TAKING THE HIT: WHY LEXICAL AND PHONOLOGICAL PROCESSING SHOULD NOT MAKE LEXICAL ACCESS TOO EASY

E. G. Bard, C. Sotillo, M. P. Aylett
Dept of Theoretical and Applied Linguistics and HCRC
University of Edinburgh, Edinburgh, Scotland

ABSTRACT

Variations in word pronunciation are usually assumed to be overcome by collaboration between acoustic phonetic and lexical processing. This paper summarizes evidence from studies of the production and perception of spontaneous speech which show that this is unlikely. Natural phonological reductions hamper but do not block lexical access in cross-modal repetition priming and in ultimate recognition of excerpted words, but these phonological changes are not constrained in any simple way. Nor are common changes in length and intelligibility with coreference often linked to a simple accented-deaccented contrast. Nonetheless, difficulty in recognizing spoken words primes prominent items in a model of the discourse. The pattern of results seems to require a stage where lexical and sub-lexical processes interact with discourse representations.

1. INTRODUCTION

It seems to be assumed that acoustic phonetic processing, phonological representation in the mental lexicon, and lexical effects on phonological representations of speech input share the burden of overcoming variations in natural pronunciation, producing a percept of roughly equal quality whatever the input. This paper summarizes work which shows why that outcome would be counterproductive.

In fact, naturally occurring variations in word pronunciation are not noise, but information [4, 7, 9]. Changes in duration or prosodic prominence, vowel centralization, phonological reductions and assimilations can indicate, for example, whether a word stands alone or forms part of an utterance [5], if the latter, whether the word lies at the boundary of a major prosodic unit or is closely related to words on either side, whether the word is predictable in its context, and whether a nominal refers to a Given or a New entity [7]. If phonetic and lexical processing iron out this variation, either the information it carries is unusable or its value must be realized via additional mechanisms which run parallel or logically subsequent to lexical access. We report a series of studies on spontaneous speech, where these variations most naturally occur, which show that (1) Natural variations in word pronunciation do affect lexical access; (2) Carrying their information by other means would not be simple; (3) There is a prima facie case for a relationship between difficulties in lexical access and the comprehension of discourse.

2. PHONOLOGICAL REDUCTIONS AND LEXICAL ACCESS

Established effects of phonological assimilations [e.g. 8] depend on read items whose general clarity may make a marked assimilation unnatural. This study compared the effects on lexical access of two kinds of spoken word tokens, those read in a list, and those taken from spontaneous running speech uttered during a route communication task [1].

2.1. Method

All spoken words were names of schematic map landmarks which invited a phonological reduction or assimilation (weak vowel reduction: canoes, elephants; nasal place-of-articulation assimilation, pine grove, deleteion, old mill). Test (48) and control primes (48) were matched as far as possible in length and frequency, with half of each from list-read and half from spontaneous running speech. The running speech primes (and their critical segments) were shorter than the read, less accurately recognized out of context; and judged more assimilated by a panel of phonicians.

Twelve groups of 3 subjects each heard primes binaurally over headphones and made lexical decisions on immediately following letter strings of 4 kinds: the identical word (elephants), a non-word differing from the auditory word in one letter (elephants); a control word (waterhole); or a nonword formed from the control (waerhole). Each subject heard 5 spontaneous and 5 read primes in each of these conditions and 40 fillers, of the same types. Experiment 1 fillers were from spontaneous speech; Experiment 2 used read list forms of the same words.

2.2. Results

Table 1 shows robust facilitation (E1: MinF' (1,98) = 22.67, p < .005; E2: F1 (1, 47) = 26.21, p < .0001; F2(1, 52) = 30.91, p < .0001) both by less intelligible spontaneous forms (t(70) = 2.95, p < .005) and by more intelligible citation forms (t(70) = 7.786, p < .001) with enhanced priming for the latter (F1 (1, 71) = 8.39, p < .005; F2 (1, 55) = 4.83, p < .04; EII: F1 (1,47) = 6.72, p < .02; F2 (1,52) = 1.79, n.s.).

