic & phonotactic structure,
- the patterning of diachronic and synchronic lenition and fortition processes;
- shaping the system-dependent selection of phonetic values in segment inventories and
- several other topics.

Project goal: Can the energy costs of speech movements be measured?

However, at best the above suggestions currently represent no more than speculative possibilities. What is clearly needed is some solid empirical work that would give us an indication of (1) whether energy costs can be measured for speech movements; and (2) if energy costs play any role at all in shaping phonetic and phonological patterns. The present study was initiated to address the first of these questions in a preliminary way.

A pilot experiment: Design and procedures

We used an experimental set-up similar to that commonly used in medical testing and exercise physiology to measure the energy consumption of whole body movements such as locomotion [7]. The equipment used was made available to us by Dr Robert Spina. For reasons of time and space we were limited to running a single subject.

To measure inspiratory and expiratory air flow and oxygen consumption and carbon dioxide production, we used the Physio-Dyne Max-1 system of Dr Spina's lab. The tracking of ventilatory parameters was performed by means of a flow transducer, amplifier and Hans Rudolph pulmonary pneumotach, components which make up the Physio-Dyne FLO-1B unit. The oxygen analyzer was an OXY-32. Devices of this sort operate on a paramagnetic principle. Records of CO_2 contents were obtained using a CO_2-44 infrared analyzer. All input channels were connected on-line to a PC computer. They were calibrated and processed by software supplied by Physio-Dyne. The subject was connected to this instrumentation by way of a face mask (Hans Rudolph model # 8920, a nasal & mouth breathing face mask two-way non-rebreathing) with appropriate adapters.

The audio signal was captured with two microphones. One microphone (Audio-Technica AT831 unidirectional lavalier microphone) was tightly custom fitted (by Hans Rudolph, Inc.) into the mask so that there might be no air passing through the connection. It was located on the left side of the nose area to prevent it from picking up the direct air stream from the mouth. The other microphone (B&K microphone type 4189 with preamplifier type 2669L and Nexus power supply type 2690A0S2) was located in front of the subject at 11" distance. Right beside the microphone was located a sound level meter (B&K type 2232) to obtain estimates in absolute dB of the subject's vocal effort. The average SPL values for the three conditions were 67 dB (normal), 77 dB (loud) and 61.5 dB (soft) as measured 11" from the subject's lips. The audio signal from each microphone was recorded into separate channels using Marantz PMD430 tape recorder. This particular machine was chosen to utilize its easy-to-read VU meters as a way to keep the subject's vocal efforts at desirable levels.

In planning the present experiment our aim was to replicate the work of Russell et al. [12]. The following overall outline is similar to that used by these authors.

Rest	7 minutes	
Speech	7 minutes	Normal vocal effort
Rest	7 minutes	
Speech	7 minutes	Loud vocal effort
Rest	7 minutes	
Speech	7 minutes	Soft vocal effort
Rest	7 minutes	

The same task was used in all three speech conditions: Counting from one through eight in Swedish (the subject's native language) which would, it was argued, minimize cognitive load and set up an "*automatic speech mode*". All digits were in their monosyllabic form:

1	en	[ˈɛn]
2	två	[ˈtvo:]
3	tre	[ˈtre:]
4	fyr	[ˈfy:z]
5	fem	[ˈfɛm:]
6	sex	[ˈsɛks]
7	sju	[ˈɧu:β]
8	ått	[ˈct:]

