FROM BAD TO BED:
THE RELATIONSHIP BETWEEN PERCEIVED AGE AND VOWEL PERCEPTION IN NEW ZEALAND ENGLISH[1]

Katie Drager: *Department of Linguistics, University of Canterbury (Private Bag 4800, Christchurch, New Zealand) <kkd14@student.canterbury.ac.nz>*

Abstract

Recent research has provided evidence of a link between social information and speech perception. This study investigates the relationship between vowel perception and age in New Zealand English, and it demonstrates that perceivers may use the apparent age of a voice to identify vowels in the context of a change in progress.

1. Introduction

It has been known for some time that visual stimuli, such as movement of the lips and jaw, can affect how sounds are perceived (McGurk and McDonald 1976), but recent studies suggest that how an individual perceives a word is also influenced by social information. Niedzielski (1999) found that the expected dialect of a speaker could influence the perception of a vowel, and research by Johnson, Strand, and D'Imperio (1999) suggests that vowel perception can be influenced by apparent gender. This paper examines the extent to which the social factor of age can influence the perception of vowels in the context of a chain shift in New Zealand English.

New Zealand English provides a promising test bed in which to investigate the effects of perceived age on vowel perception because the short front

 © Linguistic Society of New Zealand (inc.)

vowels, which will be referred to using the lexical set labels TRAP, DRESS, and KIT (Wells 1982), have taken part in a push chain that appears to be continuing, so that TRAP and DRESS are raised for younger speakers (Maclagan and Hay, forthcoming; Gordon, Campbell, Hay, Maclagan, Sudbury and Trudgill 2004).

Previous work on speech perception in New Zealand English has largely focused on speakers' attitudes toward different dialects (Bayard 1990). Bayard (1990) inspected New Zealanders attitudes toward their own dialect as well as other dialects of English by examining perceived social characteristics. This study takes Bayard's work a step further by investigating the relationship between perceived social characteristics and the perception of vowels in New Zealand English. Specifically, if speakers use their social expectations of age to determine which vowel was heard, a vowel could be perceived differently depending on the apparent age of a voice because the vowels of older speakers are different than the vowels of younger speakers. Are speakers of New Zealand English either consciously or subconsciously aware that the vowels of older speakers differ from those of younger speakers? And could their expectations regarding the age of a speaker determine how they hear a vowel?

2. The Experiment

The objective of the experiment was to shed light on whether different individuals hear vowel boundaries in different places as well as to determine the extent to which the social factor of age can influence speech perception. In order to test the hypothesis that age could influence how a vowel was perceived, a vowel continuum was developed. The continuum was designed so that all speakers of New Zealand English would agree that the token on one end of the continuum was DRESS and the token on the opposite end was TRAP, while the differences between the tokens connecting the two opposite ends of the continuum were so small that, potentially, the vowels might be perceived differently by different people.

2.1 The stimuli

Four voices were resynthesized from previously recorded words from wordlist data in the Canterbury Corpus, recordings in the Origins of New Zealand English archives held at the University of Canterbury, Christchurch.[2] Vowels from the four voices, two male and two female, were resynthesized using the program Klattworks, parameter-manager software that allows a user to independently set different parameters of a sound file, allowing for intricate manipulation of sounds (McMurray, in preparation).[3]

The vowels which were used during resynthesis had previously been spliced from the word *bad* using Praat (Boersma and Weenink 2004) and were then transported into Klattworks in order to be resynthesized. A 10-step continuum was created for the vowels in *bed* and *bad* using the Klattworks command 'Setto'. The continuum between DRESS and TRAP was created by manipulating the first and second formant values (F1 and F2), so that the values of F1 and F2 differed for the male and female voices but were identical for the two male voices and for the two female voices. The formants were set so that they had a flat formant trajectory in the centre of the vowel, but the first 15 ms of F2 was designed to rise by 200 Hz and the last 10 ms of F2 was designed to lower by 100 Hz. The formant transitions were intended to clarify the perception of the initial burst /b/ and final release /d/ which were attached to the vowels following resynthesis. The result of the resynthesis was a continuum ranging from the word *bad* to the word *bed* for four different voices.

It was discovered after the experiment was run that the formant values of the female voices were in fact not equal to each other, perhaps due to the low resampling rate of 10,000 Hz assigned to all of the voices after resynthesis. This discrepancy in the stimuli, along with the fact that when questioned two weeks following the study some participants indicated that they had thought one of the female voices was male, has complicated the data set regarding the females so the results for the female voices will not be discussed in this paper. Instead, the paper concentrates on the results from the male voices which were less problematic.

