

Northeastern University

**Psychology Faculty Publications** 

Department of Psychology

October 01, 2002

## Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues

Joseph S. Perkell Massachusetts Institute of Technology

Majid Zandipour Massachusetts Institute of Technology

Melanie L. Matthies Massachusetts Institute of Technology

Harlan Lane Northeastern University; Massachusetts Institute of Technology

#### **Recommended** Citation

Perkell, Joseph S.; Zandipour, Majid; Matthies, Melanie L.; and Lane, Harlan, "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues" (2002). *Psychology Faculty Publications*. Paper 6. http://hdl.handle.net/2047/d20000875

This work is available open access, hosted by Northeastern University.

### Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues

Joseph S. Perkell,<sup>a)</sup> Majid Zandipour, Melanie L. Matthies,<sup>b)</sup> and Harlan Lane<sup>c)</sup> Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 6 March 1998; revised 8 June 2002; accepted 16 July 2002)

This study explores the hypothesis that clear speech is produced with greater "articulatory effort" than normal speech. Kinematic and acoustic data were gathered from seven subjects as they pronounced multiple repetitions of utterances in different speaking conditions, including normal, fast, clear, and slow. Data were analyzed within a framework based on a dynamical model of single-axis frictionless movements, in which peak movement speed is used as a relative measure of articulatory effort (Nelson, 1983). There were differences in peak movement speed, distance and duration among the conditions and among the speakers. Three speakers produced the "clear" condition utterances with movements that had larger distances and durations than those for "normal" utterances. Analyses of the data within a peak speed, distance, duration "performance space" indicated increased effort (reflected in greater peak speed) in the clear condition for the three speakers, in support of the hypothesis. The remaining four speakers used other combinations of parameters to produce the clear condition. The validity of the simple dynamical model for analyzing these complex movements was considered by examining several additional parameters. Some movement characteristics differed from those required for the model-based analysis, presumably because the articulators are complicated structurally and interact with one another mechanically. More refined tests of control strategies for different speaking styles will depend on future analyses of more complicated movements with more realistic models. © 2002 Acoustical Society of America. [DOI: 10.1121/1.1506369]

PACS numbers: 43.70.Aj, 43.60.Bk [AL]

#### I. INTRODUCTION

In order to improve models of speech motor control, it is important to characterize the various constraints under which speech production operates. The requirement for intelligibility imposes constraints on the acoustic characteristics of the signal that are related to clarity (cf. Picheny et al., 1986; Moon and Lindblom, 1994; Moon, 1991). Clarity constraints may vary according to environmental conditions that require the speaker to use different styles—such as speaking clearly in a noisy environment or when the listener has a hearing loss. In comparison to normal (citation or casual) speech, clear speech has been shown to be characterized by greater intelligibility, greater intensity (by 3–5 dB in vowel nuclei), longer sound segments, an expanded vowel space, tighter acoustic clustering within vowel categories, greater distinctiveness of VOT between voiced and voiceless stop consonants and released word-final stops (cf. Picheny et al., 1986; Moon and Lindblom, 1994; Moon, 1991). Such clear speech might be produced with movements that are larger, slower, more precise, and possibly more effortful. Lindblom (1990) has hypothesized that there is a trade-off between clarity and economy of effort that occurs with changes in speaking

<sup>a)</sup>Also at the Department of Brain and Cognitive Sciences, MIT and the Department of Cognitive and Neural Systems, Boston University. Electronic mail: perkell@speech.mit.edu styles: clear speech should be produced with greater articulatory effort than normal speech. In the current study we test the hypothesis that speakers will exert more effort when asked to speak clearly than when they speak normally. We test this hypothesis by examining a relative measure of effort in the production of speech movements in various speaking conditions. For this purpose, we define "economy of effort" as a strategy in which the motor control system attempts to minimize the physical "cost" of making articulatory movements. Economy of effort appears to be a characteristic of movement in general, and it is a principle that guides speech movement planning in the control model of Guenther (Guenther, 1995; Guenther *et al.*, 1998; Perkell *et al.*, 2000).

To compare a measure of articulatory effort across different speaking conditions, the study uses peak movement speed, which is an approximation based on a cost optimization analysis of a dynamical model of single-axis frictionless movements (Nelson, 1983). A second objective of this study is to consider the extent to which such a simple model may be valid for analyzing complicated speech movements.

#### **II. BACKGROUND**

In the first study that directly addressed this issue, Adams (1990) reported on tongue-blade opening movements for the word "tad" as produced in casual and clear conditions by five speakers. "The clear speech condition was associated with longer movement durations and larger maximum dis-

<sup>&</sup>lt;sup>b)</sup>Also at the Department of Communication Disorders, Boston University.

<sup>&</sup>lt;sup>c)</sup>Also at the Department of Psychology, Northeastern University.

placement and peak velocity values relative to the casual condition in some subjects." (p. iii). In order to develop a rationale for a more thorough investigation that includes an analysis of articulatory effort, it is helpful to consider observations of articulatory kinematics that have been made in studies of other factors, such as speaking rate, speech tempo, stress, and vowel quantity.

In the Kuehn and Moll (1976) kinematic study of speaking rate, it was found that in order to increase speaking rate, some speakers increased articulatory velocities and produced little articulatory undershoot, while others did not increase velocities and produced more articulatory undershoot. There were positive relationships across subjects between both articulatory velocity and movement displacement and the size of the articulators, possibly reflecting a generally observed linear relationship between peak velocity and distance (cf. Ostry et al., 1987; Ostry and Munhall, 1985; Linville, 1982; Vatikiotis-Bateson and Fletcher, 1992; Flanagan et al., 1990). Sonoda and Nakakido (1986) studied the effect of speaking rate on jaw movements. Similarly to Kuehn and Moll, they observed that the increase in speaking rate was produced either with an increase in velocity and little change in distance (i.e., no undershoot) or with relatively constant velocity and a decrease in movement distance (undershoot).

In a kinematic study of tempo and prosody, Edwards *et al.* (1991) found that two of four subjects, in complying with "slow speech" instructions, decreased the velocity of a phrase-final mandible closing gesture, while the other two delayed the onset of the closing gesture without decreasing velocity. The latter two subjects had generally longer syllable durations than the former two. As an explanation for these findings, Edwards *et al.* (1991) hypothesized a lower limit on velocity that may be physiologically based or perceptually based (to preserve phonetic identity).

To help make inferences about articulatory effort and the control mechanisms that underlie kinematic observations, some investigators (cf. Munhall *et al.*, 1985; Ostry and Munhall, 1985; Hertrich and Ackermann, 1997) have adopted principles from a cost optimization analysis of single-axis movements of an undamped, linear, mass-spring model (Nelson, 1983). According to this analysis, *peak velocity* can be used as a *relative measure of the physical cost* of performing skilled movements.

