# Goldilocks and the three beers: sound merger and word recognition in NZE

Megan Rae and Paul Warren[1]

Victoria University of Wellington

> Regional 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.
>
> L. Bravery, letter to the editor, *Listener* June 9th 2001.

Comments such as the above are not unusual, and reflect a natural concern that speakers have when confronted with changes in the pronunciation of their language. When the change results in the loss of an earlier distinction, a common worry is that this will affect communicative ability in the language:

> To use the [...] examples of 'beer' and 'bare' and 'here' and 'hair', I go into this bar and say, 'Beer, please' and the barmaid, being an obliging girl, takes off her top and bra. Because I am devoutly decent, I say indignantly, 'Here! Here!' and the barmaid who knows when enough's enough whacks me with a jug of Old Dark which starts a bloody brawl.
>
> You see, the potential for misunderstanding is substantial and the consequences may be horrendous.
>
> Alex Veysey — Opinion column, *Evening Post*, October 29th 1994

Of course, this type of remark raises the question of whether communication is in fact increasingly impaired as the EAR and AIR diphthongs become merged. There are many words in the English language (and not just in New Zealand English) that are already ambiguous – homophones (two words with the same form) such as *bank*, or *write/right*. Such words are usually clear from context, and so do not seem to cause lasting problems for the listener. Why then should there be a problem with homophonous pronunciations of *beer* and *bare*? Are not the meanings of the key words in the extracts above sufficiently clear from their contexts?

In this article we present results from psycholinguistic experimentation that suggest that one reason why the EAR/AIR cases are importantly different from words like *bank* and *write/right* (over and above, that is, any emotional response to observed changes in the pronunciation of one's language) lies precisely in the fact that this is a change-in-progress. That is, not all speakers of New Zealand English (NZE) have merged EAR and AIR pronunciations. Additionally, this change-in-progress is asymmetric, because the sounds are not being merged on some ambiguous form equidistant between EAR and AIR, but are more clearly moving towards an EAR

[1] Please address all correspondence to paul.warren@vuw.ac.nz. The order of authors is alphabetical. The study reported in this paper was an honours research project undertaken by the first author under the supervision of the second author. We acknowledge help and advice received from Joel Zwartz, Amy Austin and Jen Hay, and financial support from Victoria University of Wellington under grant number URF1/11.





pronunciation. This combination of instability and asymmetry has implications for the processing of EAR and AIR words, as we will show.

## The progress of the merger

First, some brief comments on the progress of the EAR/AIR merger in NZE. One of the most recent published reports on this sound change is Gordon & Maclagan (2001), who provide data from a survey being repeated every 5 years amongst 14-15 year old college students in Christchurch. The data come from words containing EAR or AIR vowels read both in sentences and in word lists[2], and show quite clearly 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 & 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. Gordon and Maclagan's 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). There is also some evidence from the early sample, as also from work by Holmes and Bell (1992), of a merger towards AIR, albeit short-lived and more noticeable amongst certain speaker groups. The pattern of realisation of these vowels over the time period of the Canterbury survey is described (Maclagan & 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 & 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. The fact that some speakers for a short while appeared to merge on AIR 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 this sound change 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 & Gordon 1996: 131-133).

[2] While read (rather than spontaneous) materials are not ideal, Gimson (1963: 143), referring to the original study by Fry (1947), lists the vowels EAR and AIR as only the 17th and 18th most frequent out of 20 English vowels. (Not 15th and 16th of 18 vowels as indicated by Gordon & Maclagan, 2001: 219.) It therefore becomes necessary to use read materials in order to elicit sufficient tokens for analysis.

## Sound merger and word recognition

Overall, the situation with regard to these diphthongs is that the merger towards EAR is complete for younger speakers, but that older and more conservative speakers still distinguish the vowels. 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 EAR 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 *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. The next section presents some of the relevant psycholinguistic research on the recognition of homophones. 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. We return to this issue later.

## Psycholinguistic research on the processing of ambiguous words

On-line (response time) studies of lexical ambiguity show that although homophones may be ambiguous at first, context will be quickly used to select the appropriate meaning. For instance, Swinney (1979) presented subjects with spoken sentences containing ambiguous words, such as 'the man was not surprised when he found several *bugs*$_{(1)}$ in the corner of his room'.[3] At the offset of the ambiguous word (i.e. at point $_{(1)}$ in the example), the subjects were shown a word on a visual display. This word was either related to one of the meanings of the ambiguous word (e.g. *ant* or *spy*), or was unrelated (*sew*). It was found that subjects recognised the related words equally fast, and more rapidly than they recognise the unrelated word.