The formant values of the most DRESS-like token and the most TRAP-like token were loosely based on the values presented in Maclagan (1982), and are presented in Table 1 for the male voices.

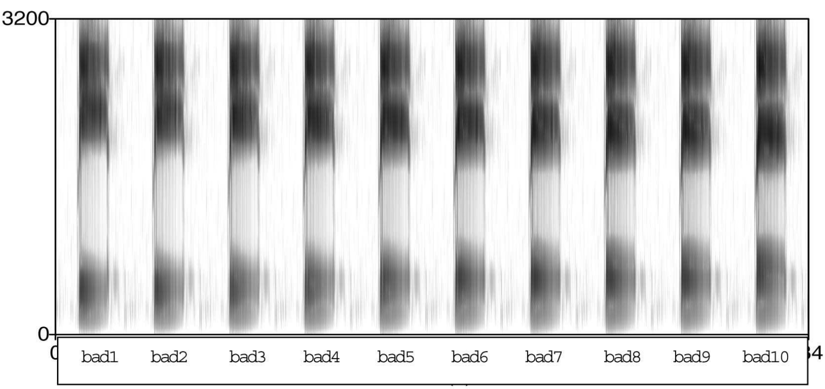
The continuum was created so that between the different tokens, formant values are evenly spaced. Perceptually, the adjacent tokens are very similar to each other. The spectrogram of the full continuum for one of the male voices, Troy, is presented in Figure 1.

In the most DRESS-like token (*bad1*), the values of F1 and F2 are further apart from each other than in the most TRAP-like token (*bad10*), and the intermediary steps link these two extreme ends in a gradually shifting continuum.

Table 1. Formant values for F1 and F2 for male voices.

TOKEN	FORMANT VALUES FOR MALES F1	F2
1	409.77	2234.1
2	415.31	2211.32
3	422.89	2141.65
4	440.63	2118.25
5	486.09	2095.7
6	513.79	2070.96
7	521.58	2043.99
8	527.59	2019.74
9	546.01	1995.74
10	590.51	1971.54

Figure 1. Spectrogram for the voice of Troy ranging from the most DRESS-like token (bad1) to the most TRAP-like token (bad10).



Because some research indicates that isolated vowels are identified less accurately than vowels in CVC contexts (Gottfried, Jenkins, and Strange 1985), vowels were played within a word rather than alone. The frame /bVd/ was used for this study because both of the tokens, *bed* and *bad*, are frequently used words with which all participants would be familiar (Johansson and Hofland 1989; Thorndike and Lorge 1944) and would therefore most likely have heard pronounced by a number of different speakers in various age groups. Additionally, the /bVd/ frame was used because a voiced final consonant would provide a longer vowel (Kluender, Diehl, and Wright 1988), and a longer vowel may help participants decide which sound they perceived.

2.2 The participants

A total of twenty-five linguistics students participated in this study, nineteen of whom were from New Zealand. Only responses of these 19 New Zealanders were analyzed for the results presented here. Of the participants, six were male and thirteen were female, with ages ranging from eighteen to thirty-six years of age.

The experiment was run in two separate groups. The first group of participants were recruited from Ling203, a second year sociolinguistics paper at the University of Canterbury, and they were not informed that the words they would hear were resynthesized speech. The second group of participants were other linguistics students who were aware that the voices had been modified.[4] The results from the two groups are similar and will be presented together.

2.3 Experiment Design

The stimuli were arranged in an order that varied the token number but maintained the speaker order. The 40 tokens, 10 for each voice, were played to the room of participants over computer speakers so that participants heard a number and then a token. The tokens were played in a fixed order which alternated between male and female voices. Participants heard each token twice and were given approximately two seconds before being expected to answer. They circled their answer on a printed answer sheet which presented the two choices: *bed* and *bad*.

After completing the perception task, participants were told that they would hear the word *bad* spoken by four different speakers. They were asked to estimate the age group of the speaker by assigning one of the age groups presented in Figure 2, age categories which were taken from Bayard (2000).

Figure 2. Question wording and age group choices used to obtain perceived talker age after hearing the most TRAP-like token for each of the four voices.