Nelson's (1983) analysis shows that minimization of energy or of jerk (the third derivative of displacement vs time) produces in each case a profile of movement velocity versus time similar to the pattern that results from an undamped linear mass-spring system with constant stiffness, in which velocity vs time for a single movement looks similar to the positive half of a sinusoid function. Figure 1(a), from Nelson (1983), shows velocity profiles for minimum energy (E), minimum jerk (J), and constant stiffness (K), all three of which look similar to those observed from speech movements. The peak velocity of the movement of the linear spring model is related to movement distance and time constraints by  $V = \pi D/2T$  (where V = peak velocity, D = distance, and T = time). Peak velocity is also equivalent to the impulse cost measure (time integral of the magnitude of the force per unit mass) in Nelson's analysis. Even though a



FIG. 1. (a) Comparison of velocity patterns of a single-axis, frictionless system for the same movement time and distance that are optimum with respect to five different objectives: A, minimum peak acceleration (solid line); E, minimum energy (dashed line); J, minimum jerk (solid); K, constant stiffness (dotted); and V, minimum peak velocity; or impulse (solid) (Fig. 3 in Nelson, 1983). (b) Curves of minimum percent cost as a function of movement time for fixed distance, D, and acceleration limit U. V, peak velocity (impulse) cost; A, peak acceleration cost; E, energy cost; and J, jerk cost (Fig. 6 in Nelson, 1983).

minimum-impulse solution produces a trapezoidally shaped velocity profile [V, in Fig. 1(a)] that is less like those of speech movements, it is possible to use peak velocity as a measure of *relative* effort, because all of the minimum-cost solutions produce similarly shaped cost functions, shown in Fig. 1(b) (when percent cost is plotted as a function of movement time—from Nelson, 1983). As Nelson points out, a single criterion is "generally insufficient to encompass what we mean by optimum" (p. 140), and skilled movements reflect a compromise or trade off among competing objectives, one of which, in the case of speech, is producing an intelligible sound sequence. As explained below, the current study

#### Perkell et al.: Economy of effort in speech

is of nearly linear two-dimensional movements, so we use peak movement speed (the tangential velocity maximum) as a relative measure of effort. We also examine additional parameters to investigate other aspects of the movements.

The ratio of peak velocity to average velocity,  $c = V/V_a$  (where  $V_a$  = average velocity, D/T), provides an index of velocity profile shape (Nelson, 1983; Ostry and Munhall, 1985). To some extent, the velocity profile shape (as indexed by the value of c) can reflect the selection of a particular cost optimization criterion (e.g., energy, jerk, impulse). For model velocity profiles that are symmetrical (equal durations of acceleration and deceleration phases), unimodal (smooth, with one acceleration peak in the first half of the movement and one deceleration peak in the second half) and have velocity values of zero at movement beginning and end, a value of c = 1.0 would correspond to a rectangular profile and a value of 2.0 would correspond to a triangular profile [profile (A) in Fig. 1(a)]. A rectangular profile would be produced by an acceleration impulse at movement beginning and a deceleration impulse of equal magnitude at movement end. A triangular profile would be produced by an acceleration pulse for half of the movement followed by a deceleration pulse of equal magnitude for the second half of the movement. Thus, these two profiles represent theoretical, physically unrealizable limits; many actual movements may fall between the two patterns. On the other hand, if actual movements have velocity profiles that do not meet the above criteria (unimodal, symmetrical, and zero velocity at movement beginning and end), values of c can exceed 2.0.

Another parameter, the ratio of peak velocity to distance, has been considered to reflect actuator "stiffness," if the system can be represented by a second-order damped dynamical model (cf. Nelson, 1983; Ostry *et al.*, 1983; Ostry and Munhall, 1985).<sup>1</sup> The level of stiffness may be thought of as a relative index of the level of muscle activity that is used to produce a movement.

In the above-referenced rate and clarity study (Adams, 1990), normal and faster-than-normal speech was produced with unimodal velocity profiles, while the slower-thannormal speech had multipeaked (i.e., less smooth) velocity profiles. Values of *c* were found to decrease with increases in speaking rate; they approached  $\pi/2$  (1.57), the value that is characteristic of the sinusoidal velocity profile of the frictionless mass-spring model.

Hertrich and Ackermann (1997) measured acoustic and labial kinematic variables in a study of vowel quantity in German. They found intersubject differences for several measures, and interactions among the measures. The results included: distinct linear peak velocity-distance relationships for each quantity class, an influence of vowel quantity on the scaling of velocity and amplitude in oral opening movements, more peaked velocity profiles for long than short vowels, and differential effects of vowel quantity on the symmetry of velocity profiles in opening versus closing movements. Values of parameter c were consistently greater than  $\pi/2$ . Closing gestures were characterized as fast and ballistic, and opening gestures were more sensitive to phonetic timing. Among other things, Hertrich and Ackermann concluded that durational information was conveyed more consistently by acoustic results than by movement durations.

In sum, previous studies of articulatory kinematics invariably have found intersubject differences. They also have shown systematic relations among movement parameters, such as velocity versus distance, that are characteristic of other types of movements and may be used to make some inferences about aspects of the underlying control. According to a cost analysis of uniaxial frictionless movements, peak velocity may be used as a relative index of effort. The ratio of peak velocity to distance may be used to indicate relative levels of muscle stiffness underlying the movements. The ratio of peak velocity to average velocity (c) can also reflect relative effort, but only if the movements being compared have smooth, symmetrical velocity profiles and have the same duration. On the other hand, if movements are not smooth or symmetrical (i.e., have acceleration and deceleration phases of different durations) and they have relatively high values of c (approaching and exceeding 2.0), the simple modeling framework may not be an entirely suitable tool of analysis.

#### **III. METHODS**

Based on the preceding background, a study was conducted of measures of effort and other movement characteristics in six speaking conditions, using data from utterances elicited from seven subjects in those conditions.

#### A. Subjects, speech materials, and data acquisition

The subjects were seven young adult speakers of American English without speech or hearing deficits or pronounced regional dialect, three females and four males.

The subjects read short sentences containing CVC "test" words in six different speech conditions. Utterances were of the form "say  $C_1VC_2$  again," where  $C_1VC_2$  is "bob," "dod" or "gog," with stress on the CVC word. The three test words were selected to investigate the effect of articulator (lower lip, tongue blade, tongue body) on the movements. (It is acknowledged that the movements of all three articulators are influenced by mandible movements: the lower lip is the most influenced and the tongue body is the least influenced. However, these influences of the mandible are not examined in the current study.) Both the opening movement toward the V1 target and the closing movement toward C2 were examined to investigate the effect of movement type.<sup>2</sup> Normal speech was elicited by asking the subjects to pronounce the utterances at a "conversational" pace. Fast speech was elicited by asking the subjects to pronounce the utterances at what they perceived as twice their normal rate. Slow speech was elicited by asking the subjects to pronounce the utterances at what they perceived as half their normal rate. Clear speech was elicited by telling the subjects that someone in the next room was checking their pronunciation and they would be rewarded according to the number of utterances pronounced correctly. (Speaking louder was discouraged; articulating clearly was encouraged.) A rapid + clear condition rewarded the subjects for a combination of brevity and number of correctly pronounced tokens. Finally, in a casual condition, called "informal," subjects were

J. Acoust. Soc. Am., Vol. 112, No. 4, October 2002

shown a number of  $4 \times 4$  matrices (one at a time), with each cell containing a test word and a number (1-4), and the columns labeled A-D. The subject was asked to tell a listener how to number a similar (un-numbered) matrix by saying, for example, "One is the 'bob' in *A*, the 'dod' in *C*, the 'gog' in *D*.... Two is the gog in *B*, the dod in *D*,...[etc.]." In this condition the subject was not rewarded for clarity and was told not to worry about mistakes; the resulting utterances were judged informally to be quite casual in nature.

