Cognitive Psychology 59 (2009) 203–243

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Recognizing speech under a processing load: Dissociating energetic from informational factors

Sven L. Mattys [a,*], Joanna Brooks [b], Martin Cooke [c,d]

[a] Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK
[b] Psychology Department, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
[c] Ikerbasque (Basque Science Foundation), Alameda Urquijo, 36-5 Plaza Bizkaia, 48011 Bilbao, Spain
[d] Departamento de Electricidad y Electrónica, Facultad de Ciencias y Tecnología, Universidad del País Vasco, 48940 Leioa, Spain

ARTICLE INFO

Article history:
Accepted 6 April 2009
Available online 6 May 2009

Keywords:
Psycholinguistics
Spoken-word recognition
Speech segmentation
Processing load
Energetic masking
Informational masking

ABSTRACT

Effects of perceptual and cognitive loads on spoken-word recognition have so far largely escaped investigation. This study lays the foundations of a psycholinguistic approach to speech recognition in adverse conditions that draws upon the distinction between energetic masking, i.e., listening environments leading to signal degradation, and informational masking, i.e., listening environments leading to depletion of higher-order, domain-general processing resources, independent of signal degradation. We show that severe energetic masking, such as that produced by background speech or noise, curtails reliance on lexical-semantic knowledge and increases relative reliance on salient acoustic detail. In contrast, informational masking, induced by a resource-depleting competing task (divided attention or a memory load), results in the opposite pattern. Based on this clear dissociation, we propose a model of speech recognition that addresses not only the mapping between sensory input and lexical representations, as traditionally advocated, but also the way in which this mapping interfaces with general cognition and non-linguistic processes.

1. Introduction

Most psycholinguistic theories of spoken-word recognition are built upon evidence gathered from tasks performed on carefully recorded speech and under conditions of undivided attention. However, such idealized conditions are likely to misrepresent the processes operating in everyday

* Corresponding author.
E-mail address: sven.mattys@bris.ac.uk (S.L. Mattys).

0010-0285/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2009.04.001

circumstances, wherein the speech signal is often experienced under a processing load of some sort—perceptual, attentional, mnemonic. In those more realistic conditions, understanding how we process speech is not so much a matter of modeling how we perceive sounds, access lexical representations, and parse sentences, but of how we do so while coping with a degraded signal or a distracting input.

Of particular interest is an account of the impact of processing load on the recognition of connected speech, in which resources available to listeners often include both sublexical information (acoustic-phonetic, phonotactic, and prosodic regularities) and higher-order knowledge (lexical-semantic and sentential-semantic inferences, referred to as "lexical-semantic knowledge" in this study). The topic of speech segmentation has been abundantly researched in the past two decades and is often used as an empirical bridge between disciplines because of its key theoretical status (psycholinguistics, phonetics), its clinical relevance (hearing sciences), and its practical implications (engineering). We therefore use the issue of speech segmentation as a conduit to understanding the effect of processing load on speech recognition more generally.

Research drawing upon load-free tasks has shown that listeners confronted with connected speech achieve segmentation by relying primarily on lexical-semantic knowledge and paying less attention to sublexical cues (e.g., Gow & Gordon, 1995; Mattys, White, & Melhorn, 2005; Norris, McQueen, & Cutler, 1995; Tabossi, Burani, & Scott, 1995; White, Melhorn, & Mattys, in press). Thus, in highly intelligible and contextualized speech, listeners favor segmentation solutions that align with lexical-semantic knowledge even when this information somewhat conflicts with sublexical cues, e.g., hearing /blu : krʌst/ as "blue crust" even when acoustic detail suggests that it might be segmented as *"bluke rust" (with * henceforth denoting a lexically unacceptable segmentation solution). Where lexical-semantic knowledge is unhelpful or ambiguous, sublexical cues become relatively more important.

Whether and how the relative weights of lexical-semantic and sublexical cues are affected by a processing load is largely unknown. In this study, we focus on two types of processing load often encountered in daily communication, broadly labeled perceptual and cognitive. We define perceptual load as any alteration to the signal leading to diminished acoustic integrity—e.g., overlaid noise or speech—and cognitive load as any load whose effect arises not from a distortion of the signal but from the recruitment of central processing resources due to concurrent attentional or mnemonic processing.

2. Processing load and speech segmentation: Synopsis of current knowledge

The literature on the interaction between processing load and speech segmentation is sparse and focuses primarily on loads of a *perceptual* kind, mainly broadband noise. Perhaps the clearest finding is that not all sources of information for word boundaries are equally affected by noise. While juncture-related prosodic cues, such as stress and *F*0 movements, are resilient to relatively high levels of noise, e.g., -5 to -10 dB signal-to-noise ratios, SNR (e.g., Mattys, 2004; Mattys et al., 2005; Smith, Cutler, Butterfield, & Nimmo-Smith, 1989; Welby, 2007), coarticulatory cues, transitional probabilities, and lexical-semantic knowledge show greater vulnerability (e.g., Fernandes, Ventura, & Kolinsky, 2007; Mattys et al., 2005). Among the latter cues, sensitivity to transitional probability between syllables, where low probabilities are treated as likely word boundaries, survives noise better than do coarticulatory cues (Fernandes et al., 2007). Coarticulatory cues, themselves, operate well only within a narrow range of contextual and signal-quality conditions. For example, Mattys et al. showed that, in intact speech, coarticulatory cues are useful for segmentation when lexical-semantic information is unavailable, e.g., in nonword stimuli; then, they even outweigh transitional-probability cues (Fernandes et al.). However, when lexical-semantic information is available, reliance on coarticulatory cues is drastically reduced (Mattys et al.). In mild noise (5 dB SNR), where lexical-semantic information is less readily available, coarticulatory cues exert their effect again; they do not with louder noise ($\leqslant$ 0 dB SNR). Finally, Mattys et al. found that the effectiveness of lexical-semantic knowledge drops steadily as a function of noise level, probably reflecting the increasingly diffuse lexical activity resulting from inaccurately encoded sensory information.

Less is known about the effect of *cognitive* load on segmentation. Of relevance, however, are studies of the effect of attentional load on statistical learning. While the segmentation of recurrent chunks