Now you will hear some of the same recordings from the first part of this experiment. Each of the speakers will say the word 'bad'. Please indicate what you consider to be the most likely age group for each speaker by circling one of the age groups provided. Again, don't worry — there's no right or wrong answer. Just follow your first intuition.

a. 18–25	b. 26–35
c. 36–45	d. 46–55
e. 56–65	f. 65+

3. Results: Male Voices

The number of times TRAP was perceived for a voice varied across different participants, so that different people heard the vowel boundary in different places, as shown in Table 2. Mistakenly, token#9 was not played for Troy's voice and instead *bad8* was played twice. The results from the two instances of *bad8* have been averaged.

Participants responded to each of the tokens differently, indicating that the perceptual boundary varies considerably among individuals. Some people perceived a token as TRAP that others perceived as DRESS. For example, the majority of participants heard *bad6* as DRESS, but there were several people who heard TRAP.

The point at which most participants stopped hearing DRESS and began hearing TRAP differed for the two male voices although the two voices had equal formant values. For *bad7*, eighteen of the participants perceived the vowel as TRAP for one of the male voices, Rex, but only seven of the participants perceived *bad7* as TRAP for Troy's voice. This difference is not only statistically significant (Fisher Exact $p<0.001$), but it is surprising considering that the formant values for both Rex and Troy are equal. What could be influencing this difference in perception?

Table 2. The number of times the vowel was perceived as TRAP for each of the tokens, bad1–bad10, and for each of the voices.

TOKEN	REX	TROY
bad1	0	0
bad2	0	1
bad3	0	0
bad4	1	0
bad5	4	3
bad6	6	5
bad7	18	7
bad8	19	16
bad9	17	
bad10	19	17

4. Discussion: male voices

Although people complained about the difficulty of assigning age to the voices, participants were significantly consistent with their assignment of relative age for the male voices (n=19; Wilcoxon matched pairs test; p(2-tail)<0.05). By averaging the assigned ages across all participants, a perceived talker age was calculated for each voice, shown in Table 3.

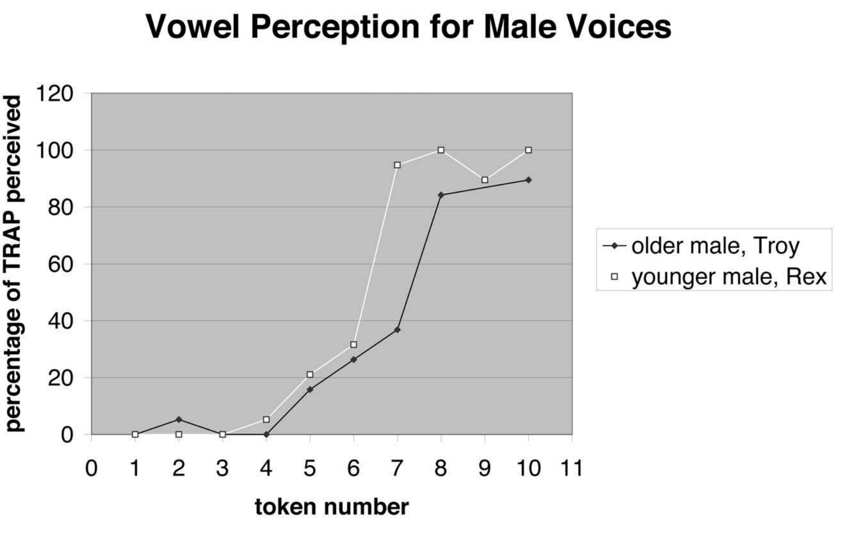
According to the age groups assigned by participants for this experiment, Troy has an older perceived talker age than Rex. Using the assigned age to classify one of the speakers as the older male and the other as the younger male, a pattern emerges as in Figure 3.

Participants consistently perceived more tokens of TRAP for the younger male voice than for the older male voice. This trend is consistent with studies on the production of DRESS and TRAP which have found that young males lead TRAP-raising over older males. It appears that not only do younger males produce a more raised TRAP than older males, but people perceive an ambiguous token, such as *bad7*, as *bed* for older males and as *bad* for younger males. This is an important finding demonstrating that people are sensitive to variation due to a change in progress and use this information in speech perception.

Table 3. Perceived voice age for each of the male voices.

	REX	TROY
PERCEIVED VOICE AGE	34.2	45.8

Figure 3. The number of times the vowel in each token was perceived as TRAP for each of the male voices, from the most bed-like token *(bad1)* to the most bad-like token *(bad10)*.