There were 15 repetitions of each utterance in each condition. Utterances containing the three CVC words were interspersed randomly with repetitions of utterances containing other CVC words and alternative stress patterns that were designed to explore the acoustic and kinematic effects of stress, vowel quality, and consonant place and manner of articulation (not covered in the current report). The full corpus included approximately 1400 utterances and took approximately 45 minutes to produce.

Recordings were made of the acoustic signal and displacement versus time signals from small (5 mm long  $\times 2.5$  mm diam) coils placed in the midsagittal plane on the lips, tongue and mandible, as transduced by an ElectroMagnetic Midsagittal Articulometer (EMMA) system (Perkell et al., 1992). The transducer coils were mounted with adhesive on the vermilion border of the upper lip (UL) and lower lip (LL), the gingival papilla between the two lower central incisors, the tongue body dorsum about 5 cm from the tip (called tongue back, TB), and the tongue blade about 1 cm from the tip (called tongue front, TF). Additional transducers, on the bridge of the nose and upper incisor, were used as a maxillary frame of reference to define the coordinate system of movements of the other transducers. A directional microphone was suspended 14 inches from the subject's mouth. Utterance materials were presented, 10 items at a time, on sheets of paper hanging in front of the subject. After a short period of adaptation, the presence of the transducers was judged aurally to have a negligible effect on the subjects' utterances.

## B. Signal processing, data extraction, and data analysis

Each channel of movement signal was digitized at 312.5 samples per second (aggregate rate for 32 channels, 10 kHz), and the speech signal was sampled at 10 kHz after being low-pass filtered at 4.5 kHz. During the subsequent signal processing, articulator x (horizontal) and y (vertical) displacements in the midsagittal plane were calculated from the EMMA output voltages (see Perkell et al., 1992). The displacement signal was low-pass filtered with an FIR filter that began to roll off at 13 Hz and was greater than 60 dB down at 22 Hz. Then, velocity and acceleration in the x and ydirections were computed by differentiating the low-pass filtered displacement vs time signals with a backward difference approximation (computing the difference between adjacent values divided by the time step, 3.2 ms). Following differentiation, the resulting velocity and acceleration signals were low-pass filtered with an FIR filter that began to roll off at 38 Hz and was greater than 60 dB down at 47 Hz.

To check the validity of the EMMA data and look for long-term trends that could include fatigue effects, we examined time-series plots of the *x* and *y* values and the EMMA misalignment correction index (Perkell *et al.*, 1992) for each transducer, extracted at the time of the beginning of each token. We also examined midsagittal x-y plots of the same data. Abrupt changes and long-term drift in the time series and outlying points in the x-y display were few in number; they were noted and the corresponding data were removed from subsequent analyses.

Figure 2 illustrates the data extraction procedures; it shows signals for a portion of the utterance "Say gog again," spoken in the normal condition. The acoustic signal [panel (A) of Fig. 2] was labeled manually in two stages: (1) identification of the tokens and (2) marking of three acoustic events: the beginning of  $C_1$  ( $C_1$ beg), the release burst for  $C_1$  ( $C_1$ rel), and the beginning of  $C_2$  ( $C_2$ beg), the same as the end of the vowel. The labeling process included the automatic extraction of vowel duration and SPL (measured from the midvowel RMS amplitude, relative to a calibration signal).

Panel (C) of Fig. 2 shows the x-y trajectory of a transducer coil on the tongue body for the utterance. Data were extracted from the C<sub>1</sub>–V opening (between 3 and 4 on the trajectory) and the V–C<sub>2</sub> closing (between 4 and 5) movements in each of the CVC words (the carrier word *say* was not analyzed). As exemplified in the figure, the movement paths were slightly curved; however, for the current purposes, it is assumed that to a first approximation, such movements can be analyzed according to the model of single-axis movements discussed above. In order to adapt the analysis framework described in the background to slightly curved, two-dimensional (2D) movements, it is assumed that *peak speed* and *distance along the path of the 2D* movements correspond respectively to *peak velocity* and *distance of single-axis movements*.

Movement speed was computed according to the formula

Speed = 
$$\sqrt{v_x^2 + v_y^2}$$
,

where  $v_x = dx/dt$  and  $v_y = dy/dt$ . The magnitude of the acceleration signal was computed according to

Acceleration magnitude =  $\sqrt{(dv_x/dt)^2 + (dv_y/dt)^2}$ .

Data were extracted algorithmically from movements of the tongue body transducer for the word "gog," the tongue front transducer for "dod" and the lower lip transducer for "bob." Movement events were identified algorithmically in the speed versus time traces, as exemplified in panel (B) of Fig. 2 for the tongue-body transducer. The vertical arrows indicate the times of the labeled acoustic events [shown in part (A)]; the asterisks show the algorithmically identified velocity peaks for the movements; and the numbered circles along the bottom axis show the algorithmically identified times of tongue movement beginning (3 and 4) and end (2, 4, and 5). [Events 2–5 are also indicated on the x-y trajectory in Panel (C).] As explained below, three movements are marked in the figure with symbols; however, only the second and third movements are analyzed and discussed.

1630 J. Acoust. Soc. Am., Vol. 112, No. 4, October 2002

Downloaded 22 Feb 2011 to 129.10.104.144. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp



FIG. 2. (A) The acoustic signal for part of the utterance "say gog..." in the normal speaking condition. (B) Speed versus time for a transducer on the tongue body. The vertical arrows indicate the times of the acoustic events; the asterisks show algorithmically identified velocity peaks for the opening and closing movements, and the numbered circles along the bottom axis show the algorithmically identified times of movement beginning and end. (C) The x-y trajectory for the tongue body transducer.