Table 1. Priming effects (ms) in cross-modal (lexical decision) by source of fillers and of prime.

<table>
<thead>
<tr>
<th>Primes</th>
<th>Fillers</th>
<th>Running Speech</th>
<th>Read in list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running Speech</td>
<td>49</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Read in list</td>
<td>99</td>
<td>88</td>
<td></td>
</tr>
</tbody>
</table>
3. COMPLEX INFORMATION

Thus, natural variation appears to affect lexical access directly, as well as affecting eventual intelligibility. Even if the intelligibility deficit could be cured with only sound and lexical information, representing and interpreting the variation in form appears to be no simple task. To support this claim, we present the results of 3 studies which show that variation in pronunciation is neither simple in form nor necessarily keyed to information which the listener has to hand.

3.1. De-accenting

An important source of variability in the pronunciation of words is the status of their referents in the discourse, with introductory mentions of New entities longer and clearer than second mentions of Given items. The former are said to be prosodically prominent, or accented, and the latter non-prominent or de-accented. If the variation in pronunciation can be represented via such a simple binary choice, then a single bit of information (± accented) could be used to aid in interpretation.

Such contrasting tokens are produced in simple laboratory tasks where subjects utter a sequence of similarly structured descriptions. In descriptions which changed instrument, action, and/or patient each time, however, repeated mentions were most likely to be deaccented when they occupied the same sentence position and played the same syntactic role as in the utterance carrying the initial mention [14]. In other kinds of repeated mention, de-accenting was rarer.

Dialogue more closely resembles those other conditions. In the route communication dialogues of the HCRC Map Task Corpus, successive utterances rarely serve the same communicative purposes: some are descriptions, some questions of various kinds, etc.,. If different purposes demand different syntactic structure, then the opportunities for deaccenting may seldom arise.

3.1.1. Method

Utterances containing 293 landmark names mentioned twice in the HCRC Map Task Corpus, were GlToBI [12] coded with mentions classed as having non-boundary accent, boundary tone or neither. Utterances were also coded for phrasal structure (sequence of NP, VP, PP etc.) and for communicative purpose (Dialogue Games and Moves [6]). Word tokens were assigned standardized durations and identified in noise within larger experiments.

3.1.2. Results

Repeated structures, which should promote de-accenting, were rare whether the mention-bearing utterances served similar purposes (35%) or not (6%). Thus, it is unsurprising that binary alternation between accented New and de-accented Given mention (18%) was significantly rarer than the 37.5% that random choice among the coding options would predict ($\chi^2$ (df = 2, n = 293) = 327.6, $p < .001$). The majority pattern, two similarly accented tokens (72%), still shows decreases with repetition in both duration ($F_2(1, 211) = 14.36, p = .0002$) and intelligibility ($F_2(1, 221) = 4.54, p < .04$). Thus the differences which could separate Given from New are appear to be continuous rather than binary.

3.2. Control by lexical competition

If variability in word pronunciation is not simple, perhaps it is predictable from information already available to the listener, and so both easy to recover from and redundant in further processing. Lindblom [11] has claimed that the words are articulated clearly enough to be decoded by listeners on the basis of the delivered acoustic pattern and the listener's other knowledge. The mental lexicon itself is one source of such knowledge: words with fewer competitors should, therefore, be less carefully delivered than those with more (see [9, 11]).

3.2.1. Method

All the words were names of landmarks of the HCRC Map Task Corpus and had forms which invited assimilation of word-final alveolar nasals toward the labial (seven beechees) or velar (pine grove) initial consonant of the following word. Both first and second mentions of the landmarks from running speech were compared to the same names read in list form by the same speakers. Nine trained phoneticians judged the degree of assimilation of all tokens [11]. Words were presented in noise to naive listeners ($n > 5$) for identification during other experiments[3].

3.2.2. Results

Table 2 shows that intelligibility was lower and judged assimilation was greater for running speech tokens than for read forms. Items judged more assimilated were less intelligible ($r(51) = -.41, p < .005$). Competitor sets, were defined as words sharing a mid-class representation of their initial syllable, (where, for example, all liquid consonants match) because this criterion includes most responses in the identification task. Table 2 shows no greater assimilation in words with smaller competitor sets.

