Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 63–64 (2014) 38–54

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Communicative function and prosodic form in speech timing

Laurence White *

School of Psychology, Plymouth University, Plymouth PL4 8AA, UK

Received 21 December 2011; received in revised form 8 April 2014; accepted 14 April 2014
Available online 9 May 2014

Abstract

Listeners can use variation in speech segment duration to interpret the structure of spoken utterances, but there is no systematic description of how speakers manipulate timing for communicative ends. Here I propose a functional approach to prosodic speech timing, with particular reference to English. The disparate findings regarding the production of timing effects are evaluated against the functional requirement that communicative durational variation should be perceivable and interpretable by the listener. In the resulting framework, prosodic structure is held to influence speech timing directly only at the heads and edges of prosodic domains, through large, consistent lengthening effects. As each such effect has a characteristic locus within its domain, speech timing cues are potentially disambiguated for the listener, even in the absence of other information. Diffuse timing effects – in particular, quasi-rhythmical compensatory processes implying a relationship between structure and timing throughout the utterance – are found to be weak and inconsistently observed. Furthermore, it is argued that articulatory and perceptual constraints make shortening processes less useful as structural cues, and they must be regarded as peripheral, at best, in a parsimonious and functionally-informed account.

Keywords: Speech timing; Prosody; Rhythm; Prosodic structure; Speech perception

1. Introduction

Speech timing appears to offer an ambiguous guide to speech structure. There are multiple potential influences on the duration of speech sounds, but the resultant variation is essentially one-dimensional: sounds can be either longer or shorter than expected. Despite this, numerous studies have shown that listeners are able to exploit durational variation in judgements of lexical and phrasal structure (e.g., Beach, 1991; Gow and Gordon, 1995; Price et al., 1991; Quené, 1992).

Classification of factors affecting speech timing is problematic, given the apparently paralinguistic nature of much durational variation. In Ladd's formulation, paralinguistic "aspects of vocal communication are clearly meaningful but not apparently organised along linguistic lines" (Ladd, 1996, p. 33); such aspects include the indication of interpersonal attitude, emotional state and formality of speech register. Ladd, however, made a constructive working distinction between *categorical* linguistic form and *gradient* paralinguistic form. As the following review aims to demonstrate, some consistent aspects of speech timing can be related to linguistic entities with categorical settings. Such linguistic influences on speech segment duration may be broadly stratified as segmental, syllabic and prosodic. At the segmental level, Klatt (1976) identified several intrinsic articulatory properties of broad classes: low vowels longer than high vowels; voiceless fricatives longer than voiced fricatives; bilabial stops longer than alveolar and velar stops. At the syllabic level, vowels are longer preceding voiced coda consonants, and consonants tend to be shorter when they occur in clusters (Klatt, 1976). Prosodic timing factors, the focus of this discussion, are the durational consequences of the organisation of syllables into words and higher-level constituents.

* Tel.: +44 (0)1752584876; fax: +44 (0)1752233176.
E-mail address: laurence.white@plymouth.ac.uk

http://dx.doi.org/10.1016/j.specom.2014.04.003
0167-6393/

A range of timing effects have been proposed to be conditioned by prosodic organisation. Considering just word-level prosody, Turk and Shattuck-Hufnagel (2000) assessed the evidence for five durational mechanisms: word-initial lengthening, polysyllabic shortening, accentual lengthening, "syllable ratio equalization" and word-final lengthening, finding support for all but the last of these. Additional processes have been associated with higher levels of prosodic structure, such as phrase-final lengthening (Wightman et al., 1992) and greater degrees of initial lengthening at higher phrase boundaries (Fougeron and Keating, 1997). There are also durational adjustments which are hypothesised to be conditioned by the composition of prominence-delimited constituents, such as the lengthening of a lexically or phrasally stressed syllable when immediately followed by another (e.g., Bolinger, 1965; Van Lancker et al., 1988), as well as trends or underlying tendencies towards isochrony of stress-delimited feet (e.g., Lehiste, 1977).

