GOLDILOCKS AND THE THREE BEERS: WORD RECOGNITION AND SOUND MERGER

Paul Warren 1, Megan Rae 1, & Jen Hay 2

1 School of Linguistics and Applied Language Studies, Victoria University of Wellington
2 Linguistics Department, University of Canterbury

ABSTRACT: The consequences for the process of spoken word recognition of changes-in-progress in the sound system of New Zealand English are investigated in a semantic priming experiment. For example, semantic associates of both beer (ie. wine) and bear (ie. hug) are facilitated after subjects have heard [biː], while only hug is primed by [beɪ]. This result reflects the asymmetry found in speech production studies in New Zealand, which show a progressive merger towards the [iː] form for words with these diphthongs.

BACKGROUND

A commonly heard complaint when language change is in progress is that speakers are debasing the language or are unable to communicate clearly. Thus one recent correspondent wrote that “[r]egional accents are wonderfully enriching contributions to the English language – but an accent becomes degenerate when the spoken word cannot differentiate between totally different meanings. Too often we hear of people crossing on the Cook Strait ‘fairy’ and flying ‘Ear’ New Zealand” (Bravery 2001).

Such comments reflect a natural concern that the loss of an earlier distinction will result in confusion, as illustrated in another example from a New Zealand newspaper (Veysey 1994). Here a scenario is constructed in which a barmaid obligingly removes her clothing when a customer asks “Beer, please”. The writer of this piece assumes that the barmaid is one of an increasing number of speakers of New Zealand English who do not distinguish between the [iː] and [eɪ] vowels, and that she thus hears [biː] as “bare”. Importantly, though, she is assumed not only to have heard “beer” as indistinguishable from “bare”, but also to have interpreted the word in a way that would appear inappropriate given the probable constraints of context. There are of course many cases in English as in other languages where, to quote one of the correspondents above “the spoken word cannot differentiate between totally different meanings”, as long as, that is, contextual constraints are ignored. We regularly encounter in our every day use of language both homophones (lexically distinct words with the same pronunciation, such as write-right or river bank – (financial) bank) and polysemes (words which have distinct but related senses, such as point (of an argument, of a headland)). It is relatively unusual, however, for such ambiguous forms to result in confusion or miscommunication. Psycholinguistic research on word recognition and interpretation suggests that this is because the multiple meanings of an ambiguous word can be accessed automatically on encountering the word, with the correct meaning then being selected via a rapid assessment of the accessed meanings against the developing interpretation of the utterance (Swinney 1979; Onifer and Swinney 1981). Moreover, if the context is suitably constraining, and/or if one meaning is particularly dominant, then it is possible that just one meaning – the one which is most appropriate or likely – will be selectively accessed (Duffy et al. 1988). It would thus seem quite unlikely on the basis of experimental psycholinguistic evidence, as well as on common-sense grounds, for the above-mentioned barmaid to be confused about the request for [biː], even if “beer” and “bare” are truly homophonous.

There is however another aspect to ongoing changes such as the so-called “EAR/AIR merger” in New Zealand English (NZE) that makes them of great interest to psycholinguistic research in speech perception and word recognition. This is the fact that such changes, by virtue of being changes-in-progress, are incomplete across the population as a whole. For instance, the NZE speech community contains speakers for whom the merger of the EAR and AIR diphthongs is complete, and others who maintain a consistent distinction between the two, as well as those who are inconsistent in their
realisations. As is explained in our review of this particular merger in the next section, the distinction between EAR and AIR is maintained by older speakers and is largely lost amongst younger speakers of NZE. This means that members of this speech community may be in a state of some uncertainty about the signal value of a particular diphthong, uncertainty which may in fact only be resolvable by appeal to demographic information (e.g., the age of the speaker). Additionally, the merger in NZE appears to be asymmetric, towards EAR. Taken together with the demographic spread of the merger, this asymmetry may mean that younger listeners treat forms with the [iə] diphthong as homophones, so that [bɪə] is either beer or bare, whereas older listeners are more inclined to treat [bɪə] as beer and [bɛə] as bare. This and other issues are addressed in the experimental section of this paper.