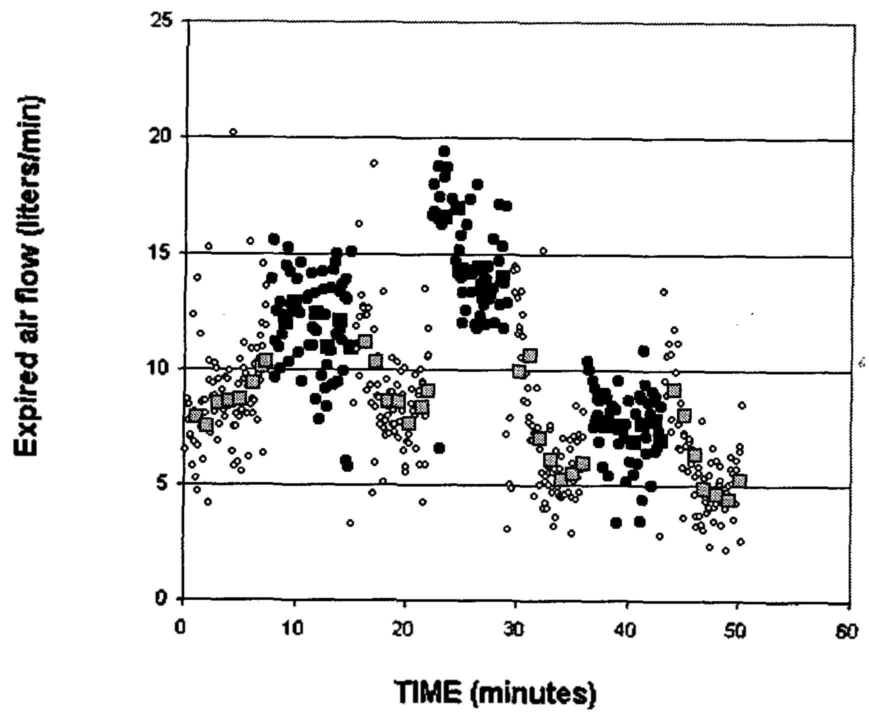
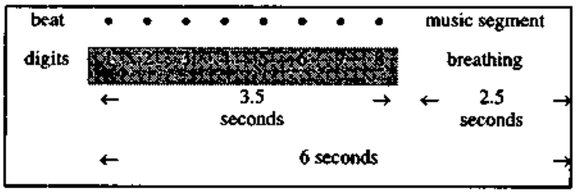


Figure 1. Minute-by-minute ventilation and breath-by-breath ventilation.

Through ear-phones the subject listened to a pre-recorded metronome beat presented in a loop mode. His task was to synchronize each syllable with a beat.



Beats occurred with a spacing of 500 ms and a pause of 2.5 seconds for breathing. The series of digits was repeated regularly once every 6 seconds which produced a regular breathing frequency of 10 breaths per minute. The identical format was used for all three conditions so as to produce comparable speaking tempos and a fixed number of stressed syllables. The subject was asked to monitor his SPL by making the needle of the tape recorder VU meter deflect to a pre-determined level.

The data were recorded in breath-by-breath mode and saved as an ASCII file on the PC hard disk.

Results

Figure 1 provides a view of the subject's respiratory behavior during the experiment. It shows the '*minute ventilation*', that is the flow of expired air in liters per minute.

Circular data points represent breath-by-breath samples. Squares are average values calculated with a 60-second window. Rest breathing is indicated by the unfilled circles and lighter squares. The normal, loud and soft conditions have filled somewhat larger circles.

The overall impression of this plot is that the data points show a falling trend with speech conditions showing higher levels relative to the surrounding rests. Average data on pulmonary ventilation [7: p228] indicate that normal adult males take about 12 breaths per minute and inhale about 500 ml per breath. That implies a minute ventilation of 6 liters per minute. Initially the values for the first rest period in Figure 1 are somewhat in excess of that number. A possible interpretation is that in anticipation of the task the subject responds by hyperventilating somewhat.

The rest periods exhibit similar concave-upwards curve shapes. Values go up before speech conditions and then fall. At the onsets of speaking deeper breaths are taken. Absolute values are highest for loud speech and lowest for soft speech. However, in view of the subject's tendency to hyperventilate initially, ventilation during tasks should be related to the surrounding rests.

In a sense, the patterns presented in Figure 1 provide a picture of the advance estimates made by the subject's

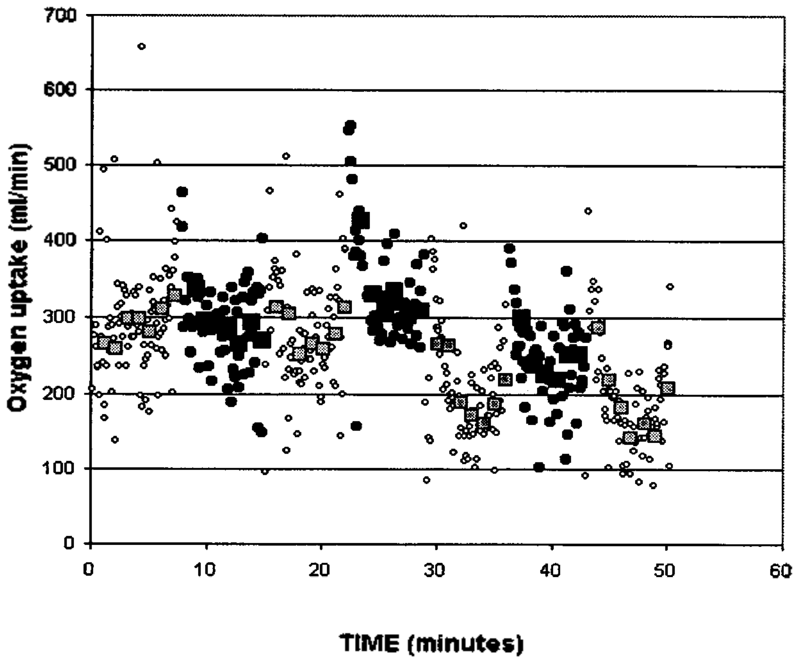


Figure 2. Oxygen consumption: breath-by-breath (circles) and minute-by-minute (squares).

brain about oxygen needs. The O_2 actually used is shown in Figure 2.

