Journal of Phonetics 39 (2011) 330–343

Contents lists available at ScienceDirect

Journal of Phonetics

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

The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing

Esther Janse [a,b,*], Mirjam Ernestus [b,c]

[a] Utrecht Institute of Linguistics OTS, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, The Netherlands
[b] Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands
[c] Centre for Language Studies, Radboud University, PO Box 310, 6500 HD Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 1 July 2009
Received in revised form
23 March 2011
Accepted 25 March 2011
Available online 4 May 2011

ABSTRACT

Highly reduced pronunciation variants, such as something like '*yeshay*' for *yesterday*, are abundant in conversational speech. Previous research has shown that listeners understand such pronunciation variants only in their sentential contexts. This study further investigates the roles of the acoustic properties of reduced words themselves and of semantic/syntactic and acoustic information in their contexts. We report five experiments in which participants were tested on the recognition of reduced words originating from a corpus of conversational Dutch. Experiment 1, a visual cloze test, demonstrates that our set of reduced words cannot be guessed just on the basis of their semantic/syntactic context. Experiment 2 replicates the earlier finding that reduced words can only be recognised in their contexts. Experiment 3 shows that the reduced words were less well recognised if the context is presented visually in the form of orthographic transcriptions. These experiments suggest that listeners need some acoustic properties of reduced words themselves, together with semantic/syntactic *and* acoustic information in their contexts, to recognise reduced words. Experiments 4 and 5 confirm the importance of acoustic information for word recognition by showing that high-frequency hearing loss hinders both the interpretation of the target words' acoustic properties, and the use of neighbouring context.

1. Introduction

In conversational speech, words can be reduced quite drastically; for instance, the four-syllable word *hilarious* can be pronounced as /hlɛrɛs/ (Johnson, 2004). Importantly, such reductions occur very frequently, especially in highly frequent words or fixed expressions, where three-to-five syllables can be reduced to only one (Ernestus, 2000). Johnson (2004) concluded that over 60% of the words in his corpus of casual American English speech deviate from their citation forms on at least one segment, and that several segments are absent in 28% of the word tokens.

Many people are not aware of such reduced pronunciation variants, either in their own speech or in speech of others. They do not notice that segments are missing and tend to report that utterances sound intact. In Kemps, Ernestus, Schreuder, and Baayen (2004) Dutch participants listened to sentences and were asked to monitor for the phoneme /l/. Listeners often responded wrongly to reduced forms such as 'eik' /ɛik/ (a reduced form of Dutch /ɛixələk/ *eigenlijk* 'actually'), thereby filling in segments that were not part of the speech signal. Importantly, this happened only if the reduced form was presented in sentence context.

Even though most listeners are not aware of reduced pronunciation variants, there is evidence that reduced speech is more difficult to process than clear speech. Relatively mildly reduced forms (e.g., variants with missing /t/'s or schwas, or variants with flapped /t/'s) produced in isolation are recognised more slowly than their unreduced counterparts (Ernestus & Baayen, 2007; Janse, 2004; Janse, Nooteboom, & Quené, 2007; Ranbom & Connine, 2007; Tucker & Warner, 2007).

The presence of context has been shown to facilitate the perception of reduced speech substantially. Ernestus, Baayen, and Schreuder (2002) showed that highly reduced word forms, cut from a speech corpus of casual Dutch, were poorly recognised by young normal-hearing listeners when presented as isolated fragments (52% correct). They also tested identification of the reduced words if they were presented in their 'Limited context': the target word's adjacent vowels and intervening consonants. Third, identification of the same target forms was also investigated when presented in their full context. The results showed that identification of the highly reduced target words went from 52% correct in the isolation condition to 70% in the Limited

* Corresponding author at: Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands. Tel.: +31 24 3521432; fax: +31 24 3521213.
E-mail address: Esther.Janse@mpi.nl (E. Janse).

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.03.005

context condition, to 92% in the Full context condition. Similarly, Pickett and Pollack (1963) showed that very short excerpts from connected speech were often entirely unintelligible, but the same excerpts became intelligible when presented re-embedded in the original context.

The question now arises as to which properties of the context are beneficial. There is no doubt that semantic or syntactic information can aid spoken-word recognition (e.g., McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, & Cutler, 2000; Swinney, 1979). Further, several studies have shown that acoustic context cues facilitate recognition of words. Some of these contextual cues are relatively local in the speech signal: if a word's acoustic features spread through coarticulation or assimilation, this will mainly affect its immediately neighbouring segments (Ellis & Hardcastle, 2002; Gaskell & Marslen-Wilson, 1998; Gow, 2001, 2002; Nolan, 1992). The same holds for durational cues indicating the location of word boundaries (Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003), or cues indicating grammatical function (Local, 2003).

There are also acoustic context cues that are located at a greater distance from their source. Cues to the presence of /r/ in a word can be found a number of syllables before or after the word containing this phoneme (Heid & Hawkins, 2000; Kelly & Local, 1986; Local, 2003; Tunley, 1999; West, 2000). Similar long-distance effects were found for the feature [anterior] (Coleman, 2003). Hawkins (2003) argues that this spreading of features over a target word's neighbouring syllables may make the speech signal perceptually coherent.