PRODUCTION STUDIES OF THE EAR/AIR MERGER

One of the most recently published reports on this sound change is Gordon and Maclagan (2001), who provide data from a survey being repeated every 5 years amongst 14-15 year olds in Christchurch. The data come from words containing EAR or AIR vowels read both in sentences and in word lists, and reveal that the diphthongs, both still widely present in the initial survey in 1983, have become almost completely merged on EAR. A comparison of two age groups recorded in 1994 (Maclagan and Gordon 1996) confirms a more complete merger for younger speakers (20-30 years old) than for older speakers (45-60 years old), with the shift again in the direction of EAR. These studies also show that there is a great deal of variability in the diphthongs of many speakers, with some speakers showing no clear pattern of merger towards either EAR or AIR. For instance, some 62% of students recorded in 1983 showed this kind of variability (while 13% of speakers kept the diphthongs distinct and 25% collapsed the distinction in one direction or the other). The pattern of realisation of these vowels over the time period of the Canterbury survey is described (Maclagan and Gordon 1996: 144-5) as part of the chain-shift raising of the short front vowels of NZE (whereby the vowel of pat has raised to the position previously taken by that of pet, and pet to pit, etc.). Thus the starting point of the AIR vowel has also been raised towards that of EAR. Gordon and Maclagan (2001:232) conclude that the change is most likely a “merger of approximation” rather than a “merger of expansion” (Labov 1994:321), i.e. the two sounds are collapsing on a single form, in this case the closer EAR pronunciation, rather than continuing to use the whole range of pronunciations previously available to both EAR and AIR. It should also be noted that there is some evidence from the early sample, and from work by Holmes and Bell (1992), of a merger towards AIR, albeit short-lived and more noticeable amongst certain speaker groups. This is explained as hypercorrection, i.e. the merger on EAR became stigmatised (perhaps because it was seen as representative of speakers from lower socio-economic groups) and more conservative speakers responded by moving some of their EAR vowels towards AIR. A further claim with regard to the overall merger on EAR is that it has progressed through NZE by a process of lexical diffusion, i.e. it has affected some words before others, and has then spread through the inventory of relevant words (Maclagan and Gordon 1996: 131-133).

The picture, then, is of a sound change that is progressively more complete as we move from older to younger speakers, and which is asymmetric in that the merger is towards one of the two forms distinguished by the older speakers. Thus all NZE speakers are still likely to hear both vowels (also through hearing other varieties through the media etc.), but many younger speakers may produce only the [iə] vowel sound in both EAR and AIR words. What is of interest is how this state of affairs may affect the ability to accurately recognise words containing these vowels. If the merger were complete across all speakers, then words like beer and bare or cheer and chair would be homophones, just like right and write or bank (the financial institution) and bank (the edge of the river), and presumably would therefore be recognised in similar manner. However, since the EAR/AIR merger is not complete in NZE, we might expect the processing of words like cheer and chair to differ somewhat from that of true homophones.

EXPERIMENT

The current study employs a semantic priming paradigm to investigate the recognition of EAR and AIR words in NZE. The primes are members of the near-homophone pairs (i.e., beer or bear, fear or fair, rear or rare etc.) and the probes are words associated in meaning to one member of the pair (such as shout or sit associated to the cheer/chair word pair). In a population where the EAR and AIR
words are totally distinct, i.e. have separate pronunciations and separate mental representations, then (for example) one would expect [t\textipa{i}\textipa{a}] (the phonetic form for cheer with an EAR vowel) would prime a semantic associate such as shout, and [t\textipa{i}\textipa{a}] (that is, chair with an AIR vowel) would prime sit, but that there would be no clear priming (other than some that may result from phonetic overlap of the two forms) of sit by [t\textipa{i}\textipa{a}] or of shout by [t\textipa{i}\textipa{a}]. This situation is illustrated in panel a of Figure 1. (Square brackets are used in this figure to indicate a pronunciation form, and braces to indicate a word in the listener’s mental dictionary.) However, if the merger on EAR is complete and cheer and chair have the homophonous form [t\textipa{i}\textipa{a}] then hearing this form will lead to the retrieval of both words from the mental dictionary (as in Figure 1d). In this case, we might expect the presentation of the homophone in an unbiased context to result in equal priming of a word related to the cheer meaning (e.g. shout) and of one related to the chair meaning (e.g. sit). What then is the situation for a merger-in-progress? If the merger means that two forms are not clearly distinguished form one another, then we might expect both [t\textipa{i}\textipa{a}] and [t\textipa{i}\textipa{a}] to prime both shout and sit, as in Figure 1b. Alternatively, given the instability of the merger pointed out in the discussion of the production data, we might predict a quite different result – while younger speakers as an isolated group might treat words like cheer and chair as homophones, the presence of non-merged forms amongst older speakers means that this might not always result in successful recognition. Given the asymmetry in the merger, we might expect that when listeners hear an EAR form they retrieve both EAR and AIR words, but that when they hear an AIR form they access only the AIR word, as illustrated in Figure 1c. These possibilities, based on possible historical developments of the merger, are examined in the experiment reported below.