<table>
<thead>
<tr>
<th>Set:</th>
<th>Intelligibility</th>
<th>Assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>large</td>
<td>small</td>
</tr>
<tr>
<td>Mention:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>.28</td>
<td>.17</td>
</tr>
<tr>
<td>Second</td>
<td>.43</td>
<td>.28</td>
</tr>
</tbody>
</table>

3.3. Control by Listener Knowledge

Speakers' articulation is said to be constantly constrained by the other information which listeners can bring to bear on the process of recognizing spoken words.[11]. In essence the speaker bears the computational burden of
Table 3. Summary of speaker and listener knowledge conditions under which repeated mentions fell in standardized duration and intelligibility. +: condition holds; -: condition does not hold; /+: or

<table>
<thead>
<tr>
<th>Experiment</th>
<th>How Given status achieved</th>
<th>By speaker</th>
<th>By listener</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Said</td>
<td>Sees</td>
<td>Heard</td>
</tr>
<tr>
<td>I: repeated intro</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>II: /+ deny</td>
<td>+</td>
<td>+</td>
<td>- /+</td>
</tr>
<tr>
<td>III: self / other</td>
<td>-/+</td>
<td>+</td>
<td>- /+</td>
</tr>
<tr>
<td>IV: -/+ shared</td>
<td>-</td>
<td>/+</td>
<td>+</td>
</tr>
</tbody>
</table>

continuously updating a model of the listener’s knowledge. This onerous task is ignored in other domains [10]. We again used the running speech of the HRC Map Task Corpus to examine the duration and clarity of words in contexts where speakers’ knowledge was known to differ from listeners’.

3.3.1. Method
Words from first and second mentions of landmark names were excerpted from speaker-per-channel DAT tape recordings, measured for duration and presented to groups of naive listeners for identification [3]. Experiments contrasted introductory mentions to a first and a second listener (I); with or without intervening negative feedback on the visibility of the introduced landmark to the listener (II), by the same and different speakers (III), when the landmark was or was not visible to the repeater (IV).

3.3.2. Results
Table 3 summarizes findings. The usual effect of repeated mention was found throughout, regardless of the manipulations of listener knowledge recorded in the rightmost columns of the table. All manipulations, however, shared the condition in the shaded column: if the speaker has heard the prior mention, Given status is indicated by a reduction in standardized duration and in effective intelligibility.

4. DISCOURSE COMPREHENSION
If spoken words present perceptual difficulties which are not resolved in a binary way or with readily available information, the listener has a delicate perceptual task for which the key is the speaker’s internal representation of the discourse. But given that less clear words accompany more speaker knowledge, what the listener needs to do is to recruit an account of the speaker’s knowledge where intelligibility is lowest. In effect, we are predicting that difficulty in recognizing spoken words should enhance discourse function.

To test this prediction we [4] used a variant of the primed probe recognition task of [7], in which a spoken monologue is interrupted by a pair of excerpted words which must be classed as having occurred previously or not. In the critical cases, the probe is related to the first token of the prime word and the probe site is close to that first mention but before the second. In [7], faster probe RTs followed the shorter, less clear second token primes. Here we contrast co-referential with non-coreferential repeated tokens, and we use the materials as recorded and when overlaid with noise. Our suggestion predicts greatest benefits to probe reaction time when noise is added to the co-referential case, for here the added difficulty should permit the involvement of a representation of the discourse which allots the named entity a prominent role.

4.1. Method

Materials. Stimuli came from digital recordings of business dictations made by 24 speakers in their workplaces. From these, 25 non-coreferential repeated uses of a word were selected, each one an initial mention of a different entity, and 23 coreferential repetitions, all first or second mentions of a single entity. Panels of judges who were asked to mark on transcripts any word they thought would have to be mentioned again in the document. First mentions of the Coreferential items were selected more often (60%) than Non-coreferential (35%) \( (F_{1, 26} = 47.57, p < .0001; F_{1, 46} = 7.94, p < .008) \). Coreferential pairs differed in length \( (614 \pm 551 \text{ ms}, t(22) = 2.61, p < .01) \) unlike non-coreferential \( (613 \pm 585 \text{ ms}) \). In pretests, stimuli were recognized more often in their original ‘clean’ form (87.2% for first tokens, 81.5% for second) than with noise occupying 30% of the signal amplitude. \( (67.9\% \text{ and } 63.7\%) \) \( (\text{Min } F_{1, 103} = 17.69, p < .005) \).