1.1. A functional approach to prosodic speech timing

A functional approach is here proposed to cut through the thicket of putative prosodic timing effects in English. The intention is to identify those effects that have a linguistic communicative function and as such are integrated into the speech planning process, rather than arising from physiological constraints or from transient performance factors.

To this end, speech must be considered in communicative context – the interactive exchange of information between two or more interlocutors. Speakers use variation in the signal to encode information; furthermore, information is only transmitted if this variation is not predicted by the listeners. In theory, an auditorily-based communication system could encode information through spectral variation alone; however, over the development of human languages, temporal variation has been adapted or exapted to communicative ends.

The glossogenetic origins of the temporal encoding of information may sometimes relate to durational consequences of articulation. Thus, the greater length of vowels before voiced consonants than before voiceless consonants is a natural corollary of the cessation of voicing during the transition to unvoiced sounds (Klatt, 1976). In languages such as English, this seems to have been exaggerated and systematised so that vowel duration is a robust cue to consonant voicing (Klatt, 1976; Raphael, 1972). At the prosodic level, the lengthening of segments at the end of phrases and utterances has often been described as arising from a universal, non-linguistic tendency, for example: "related to the general deceleration of motor activity" (Klatt, 1976, p. 1212) or reflecting "the braking that inertial systems show generally as they stop gently" (Fowler, 1990, p. 205). This parallels Gussenhoven's (2002) identification of the "Production Code" in pitch variation, with greater subglottal pressure at the start of the utterance associated with higher pitch, a tendency which has become integrated into intonational phonology so that low final pitch is associated with topic closure whilst high final pitch implies continuation (and *vice versa* for initial pitch). Similarly, as the review below indicates, if final lengthening does have a non-linguistic, motoric origin, it appears to have become systematised so that it is characterised by a phonologically-defined locus, whilst lengthening effects within other distinct loci may serve as cues to the onset of a word or phrase or to identify the strongest element therein. Thus, one pressure for systematisation of natural speech timing tendencies is disambiguation: given the multiple influences on segment duration, it is argued here that differential distributions of distinct prosodic timing effects serve to indicate their function. (Arguing somewhat against the non-linguistic final lengthening hypothesis, it may be noted that Snow, 1994, found that phrase-final intonation patterns are acquired by infants at a younger age than consistent final lengthening, which he argues is therefore an acquired skill rather than a product of articulatory constraints.)

Temporal coding of information in speech is limited, however, by the number of discrete durational distinctions that speakers can produce and listeners can interpret. Many phonological distinctions are binary, including phonemic vowel length in almost all studied languages (e.g., Bye, 1997; Chomsky and Halle, 1968). The evidence reviewed below suggests that temporal coding of prosodic distinctions may also be binary, at least to the extent that deviations from expected timing are only exploited as cues to structure when segments are *lengthened* rather than shortened. Whether listeners distinguish multiple levels of lengthening is considered below.

There are several pressures conspiring towards the use of lengthening rather than shortening as a prosodic cue. Firstly, there is a temporal asymmetry intrinsic in speech production: the maximum duration of continuant sounds is primarily limited by the respiratory cycle and vowels lasting several seconds are easily achievable. However, there are obvious articulatory limits to the shortening of speech sounds at typical speech rates. At extremes of shortening, sounds are no longer realised as recognisable exemplars of underlying phonemes. For example, in Dinka, a Nilo-Saharan language providing a rare demonstration of a three-level phonemic length contrast, the shortest vowel is more centralised than the two longer ones, as there is insufficient time for full articulation; thus, it is argued that ternary quantity contrasts are unstable and tend to revert to quality contrasts (Remijsen and Gilley, 2008). Clearly, in languages with phonemic vowel length contrasts, the scope for suprasegmental lengthening *and* shortening is particularly restricted.

Secondly, there may be perceptual constraints on the processing of lengthening and shortening cues by listeners. The Effort Code interpretation of intonational patterns states that there is a universal tendency for high pitch and high pitch range to be associated with high information load (Gussenhoven, 2002). Lengthening of segments likewise implies articulatory care and should similarly be associated