Linear and Logarithmic Speed-Accuracy Trade-offs in Speech Production

Mark L. Latash[1] and Irina L. Mikaelian[2]
([1] Department of Kinesiology, The Pennsylvania State University, University Park, PA 16802, USA; mll11@psu.edu [2] Department of Germanic and Slavic Languages and Literatures, The Pennsylvania State University, University Park, PA 16802, USA; ixm12@psu.edu)

Abstract

We used picture description tasks to explore the scaling of speech time (ST) with task difficulty manipulated by changing the number of objects and number of their characteristics. In Experiment-1 linear scaling of ST with the index of difficulty (ID) was observed. The scaling coefficients were larger in subjects who spoke non-native English as compared to native both English and non-English. Experiment-2 explored a broader range of ID and resulted in logarithmic scaling of ST with ID. Chinese speakers were much faster than speakers of Indo-European languages. We interpret the findings as reflections of a basic scaling of action time with ID common across motor tasks (as in Fitts' law) and speech tasks. Scaling of ST with ID is sensitive to both native vs. second language and to cross-linguistic comparisons.

1 Introduction

Speech and coordinated voluntary limb movements share a number of common features. In particular, the production of both coordinated limb/body movements and grammatical phrases has been viewed as controlled in a hierarchical way. Both movement tasks and speech tasks are commonly characterized by abundance of solutions: A task typically imposes insufficient constraints such that numerous seemingly equivalent solutions can be used. As a first step towards approaching speech as a time-varying biological process, this study explores applicability of the classical speed-accuracy trade-off [1, 3] to speech. To our knowledge, speed-accuracy trade-off has not been studied with respect to the whole act of production of natural speech. The classical speed-accuracy trade-off in voluntary movements reflects scaling of movement time (MT) with task difficulty quantified using an index of difficulty (ID), a log-transformed ratio of movement distance (D) to target width (W): MT = $c + d \cdot$ID; where ID = $\log_2$(D/W), c and d are constants. Fitts' law states that it takes longer to move to a small and distant target than to larger/closer targets, and this dependence is non-linear.

Is Fitts' law valid for speech production? To answer this question, speech metrics has to be introduced equivalent to target width and distance. We assume that the main goal of speech production is to express a meaning, which we cannot define exactly. In this study, we have tried to standardize the loosely defined *meaning* using picture description tasks.

The described studies may be viewed as a natural development of a series of classical studies by Sternberg and colleagues [4-6]. In those studies, the subjects were required to produce an utterance as quickly as possible after an imperative signal. The utterance could consist of words, numerals, or non-words presented to the subject in advance. So, the subjects did not need to construct a phrase. Close to quadratic relationships between the number of words and the speech time were found corresponding to a linear increase in time per word with the number of words. The latter linear relationship shifted upwards (longer times per word) with an increase in the number of syllables per word.

Our method of picture description is much less constraining: It allows the subjects to use any word combinations to create a picture description in real time. Hence, the task does not prescribe using specific words or even the total number of words and syllables. In fact, we have been interested in regularities of the relationships between movement time and features of the task that would be independent of the exact words used by the subjects thus reflecting the natural variability of speech. Besides, we did not force our subjects to utter as quickly as possible a word combination presented in advance but asked them to be "fast and accurate" – a typical instruction for Fitts'-type tasks. The experimental design led to our subjects speaking much slower (see Results) than in the cited studies. So, in contrast to the studies by Sternberg and colleagues, we asked a different question: How long does it take to describe a set of 'n' objects with each object having 'm' characteristics?

2 Experiment-1: Methods

Two experiments were performed. In Experiment-1, two groups of subjects, six native English speakers and six bilinguals, looked at pictures presented on the computer screen and were instructed to describe them "as quickly and accurately as possible". The bilingual subjects performed the test twice, in English and in their native language. The pictures could contain one to six objects, all of the same kind (sets of same objects) and all different (sets of different objects). There were four series of trials, 36 trials in each series. The series differed in the number of characteristics of the objects the subjects were instructed to describe: Type of objects, their color, size, and relative location.