The faster recognition of the related word in Swinney's experiment reflects what is known as 'semantic priming' — the meaning of one word primes or facilitates the recognition of another word. Semantic priming was earlier demonstrated by Meyer and Schvaneveldt (1971) in a lexical decision experiment with visually presented words. In this task, participants have to indicate — by pressing one of two buttons — whether a stimulus is or is not a word of their language. Meyer & Schvaneveldt (1971) showed that the identification of a target word (referred to as a *probe*) is made easier if the word immediately preceding it (referred to as a *prime*) is related in meaning. For example, *nurse* will have a faster response time (RT) if it is preceded by a word related in meaning such as *doctor* than if it is preceded by a word unrelated in meaning such as *butter*. In other words, *doctor* is said to prime *nurse* (Harley 1995).

In his use of semantic priming, Swinney (1979) also tracked the time-course of listeners' access to the meanings of ambiguous words. In another part of his study, subjects heard the sentence: 'the man was not surprised when he found several spiders, roaches and other *bugs*$_{(1)}$ in the corner$_{(2)}$ of his

[3] Technically, *bugs* is probably a polyseme (a word with more than one related sense) rather than a homophone. However, the general treatment of ambiguous words is what is at issue here.



room', where the context biases interpretation towards one of the meanings of *bugs*. The words *ant*, *spy* and *sew* were presented (to different participants) at both points $_{(1)}$ and $_{(2)}$. At point $_{(1)}$, both *ant* and *spy* were recognised quickly in comparison to *sew*. However, at point $_{(2)}$, only *ant* was recognised quickly, and both *spy* and *sew* were slow to be recognised. What this study shows is that when we come into contact with a homophone, no matter whether the context is ambiguous or not, all meanings of the homophone are initially accessed. The context then quickly forces a decision between the meanings and the appropriate one is selected.

## Priming and change-in-progress

The current study employed a similar psycholinguistic paradigm to that used by Meyer and Schvaneveldt to investigate the recognition of EAR and AIR words in NZE. Whereas Meyer and Schvaneveldt used visually presented words, the experiment reported here uses spoken words, both as primes and as probes. 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. separate pronunciations and separate mental representations, then (for example) one would expect that [tʃiə] (the phonetic form for *cheer* with an EAR vowel) would prime a semantic associate such as *shout*, and [tʃeə] (that is, *chair* with an AIR vowel) would prime *sit*, but that there would be no clear priming of *sit* by [tʃiə] or of *shout* by [tʃeə][4], as illustrated in panel a. of Figure 1 below. (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 *cheer* and *chair* are homophonous and both pronounced [tʃiə], then hearing this form will lead to the retrieval of both words from the mental dictionary (as in Figure 1b). 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*). Finally, 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 will retrieve both EAR and AIR words, but that when they hear an AIR form they will access only the AIR word, as illustrated in panel c of Figure 1. This possibility is examined in the experiment reported below.

[4] There may, of course, be a certain amount of priming in these latter conditions, simply because [iə] and [eə] are similar forms and may prime one another on the basis of their phonetic overlap, even in a dialect without EAR/AIR merger.

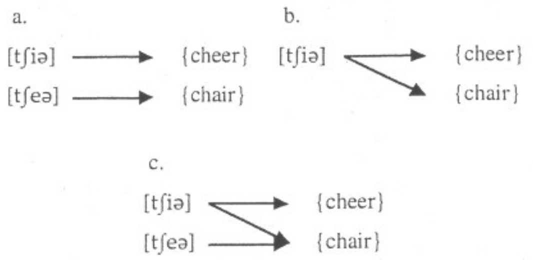



*Figure 1 – hypothetical mappings from phonetic input onto lexical form for different states of merger of EAR and AIR (see text for details).*

## Method

The overall design consisted of 4 experimental conditions and 2 control conditions:

EAR prime followed by EAR probe (ear_EAR – e.g., *cheer, shout*)

AIR prime followed by AIR probe (air_AIR – e.g., *chair, sit*)

EAR prime followed by AIR probe (ear_AIR – e.g., *cheer, sit*)

AIR prime followed by EAR probe (air_EAR – e.g., *chair, shout*)

Control prime followed by EAR probe (control_EAR – e.g., *bee, shout*)

Control prime followed by AIR probe (control_AIR – e.g., *toe, sit*)

The four test conditions form two main conditions used for the analysis: *appropriate*, where the prime is followed by a semantically related probe (i.e., ear_EAR and air_AIR), and *inappropriate*, where the prime is followed by a probe to which it is not semantically related (i.e., ear_AIR and air_EAR).