![Figure 1](image-url)  
Figure 1 – hypothetical mappings from phonetic input onto lexical form for different states of merger of EAR and AIR (see text for details).

Method

The overall design consisted of 4 test conditions and 2 control conditions, as in Table 1. As the examples show, in appropriate conditions the prime is followed by a semantically related probe, and in inappropriate conditions the prime is followed by a probe to which it is not semantically related. (These relationships are of the probe to the relevant prime in an unmerged system.) Control conditions allow us to obtain a base-line response time for each probe word, relative to which we can measure facilitation of the test words.

![Table 1](table-url)  
Table 1. Conditions used in semantic priming experiment, with illustrative examples of materials.

Materials

12 monosyllabic EAR/AIR word pairs were selected on the basis of similarity of relative frequency and familiarity. For each of these 24 EAR/AIR words a semantic associate was selected using the Edinburgh Associative Thesaurus, the Florida Free Association Norms (see reference list for URLs) and the Birkbeck Association Norms (Moss and Older 1996). The resulting test materials (see first
four lines of Table 1) were arranged in four lists, so that each list contained a prime from each EAR/AIR pair, immediately followed either by its semantic associate (an appropriate probe) or by the semantic associate of its paired EAR/AIR word (an inappropriate probe). Conditions were counterbalanced across lists so that each list had an equal number of items (i.e. six) in each of the four test conditions. In addition, 24 control primes were selected, semantically unrelated to the probes, and these were included in a separate list, so that each probe word was immediately preceded by its unrelated prime. Each test list also included 20 real word fillers, as well as 44 non-words constructed by altering one phoneme in existing words – e.g., jatt from cat. As well as the control prime-probe combinations (48 items), the control list contained 26 real word fillers and 74 non-words. The word and non-word fillers reduced the number of sequences of unrelated words, reducing expectations that a word would be followed by another semantically related word. All items were monosyllabic; the real words were approximately matched for frequency.

Procedure

The 16 subjects were linguistic undergraduates at Victoria University who volunteered as a part of their course of study. They were all fluent native NZE speakers, aged 18-25. All subjects were tested individually and heard each of the lists described above, with list order rotated across subjects. The experiment was presented over two sessions, one week apart. Two lists were presented in each session, separated by unrelated tasks to reduce priming effects across the session. The reason for presenting all items in each condition to every subject was to allow comparison of responses to items in the four conditions to the same subjects’ own productions of the test words. In effect, the design appeared unable to eliminate priming effects between lists within sessions, and so the results below are for the first list only in each session. The experiment was administered using Psyscope (Cohen et al. 1993). In each session, subjects first read the experiment instructions, then listened over closed-ear headphones to each item. They indicated whether the item was a real word or a non-word by pressing one of two buttons on a millisecond response timer. They were encouraged to respond as quickly as possible while not sacrificing accuracy. The interval between the onset of each stimulus was 1400 milliseconds, which had been determined as being long enough to allow subjects to respond, but short enough to maintain a fast response rate.