3 Experiment-1: Results

Reaction time showed a modest, close to linear scaling with the number of objects (N_O). It was significantly longer for the bilingual subjects performing in English as compared to their performance in the native language and to the English speaking subjects. This difference in reaction time did not depend on N_O.

Speech time (ST) showed a close to linear scaling with N_O within each of the four series (Figure 1). The linear regression coefficient between ST and N_O showed a linear increase with the number of characteristics (N_C) of the objects across the four series for both sets of same objects and sets of different objects (Figure 2). Bilingual subjects showed longer ST when performing in English as compared to both their performance in the native language and to the performance of the native English speakers. These differences scaled with both N_O and N_C.

These results suggest that there is a speed-accuracy trade-off relating speech time to an index of transmitted information (equivalent to ID) computed as the product of the number of objects and number of their characteristics. This trade-off is similar to Fitts law although it is linear while Fitts law is logarithmic. We suspected that this experiment explored a relatively small, linear portion of a non-linear function corresponding to very simple tasks with small numbers of object characteristics.

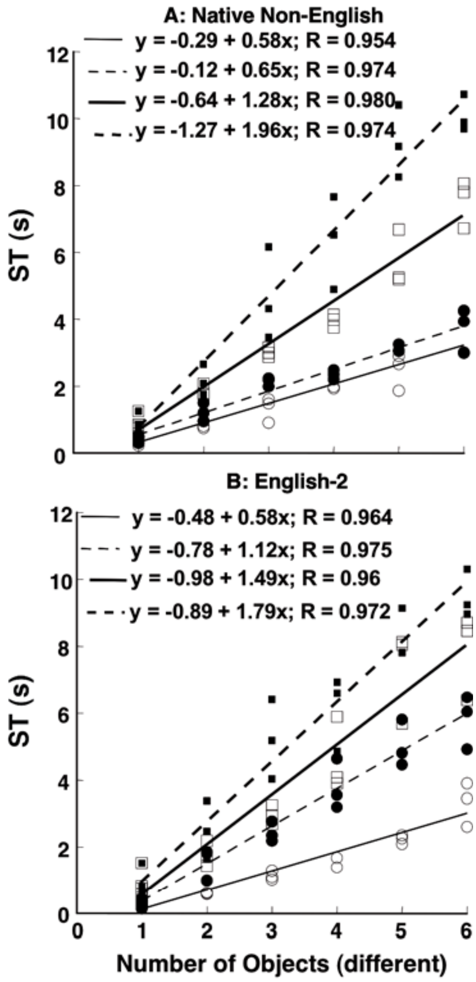

Figure 1: *Dependences of speech time (ST) on the number of different objects for the native non-English (A) and English-2 (B) with regression lines and equations for the four series.*

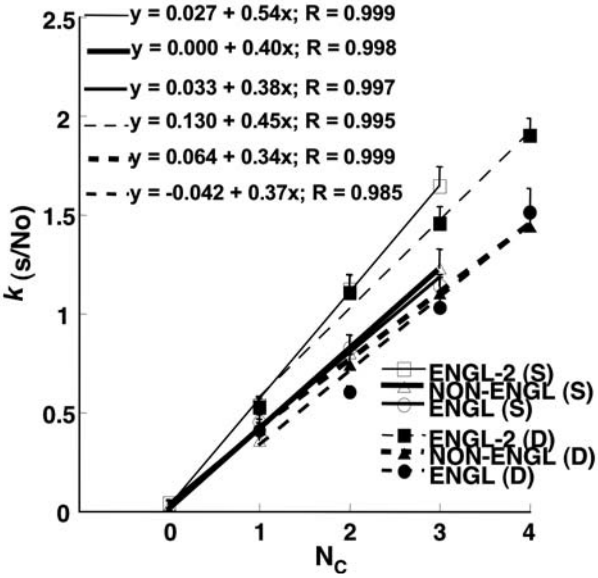

Figure 2: *Dependences of the regression coefficients (k) between speech time ST and number of objects on the number of object characteristics (N_C) for similar (S) and different (D) objects and for English-native (ENGL), non-English native (NON-ENGL), and English-second (ENGL-2) with regression lines and equations.*