The test materials consisted of 12 EAR/AIR word pairs, giving 24 primes, for which there were 24 semantic associates used as probes. There were also 24 control primes, which were semantically unrelated to the probes. The purpose of these primes was to give a 'baseline' response time for each probe. In addition, the subjects heard further sequences of unrelated words, so that they did not have any strong expectation that a word would be followed by another semantically related word. All the words used were monosyllabic and approximately matched for frequency. The experiment also included as many non-words as there were real words. The non-words were made by altering one phoneme in existing words — e.g., *jatt* from *cat*.

Published association norms were used to find the semantic associates to be used as probes for each of the EAR/AIR primes.[5] We ensured that the associates, like the other test words, were monosyllabic, and also that the two associates for each pair of EAR/AIR words were well matched for lexical frequency.

[5] Three sets of association norms were consulted: the Edinburgh Associative Thesaurus, the Florida Free Association Norms (see reference list for URLs) and the Birkbeck Association Norms (Moss & Older 1996).



## Procedure

The 16 subjects were all linguistic undergraduates at Victoria University who volunteered as a part of their course of study. They were all fluent native NZE speakers.

The experiment was presented over two sessions, one week apart. The two control conditions above were combined into one block. The four test conditions were distributed over four blocks to get an even number of conditions in each block, following the pattern below:

| Word pair | Block 1 | Block 2 | Block 3 | Block 4 |
|---|---|---|---|---|
| *beer/bear* | ear_EAR | air_AIR | ear_AIR | air_EAR |
| | *beer, wine* | *bear, hug* | *beer, hug* | *bear, wine* |
| *fear/fair* | ear_AIR | air_EAR | air_AIR | ear_EAR |
| | *fear, ground* | *fair, fright* | *fair, ground* | *fear, fright* |
| *spear/spare* | air_AIR | ear_EAR | air_EAR | ear_AIR |
| | *spare, tyre* | *spear, sword* | *spare, sword* | *spear, tyre* |
| *rear/rare* | air_EAR | ear_AIR | ear_EAR | air_AIR |
| | *rare, end* | *rear, steak* | *rear, end* | *rare, steak* |

All 16 subjects heard each block, but with the test blocks in different orders, over the different sessions. The first session consisted of a practice block, two test blocks (either 1 and 2 or 3 and 4 in the table above) and the control block. The second session consisted of a reduced practice block and the other two blocks that the subject had not already heard.

Each block was separated by a filler task that was not related to the aim of the experiment. The overall design attempted to break up the blocks as much as possible to reduce *long-term priming* effects. For example, on hearing *cheer* then *shout* in a task and subsequently hearing *chair, shout* later on in the task, the second occurrence of *shout* may result in a faster reaction time not because *chair* primes it, but because the word {shout} is still activated from the first *shout* (i.e., *shout* is priming itself over the course of the experiment).

The experiment was administered using Psyscope (Cohen et al. 1993). Subjects sat in front of a computer with headphones and read the instructions before responding accordingly.

## Predictions

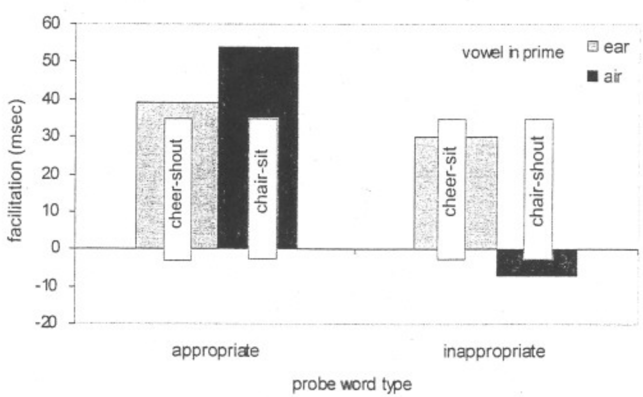
Most recent evidence on the EAR/AIR merger suggests that it is proceeding towards the closer EAR variant (see Rae & Warren, in prep., for evidence concerning Wellington NZE). Older speakers still distinguish EAR and AIR words. The young speakers making up our subject group will have exposure both to speakers who merge the vowels and to those who keep them distinct. We therefore predict that the lexical decision experiment will result in asymmetric priming, reflecting the hypothetical situation in panel c of Figure 1 above.

## Results

Response times (RTs) for each item were recorded in milliseconds (ms). Initially, one EAR/AIR set was excluded — *fear/fair* — due to an error in the experiment design. Missing data and incorrect responses were excluded from the analysis, affecting 3.69% of the data.