Results

Response times (RTs) for each item were recorded in milliseconds. One item set – fear/fair – was excluded due to an error in the experiment design. Missing data and incorrect responses were excluded from the analysis, affecting 3.69% of the test and control data. Facilitation times (FTs) were calculated by subtracting a subject's RT for a probe in a test condition from that subject's RT to the same probe in the control condition. Average FTs in each of the four test conditions were calculated for each subject, and the resulting averages were subjected to Analysis of Variance, with Facilitation Time as the dependent variable and Prime (EAR word vs AIR word) and Probe (appropriate or inappropriate – see Table 1) as independent variables. In addition, planned comparisons of probe types following each kind of prime were used to test the predictions illustrated in Figure 1 above. Overall averages for the four conditions are shown in Figure 2.

The statistical analysis confirmed that appropriate probes received greater priming than inappropriate probes (F(1,15)=5.85, p <0.03). In addition, there was a significant interaction of prime and probe type (F(1,15)=4.90, p<0.05). In the planned comparisons, AIR primes facilitated appropriate probes more than inappropriate probes (chair-sit > chair-shout; F(1,30)=6.10, p <0.02), while there was no difference in the priming of appropriate and inappropriate probes by EAR primes (cheer-shout = cheer-sit) (F(1,30)=0.23, p >0.60).

DISCUSSION

The results of the Analysis of Variance show strong support for our predictions of an asymmetry in recognising the EAR and AIR vowels, consistent with the main trends shown in the production studies reviewed above, and supporting the hypothetical position displayed in panel c of Figure 1. In other words, on hearing words with an [ɔː] vowel, our young subjects accessed both EAR and AIR words, just as the different meanings of bank are both automatically accessed on hearing the homophone
(Swinney 1979). At the same time, however, the [eә] form leads to access of the AIR word (and not the EAR word), reflecting the fact that the phonetic [eә] form is still heard in the subjects’ environment.

Figure 2. Average facilitation times (in milliseconds, relative to the control condition) for appropriate and inappropriate probes following ear and air primes (eg., from left to right in the figure, cheer-shout; cheer-sit; chair-sit; chair-shout).

Note though from Figure 2 that the level of facilitation for both probe words following the EAR prime is less than that found for the appropriate probe after the AIR prime. A possible explanation of this result, which requires further experimental exploration, is in terms of recent findings by Rodd and her colleagues (Rodd et al. 2002). In a series of lexical decision experiments with both visual and auditory stimuli, these researchers find that while polysemous words with multiple senses (such as slide) are responded to more rapidly than those with few senses (shirt), homophonous words (bark) show a processing disadvantage relative to unambiguous words. This leads them to propose that the processing of ambiguous words involves competition between distinct lexical representations. In our experiment, this would be competition between {cheer} and {chair} on the basis of the phonetic input [tʃiә] (Figure 1c). Note though that compared to the control condition there is still facilitation of probe words related to both meanings, so that any disadvantage is relative to the priming of unambiguous words, rather than a claim that uncertainty about the lexical identity of [tʃiә] results in inhibition of lexical representations.

In our future research into the EAR/AIR diphthongs in NZE, we will also explore the possibility that older subjects, who feature both vowels more consistently in their productions, do not show the same asymmetry in the experiment, or at least show a smaller effect. In addition, we wish to test whether age information about the speaker will affect the interpretation of EAR and AIR forms by subjects in the response time task. In other words, will subjects who are given different age information about the speaker adjust their interpretation of the same phonetic forms in a way that reflects the attested age-related differences in the extent of the merger?

Finally, as noted earlier, our experimental sessions also involved the collection of production data from our participants, as well as discrimination scores for EAR/AIR pairs. Further investigation of these sources of data will explore the relationship between subjects’ own phonetic realisations of EAR and AIR words and their interpretation of the experimental materials, and also the issue of whether different lexical pairs are responded to in a way that reflects the relative distinctions made between them in production.
ACKNOWLEDGEMENTS

The semantic priming study reported in this paper was an honours research project undertaken by the second author under the supervision of the first author. We acknowledge help and advice received from Joel Zwartz and Amy Austin, and financial support from Victoria University of Wellington to the first author under grant number URF1/11.

REFERENCES


URLs

Edinburgh Associative Thesaurus: http://monkey.cis.rl.ac.uk/Eat/htdocs/eat_old.html

Florida Free Association Norms: http://w3.usf.edu/freeassociation