However, when using two different voices for speech perception tests, there is some possibility that other influences inherent in the voices could also influence perception. Three factors in particular could have influenced perception for this study, namely vowel length, integration of F2 and F3, as well as normalization based on what Ainsworth (1975) refers to as intrinsic rather than extrinsic factors.

While the literature suggests that normalization possibly occurs as a combination of both intrinsic factors, such as F0, and extrinsic factors, such as the relative values of F1 and F2 across different vowels of a single speaker (Nearey 1989; Strange 1989; Johnson, in press), normalization on the basis of extrinsic factors could not have played a role in this experiment. Participants heard only a continuum between two vowels so there would have been no basis for comparison. What remains is speaker normalization using pitch or higher formants such as F3.

Evidence suggests that there may be a loose correlation between F0 and F1 (Nearey 1989), meaning that F0 may influence the perceptual height of a vowel. A higher pitch for a certain voice could cause a listener to compensate for the high pitch, and presumably shorter vocal tract, and perceive the vowels for that voice as lower than a different voice with a lower pitch. If F0 played a role in speaker normalization in this experiment, participants could have been inclined to perceive a more DRESS-like token for Troy because Troy's voice has the higher pitch. This may explain the tendency for participants to hear more tokens of TRAP for Rex than for Troy.

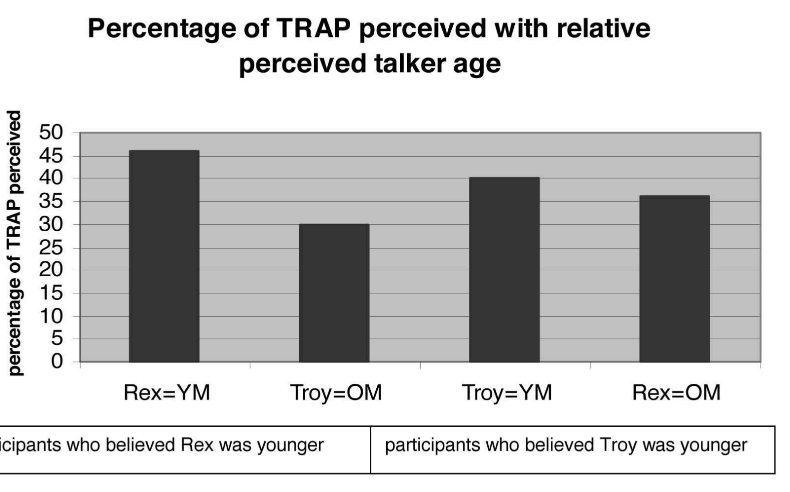
Because the value of F3 varies across speakers but varies little across vowels of a single speaker, the third formant as well as other higher formants may be used in order to determine which sound was perceived (Fujisaki and Kawashima 1968), perhaps by being integrated into F2 for front vowels when values of F2 and F3 are within 3 Bark of one another (Johnson 1989). A higher value for F3 could potentially influence the perceptual frontness of a vowel. Because the value of F3 for Troy's voice is lower than the value for Rex's voice, participants may have perceived Troy's voice as being more TRAP-like. This does not reflect the results of the experiment and therefore lends support to the claim that perceived talker age influenced perception.

While the value of F0 may cause Troy's tokens to sound more DRESS-like, the value of F3 may make them sound more TRAP-like. It is difficult to determine whether participants are likely to pay more attention to the vowels' perceptual height, influenced by F0, or perceptual frontness, affected by F3.

Another possible influence on perception is vowel length. The length of the vowels was not controlled, and Rex's vowels, which were 0.23353 seconds long, are longer than Troy's vowels of 0.174086 seconds. In most varieties of English, DRESS tends to be shorter than TRAP, so Troy's shorter vowels may sound inherently more *bed*-like than Rex's. However, Langstrof (2004) has shown that vowel length for speakers of New Zealand English born between 1890 and the 1930s differs from that in other dialects, such that the difference in vowel length between DRESS and TRAP is statistically insignificant and is almost identical when preceding a voiced consonant. If DRESS and TRAP continue to be the same length in contemporary New Zealand English, then it is likely that in this experiment vowel length did not play a large role in determining which vowel was perceived.

While it is possible that the combined effect of pitch and vowel length is partially responsible for the perception of Troy's voice as more DRESS-like, further analysis of the results support the hypothesis that participants were also relying on age. While most of the participants in Experiment 1 indicated that Troy was the older male voice, there were three individuals who felt that Rex was older and one participant who indicated that the two male voices were the same age. If the results of individuals are divided into two groups so that one group contains those who indicated that the older voice was Troy and the other group consists of those who specified that the older voice was Rex, an interesting pattern emerges, as shown in Figure 4.

