Listener Variability and Multiple Perception Processes

Stephanie Lindemann

Program in Linguistics, University of Michigan, Ann Arbor, MI 48109

Abstract: The relationship between the perception of individual phonemes and the perception of words is investigated to determine if these are similar tasks for listeners. Since listeners may use top-down processes in full words, it is argued that, at least for some, the task of identifying isolated phonemes may be very different from the task of identifying a full word. Native speakers of Sudanese Arabic were tested on identification of English phonemes in word context and the same instance of these phonemes extracted from the words. Individuals’ overall agreement scores on corresponding items on the two tests varied substantially, with only 4 of the 15 significantly different from chance. Also, not all subjects performed better on the full word task, suggesting that they do not all use higher-level information. It is argued that studies of phoneme-level perception are more relevant to some listeners’ word perception than to others’.

INTRODUCTION

Few speech perception studies have investigated the relationship between the perception of individual phonemes and the identification of entire words. One possible difference between perception of phonemes and perception of words is that perception of words is more likely to involve ‘top-down’ processes including lexical effects. For example, a frequency effect would allow high-frequency words to be identified more quickly than low-frequency words, as described in (1). While the existence of ‘top-down’ effects on the perception of phonemes is still controversial, it seems reasonable to suppose (2) that listeners use any information that may be in the signal to understand a speaker. It is not yet clear whether such ‘top-down’ processes may be as relevant for normal speech perception as are ‘bottom-up’ processes, or if they are equally relevant for all listeners.

This study investigates how lower and higher levels of perception may interact by comparing within-listener performance in identifying a particular phoneme when it is in the context of an entire word versus when it is extracted from the word. Assuming that there are listeners who perceive phonemes poorly and others who perceive them comparatively well, the question is how their perception of phonemes matches their perception of whole words. In other words, are poor perceivers of phonemes necessarily also poor perceivers of words?

As variable inter-listener performance on speech perception tasks by nonnative listeners has already been documented (e.g. 3), this group is particularly appropriate for investigating within-listener variation correlation in performance on two different perception tasks. At issue is whether performance on one task will help predict performance on another, or if the two tasks appear to be independent of each other. In this study I therefore look at the perception of English words and phonemes by native speakers of Sudanese Arabic, looking at the performance of English native speakers on the same tests for purposes of comparison.

METHOD

Ten English phones with varying phonemic status in Arabic ([b, p, v, f, s, j, ð, 8, g], and [z]) were chosen for the identification tasks; three words were chosen for each of the ten phones. In each case the target sound was intervocalic in a two-syllable word with stress on the first syllable, so that the phones occurred in reasonably equivalent contexts. In addition, words were chosen such that the phone being tested is at the uniqueness point of the word, lessening the possibility that the phone could be guessed without it being perceived at all, as it easily could be if it occurred after the uniqueness point and thus were completely predictable from the part of the word preceding it. Words were excised from a colloquial-style short story read by a native speaker of American English; the target consonants were then excised from these words, including at least 25 ms of the following sonorant. An additional 29 two-syllable words and 22 CVs with 9 control phones were excised from the passage and added as controls. A 59-item word test and a 52-item phoneme test were then created from these stimuli; these tests were presented over headphones in a quiet room to 12 native speakers of English and 15 native speakers of Sudanese Arabic with a good command of English. Subjects were asked to write down the initial consonant sound for each CV and the word (or whatever it sounded like to them) for each word. Answers on the word task were counted as correct if the medial consonant was correct (e.g., the answer “topics” for the test item *tropics* was counted as correct, whereas “traffics” was counted as incorrect). The order of the two tests was counterbalanced across subjects.
RESULTS

Average word, phoneme, and agreement scores are presented in Table 1. For the agreement score, each item on the word test was compared with the corresponding item on the phoneme test for each subject in order to determine whether successful identification of the sound in the context of the entire word was related to identification of the sound without the full-word context. Answers were considered to agree on the two tests if the subject either identified the target sound correctly on both tests or misidentified it on both tests.

<table>
<thead>
<tr>
<th>Score</th>
<th>Native Speakers</th>
<th>Nonnative Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word test</td>
<td>95% (SD = 3)</td>
<td>80% (SD = 9)</td>
</tr>
<tr>
<td>Phoneme test</td>
<td>64% (SD = 6)</td>
<td>54% (SD = 11)</td>
</tr>
<tr>
<td>Agreement</td>
<td>64% (SD = 7)</td>
<td>59% (SD = 8)</td>
</tr>
<tr>
<td>Agreement range</td>
<td>53% - 77%</td>
<td>47% - 74%</td>
</tr>
</tbody>
</table>

Control-group individuals' differences in overall scores on the two tests fell between 20% and 40% favoring the word test; nonnative group differences were between 48% favoring the word test and 8% favoring the phoneme test. According to a binomial test, only the 42% of native subjects and 27% of nonnative subjects with agreement scores of 67% or greater had agreement significantly different from chance. In those cases where the subjects' performance on test items did not agree, the phoneme was misidentified in the full-word context 6% of the time by the native group and 18% of the time by the nonnative group. Two nonnative subjects had a noticeably higher number of mismatches between the two tests where the phoneme was heard correctly when not in word context; when these two subjects are considered separately, the remaining nonnative subjects' in-word-incorrect out-of-word-correct responses make up only 13% of the mismatches. In this type of mismatch, word context appeared to actually mislead subjects, e.g., subjects identified tropics as some form of the more common word ‘traffic’.

DISCUSSION

Although some subjects' phoneme test score is related to their word test score, performance on one cannot generally be used to predict performance on the other. In other words, those who perceive isolated phonemes poorly are not necessarily also poorer perceivers of words. Some subjects are aided by lexical information more than others, and in some cases lexical information can even be a distraction.

Ironically, the strongest evidence that some sort of 'top-down' strategy is being used are cases in which the listener misidentifies the word (and the target sound within that word), but identifies the phoneme correctly outside of the word context. In these instances, it is likely that the similarity to a more familiar word (e.g., “traffic” instead of tropics, “super” for sober) would cause the listener to disregard the acoustic information to some extent.

Nonnative listeners differed from the control group not only in terms of overall scores on each of the two tests, but also in the within-listener differences between scores on the two tests. In both cases a minority of individuals' agreement scores were different from chance, but while native speakers always had higher scores on the word task than on the phoneme task, this was not the case for the nonnative subjects, although phoneme task scores were never substantially higher than word test scores. This may be due to the fact that 'top-down' processing was more regularly available to native speakers, who presumably knew all or nearly all of the stimulus words. The variation in listeners' use of different-level processes for perceiving phonemes and words has implications for our understanding of work in both phoneme perception and lexical access, in that studies of phoneme-level perception appear to be more relevant to some listeners' word perception than to others'.

REFERENCES