Average data on oxygen consumption indicate that normal adult males use about 250 ml/min during rest. Initial values for the first rest period in Figure 2 are close to that figure. However, the overall pattern of O_2 uptake shows systematic deviations from that baseline value. The differences are similar to those in Figure 1 in that: (1) The data points show a declining trend during the course of the experiment. (2) They fall and rise in between speech tasks. And (3), values during speaking appear somewhat elevated relative to the surrounding rests - an effect which perhaps most clearly seen for 'loud' and 'soft'.

Preliminary interpretations

One of the questions that we are trying to address is whether the present method (an adaptation of a standard procedure in exercise physiology) is sensitive enough to record excess oxygen uptake associated with speech activity superimposed on basic metabolic processes. In other words: 'Can energy costs for speech movements be reliably measured?' Do the present measurements indicate speech-specific effects?

For a preliminary answer, we propose the following interpretation of the data of Figures 1 and 2.

To determine whether there are differences between O_2 behavior during quiet breathing and speech we first need to define a "baseline". Again we note that the hyperventilation pattern produces a declination of the baseline both for minute ventilation and for oxygen use and that there is a tendency for the data to reach minima during rest periods. The approach chosen here was to determine the minimum values for each rest period in terms of the 60-second averages and then interpolate linearly between those points.

Second, we should consider the shape of the curve during the speech conditions. Figure 2 indicates initial 'overshoots' in all three cases but that the last four minutes of each speech task tends to be more stable showing a 'steady-state'. We assume that the first three minutes show the effect of the system adapting to the task. They will therefore be disregarded.

These two observations underlie the presentation in Figure 3. It shows the same data points as in Figure 2 except that only the averages for the last four minutes of the speech conditions are plotted (darker squares). Also we have added a "baseline" for quiet breathing derived as described above.

We took the differences between the average O_2 levels and the corresponding baseline values and added them up over the last four minutes. We then obtained the numbers shown in Table 1 below.

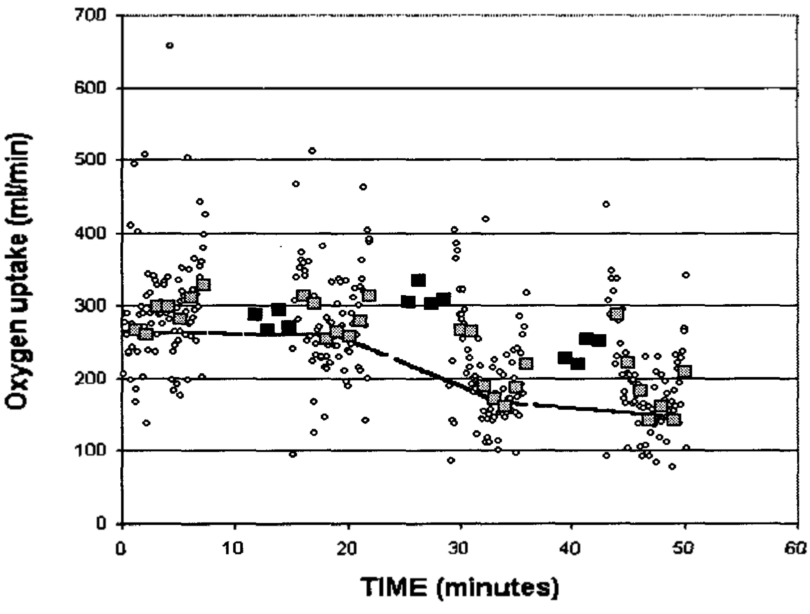


Figure 3. Oxygen consumption: steady-states and baseline.

Table 1. Amount of excess oxygen consumed during the last 4 minutes of each speech condition.

	Amount of excess oxygen (ml) during 4-minute period
Soft	320
Norm	79
Loud	414

To translate the observations of Table 1 into energy and rate of energy use we first note that, in a first approximation, 1 cm^3 of oxygen used corresponds to 20 joules of energy [6: p.65]. The rate at which energy is used is usually expressed in watts (1 *watt* = 1 *joules/second*).

