ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2pSCb: Speech Intelligibility (Poster Session)

2pSCb1. Effect of speech clarity on perception of interrupted meaningful and anomalous sentences

Rajka Smiljanic*, Stanley Sheft, Bharath Chandrasekaran and Valeriy Shafiro

*Corresponding author's address: Linguistics, University of Texas at Austin, Austin, Texas 78712, rajka@mail.utexas.edu

The influence of speech clarity on the perception of interrupted speech was examined for sentences distinguished by the presence of semantic-contextual cues. Semantically meaningful and anomalous sentences produced in either conversational or “clear” speech were periodically interrupted at gating rates ranging from slow (0.5 Hz) to fast (24 Hz) and presented to 26 native English listeners. At slow rates, speech perception may be based on integration of whole syllables and words, with “glimpsing” of (sub)phonemic segments playing a role at faster rates. Our results show that semantic context and speech clarity had a significant rate-dependent impact on the intelligibility of interrupted speech. At the lowest rates, intelligibility differences between conditions were minimal. Overall, interruption was most deleterious for anomalous conversational sentences. Such effects were seen even at the highest gating rate of 24 Hz for which interruption effects are generally minimal. The magnitude of the clear-speech benefit varied with gating rate for the two types of sentences, starting at 1 Hz for meaningful and 2 Hz for anomalous sentences. Acoustic-phonetic enhancements of clear speech thus “shifted” contextual benefit to lower gating rates. The implications of these results for our understanding of different time scales of speech processing will be discussed.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Speech perception requires processing of multiple temporally distributed acoustic cues to integrate the information into phonemes, syllables and words. This process is remarkably rapid for intact speech signals. Even for interrupted speech in which portions of the signal are removed, intelligibility remains high if the interruption or gating rate is higher than 6 Hz (Miller and Licklider, 1950). It is likely that perceptual processing of interrupted speech differs across rates depending on the characteristics of the removed and remaining speech fragments. At slow interruption rates, the duration of the available speech segments approaches that of whole words and allows for higher-level linguistic knowledge to aid in filling in the missing information. At high interruption rates, the brief duration of the missing (sub)phonemic fragments of speech may lead to temporal smoothing over the short gaps in the signal. In other words, at slow rates, speech perception may be based on integration of multiple syllables and words, with “glimpsing” of (sub)phonemic segments playing a role at faster rates. Consequently at both interruption rates, intelligibility can remain high despite differences in the underlying mechanisms that provide and maintain the speech information. At intermediate interruption rates (2-4 Hz), however, the roughly syllable-size duration of available speech segments may allow for neither rapid lexical access nor glimpsing of acoustic cues, resulting in a decrease in intelligibility.

The present study examined the influence of conversational-to-clear speech modifications on the perception of interrupted speech for sentences distinguished by the presence of semantic-contextual cues. Our previous work with the same stimuli showed that the overall probability of correct sentence-in-noise recognition was significantly higher for meaningful vs. anomalous sentences and for clear vs. conversational speech (Van Engen et al., 2012). Furthermore, the clear-speech benefit was larger for meaningful sentences compared to anomalous sentences. Thus in this study, it was expected that at rates where information is integrated over words, the meaningful sentences would be more intelligible compared to anomalous sentences. Furthermore, more robust acoustic-phonetic cues in the clear speech signal would lead to the use of contextual cues at lower interruption rates. To that end, we examined intelligibility of semantically meaningful and anomalous sentences produced in either conversational or “clear” speech that were interrupted periodically at gating rates ranging from slow (0.5 Hz) to fast (24 Hz).

METHODS

Participants

Twenty-six normal-hearing adults (14 females) between the ages of 18 and 34 participated in the listening tests. They all passed a hearing-screening test (1000, 2000, and 4000 Hz at 25 dB). All subjects were native, monolingual American English speakers. All were undergraduate students at the University of Texas at Austin. Participants provided written informed consent and received course credit.

Stimuli

The stimuli consisted of 80 semantically anomalous sentences (e.g., The wrong shot led the farm) from the Syntactically Normal Sentence Test (SNST) (Nye & Gaitenby, 1974) and 80 semantically meaningful sentences (e.g., The grey mouse ate the cheese) derived from the Basic English Lexicon (BEL) (Calandruccio & Smiljanic, 2012) sentences to match the SNST sentences in syntax, length, and amount of repetition within the set. All sentences were read in conversational and clear speaking style by a 26-year old female native speaker of American English. Following previous protocol to elicit the speaking style changes (Smiljanic & Bradlow, 2005, 2009), the stimuli were recorded in a sound-attenuating booth. Individual sentences were segmented from the long recording and equalized for RMS amplitude using Praat (Boersma & Weenink, 2009). For the current experiment, target sentences were periodically interrupted at seven gating rates ranging from slow (0.5 Hz) to fast (24 Hz). These rates were selected to include the range of speech-on intervals, from (sub)phonemic to word intervals, expected to play a role in perceptual processing on different time scales.
Procedure

Listeners were seated in front of a computer and heard target sentences over Sennheiser HD570 or Sony MDR-CD780 headphones. They were asked to write down what they heard after each sentence using the keyboard. Order of the 8 conditions (7 with gating and a control in which sentences were not interrupted) was randomized across listeners. Each listener heard 20 randomly selected sentences (10 anomalous + 10 meaningful) per gating condition in only one speaking style for a total of 160 sentences per listener. Listeners never heard the same sentence twice. Written sentence output was scored for keyword correct score. Each sentence contained 4 target keywords for a total of 40 per gating condition/sentence type/speaking style.