To identify each movement, its speed peak was found in relation to a nearby acoustic event; then movement beginning and end points were identified as the minima immediately preceding and following the peak. By definition, this approach yielded one speed peak per movement. In some cases, the end of one movement was the same event as the beginning of the next. This is illustrated in panel (B) by the circle numbered 4, which is at the end of the opening movement and at the beginning of the closing movement. In other cases the two events were different: the end of the closing

movement for  $C_1$  (shown by circle 2) occurs prior to the beginning of the opening movement for the vowel (circle 3). The time interval between events 2 and 3 is called an "intermovement interval." It contains small, low-speed movements of the transducer that occur during consonant closure. Although an example is not shown in the figure, intermovement intervals also occurred at maximum vowel opening, when the articulator paused briefly between the opening and closing movements. Both types of intermovement intervals

#### J. Acoust. Soc. Am., Vol. 112, No. 4, October 2002

#### Perkell et al.: Economy of effort in speech 1631

(during  $C_1$  closure and during maximum V opening) occurred more often at low speaking rates. Since speed includes x and y components of velocity, it almost never reaches a zero value (cf. Mooshammer, Hoole, and Kühnert, 1995).

From examining a number of x-y trajectories marked as in panel (C), Fig. 2, it was inferred that the end of the closing movement for the preceding consonant,  $C_1$  (event 2), corresponded approximately to the time that the tongue body or blade collided with the hard palate (or the two lips collided with one another). The beginning of the opening movement for the vowel (event 3) corresponded approximately to when the articulators were breaking contact at consonant release.

As the algorithm was being run, it displayed each speed trace with decisions marked on the computer screen [similar to panel (B) in Fig. 2, but with 16 tokens to a screen]. Visual inspection revealed ubiquitous nonzero speed values at movement beginning and end points and frequent intermovement intervals, as mentioned above. There were also some more variable traces for which the algorithm failed; therefore, the experimenter observed every decision and noted the tokens in which the extraction was not successful. Those tokens were later eliminated from further analysis. [The original data extraction also included the  $C_1$  closing movement, between circles 1 and 2 in Fig. 2(B). Because of context-related variability in this movement, it is not included in the current report. However, a failure to correctly extract data from any of the three movements was cause for rejection of a token.] Usually, 13 to 15 of the 15 tokens in each condition were analyzed successfully; the minimum was nine.<sup>3</sup>

The following parameters were extracted and calculated for each opening and closing movement:

- (1) movement duration,
- (2) distance along the x-y path,
- (3) peak speed (maximum speed reached during a movement)—a relative measure of effort,
- (4) peak speed/distance—a relative index of the "stiffness" of underlying muscle contraction,
- (5) c = peak speed/average speed, where average speed = distance/duration,
- (6) number of peaks in the acceleration magnitude signal—an index of the lack of movement smoothness,
- (7) symmetry of the speed profile, measured by the percentage of the movement duration spent in acceleration (where 50% represents true symmetry),
- (8) movement curvature ratio (distance along the actual trajectory/straight-line distance between the movement end points).

To investigate the main hypothesis, clear-condition speech is produced with greater effort than normal-condition speech, the data were analyzed for each subject individually in the following ways.

(1) Six three-way repeated-measures ANOVAs were performed for each subject. In each ANOVA, the main effects tested were for speech condition (clear versus normal), movement (opening versus closing), and articulator (tongue back, tongue front, and lower lip) and their interactions. The six dependent variables were acoustic vowel duration, SPL, and the first four movement parameters listed above.

In order to compare the strengths of the main effects in the ANOVAs, it was necessary to correct the values of *F* for their variable degrees of freedom. In general, the numerical value of a test of significance reflects the product of the size of the effect and the size of the study. Specifically, we used the measure eta-square (Young, 1993):  $F = (\eta^2/1 - \eta^2)^*$  (*df* error/*df* means)

(2) The mean values are compared in bar plots, with significant differences indicated by showing the values of  $\eta^2 \times 100$ , which indicates the percentage of variance accounted for in the comparison.

(3) Data from all of the conditions are examined graphically in a peak speed, distance, duration "performance space," in which bounding parameter values are determined by the above-mentioned second-order model (Nelson, 1983; Nelson *et al.*, 1984).

To explore the validity of using the undamped linear spring model, values of movement parameters 5-8 (above), as well as occurrences of intermovement intervals are considered in relation to the model's underlying assumptions.

#### **IV. RESULTS**

The main focus of this study is the difference in kinematic parameters between normal and clear speech. These differences are analyzed most extensively and are then compared with kinematic parameters from other speech conditions in a more limited analysis.

## A. Differences between the clear and normal speech conditions, opening and closing movements and articulator

The top half of Fig. 3 is a plot of mean values of vowel acoustic duration (in seconds) for the seven subjects, comparing the clear condition (unfilled bars) with the normal condition (shaded bars). The error bars show one standard error about the mean. The data for each bar are averaged across all repetitions of the tokens in the condition for the three test words, bob, dod, and gog. For each significant main effect in the ANOVA ( $p \le 0.05$ ), the percentage of variance accounted for by the effect is shown above the pair of bars. The range of mean duration values across subjects and conditions is about 0.11 s to 0.22 s. The figure shows that for subjects 1, 2, 3, 6, and 7, the clear condition.

The bottom half of Fig. 3 shows mean values of SPL (in dB), plotted in the same way as in the top half of the figure. The range of means across subjects and conditions is about 68 to 77 dB. The figure shows that subject 5 spoke with more volume in the clear condition (in spite of the instruction to avoid speaking louder). Subjects 1 and 3 actually spoke more softly in the clear condition (perhaps because of the instruction to avoid speaking louder).

Mean peak speed values ranged from about 14 to 35 cm/s. The top panel of Fig. 4 shows mean peak speed (in cm/s) for the clear versus normal condition, averaged across movement and articulator and plotted as in Fig. 3. It shows

Downloaded 22 Feb 2011 to 129.10.104.144. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp



FIG. 3. Top half: Mean values of vowel duration (in seconds) for the seven subjects, comparing the clear condition (unfilled bars) with the normal condition (shaded bars). The error bars show one standard error about the mean. The data for each bar are averaged across movement and articulator. For each significant main effect in the ANOVA ( $p \le 0.05$ ), the percentage of variance accounted for by the effect is shown above the pair of bars. Bottom half: Mean values of SPL (in dB), plotted in the same way as in the top half.

that subjects 3, 4, and 6 used faster movements in the clear than in the normal condition. The middle panel shows mean peak speed values for opening (shaded bars) vs closing (clear bars) movements, averaged across condition and articulator. It shows that subjects 1, 2, 3, 6, and 7 had faster closing than opening movements and subjects 4 and 5 had faster opening than closing movements. The bottom panel compares mean speed values for the lower lip (dark bars), tongue front (light bars), and tongue back (clear bars) averaged across condition and movement type. It shows that tongue back movements were fastest for subjects 1, 4, and 7; tongue front movements were fastest for subjects 5 and 6 and lip movements were fastest for subjects 2 and 3.

Figure 5 shows mean values of movement duration (s), plotted in the same way as in Fig. 3. Movement duration means ranged from about 0.07 s to 0.17 s. The top panel shows that subjects 2, 3, 4, 6, and 7 produced the clear condition with longer duration movements than the normal condition and subject 5 produced the clear condition. The middle panel shows that opening movements were longer than closing movements for all the subjects, and the bottom panel shows intersubject differences in the ordering of movement duration by articulator.

