Perception of Sung and Spoken Vowels in New Zealand English

Andy Gibson
Institute of Culture, Discourse and Communication, AUT University, Auckland, New Zealand; andy.gibson@aut.ac.nz

This research uses a phoneme boundary experiment to examine the perception of singing and speech in New Zealand. Several previous studies have found that speech perception can be affected by listener expectation. Ladefoged and Broadbent (1957) found that by altering the formant frequencies of an introductory sentence the perception of the following word could also be altered. Hay, Drager, & Warren (2006) found that NZE-speaking participants were better at assigning NEAR and SQUARE\textsuperscript{1} words to their lexical sets if the experiment's instructions were given in an RP accent instead of a NZE accent. Recent studies have found that even without preceding linguistic material, listeners' expectations about the social characteristics of a speaker can affect perception (e.g. Niedzielski, 1999; Hay, Nolan, & Drager, 2006; Johnson, Strand, & D’Imperio, 1999; Drager, 2006). These results have been interpreted in an exemplar theory framework by proposing that remembered exemplars of language may be indexed or tagged with relevant social information. The present research will propose that exemplars may also be indexed with salient contextual information. Particularly, it is hypothesized that language which is sung may be indexed differently to language which is spoken. Most people who listen to popular music are exposed to singing in accents of English different to their own speech. This is especially so for people whose spoken variety is rarely used in singing, such as New Zealanders. The sum of such a person’s exposure to a given vowel sound is likely to be systematically different depending on whether the vowel occurred in singing or in speech.

Research conducted into pronunciation in popular singing (e.g. Trudgill 1983, Simpson 1999, Carlsson 2001) has shown that people sing differently than they speak. These singing accents often exhibit features of American English. Coddington (2004) found that even New Zealand artists who expressed a desire to sing in New Zealand English still consistently used a range of American English features in their singing. These singers must navigate a path between a desire to express their own identity (e.g. through the use of their regional dialect), and a compulsion to use the pronunciation most often heard in the genre of music being performed.

If remembered exemplars of sung language are stored differently to exemplars of spoken language, it is likely that in production of singing the sung exemplars are more highly activated and have more influence over the vowel sound to be produced. And so, New Zealanders sing the way they have heard other people sing, and in turn, their listeners build up stronger associations between singing and non-New Zealand English accents. The aim of this research is to test whether New Zealanders’ expectations about accents in singing may affect their perception of ambiguous vowel sounds.

The study focuses on the NZE DRESS and TRAP vowels which are raised and fronted compared to other dialects of English. Drager (2006) conducted a phoneme boundary experiment where listeners were asked to decide which of two words they heard. The stimuli were a set of words which ranged on a continuum from sounding like the word ‘bed’ to sounding like the word ‘bad’. It was found that the perceived age of the speaker affected the perceived position of the phoneme boundary between the DRESS and TRAP vowels.

The present study tests whether listeners perceive the boundary between DRESS and TRAP differently when they are listening to sung language as opposed to spoken language. The starting assumption is that in the lexicon of a New Zealander, the distribution of sung exemplars of DRESS is lower and backer in auditory/acoustic space than the centre of the DRESS exemplar cloud overall. This would be caused by repeated exposure to singing accents with open DRESS vowels and little or no exposure to singing with NZE-like raised DRESS. It is likely that there is considerable acoustic overlap between spoken exemplars of NZE TRAP and sung exemplars of DRESS. This situation could result in listeners perceiving the phoneme

\textsuperscript{1} These vowel names (along with DRESS and TRAP below) are based on the lexical sets outlined by Wells (1982).
boundary between the two vowels differently depending on whether they are listening to spoken or sung language.

The methodology is based on that used in Drager (2006). A 12-step continuum was created, with the first two formant frequencies manipulated to create stimuli which sound like ‘bed’ (F1=400; F2=2150) through to those that sound like ‘bad’ (F1=765; F2=1820). The continuum was created from one word recorded by the author with formants roughly in the centre of the continuum. Once the continuum was completed, the ‘sung’ and ‘spoken’ conditions were created. The ‘sung’ stimuli were presented to participants in the context of a background musical accompaniment consisting of a recording of guitar, keyboard, bass and drums cycling around the chord progression | G | Em | C | D7 |. In order to give the impression that the voice is actually singing, half of the stimuli in the singing condition were pitch-shifted up one semitone. This difference is small enough not to greatly affect spectral structure but large enough to allow the sung stimuli to be fitted to the musical chord progression. This gives an impression of melodic movement despite the stimuli being relatively fixed in their actual pitch. Each word is presented twice, on the first and third beat of the bar. There is then a gap of one bar for the participant to respond to the task by circling either the word ‘bed’ or the word ‘bad’ on a response sheet. In the spoken condition, the stimuli are presented at the same rate but without any background music, and all at the original pitch. The spoken and sung stimuli are presented in separate blocks. In each block, every word from the continuum is played twice in randomized order. This order is always the same in both conditions for any given participant. Half of the participants heard singing first and half heard speech first.

It was hypothesized that in the sung condition a higher percentage of the stimuli would be categorized as ‘bed’ than in the spoken condition. This hypothesis is based on the idea that a New Zealander’s expectation, based on past experience, is to hear lowered DRESS in singing and raised DRESS in speech. These expectations will then affect perception of ambiguous stimuli. Initial results support this hypothesis. Steps 5-8 of the continuum all received a mixture of ‘bed’ and ‘bad’ responses. Of all the responses for these ambiguous stimuli, 48% were categorized as ‘bed’ in the sung condition while only 38% were perceived as ‘bed’ in the spoken condition. These results provide support for the argument that remembered exemplars of sung language are indexed differently from exemplars of normal speech. Further development of this theory may help to explain why people use different accents in their singing than in their speech.

References