4 Experiment-2: Methods

In Experiment-2, we fixed the number of objects (N_O=2). One of the objects was always a plate shown in the center of the screen, while the other one could be a stick, a fork, or a knife. Note that the stick was an elongated symmetrical object, the fork had two different ends, and the knife had two different ends and two different sides. The second object could be located above, below, to the left, to the right, and at the corners with respect to the plate. It could be oriented vertically, horizontally, or tilted slightly or by 45 degrees. Index of difficulty (ID) was computed as the product of difficulty indices assigned to the object (ranging from 1 to 4), its location (1 or 2), and its orientation (1 or 4). Twelve subjects were tested in their native language, English (3), Russian (3), and Chinese (6). Each subject performed 72 trials in a self-paced manner, presented in three blocks in a randomized fashion.

5 Experiment-2: Results

Reaction time (RT) was significantly shorter in the Chinese speaking subjects (C-group, on average, 0.97 s) as compared to the English and Russian speaking subjects (IE-group, on average, 1.13 s). The C-group did not show significant RT changes with ID. In contrast the IE-group showed a close to linear increase in RT with ln(ID) that was modest in magnitude and significant only in the averaged across subjects data (R^2=0.53; $p < 0.05$). In contrast, ST showed strong scaling with ID in all subjects: ST = a + b*ln(ID), where a and b are constants. After the data were collapsed across all trials with the same ID values, linear regression coefficients between ST and ln(ID) computed for each subject separately ranged from 0.926 to 0.996 (median value 0.989), $p < 0.01$ for each subject. The Chinese speaking subjects showed significantly faster ST values across the range of ID values. Their regression line slopes were, on average, about 50% lower than those of the other subjects who spoke Indo-European languages, 0.64 and 1.43 s/ID, without a difference in the intercept, 1.47 and 1.42 s.

6 Discussion and Conclusions

The results of the second experiment suggest that there is indeed a strong logarithmic relation between task difficulty (as reflected by ID) and speech time. On the one hand, these results corroborate the earlier assumption that Experiment-1 might have failed to reveal a non-linear speed-accuracy trade-off in speech production due to the narrow range of difficulty associated with each of the presented objects. On the other hand, Experiment-1 used up to six objects while Experiment-2 was limited to only two objects; nevertheless, nonlinear relations between ST and ID were seen only in Experiment-2. The two studies suggest different effects of two contributors to task difficulty: (1) The number of objects seems to affect ST in a linear fashion, maybe because of a particular strategy of describing the objects one-by-one that could be used in the tasks; and (2) Difficulty in describing each object affects ST in a logarithmic fashion, possibly because various object characteristics are integrated during the generation of a verbal description.

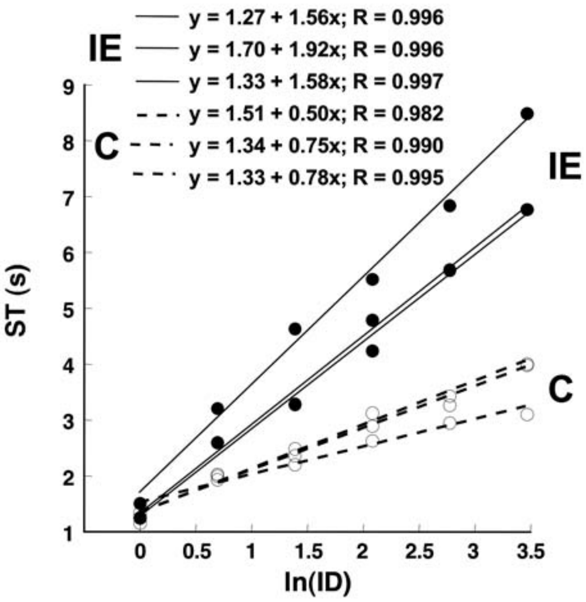

Figure 3: *Dependences of speech time (ST) on log-transformed index of difficulty (ID) for three subjects speaking Indo-European languages (IE) and three subjects speaking Chinese (C).*