Figure 6 shows mean values of movement distance (cm), plotted as in Fig. 4. Mean values of distance ranged from about 1 to 2 cm. The top panel shows that subjects 3, 4, and 6 produced the clear condition with larger movements than the normal condition. The middle panel shows that all subjects used larger opening than closing movements, and the



FIG. 4. Top panel: Mean peak speed (in cm/s) for the clear versus normal condition, averaged across movement and articulator and plotted as in Fig. 3. Middle panel: mean peak speed values for opening (shaded bars) versus closing (clear bars) movements, averaged across condition and articulator. Bottom panel: Mean peak speed values for the lower lip (dark bars), tongue front (light bars), and tongue back (clear bars) averaged across condition and movement type.

bottom panel shows subject-specific differences in the ordering of distance by articulator. Although each C1VC2 is symmetric, the movement paths for  $C_1V$  and  $VC_2$  are different and do not overlap [e.g., Fig. 2(C)]. This is presumably due to the anatomical arrangement of the different muscle groups that are used for opening and closing movements, as well as some muscle interaction (co-contraction). For example, during the production of the CVC, "gog," the anterior genioglossus and hyoglossus muscles depress the tongue body for the opening movement and the styloglossus, posterior genioglossus and mylohyoid muscles raise the tongue body for the closing movement toward the hard palate (Maeda and Honda, 1994). Based on the anatomy and modeling work (Perkell, 1996), the directions of the resultant force vectors for the tongue lowering and tongue raising muscle groups cannot be equal and opposite to one another over the courses of the lowering and raising movements. Velar consonants, as illustrated in Fig. 2(C), are almost always produced with some sliding contact in which the tongue body moves in the anterior direction (Mooshammer et al., 1995). A comparison of the bottom panels in Figs. 4 (peak speed) and 6 shows that the ordering of peak speed by articulator parallels that of



FIG. 5. Top panel: Mean values of movement duration (s), plotted as in Fig. 4. Middle panel: Mean durations for opening (shaded bars) versus closing (clear bars) movements, averaged across condition and articulator. Bottom panel: Mean durations for the lower lip (dark bars), tongue front (light bars), and tongue back (clear bars) averaged across condition and movement type.

movement distance for the different subjects, reflecting the commonly found linear relationship between velocity and distance.

Figure 7 shows mean values of peak speed/distance  $(s^{-1})$ , a relative indicator of muscle stiffness, plotted as in Fig. 4. The top panel indicates that subjects 2, 6, and 7 used less muscle stiffness in the clear than the normal condition although the effect was strong only for subject 6. The middle panel shows that muscle stiffness was greater for closing than opening movements for all the subjects, and bottom panel shows intersubject differences in the ordering of stiffness by articulator.

Table I summarizes the observations made from Figs. 3–7 about the differences between the clear and normal conditions. It shows the percent of variance accounted for by significant main effects in the ANOVAs, in which the clear condition had greater mean parameter values than the normal condition. The rows correspond to: peak speed, movement duration, movement distance, peak speed/distance, vowel duration and SPL. The columns correspond to subjects. Percentage values of 80 or higher are shown in boldface. For cases in which the mean parameter value was greater in the normal than in the clear condition, the percent value is enclosed in braces; all but one of these main effects were rela-



FIG. 6. Top panel: Mean values of movement distance (cm), plotted as in Fig. 4. Middle panel: Mean distances for opening (shaded bars) versus closing (clear bars) movements, averaged across condition and articulator. Bottom panel: Mean distances for the lower lip (dark bars), tongue front (light bars), and tongue back (clear bars) averaged across condition and movement type.

tively weak. The table shows that for the clear condition, Subjects 3, 4, and 6 increased peak speed, movement duration and movement distance (also see Adams, 1990). It is likely that the co-occurrence of changes in these three parameters reflects the commonly found relations among pairs of these parameters (see Background). Subject 6 had the largest number of significant parameter changes, including the only strongly significant change in peak speed/distance, a decrease in the clear condition. Subjects 1, 2, and 7 mainly employed longer vowel duration for the clear condition and subjects 2 and 7 also lowered peak speed/distance slightly. Subject 5 increased SPL. Thus, there were substantial differences among the speakers in the way they produced the test utterances in the clear condition compared to the normal condition.

# B. Examination of data from additional conditions in a peak speed, distance, duration performance space

In order to gain further insight into whether the above observations reflect increased effort in the clear condition for subjects 3, 4, and 6, and to compare the normal and clear



FIG. 7. Top panel: Mean values of peak speed/distance, plotted as in Fig. 4. Middle panel: Mean values of peak speed/distance for opening (shaded bars) versus closing (clear bars) movements, averaged across condition and articulator. Bottom panel: Mean values of peak speed/distance for the lower lip (dark bars), tongue front (light bars), and tongue back (clear bars) averaged across condition and movement type.

conditions with the other speech conditions, the data are examined in a peak speed, distance, duration "performance space."

Figure 8 plots peak speed, distance and duration for tongue blade opening movements for subject 6. Each symbol in the plot represents a single movement; different symbols identify data from the speech conditions: N for normal, C for clear, S for slow, F for fast, R for rapid+clear, and I for informal. (Each data value lies at the lower left corner of the symbol.) The concave surface represents a limiting bound for movements of the one-dimensional, frictionless dynamical system with an acceleration limit of 1.5 g;<sup>4</sup> optimized to

S6: dod: Opening Movement, U=1.5,b=0.1



FIG. 8. Values of peak speed, distance and duration for tongue blade opening movements for subject 6. Each symbol represents a single movement; different symbols identify data from the speech conditions: N for normal, Cfor clear, S for slow, F for fast, R for rapid+clear, and I for informal. The concave surface represents a minimum "effort" (peak speed) bound for a one-dimensional, frictionless dynamical model with an acceleration limit of 1.5 g.

minimize the impulse cost (Nelson, 1983). Parameter values that define the surface are calculated according to the equation that describes minimum-impulse movements:

$$V_m = (TU/2) - \sqrt{(TU/2)^2 - DU}$$

where T and D are the movement duration and distance respectively, and U is the maximum acceleration limit [Eq. (10), p. 138; Nelson, 1983]. As U increases, the height of the surface decreases. A value of 1.5 g for U is adequate to include speech movements with the highest acceleration values measured in the current study (see also Nelson *et al.*, 1984). All actual data points must lie above the bounding surface because they cannot be frictionless and generally are not of the minimum-impulse form.

For slow-condition movements, there is little change of distance with a change in duration (time); however, for faster movements in the other conditions, the data are distributed in a roughly linear fashion above a "knee" in the bounding surface, beyond which the effort gradient begins to increase sharply. According to Nelson (1983, p. 142), if "movements can be characterized as having an economy of effort as well as time, they should be located above the knee region of this surface, where a reasonable trade-off between effort and time is possible."

TABLE I. Percent of variance accounted for by significant effects in ANOVAs ( $p \le 0.05$ ).