RESULTS

The mean intelligibility scores for each gating condition, sentence type and speaking style are shown in Table 1 and Figure 1 (note that 0 rate indicates uninterrupted sentences, i.e., sentences presented in their entirety). A split-plot analysis of variance (ANOVA) on the within-subjects effect of gating rate (0, .05, 1, 2, 4, 8, 16 and 24) and sentence type (anomalous vs. meaningful) and the between-subjects effect of speaking style (conversational vs. clear) revealed significant main effects of gating rate [F(1, 24) = 284.220, p<.001], sentence type [F(1, 24) = 138.297, p<.001] and speaking style [F(1, 24) = 102.229, p<.001]. The gating rate by style [F(1, 24) = 33.521, p<.001], sentence type by gating rate [F(1, 24) = 27.198, p<.001] and sentence type by gating rate by style [F(1, 24) = 15.938, p<.001] interactions were significant. The results showed that semantic context and speech clarity had a significant rate-dependent impact on the intelligibility of interrupted speech. Overall, interruption was most deleterious for anomalous conversational sentences. At the lowest gating rates, intelligibility differences between sentence type conditions were minimal. In the conversational speaking-style condition, .05 and 1 Hz interruption rates affected speech intelligibility for both sentence types equally. Starting at the 2 Hz gating rate, intelligibility of conversational sentences was enhanced through the presence of context. That is, meaningful sentences were more intelligible compared to anomalous sentences at each subsequent gating rate. The beneficial effect of context remained even at the highest gating rates of 16 and 24 Hz for which interruption effects are generally minimal. The deleterious effect of interruption was less severe for clear speech than conversational sentences at most gating rates. The magnitude of the clear-speech benefit varied with gating rate for the two types of sentences, starting at 1 Hz for meaningful and 2 Hz for anomalous sentences. Similar to the conversational speaking style condition, the interruption effect was larger for anomalous sentences compared to meaningful sentences at all gating rates. Compared to the conversational style conditions, the intelligibility difference between anomalous and meaningful sentences was smaller in the clear speech conditions at the higher gating rates (16 and 24 Hz).

<table>
<thead>
<tr>
<th>Interruption rate (Hz)</th>
<th>Anomalous</th>
<th>Meaningful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conversational</td>
<td>Clear</td>
</tr>
<tr>
<td>0</td>
<td>94.8 (3.9)</td>
<td>98.6 (2.4)</td>
</tr>
<tr>
<td>0.5</td>
<td>40.8 (6.1)</td>
<td>40.2 (8.8)</td>
</tr>
<tr>
<td>1</td>
<td>34.2 (8.4)</td>
<td>36.7 (8.9)</td>
</tr>
<tr>
<td>2</td>
<td>28.1 (9.0)</td>
<td>50.3 (14.1)</td>
</tr>
<tr>
<td>4</td>
<td>30.8 (12.1)</td>
<td>67.5 (14.5)</td>
</tr>
<tr>
<td>8</td>
<td>53.9 (14.6)</td>
<td>82.2 (8.5)</td>
</tr>
<tr>
<td>16</td>
<td>69.8 (15.3)</td>
<td>88.9 (6.8)</td>
</tr>
<tr>
<td>24</td>
<td>76.1 (11.7)</td>
<td>94.4 (6.9)</td>
</tr>
</tbody>
</table>
The results of the current study demonstrated that the intelligibility of gated speech was influenced by both higher level cognitive factors represented by sentence type and lower level acoustic-phonetic factors represented by speaking style adaptations. Previous work has shown that both semantic and acoustic-phonetic enhancements significantly improve speech recognition in noise. Here, we demonstrated that the same enhancements also improve intelligibility of interrupted speech at gating rates ranging from slow (0.5 Hz) to fast (24 Hz). However, the effects of semantic and acoustic-phonetic cues are rate-dependent, being considerably greater at rates of 1 Hz and higher, corresponding to speech fragments of approximately syllable size or smaller. Furthermore, the semantic and acoustic-phonetic cues can have additive effects on intelligibility of gated speech as evidenced by the superior intelligibility of meaningful clear sentences to all other conditions. The semantic and acoustic-phonetic cues thus both contributed to the ability of listeners to integrate temporally distributed audible speech fragments into coherent percepts. The magnitude of these contributions, however, is related to the duration of the missing and available speech fragments. Importantly, acoustic-phonetic enhancements introduced by clear speech “shifted” contextual benefit to lower gating rates. The contextual effect only became evident at 2 Hz for conversational sentences, while in clear speech, semantic context provided a large intelligibility benefit already at the 1 Hz interruption rate. This speaking style by sentence type interaction suggests that enhanced low-level acoustic cues provided by conversational-to-clear speech adjustments enable listeners to utilize higher-level semantic cues at slower interruption rates. The current findings for anomalous vs. meaningful sentences are consistent with our previous work showing that intelligibility of interrupted speech varies as a function of speech materials. Shafiro et al., (2011) found that the cross-rate intelligibility gating functions differed considerably between random word strings vs. meaningful sentences. The present results reinforce past speculations that speech gating at different rates taps into different time scales and separate perceptual processes operating on the available speech information. Perceptual integration of missing information thus depends on successful integration of low-level acoustic information and high-level linguistic knowledge.

REFERENCES