It is tempting to link ST to such characteristics as the number of words or the number of syllables. In particular, in the studies of very quick speech production by the group of Sternberg (Sternberg et al. 1978, 1988), a number of potential "action units" (elements) have been considered such as words, syllables and stress groups (see also Fowler 1981). Attempts to link variations in ST in our experiments to variations in the number of words, syllables or stress groups might be an interesting development, which may be pursued in future studies. For the purposes of this study, we limit ourselves to relationships between task characteristics and ST. We would like to mention, however, that the consistency of findings over the performance in different native languages speaks against such simplistic attempts at analyzing the speed-accuracy trade-off in natural speech. Indeed, most words that described the objects and their characteristics have more syllables in Polish and Russian as compared to English, and no clear differences have been seen across the performances in these languages by the native speakers.

The two studies demonstrate that picture description tasks may be used to generate speech characteristics that show lawful relations to task features reflecting its formal meaning. Task performance is also sensitive to differences across languages (Chinese vs. Indo-European) and between native and non-native languages.

In Experiment-1, performance of the bilingual subjects in English was significantly different (slower) from their performance in the native language as well as from the performance of native English speakers. The slowness had two components to it, the bilingual subjects performing in English showed longer RT and longer ST. These findings may be viewed as intuitively predictable. Note, however, that in Experiment-1 we found different patterns of slowing for RT and ST, which are not intuitively obvious: RT showed a difference that did not depend on the number of objects while ST showed a difference that scaled with the number of objects.

Experiment-2 showed much shorter ST values for Chinese speakers, which may be related to the difference in using prosody for information transmission between the tone and stress languages [7] as well as to other factors that are beyond the scope of this paper.

In both studies, we observed weak dependences RT(ID) that were qualitatively similar to those of ST(ID). Moreover, they showed similar differences between the groups. In particular, in the first experiment, RT was longer in subjects performing in their second language as compared to their performance in the native language. The same was true for ST. In the second experiment, RT was shorter in the C group as compared to the IE group. The same was true for ST. We suggest that these similarities are not coincidental but reflect common rules that define RT(ID) and ST(ID) relations. We propose that, under the instruction to describe a picture "as fast and accurate as possible", a cognitive process of task analysis and phrase construction starts before RT and continues until the end of the utterance. If there are differences in the speed of this process between subjects, the differences start to accumulate from the initiation of the process such that stronger differences are accumulated over ST as compared to RT.

References

[1] P.M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47: 381-391, 1954.

[2] C.A. Fowler. A relationship between coarticulation and compensatory shortening. *Phonetica* 38: 35-50, 1981.

[3] R. Plamondon, Alimi A.M. Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences* 20: 1-31, 1997.

[4] S. Sternberg, Knoll R.L., Monsell S., Wright C.E. Motor programs and hierarchical organization in the control of rapid speech. *Phonetica* 45: 175-197, 1988.

[5] S. Sternberg, Monsell S., Knoll R.L., Wright C.E. The latency and duration of rapid movement sequences: Comparisons of speech and typing. In: G.E. Stelmach (Ed.) *Information Processing in Motor Control and Learning*, p. 117-152, Academic Press: New York, 1978.

[6] S. Sternberg, Wright C.E., Knoll R.L., Monsell S. Motor programs in rapid speech: Additional evidence. In: R.A. Cole (Ed.) *Perception and Production of Fluent Speech*, p. 507-534, Lawrence Erlbaum Associates: Hillsdale, NJ, 1980.

[7] Q. Zhang, Yang Y. Electrophysiological estimates of the time course of semantic and metric encoding in Chinese speech production. *Neuroscience* 147: 986-996, 2007.