Table 2. The cost of different vocal efforts in watts.

	Amount of excess oxygen (ml)	Energy expenditure (watts)	
Soft	320	20*320/4*60	27
Norm	79	20*79/4*60	7
Loud	414	20*414/4*60	35

Table 2 and Figure 4 show the result of carrying out those steps.

Limitations of present study

One of the aims was to get a preliminary indication whether a method widely used in exercise physiology could be adapted to determine the energy costs of phonetic movements by measuring excess oxygen uptake due to speech. Some limitations will first be mentioned before we present the conclusions.

The measurements are limited to data from a single subject. This calls for caution in generalizing the results. The experiment investigated three degrees of vocal effort. The subject was instructed to count from 1 to 8 in Swedish in step with a metronome signal presented through ear-phones. This procedure made the 'normal', 'loud' and 'soft' conditions perfectly comparable in terms of speaking tempo and the number of stressed syllables produced per breath.

However, a *single order* of the three tasks was used: (1) 'normal', (2) 'loud', (3) 'soft'. Ideally, the experiment should have included all possible orders of those conditions. It would then have been possible to determine if certain aspects of the data are genuine, or simply order effects. A case in point is the downdrift of both minute ventilation (Figure 1) and oxygen consumption (Figure 2). The 'baseline' reached during rest periods is seen to exhibit a gradual declination during the course of the experiment.

The *'baseline' for rest* was defined in terms of the minimum values reached during each rest interval (Figure 3). This approach can certainly be questioned. It needs further investigation and justification. The values in Tables 1 and 2 and in Figure 4 were derived on the basis

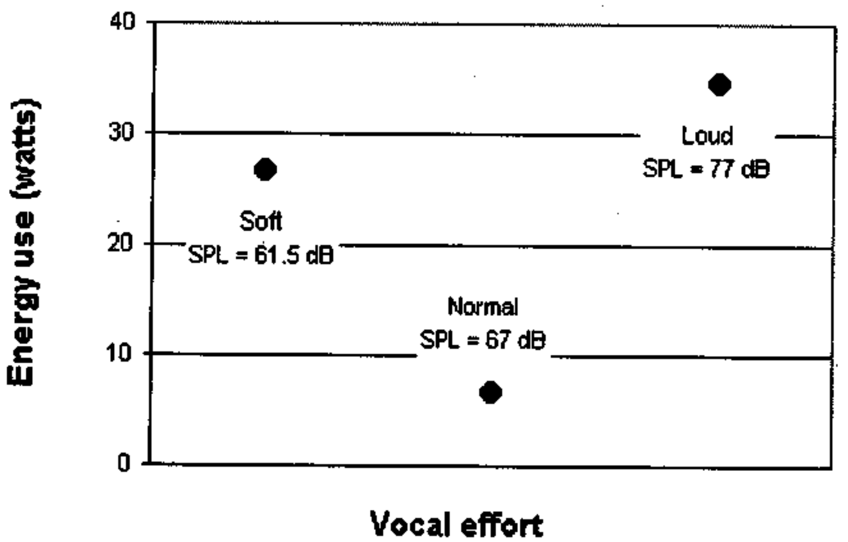


Figure 4. Power consumption at different vocal efforts.

of that definition. They should therefore be used with caution.

Another step in the presentation is the *omission of the first three minutes* of each condition. In the minute ventilation and the oxygen data, the talker tends to show an initial 'overshoot' that changes into a steady-state pattern after approximately three minutes. We would justify this elimination by suggesting that, near the onset, each task gave rise to hyperventilation.

The experiment did not measure CO_2 levels at the end of each breath. Possibly, tracking this parameter might throw some further light on the data.

The preceding points limit the conclusions of the study. However, they do not seem irremediable. It should be possible to run more subjects, use a more varied and properly counterbalanced design and to examine alternative ways of processing the data. Additional parameters should be measured (PET CO_2, heart rate, subglottal pressure, ...).

Tentative conclusions

The above remarks notwithstanding, it does appear possible to make a few positive observations.

The general pattern of the data indicate that, both for minute ventilation and oxygen consumption, the values for the initial rest period approach standard population level averages. The hyperventilation effects (both overall and at the onset of each speech task) are also compatible with normal behavior during exercise.

Speaking does seem to modify both the breathing and the oxygen uptake. Most clearly we see this for the 'loud' condition, but also to some extent in 'soft' speech.

For comparison, we present Table 3 with data from the study by Russell et al. [12]. It shows the average oxygen consumption for 12 adult male speakers during the last 2 minutes of 7 minutes long speech tasks.