Crucially, most studies on the role of context, regardless of whether this context mainly contained semantic/syntactic constraints or acoustic cues to the target word, have used carefully lab-recorded speech. The present study investigates the importance of acoustic properties of the word itself, and of information contained in a word's context in the perception of *spontaneous* rather than careful *read* speech. Though context plays a crucial role in understanding conversational speech, the very reduction of the context as well as the target that is characteristic of conversational speech may cause specific contextual influences to be less powerful in conversational than in read speech. Conversational speech is often faster than read speech, particularly if we focus on speech fragments containing reduced pronunciation variants. Listeners may not have the time to fully process all information contained in these speech stretches. This may hold for processing of both semantic/syntactic and acoustic information. A second issue is that acoustic cues signalling a target word in conversational speech could be reduced, relative to carefully read speech. Acoustic context cues in casual speech could then be too subtle to be picked up by listeners. The results of Pickett and Pollack (1963) suggest that listeners are able to pick up contextual information in speech that is read aloud fast, but their study did not specify how reduced the fast articulated target words or their contexts were, or which type of information was contained in the context. We specifically focus on stretches of conversational speech containing (strongly) reduced speech, and on the roles of different sources of information in the recognition of reduced words. First, we investigate which information in the words' contexts is used to facilitate target word recognition. Second, we investigate the importance of acoustic properties of the reduced words themselves for spoken-word recognition.

In order to investigate the roles of acoustic (bottom-up) information and semantic/syntactic context (top-down) information in the recognition of reduced words, we followed up the study of Ernestus et al. (2002). We first set out to investigate how semantically/syntactically constraining the context fragments of their (2002) study were to see which type of context in the (2002) study was particularly beneficial to recognition of the target words. The results of a cloze test on written materials (our Experiment 1) suggest that their 'full context' did not contain many syntactic/semantic cues. In Experiment 2, we replicated the Ernestus et al. (2002) context benefit results with a number of improvements in the experimental design, including only context fragments that are not strongly semantically or syntactically biasing. In Experiment 3, we presented the auditory target words together with a (visual) orthographic transcription of their sentential contexts. This was done to investigate whether the combination of the acoustic properties of the reduced words and the semantic/syntactic information contained in their contexts would facilitate word recognition as much as presenting the targets in their auditory contexts (as was done in Experiment 2). The results showed that providing the visual context also boosted recognition of the (auditorily presented) reduced words, but to a lesser extent than presenting the reduced words in their auditory contexts. Together, these three experiments show that the context effects on target word recognition in Ernestus et al. (2002) should be attributed to the interaction between the acoustic properties of the target word and information contained in the word's context (being both semantic/syntactic constraints and acoustic cues to the target word).

Experiments 4 and 5 provide initial data about the relevance of the high-frequency part of the spectrum for recognition of target words in and out of their natural contexts. Several studies have shown that while young normal-hearing listeners are hardly aware of the massive reduction in conversational speech style, reduced articulation is much more evident to people with hearing loss, which often affects hearing in the high frequencies in particular (Payton, Uchanski, & Braida, 1994; Picheny, Durlach, & Braida, 1985, 1986, 1989; Uchanski, Choi, Braida, Reed, & Durlach, 1996). This suggests that high-frequency acoustic information may be important for the recognition of conversational speech. Experiment 4 tests this hypothesis directly. In Experiment 4, young normal-hearing listeners were presented with the speech materials as in Experiment 2, but these materials were filtered with a downward sloping filter, attenuating high frequencies more than low frequencies.

Experiment 5 investigates the recognition of reduced target words in and out of their contexts by older adults. Many older adults suffer from (age-related) hearing impairment that particularly affects the high frequencies. Unlike the young adults tested in Experiment 4, who suddenly had to deal with a speech signal with attenuated high frequencies, older adults may have gradually developed strategies to cope with a more impoverished speech signal. Such strategies may involve a stronger reliance on low-frequency information (e.g., Stelmachowicz, Lewis, Kelly, & Jesteadt, 1990), or on semantic context (Nittrouer & Boothroyd, 1990; Pichora-Fuller, Schneider, & Daneman, 1995). Experiment 5, therefore, shows how indispensable high-frequency information is for the recognition of our reduced speech materials.

In short, the present study investigates how bottom-up acoustic information in a target word and in its context and top-down semantic/syntactic information in its context contribute to the recognition of that word in conversational speech. These issues are addressed in five experiments: one visual cloze test to test whether our targets' contexts were semantically/syntactically constraining, one study in which auditory presentation of the reduced words was supplemented with a visual (orthographic) presentation of their contexts, and three auditory recognition studies testing how well different groups of listeners recognise the words in and out of their contexts.

2. Experiment 1: Cloze procedure on sentence contexts

Experiment 1 investigated whether the context effects on target word identification reported in Ernestus et al. (2002)