Speech tempo

Introduction

This article presents an overview of the literature on speech tempo and the topics dealt with. Tempo is not one of the more frequently explored areas of speech research, and any possible consequences of tempo variation for other phonetic phenomena have all too often been taken for granted. Unfortunately, this field is complicated by pitfalls of definition and hazards of numerical treatment.

A glance at some of the elementary textbooks yields the following. D. Jones (An Outline of English Phonetics, Cambridge 1967, §43) put the average conversational rate of native English speakers at 300 syllables/minute and recommended this as a convenient target for foreign learners. A. Gimson (An Introduction to the Pronunciation of English, London 1962, p. 25) made several brief observations in one short paragraph in a discussion of quantity and duration: (i) "the absolute duration of sounds or syllables will, of course, depend on the speed of utterance", (ii) "an average rate of delivery might contain anything from 6 to 20 sounds per second", and (iii) "lower and higher speeds are frequently used without loss of intelligibility". These simple statements alone disclose a number of fundamental problems, such as how is speaking rate to be measured, what is the range of variation of speaking rate, how far do durations of other physical or physiological phenomena depend on speaking rate? Or, conversely, how far is speaking rate a disturbing factor in investigations of physical properties of speech, and how are speaking rate and intelligibility related? D. Abercrombie (Elements of General Phonetics, Edinburgh 1967, p. 46) had the following to say about speech rate: (i) tempo (speed of speaking) is best measured by rate of syllable succession, (ii) tempo is variable, and (iii) "everyone who starts learning a foreign language has the impression that its speakers have an exceptionally rapid tempo". This indicates a third area of interest, what is perceived tempo and what factors does it depend on? Yet another area is revealed by R. Heffner (General Phonetics, Madison 1960, §8.1), who discussed in particular the maximum rate of articulation available to man.

Tempo is not one single, unambiguous concept, but has in fact been used to denote the rate of several different processes in speech production. Note also that the word speed has a special meaning when applied to speech tempo, referring to frequency of repetition measured as the number of units (words, morphemes, syllables, phonemes, gestures etc.) in a period of time. It does not refer to velocity. Frequency and velocity must be distinguished in the lab., precisely because they are fuzzy in everyday speech. For example, there is the issue of whether articulator velocities vary with speech rate variation.

It is customary to distinguish between gross rates based on the total time of speaking, (i.e. including pauses) and net rates based on the periods of actual utterance (i.e. excluding pauses). These two fundamental measures have received various names. F. Goldman-Eisler (in a series of articles published in Language and Speech from 1958 to 1961) referred to talking rate as a measure of the entire cognitive and articularatory activity involved in the production of an utterance and articulation rate for the amount of speech produced in the time actually taken to articulate it. J. Kelly and M. Steer (Revised concept of rate, Journal of Speech and Hearing Disorders 14, 222-226, 1949) had over-all rate (comprising "intentional pauses and unintentional pauses as well as meaningful words spoken in the elapsed time") and phrase by phrase sentence rate excluding pauses. A decision on meaningfull words highlights the difficulty of what to do with the hesitant murmurs and uhms and ahs of filled pauses in spontaneous everyday speach. T.
Clevenger and M. Clarke (Coincidental variation as a source of confusion in the experimental study of rate, Language and Speech 6, 144-150, 1963) defined three measures based on total time, phrase time and pause time. In addition to gross rate (total time of phrase time and pause time together) and intra-phrase rate (phrase time only), they suggested that percentage of pause (pause time as a proportion of total time) was a useful measure in the study of rate.

The difference between these measures can be exemplified with same data from one of my informants, a speaker of West Greenlandic Inuit, who read a page from a novel. His style was fairly casual. He uttered 333 syllables in 31 phrases in a total time of 74 seconds, a gross (talking) rate of 4.5 syllables/second. This indicates how fast he was communicating (i.e. composing and transmitting his message) but tells nothing of how fast he was articulating speech (which might relate to the load on the articulators from coarticulation and reduction). He articulated the 333 syllables in 50 seconds, an average net articulation rate (or intra-phrase rate) of 6.7 syllables/second. From phrase to phrase he varied from 4.7 to 8.7 syllables/second.

The expressions tempo and speech rate are used in different meanings by different authors. V. Kozhevnikov and L. Chistovich (Speech, Articulation and Perception, translated by Joint Publications Research Service, Washington, 1965) first defined tempo as the rate at which an articulatory program is accomplished (p.77) and subsequently as the rate of succession of individual commands as distinct from the rate of individual movements (p. 90). They accused R. Stetson, C. Hudgins and E. Moses (Contribution à l'étude de la vitesse du débit et de la lecture dans le néerlandais, Archives Néerlandaises de Phonétique Expérimentale 14, 103-116, 1940) of confusing the issue by failing to observe this distinction. These three authors were studying the ranges of temporally constrained gestures, their aim being to discover factors influencing palatographic records as an aid to understanding them. This is a very different area from Kozhevnikov and Chistovitch's focus on the programming of speech articulation. In yet another area, H. Karlgren (Speech rate and information theory, Proceedings of the 4th International Congress of Phonetic Sciences, 1962) applied information theory to speech in order to study the rate of transmission of the content of the underlying message.

It is tempting to regard speech rate as a source of interference that distorts the flow of speech. An inability of speech articulators to function adequately when temporally constrained might be reflected in the apparent duration dependency of many speech gestures or acoustic features as reported by, for example, Stetson et al. (1940), B. Lindblom (Spectrographic study of vowel reduction, Journal of the Acoustical Society of America 35, 1773-1781, 1963), or T. Gay (Effect of speaking rate on diphthong formant movements, Journal of the Acoustical Society of America 44, 1570-1573, 1968). But there is an alternative view. Kozhevnikov and Chistovitch explicitly excluded such an interpretation of speech rate, and Karlgren postulated that the reductions associated with more rapid speech are a measure of coding efficiency. A. Lieberman et al. (Perception of the speech code, Psychological Revue 74, 431-461, 1967) have also emphasized the necessity for restructuring phonemes in order to overcome the inability of the human hearing system to resolve discrete elements arriving at the rates of phoneme flow customary in speech (less than 20/second) and the inability of the articulators to produce separate discrete gestures at such rates. They suggested that "dividing the load among the articulators allows each to operate at a reasonable pace, and tightening the code keeps the information rate high; it is this kind of parallel processing that makes it possible to get high speed performance with low speed machinery".

Yet the range of possible definitions outlined above represents only some of the possibilities. The treatment of pauses requires careful consideration since this determines the duration measured for
the speech sample. Similarly, the speech units counted can be concrete or abstract in various degrees. Care must be given to reduced segments. There is wide freedom for combining decisions on just these factors. Neither of the two entities involved in the computation of speech rate - duration and amount of speech - is defined a priori and the number of possible definitions of speech rate becomes very large. Then add in attempts to handle acceleration and retardation of speech rate.

The literature reviewed in this overview is not exhaustive but represents what has been accessible so far. The topics dealt with appear to fall into the following areas and will be presented in the same order: measurement of duration, suitable quanta of speech, estimates of normal tempo, cognitive planning activity, causes of tempo variation, consequences of tempo variation, information theory aspects of tempo, the perception of tempo, experiment design.