Effect (boldface: %≥80)	Subject							
Clear>Normal {opposite}		1	2	3	4	5	6	7
Peak speed	(Fig. 4)			83	98		93	
Movement duration	(Fig. 5)		75	77	64	{40}	94	71
Movement distance	(Fig. 6)			84	97		95	
Peak speed/distance	(Fig. 7)		{61}				<b>{95}</b>	{44}
Vowel duration	(Fig. 3)	90	89	71			96	51
SPL	(Fig. 3)	{44}		{66}		86		



FIG. 9. Distance versus duration plots of tongue-front (blade) opening movements for the seven subjects in "top-down" views of the 3D performance space shown in Fig. 8. In each plot the solid curve on the left represents a theoretical minimum-time (maximum acceleration) limit of 1.5 g. The straight lines radiating from the minimum time limit show "iso-effort" (iso-peak speed) contours that represent the height of the bounding surface in Fig. 8.

Figure 9 shows distance vs duration plots of tonguefront (blade) opening movements for the seven subjects. These are "top-down" views of the 3D minimum-impulsecost performance space that is exemplified in Fig. 8. In each plot the solid curve on the left represents a theoretical minimum-time (maximum acceleration) limit of 1.5 g. It corresponds to a top-down view of the left edge of the 3D surface where the surface becomes virtually vertical. The minimum-time solution is defined by

$$T_{\min} = 2\sqrt{D/U}.$$

The straight lines radiating from the minimum time limit show "iso-effort" (iso-peak speed) contours that represent the intersection of the bounding surface shown in Fig. 8 with a horizontal plane at different heights (values of peak speed). Corresponding to the increasing height of the surface in Fig. 8, these iso-effort contours show a gradient that increases sharply as the minimum-time bound is approached, reflecting

#### 1636 J. Acoust. Soc. Am., Vol. 112, No. 4, October 2002

#### Perkell et al.: Economy of effort in speech

increasingly larger levels of effort. Actual movement points have to lie to the right of the bound. In general, the slow condition data (S) lie in separate groups to the right of the other data, indicating the longest durations. Compared to the other data, the slow condition data also appear to show less variation in distance with variation in duration. Post-hoc planned contrasts showed significant differences between the slow condition data and all of the other conditions grouped together for almost every parameter and subject. In addition, the slow movements had approximately twice as many acceleration peaks as the other movements, i.e., they were less smooth than the other movements (also see Weineke *et al.*, 1987).

Movements in the other conditions have shorter durations; their data tend to be distributed along a bounding isopeak-speed (effort) contour, the level of which differs among the subjects. Along the contour, movement duration tends to vary linearly with movement distance. The level of the limiting peak speed contour shown in Fig. 9 ranges from a low of 60–80 mm/s for subjects 3, 1, and 2 to a high of about 180 mm/s for subjects 5 and 4, with subjects 6 and 7 falling in between. There is a great deal of overlap among the data from the different conditions, except for slow.

Such plots were similar in form for the other articulators and for closing movements (not shown), although the data values differed somewhat (as demonstrated above in Sec. IV A).

Figure 10 shows the same kind of plots for tongue front opening movements as in Fig. 9, but the data for each condition are represented by a single symbol at the centroid of the distribution for the condition. A convex hull shows the outer limits of all of the underlying individual data points. The overall amount of variation in the data differs among the subjects; subject 6's data have about twice the range of distance and duration as subject 3's data. For subjects 3, 4, and 6, the centroid of the clear condition data (C) lies at a somewhat higher iso-effort level than the normal condition data (N). This result is consistent with the observations in Sec. IV A of higher values of peak speed, duration, distance and parameter c for these subjects. For all except subject 3, the centroid of the slow condition data lies at a lower iso-effort level than the other data. For all except subject 5, the centroids for the fast (F) and/or rapid+clear (R) data lie furthest to the lower left of the plots; however there are only two examples in which iso-effort levels are obviously higher for the F or R centroids than for the other conditions (F for subject 1, R for subject 6). Consistent with the observations made in Sec. IV A, subject 5's movement data do not differ among conditions aside from the slow condition.

Table II summarizes observations from plots like Fig. 10 for all three articulators and opening and closing movements. A + sign indicates that the clear condition centroid was at least one iso-effort level (20 mm/s) higher than the normal condition centroid. A + in parentheses indicates that the clear condition was about one-half an iso-effort level (10 mm/s) higher than the normal condition centroid. The table shows that the observations made from Fig. 10 for tongue front opening movements are largely representative of the data for the other articulators and closing movements, with

subject 4 showing the most consistent effects across movement type and articulator.

## C. Movement characteristics that differ from assumptions of the model

As observed above for subjects 3, 4, and 6, the pattern of differences for the acoustic and kinematic measures is generally consistent with the hypothesis that movements in the clear condition (with respect to the normal condition) are characterized by higher peak speed, indicating greater effort. The movements for these subjects also have longer movement durations and greater movement distances. However, since the measures are made from 2D movements of points on very complicated 3D structures that are interacting mechanically with one another, it is necessary to be cautious about the use of a measure of effort that is based on a relatively simple model. As is shown below, further analyses of the movements indicate that a number of them fail to meet criteria that are assumed by the model. Specifically (a) many movements are not smooth (with simple acceleration and deceleration phases), (b) their velocity profiles are not symmetrical (with equal amounts of time spent in acceleration and deceleration), and (c) as discussed above, their paths are not straight. These factors are examined in the following analyses, with the exclusion of data from the "informal" condition (which was not produced by one subject) or the "slow" condition (in which movements had about twice as many acceleration peaks as in any of the other conditions).

To indicate movement smoothness, Fig. 11 shows the distribution of the number of peaks in the acceleration magnitude signal (representing both acceleration and deceleration phases of the movements). Data from all seven subjects for opening and closing movements are grouped together. Smoother movements have lower numbers of acceleration peaks. About 60% of the movements have two acceleration magnitude peaks, which is expected for smooth movements with single acceleration and deceleration phases and is assumed by the model. About 12% of the movements have only one acceleration magnitude peak. These occurred primarily in the fast and rapid+clear conditions, most likely because the acceleration peak responsible for the acceleration phase occurred just before the algorithmically identified movement beginning. The remaining 28% of the movements had more than two acceleration magnitude peaks, which were related to small inflections in the movement and velocity signals. Similar plots were examined by subject, condition, and movement type. The distribution shapes differed only slightly among the subjects; the clear condition had a somewhat larger proportion of movements with more than two acceleration magnitude peaks than the normal, fast or rapid+clear conditions; and opening movements showed a somewhat larger proportion of movements with more than two peaks than closing movements. In sum, a significant number of movements (even excluding slow movements) were not entirely smooth, contrary to one of the assumptions of the undamped linear spring model.