Figure 4. Percentage of TRAP perceived for the voices when participants indicated that a particular voice was younger (YM) or older (OM).



Relative perceived talker age was determined by comparing the individually assigned age of each voice so that one voice is older in relation to the other voice. When comparing the voices of Rex and Troy within the same relative perceived talker age, it becomes evident that participants consistently perceived more TRAP for Rex than for Troy. When the voices were relatively younger (YM) participants heard more TRAP for Rex, and when the voices were relatively older (OM) participants also heard more TRAP for Rex. This indicates that participants were sensitive to inherent cues within the voice other than age, such as the value of F0.

However, age also appears to have some effect on perception. Those participants who believed Rex was younger than Troy perceived more instances of TRAP for Rex, and those participants who indicated that Troy was younger perceived a higher percentage of TRAP for Troy. In both groups, TRAP was perceived more often for the voice with the younger relative perceived talker age. Although these results are statistically insignificant, the trend lends considerable support to the argument that how old a voice sounds can influence the perception of vowels in the context of a chain shift in progress.

5. Results and discussion: Difference across gender

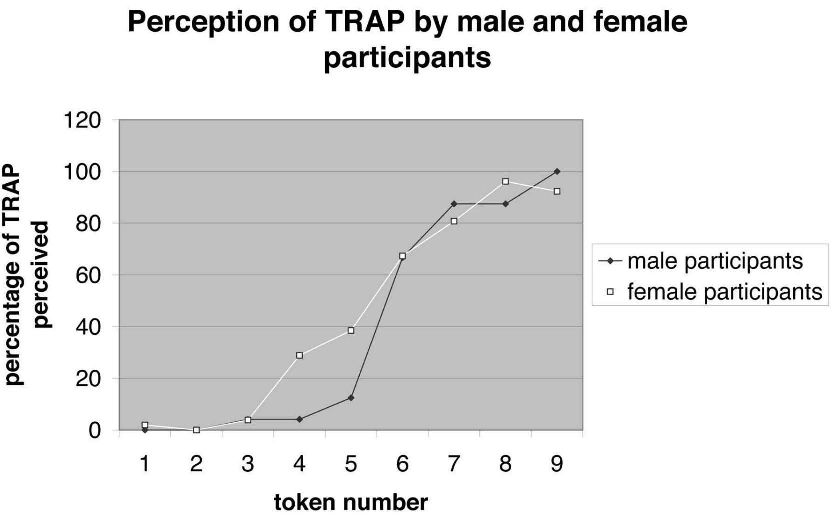
Another pattern emerged from results of this experiment. It seems that female and male participants may perceive DRESS and TRAP differently. As shown in Figure 5, the male subjects who took part in this experiment indicated hearing fewer tokens of TRAP than the female participants.

Past research has shown that females produce more raised tokens of TRAP than males do (Maclagan, Gordon, and Lewis 1999; Gordon, Campbell, Hay, Maclagan, Sudbury, and Trudgill 2004), and the results presented in Figure 5 suggest that females not only produce a more-raised variant of TRAP, but they also perceive one. However, only six males participated in the experiment and further investigation is needed to clarify the pattern.

6. Conclusion

While it is likely that a number of factors influenced the perception of the vowels in this experiment, there is evidence that the apparent age of a voice can affect how a vowel is perceived. If participants believe a voice sounds

Figure 5: The perception of TRAP divided by gender of participants.



older than another voice, they are more likely to perceive a vowel as DRESS because an older speaker of New Zealand English is less likely to produce a variant of TRAP that is as raised as for younger speakers. In the context of a vowel change in progress, the apparent age of a speaker can affect how a vowel is perceived.

In addition to providing data for female voices, ongoing research will help to determine the extent to which social factors affect vowel perception and will help to clarify the patterns which have emerged in this experiment.

Notes

1. I would like to thank Dr. Jen Hay for her insightful suggestions, patience and never-ending source of inspiration. I would also like to acknowledge the useful feedback from audiences at the University of Canterbury, the Postgraduate Conference at Victoria University 2004, and the New Zealand Language and Society Conference 2004.