Procedure. Dictations were played to listeners through headphones. At 8 points in each passage a tone warned of the imminent presentation of prime and probe (at ISI of 1500ms). In critical cases, the probe had occurred in close syntactic relationship to the initial, recently heard token of the critical prime. The prime was that first token, the second token of the word, or an unrelated recently heard word, distributed by Latin square across subject groups. Six of the 8 interruptions were filler trials counterbalancing for ‘yes’ and ‘no’ answers in both positions. In the ‘noise’ condition all primes and probes were treated with noise.

Subjects. Thirty-two members of the Edinburgh University community with no known hearing loss
participated in the ‘clean’ condition; another 42 in the ‘noise’ condition.

4.2. Results

Five subjects were eliminated from the clean condition for equipment or accuracy failure and 14 from the noise condition. Wrong answers (7.7%, 9.5%) were replaced by the cell mean and outliers by the subject or item mean ± 2 s.d. (≤ 4.8%).

Table 4 gives priming effects. Priming was found only in the coreferential cases (Priming x coreferentiality: $F(2, 102) = 4.87, p < .01$; $F(2, 76) = 6.88, p < .002$; both cell comparisons to controls $p < .05$). Added noise enhanced the priming effect for coreferential primes and decreased it for non-coreferential. ($F(1, 51) = 8.06, p < .007; F(2, 38) = 1.66, n.s.$). Multiple regression analyses were used to view design- and word-based effects. For non-coreferential materials, 32% of the variance in probe RT was accounted for by probe word characteristics (duration, distance of interruption from original word) and only 8% by prime variables; for coreferential items, 20% was associated with were prime word frequency, duration, RT, and distance from original word and only 12% with target characteristics (Mult. $R^2 = .44$ and .41 respectively, $p < .0001$). In both cases, longer target RTs accompanied longer prime RTs ($\beta = .34$ and .42, $p < .01$).

Table 4. Priming effects (ms) for probed recognition of targets after related primes - by noise condition, mention, and referential relationship between mentions

<table>
<thead>
<tr>
<th>Noise condition:</th>
<th>1st</th>
<th>2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>76</td>
<td>62</td>
</tr>
<tr>
<td>Noise</td>
<td>89</td>
<td>63</td>
</tr>
<tr>
<td>Pairs:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coreferential</td>
<td>-1</td>
<td>8</td>
</tr>
<tr>
<td>Non-coreferential</td>
<td>21</td>
<td>13</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The studies just summarized show that frequently occurring alterations from ‘canonical’ word pronunciation can affect lexical access in a graded way, are not conditioned by lexical competition, but are conditioned by the speaker’s rather than the listener’s knowledge. It would be computationally elegant to use the corresponding, graded difficulty in lexical access to further the listener’s ultimate goal, comprehending the speaker’s message. We have at least initial evidence that reader access to a model of the discourse, with important information made particularly accessible, accompanies difficulties in deciphering spoken words. If difficult stimuli yielded perfect but slower recognition via lexical competition, thereby giving discourse understanding time to proceed independently, we would expect long prime RTs to accompany short probe RTs in the final study, but they do not. A more suitable model seems to demand interaction between lexical and discourse processes.

6. ACKNOWLEDGEMENTS

This work was supported by the ESRC(UK) grant to HCRC and by EU BRA BR3175. This paper includes materials reported in [2, 3, 4, 5, 13]. Address for correspondence: Dr. E. G. Bard, Dept. of Theoretical and Applied Linguistics, University of Edinburgh, George Square, Edinburgh EH8 9LL, U.K. Email: ellen@ling.ed.ac.uk.

7. REFERENCES