To examine the symmetry of the movement speed profiles, Fig. 12 shows the distribution of values of a measure of symmetry of the speed trace, the amount of time spent in

Downloaded 22 Feb 2011 to 129.10.104.144. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp



FIG. 10. Distance versus duration plots of tongue-front (blade) opening movements for the seven subjects, as in Fig. 9, with the data for each condition represented by a single symbol at the centroid of the distribution for the condition. A convex hull shows the outer limits of all of the underlying individual data points.

acceleration as a percentage of the movement duration, for both opening and closing movements. The distribution is quite broad, with substantial numbers of movements occurring in bins that span the range from 38% to 70%. Although movements in the fast condition were more symmetrical than in other conditions, this result indicates that most of the speed profiles were far from symmetrical. The occurrence of an inter-movement interval preceding a vowel opening movement indicates that the identified beginning of the opening movement corresponded to the breaking of articulatory contact, rather than the actual onset of the underlying opening gesture. Among the 840 opening movements in the normal, clear, fast and rapid+clear conditions, 171 were preceded by intermovement intervals. In other

#### 1638 J. Acoust. Soc. Am., Vol. 112, No. 4, October 2002

#### Perkell et al.: Economy of effort in speech

TABLE II. Cases in which iso-effort for clear is greater than for normal.

		Subject						
Articulator	Movement	1	2	3	4	5	6	7
Tongue Back Tongue Front Lower lip	closing opening closing opening closing opening	+	(+) (+)	(+) + (+) +	(+) + + + + +	(+)	(+) + + +	

words, for 20% of the tokens (excluding slow), the measured characteristics of the opening movement must have been influenced by articulatory contact during the preceding consonant. It is very likely that the actual beginning of the opening movement occurs during the closure and is obscured by it. Consistent with these observations, Löfqvist and Gracco (1997) speculate that the spatial target for movements involving articulatory contact is at a virtual location beyond the place of contact. (Also see Westbury and Hashi, 1997.)

Movement curvature was quantified with the "curvature ratio," the ratio of the integrated distance along the movement path to the straight-line distance between the two movement end points. While most of the movements followed relatively straight paths, with curvature ratios less than 1.1, 17% of the movements (281 of 1680) had ratios greater than 1.1, ranging up to about 1.4.

Consistent with the preceding observations, across the normal and clear data sets for all seven subjects, mean values of parameter c ranged from about 1.6 to 2.2. The occurrence of values approaching and exceeding 2.0 is a further indication that at least some of the movements do not meet the model's assumptions. Other studies have also reported values of c that exceed 2.0 (cf. Adams *et al.*, 1993; Shaiman *et al.*, 1997), possibly for the same reasons.

These results indicate that a significant proportion of the movements do not strictly meet the criteria required for making inferences about underlying control and articulatory effort from a kinematic analysis based on the linear, frictionless mass-spring model.



FIG. 11. The distribution of the number of peaks in the acceleration magnitude signal for the movements including all the speech conditions except "informal" and "slow."



FIG. 12. The distribution of values of a measure of symmetry of the speed trace, the amount of time spent in acceleration as a percentage of the movement duration. Opening and closing movements from all subjects in all conditions but "informal" and "slow" are included.

#### V. SUMMARY AND CONCLUSIONS

Three of the seven speakers (subjects 3, 4, and 6) responded in an expected way to the instructions. In the clear condition they used higher peak speeds, longer movement durations and greater distances. The performance space analysis indicated that these three subjects used increased effort (peak speed) in the clear condition. The data from these three *S*'s therefore, support the hypothesis that clear speech is produced with greater effort than normal speech. On the other hand, subjects 1, 2, and 7 mainly used vowel duration increases for the clear condition, and subject 5 mainly used an SPL increase. Changes in peak speed/ distance (an indirect index of relative stiffness of underlying muscle contraction) were less consistent and the three significant ones were in the negative direction.

A consistent outcome was that slow-condition movements are very different from those elicited in the other conditions. According to the performance space analysis, they are produced with less effort because their longer durations are not accompanied by proportionally larger distances. They are also less smooth than movements in the other conditions. These results are consistent with others (cf. Adams, 1990) and indicate that such slow movements may be un-natural for normal speakers (Weineke *et al.*, 1987). The lack of smoothness of slow movements may reflect some physiological lower limit, although such an idea would require further investigation.

It has been speculated that articulator size would influence articulatory kinematics (see review in Perkell, 1997). For example, it might be expected that movements of the tongue body would be slower than those of the much smaller tongue blade. Contrary to this speculation, there was intersubject variation in which articulator showed the highest values of peak speed, movement duration, distance, peak speed/ distance and parameter c. It is possible that these subject by parameter interactions for articulator are conditioned primarily by subject differences in distances moved by the tongue body, tongue blade and lips. Such differences may be due to habit, as appears to be the case for readily observed crossspeaker differences in the range of vertical mandible move-

ments during speech. They may also be due to anatomy. For example, the amount of vertical tongue displacement differentiating nonlow vowels from one another can depend on the ratio of palatal height to width (Perkell, 1979). As mentioned in the introduction, Kuehn and Moll (1976) found positive relationships across subjects between both articulatory velocity and movement displacement and the size of the articulators. In the current study, the observed similar patterning of peak speed and movement distance across subject and articulator would follow from an approximately linear relation between speed and distance.

The finding that consonant-closing movements uniformly have shorter durations and higher values of peak speed/distance than vowel opening movements is consistent with the result of Hertrich and Ackermann (1997) and their suggestion that the closing movements may be more ballistic in nature. However, there are intersubject differences in whether peak speed is higher for opening or for closing movements. These findings need to be explored further in future studies.

Further generalizations from the data are hampered by intersubject differences and the fact that movement characteristics were somewhat different from those assumed by the underlying model. As mentioned above, the lip and especially the tongue movement data come from single points on the surfaces of anatomically complicated structures that are composed mostly of muscle and are interacting mechanically with other structures. The model was used originally to analyze movements of the jaw (Nelson, 1983), which is a solid, relatively massive articulator and is therefore the articulator that is most subject to dynamic constraints and effort costs. It also has a relatively clear principal axis of motion. The following companion paper (Perkell and Zandipour, 2002) represents an initial step in examining the influence of the mandible on the kinematics of the other articulators. Another complicating factor is the occurrence of inter-movement intervals, some of which are due to articulatory contact. We suggest that the complexity of articulatory structures and their physical interactions, as exemplified by contact, make it difficult to precisely quantify some important movement characteristics with current techniques and analyses.

Several additional factors could underlie the individual differences in the results. There could be differences in the way the subjects interpreted the instructions and they could differ in the way they produce and/or perceive the acoustic correlates of clear speech. An important issue, to be addressed in future work, is the manner and extent to which each subject produced acoustic correlates of clarity. The subjects could also differ in their motor performance limits. The following companion study explores this idea, but the results do not resolve the issue.