2 . This data was collected by members of the New Zealand English Class of the Linguistics Department, University of Canterbury. The work done by members of the Origins of New Zealand English Project (ONZE) in preparing the data, making transcripts and obtaining background information is also acknowledged.
3 . I would like to thank Bob McMurray for making Klattworks available.
4. I am extremely grateful to the students from ling 203 and to all of the anonymous participants who made this research possible.

References

Ainsworth, W. 1975. 'Intrinsic and extrinsic factors in vowel judgments.' In Gunnar Fant and M. Tatham (eds) *Auditory Analysis and Perception of Speech*. London: Academic Press. 103–113.

Bayard, Donn. 1990. '"God help us if we all sound like this": Attitudes to New Zealand and other English accents.' In Janet Holmes and Allan Bell (eds) *New Zealand Ways of Speaking English*. Wellington: Victoria University Press, 67–96.

Bayard, Donn. 2000. 'Evaluating English accents worldwide: Questionnaire.' http://www.otago.ac.nz/anthropology/Linguistic/Questionnaire/Questionnaire.html

Boersma, Paul and David Weenink. 2004. 'Praat: doing phonetics by computer, version 4.1.23. 1992–2004.' available at: http://www.praat.org/.

Fujisaki, Hiroya and Takako Kawashima. 1968. 'The roles of pitch and higher formants in the perception of vowels.' *IEEE Transitions on Audio and Elect., vol AU-16, no.1*. 73–77.

Gordon, Elizabeth, Lyle Campbell, Jen Hay, Margaret Maclagan, Andrea Sudbury and Peter Trudgill. 2004. *New Zealand English: Its Origins and Evolution*. Cambridge: Cambridge University Press.

Gottfried, Terry L., James J. Jenkins and Winifred Strange. 1985. 'Categorical discrimination of vowels produced in syllable context and in isolation.' *Bulletin of the Psychonomic Society* 23: 101–104.

Johansson, Stig and Knut Hofland. 1989. *Frequency Analysis of English Vocabulary and Grammar*, vol. 1. Oxford: Clarendon Press.

Johnson, Keith. 1989. 'Higher formant normalization results from auditory integration of F2 and F3.' *Perception and Psychophysics* 46: 174–180.

Johnson, Keith. in press. 'Speaker normalization in speech perception.' In D.B. Pisoni and R. Remez (eds) *The Handbook of Speech Perception*. Oxford: Blackwell Publishers.

Johnson, Keith, Elizabeth Strand and Mariapaola D'Imperio. 1999. 'Auditory-visual integration of talker gender in vowel perception.' *Journal of Phonetics* 27: 359–384.

Kluender, Keith, Randy Diehl, and Beverly Wright. 1988. 'Vowel-length differences before voiced and voiceless consonants: an auditory explanation.' *Journal of Phonetics* 16: 153–169.
Langstrof, Christian. 2004. 'How to save Labov from New Zealand English: How short are short front vowels?' Presentation given at the University of Canterbury, September 2004.
Maclagan, Margaret. 1982. 'An acoustic study of New Zealand English Vowels.' *The New Zealand Speech Therapists Journal* 37: 20–26.
Maclagan, Margaret A., Elizabeth Gordon and Gillian Lewis. 1999. 'Women and sound change: conservative and innovative behaviour by the same speakers.' *Language Variation and Change* 11: 19–41.
Maclagan, Margaret and Jen Hay. forthcoming. 'The rise and rise of New Zealand English DRESS.' Proceedings of the Australian International Conference on Speech Science and Techonology: Workshop on The Phonetics and Phonology of Australian and New Zealand English, Sydney, Dec 2004.
McGurk, Harry and John McDonald. 1976. 'Hearing lips and seeing voices.' *Nature* 264: 746–748.
McMurray, Bob. in preparation. 'Klattworks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research.'
Nearey, Terrance M. 1989. 'Static, dynamic and relational properties in vowel perception.' *The Journal of the Acoustical Society of America* 85: 2088–2113.
Niedzielski, Nancy. 1999. 'The effect of social information on the perception of sociolinguistic variables.' *Journal of Language and Social Psychology* 18: 62–85.
Strange, Winifred. 1989. 'Evolving theories of vowel perception.' *The Journal of the Acoustical Society of America* 8: 2081–2087.
Thorndike, Edward L. and Irving Lorge. 1944. *The Teacher's Word Book of 30,000 Words*. New York: Columbia University
Wells, John. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Copyright of Te Reo is the property of Linguistic Society of New Zealand and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.