To further avoid long-term priming effects, test probe RTs from the second block of each session were discounted.[6] For the remaining data, facilitation times (FTs) were calculated by taking the difference between probe RTs in the test and control conditions. For example, if a participant had RTs for *shout* of 1028 ms when it was preceded by *bee* (control_EAR) and 788 ms when preceded by *cheer* (ear_EAR), then the facilitation time (i.e. the relative increase in response speed due to priming by *cheer*) would be the difference between the two, i.e., 240 ms. Average FTs in each of the four test conditions were calculated for each subject and again for each word pair, and the resulting two sets of averages were subjected to the usual type of statistical analysis used with this kind of data.[7] Overall averages for the four conditions are shown in Figure 2.

Figure 2: facilitation of probe words in test conditions
(i.e. control RT-test RT)




The statistical analysis confirmed that 'appropriate' probes received greater priming than 'inappropriate' probes, but also that the 'inappropriate' ear_AIR

[6] These showed a dramatic levelling-out of differences between conditions compared to the first block. We had initially decided to obtain response time data in this way for all subjects in all conditions for each item, since we were interested in comparing their performance in this task to that in a discrimination task and with the same subjects own production data which were collected at the end of the second session (Rae & Warren in prep).

[7] These were subject and item ANOVAs (Analyses of Variance), with Facilitation Time as the dependent variable and Prime (EAR word vs AIR word) and Probe (appropriate or inappropriate as a semantic associate of the prime) as independent variables. All results reported in this paper as significant were so at the level of $p<0.05$. Details of the statistical analysis are provided in (Rae & Warren in prep).



condition (*cheer, sit*) resulted in more priming than the 'inappropriate' air_EAR condition (*chair, shout*), and not significantly less priming than ear_EAR (*cheer, shout*).

## Discussion

These results 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 above. In other words, on hearing words with an [iə] vowel, our young subjects access both EAR and AIR words, just as the different meanings of *bank* are both automatically accessed on hearing the homophone. At the same time, however, the [eə] form leads to access of the AIR word (and not the EAR word), reflecting the fact that the phonetic [eə] form is still heard in the subjects' environment.

Note though from Figure 2 that the level of facilitation for both probe words following the EAR prime is less than that found in the air_AIR condition. There are at least two accounts that might be suggested for this situation. One account assumes that there is a fixed amount of processing attention available, so that when a participant hears [tʃiə], this processing attention has to be shared between the *shout*-related and the *sit*-related meanings, resulting in less priming for each of these words than for *sit* when primed by [tʃeə]. If this is typical finding for homophones then it suggests that such ambiguous words result in a processing disadvantage. However, under an alternative account, there is no such limitation on the overall amount of processing attention, and both meanings of a homophone (including new homophones such as [tʃiə]) are accessed as fully as the individual meanings of unambiguous words (which are probably in any case rather the exception than the rule). The difference in priming then results from subsequent competition between the two accessed words, so that there is — over time — a reduction in the facilitation of the semantically related words. We plan to investigate these possible accounts in further research.

Our further research 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 (see footnote 6), 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 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 the production data (Rae & Warren in prep).

## References

Cohen, Jonathan D., Brian MacWhinney, Michael Flatt & Jefferson Provost 1993. Psyscope: A new graphic interactive environment for designing psychology experiments. Behavioural Research Methods, Instruments & Computers 25(2): 257-271.

Fry, Dennis B. 1947. The frequency of occurrence of speech sounds in Southern English. Achives Néerlandaises de Phonétique Expérimentale XX: 103-6.

Gimson, A. C. 1963. An Introduction to the Pronunciation of English. London: Edward Arnold.

Gordon, Elizabeth and Margaret Maclagan 2001. 'Capturing a sound change': A real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. Australian Journal of Linguistics 21(2): 215-238.

Harley, Trevor A. 1995. The Psychology of Language. Hove: Psychology Press.

Holmes, Janet & Allan Bell 1992. On shear markets and sharing sheep: The merger of EAR and AIr diphthongs in New Zealand English. Language Variation and Change 4: 251-273.

Labov, William 1994. Principles of linguistic change. Oxford: Blackwell.

Maclagan, Margaret & Elizabeth Gordon 1996. Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. Language Variation and Change 8: 125-147.

Meyer, David E. & Roger W. Schvaneveldt 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology 21: 863-883.

Moss, Helen E. & Lianne Older 1996. Birkbeck Word Association Norms. Hove, UK: Lawrence Erlbaum.

Rae, Megan & Paul Warren in prep. The asymmetrical change in progress of EAR and AIR vowels in NZE: psycholinguistic evidence. Wellington Working Papers in Linguistics.

Swinney, David A. 1979. Lexical access during sentence comprehension: (Re) consideration of context effects. Journal of Verbal Learning and Verbal Behaviour 18: 645-659.

## URLs

Edinburgh Associative Thesaurus: http://www.cis.rl.ac.uk.proj/psych/eat/eat

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

*NZEJ* 16 (2002) 42-52