It appears that more adequate tests of control strategies for different speaking styles will depend upon further developments in the physiological-biomechanical modeling of more complicated movements with more realistic boundary conditions. Physiological-biomechanical simulations could help to quantify the effects of articulatory contact and relative amounts of energy flow in the musculature. Such work should lead to more accurate measures of physical articulatory effort, which would make it possible to determine the limitations of using simple dynamical models. When more accurate effort measures are made, it may be found that physical cost is too simple a concept to account for changes in speaking style, and it may be necessary consider ideas such as motor programming complexity and psychological factors as well. Future work should also include thorough investigations of the sources of individual differences and studies of the clarity-related acoustic characteristics of the utterances and their intelligibility in the different speaking conditions. Ultimately, as suggested by Hertrich and Ackermann (1997), it may be found that the most invariant aspects of different speaking styles are in their acoustic or perceptual results.

#### ACKNOWLEDGMENTS

This work was supported by research Grant No. 5 R01 DC01925 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We are very grateful to Winston Nelson for his valuable comments, to Hermann Ackermann, Ingo Hertrich, Anders Löfqvist, David Ostry, and Gary Weismer for their helpful reviews, and to Gareth Coote, Brandy Karl, and especially Erik Strand and Jennell Vick for their help with data extraction.

<sup>1</sup>This measure of actuator "stiffness" corresponds to the square root of the stiffness factor, *k*, in Nelson, 1983 [Eqs. (12) and (13), p. 139 and Appendix E]. It also relates to the natural frequency of simple harmonic motion,  $\omega_n = 2 \pi / \tau$ , where  $\tau$  is the duration of a complete cycle,  $\tau = 2$  T.

<sup>2</sup>Although the data are examined for the effects of articulator and movement type, these effects are not the primary focus of this study. They will be dealt with in more depth in subsequent analyses of additional materials from the corpus.

<sup>3</sup>The experiment originally included an eighth subject. His movement speeds were higher than for any of the other subjects; however, his displacement and velocity profiles were very irregular, varying so much from one token to the next that they could not be analyzed with the algorithm.  ${}^{4}1 \text{ g} = 9.806 \text{ m/s}^2$ , the acceleration due to gravity.

- Adams, S. G. (1990). "Rate and clarity of speech: An x-ray microbeam study," Ph.D. thesis, University of Wisconsin, Madison.
- Adams, S. G., Weismer, G., and Kent, R. D. (1993). "Speaking rate and speech movement velocity profiles," J. Speech Hear. Res. 36, 41–54.
- Edwards, J., Beckman, M. E., and Fletcher, J. (1991). "The articulatory phonetics of final lengthening," J. Acoust. Soc. Am. 89, 369–382.
- Flanagan, J. R., Ostry, D. J., and Feldman, A. G. (1990). "Control of human jaw and multi-joint arm movements," in *Cerebral Control of Speech and Limb Movements*, edited by G. E. Hammond (North-Holland, Elsevier Science Publishers, Amsterdam), pp. 29–57.
- Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.
- Guenther, F. H., Hampson, M., and Johnson, D. (1998). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. 105, 611–633.
- Hertrich, I., and Ackermann, H. (1997). "Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures," J. Acoust. Soc. Am. 102, 523–536.
- Kuehn, D. P., and Moll, K. L. (1976). "A cineradiographic study of vc and cv articulatory velocities," J. Phonetics 4, 303–320.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–440.

Downloaded 22 Feb 2011 to 129.10.104.144. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

- Linville, R. N. (**1982**). "Temporal aspects of articulation: Some implications for speech motor control of stereotyped productions," Ph.D. thesis, University of Iowa, Department of Speech Pathology and Audiology.
- Löfqvist, A., and Gracco, V. (**1997**). "Lip and jaw kinematics in bilabial stop consonant production," J. Speech Lang. Hear. Res. **40**, 877–893.
- Maeda, S., and Honda, K. (**1994**). "From EMG to formant patterns of vowels: The implication of vowel spaces," Phonetica **51**, 17–29.
- Moon, S.-J. (**1991**). "An acoustic and perceptual study of undershoot in clear and citation-form speech," PERILUS XIV, University of Stockholm, Institute of Linguistics, pp. 153–156.
- Moon, S.-J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. 96, 40–55.
- Mooshammer, C., Hoole, P., and Kühnert, B. (1995). "On loops," J. Phonetics 23, 3–21.
- Munhall, K. G., Ostry, D., and Parush, A. (1985). "Characteristics of velocity profiles of speech movements," J. Exp. Psychol. 11, 457–474.
- Nelson, W. L. (1983). "Physical principles for economies of skilled movements," Biol. Cybern. 46, 135–147.
- Nelson, W. L., Perkell, J. S., and Westbury, J. R. (1984). "Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations." J. Acoust. Soc. Am. 75, 945–951.
- Ostry, D., Keller, E., and Parush, A. (1983). "Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement is speech," J. Exp. Psychol. 9, 141–147.
- Ostry, D., and Munhall, K. G. (1985). "Control rate and duration of speech movements," J. Acoust. Soc. Am. 77, 640–648.
- Ostry, D. J., Cooke, J. D., and Munhall, K. G. (1987). "Velocity curves of human arm and speech movements," Exp. Brain Res. 68, 37–46.
- Perkell, J. S. (1979). "On the nature of distinctive features: Implications of a preliminary vowel production study," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, London), pp. 365–480.
- Perkell, J. S. (1996). "Properties of the tongue help to define vowel categories: hypotheses based on physiologically oriented modeling," J. Phonetics 24, 3–22.

- Perkell, J. S. (1997). "Articulatory processes," in *The Handbook of Phonetic Sciences*, edited by W. Hardcastle and J. Laver (Blackwell, Oxford), pp. 333–370.
- Perkell, J. S., Cohen, M. H., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (1992). "Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. 92, 3078–3096.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (2000). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics 28, 233–372.
- Perkell, J. S., and Zandipour, M. (2002). "Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements," J. Acoust. Soc. Am. 112, 1642–1651.
- Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," Speech Hear. Res. 29, 434–446.
- Shaiman, S., Adams, S. G., and Kimelman, M. D. Z. (1997). "Velocity profiles of lip protrusion across changes in speaking rate," J. Speech Lang. Hear. Res. 40, 144–158.
- Sonoda, Y., and Nakakido, K. (1986). "Effect of Speaking Rate on Jaw Movements in Vowel Sequence," J. Acoust. Soc. Jpn. (E) 7, 5–12.
- Vatikiotis-Bateson, E., and Fletcher, J. (1992). "Articulatory correlates of linguistically contrastive events: Where are they?" in *Speech Perception*, *Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (IOS, Amsterdam), pp. 341–358.
- Weineke, G., Janssen, P., and Belderbos, H. (**1987**). "The influence of speaking rate on the duration of jaw movements," J. Phonetics **15**, 111–126.
- Westbury, J., and Hashi, M. (1997). "Lip-pellet positions during vowels and labial consonants," J. Phonetics 25, 405–419.
- Young, M. A. (1993). "Supplementing tests of statistical significance: Variation accounted for," J. Speech Lang. Hear. Res. 36, 644–656.