Table 3. Data from Russell et al (1998) on oxygen consumption in 12 adult male speakers for different vocal efforts.

	SPL (dB)	O_2 (ml/min)
quiet breathing		358
soft	88	276
normal	93	362
loud	103	428

The numbers are higher than those reported here. Since body weight influences the absolute values, it is not possible to compare the figures in detail. As suggested by the table, Russell et al. conclude that there is an increase of oxygen uptake as a function of effort. They observe that the 'soft' condition shows reduced levels compared with rest and argue that this can undoubtedly be attributed to the subjects 'holding back' and creating an 'oxygen debt' to be paid back later during the following rest. The present data seem to suggest a similar interpretation (cf. below).

Figures 4 uses a format that differs from that in Russell et al. since we are interested in determining the amount of oxygen used in excess of basic metabolic needs. Furthermore, to facilitate deciding whether the levels observed are at all reasonable the data are calibrated in watts (joules/second). As expected, the numbers are small. Normal speech appears to need less power than the light bulb in the typical bedside lamp whereas soft and loud require more. The U-shaped configuration is reminiscent of the observations reported for locomotor tasks (cf. introduction). It is also compatible with the discussion in Russell et al. who state that,

> " ... as SPL rises above or below comfortable SPL, speech breathing requires more energy". [12: p.239].

On the basis of the present limited investigation, we feel justified in concluding that:

- There is nothing in this study to discourage further attempts to investigate speech energetics.
- Rather, the use of conventional exercise physiology techniques appears feasible in studying speech movements in view of the fact that different vocal efforts were seen to cause measurable and systematic perturbations of breathing and oxygen uptake patterns.
- It would appear that speech-induced oxygen use in excess of basic metabolic levels has been preliminarily demonstrated in this pilot experiment.
- Using the observed excess oxygen consumption to estimate speech energy costs, we note that the present subject appears to have used less power for normal than for soft and loud speech (cf. the U-shaped configuration of Figure 4).
- This parallel with a large body of locomotion observations is certainly encouraging but needs to be substantiated in more comprehensive studies before solid conclusions about speech energy costs are possible

References

1. P. Passy, Étude sur les changements phonétiques, Paris: Firmin-Didot, 1890.

2. P. Kiparsky, "Phonological change", in F. J. Newmeyer(ed), Linguistics: The Cambridge Survey, Vol I, Linguistic Theory: Foundations, (University Press, Cambridge, 1988) pp. 363-415.
3. K. Kohler, "Segmental reduction in connected speech in German: Phonological facts and phonetic explanations", in W.J. Hardcastle & A. Marchal (eds), Speech production and speech modeling, (Kluwer, Dordrecht, 1990) pp. 69-92.
4. P. Ladefoged, "Some reflections on the IPA", Journal of Phonetics 18, pp.335-346, 1990.
5. J. J. Ohala, "The phonetics and phonology of aspects of assimilation", in J. Kingston & M. Beckman (eds), Papers in laboratory phonology: Vol 1. Between grammar and the physics of speech, (Cambridge University Press, Cambridge, 1990), pp. 258-275.
6. R. McNeill Alexander, The Human Machine, Columbia University Press, New-York, 1992.
7. W. D. McArdle, F. I. Katch, & V. L. Katch, Exercise physiology, 4th ed, Williams&Wilkins, Baltimore, 1996.
8. A. A. Biewener "Bio-mechanics of mammalian terrestrial locomotion", Science 250, pp.1097-1103, 1990.
9. D. F. Hoyt & C. R. Taylor, "Gait and the energetics of locomotion in horses", Nature 292, 239, 1981.
10. S-J. Moon & B. Lindblom, "Interaction between duration, context and speaking style in English stressed vowels", J Acoust Soc Am 96(1), pp. 40-55, 1994.
11. B. Lindblom, "Economy of speech gestures", in P. F. MacNeilage (ed), Speech Production, (Springer Verlag:New York, 1983), pp.217-246.
12. B. A. Russell, F. J. Cerny & E. T. Stathopoulos, "Effects of varied vocal intensity on ventilation and energy expenditure in women and men", JSLHR 41, pp.239-248, 1998.

▲ Björn Lindblom

1960 Stockholm University (Fil mag)
1963 Uppsala University (Fil lic)
1968 Lund University (Fil dr)
1994-1999, Professor of phonetics, Stockholm University
1999- present, Emeritus professor, Stockholm University
1997 - present, Visiting professor, University of Texas at Austin

▲ Seung-Jae Moon

The Journal of the Acoustical Society of Korea, Vol. 18, No.8, 1999.