# Vowel perception by listeners from different English dialects

Marzena Karpinska, Shodai Uchida, Izabelle Grenon

The University of Tokyo, Graduate School of Arts and Sciences,
Language and Information Sciences
marzenak@phiz.c.u-tokyo.ac.jp, shodai@phiz.c.u-tokyo.ac.jp, grenon@boz.c.u-tokyo.ac.jp

## ABSTRACT

Native English listeners from North America rely primarily on changes in formants, not vowel duration, when perceiving the vowel contrast in the minimal pair *bit* and *beat* manipulated from a Canadian English sample [5]. In this paper, we evaluated which cue do native English listeners from other regions use when perceiving the same North American vowel contrast. For this purpose, we used the same task and stimuli as in the study with North American listeners. Our results indicate that listeners from the UK, New Zealand, Ireland and Singapore used primarily changes in formants, a pattern similar to listeners from North America. Australian listeners, however, relied primarily on vowel duration rather than formants. The reaction time results suggest that the difference between Australian listeners and other listeners may be due to differences in the characteristics of vowels in Australian English versus North American English.

**Keywords**: speech perception, English dialects, vowel contrast

## 1. INTRODUCTION

Acoustic characteristics of speech sounds may vary across English dialects [10, 12]. Australian and New Zealand English differ, for instance, in the realization of formants when producing vowels [13]. Thus, there are considerable variations in the *production* of vowels across dialects of English. It is unclear, however, if these variations extend similarly to the *perception* of vowels.

When native English listeners perceive vowels, they may rely on vowel duration or changes in formants [1, 5, 7]. Native English listeners from North America use primarily changes in the first and second formants (F1 and F2) when identifying the high front lax vowel /I/ (as in *bit*) and the tense vowel /i/ (as in *beat*) [5]. The current study investigates whether native English listeners from regions other than North America (i.e. Australia, UK, Ireland, New Zealand and Singapore) perceive the same manipulated North American *bit/beat* contrast used in [5] in a way comparable to listeners from North America. That is, do native English listeners, irrespective of their dialect, rely primarily on changes in F1/F2 to perceive the high front tense-lax vowel contrast when produced by a speaker from a different dialect than their own?

## 2. METHODOLOGY

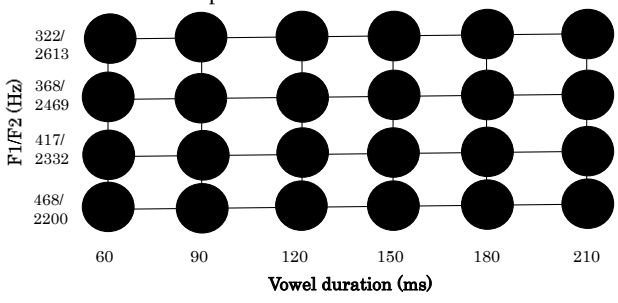
### 2.1. Participants

Twenty-three native English listeners (5 females, 18 males) from Australia (9), England, UK (3), Ireland (2), New Zealand (4), Scotland, UK (1) Singapore (3), and Welsh, UK (1) were recruited in Tokyo area. They had been staying in Japan from a few days to 21 years (mean 3.3 years). They were between 19 and 72 years old (mean 33). Since this study investigates whether there is a difference in vowel perception among native speakers of various dialects there was no particular age or region restriction. All participants reported basic or no knowledge of Japanese or other languages.

### 2.2. Stimuli

The same twenty-four *bit-beat* tokens as in [5] (manipulated from a Canadian English sample) were used. They were manipulated in terms of vowel duration and vowel quality (F1 and F2) using Praat [3]. The F1 and F2 varied in equal steps of 50 Mel, yielding four spectrally different vowels. The duration of each of the four vowels was manipulated in equal steps of 30ms, and varied from 60ms to 210ms as shown in Figure 1. All other cues remained constant.

**Figure 1**: Vowel duration and F1/F2 values for the bit/beat tokens used for the experiment.



### 2.3. Procedure

The same forced-choice identification task as in [5] was used to ensure our data were directly comparable with native North American English listeners. Participants completed a practice block with the 24 stimuli followed by the test phase consisting of 3 test blocks each containing the 24 stimuli for a total of 72 test tokens. The task was performed in a quiet room with headphones. The participants listened to the randomly presented test tokens and had to choose which word (*bit* or *beat*) they thought they heard. They were informed that their response time was measured. The experiment lasted approximately 5 to 10 minutes.

## 3. RESULTS AND DISCUSSION

In order to evaluate the exact use of formants and vowel duration by participants from each region, we first categorized individual data by using a mathematical criterion, the bias-ratio, as introduced and justified in [5]. Each of them were separated according to the acoustic cue they appear to most rely on based on the bias-ratio. Table 1 reports the number of participants from each region who exhibited a bias towards using only formants, both formants and vowel duration, or duration only. The data of North American English are cited from [5]. All English listeners showed a bias towards using at least one of the cues manipulated. Most Australian listeners (7/9 = 78%) had a bias towards using only duration when classifying the *beat/bit* contrast. Conversely, most listeners from other regions (11/14 = 79%) had a bias towards using primarily formants, sometimes in addition to using duration. Given this demarcation between listeners from Australia and English listeners from other regions, the data for the Australian participants will henceforth be presented and analysed separately.

**Table 1**: Number of English listeners from North America (NA; cited from [5]), Australia (AUS), New Zealand (NZ), United Kingdom (UK), Ireland (IR) and Singapore (SIN) exhibiting each of the possible bias pattern. The number inside parentheses corresponds to the total number of participants per group.

| | | | Other regions | | | |
|---|---|---|---|---|---|---|
| | NA (24) | AUS (9) | NZ (4) | UK (5) | IR (2) | SIN (3) |
| Formant | 19 | 1 | 2 | 4 | 1 | 2 |
| Both | 5 | 1 | 1 | 0 | 0 | 1 |
| Duration | 0 | 7 | 1 | 1 | 1 | 0 |

In Grenon's study [5], all 24 North American English listeners had a bias towards using formants, with only 5 listeners relying also on vowel duration. Hence, the overall pattern of NA listeners (figure 2) suggests a strong use of formants and use of vowel duration possibly only as a secondary cue. In figure 2, the white circle corresponds to a stimulus identified in most cases as *beat* and a black circle as *bit*.

In contrast, looking at the overall pattern of Australian listeners, they appear to rely mostly on duration (figure 3). As for native listeners from New Zealand, United Kingdom, Ireland and Singapore, they exhibit an overall pattern more similar to North American listeners [5] than to Australian listeners (current study) by relying primarily on changes in formants.

**Figure 2**: Average identification as either *beat* or *bit* by North American English listeners (N=24). The size of circle corresponds to its identification frequency. The shading of circle indicates the most frequently identified category: white for *beat* and black for *bit*. The number within each circle indicates the identification percentage for the most frequently identified category. ([5], reproduced with permission).

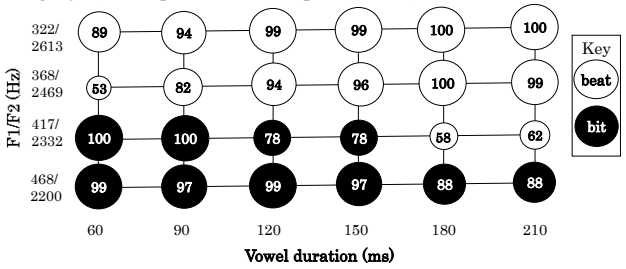



**Figure 3**: Average identification as either *beat* or *bit* by Australian English listeners (top) and English listeners from other regions (bottom).

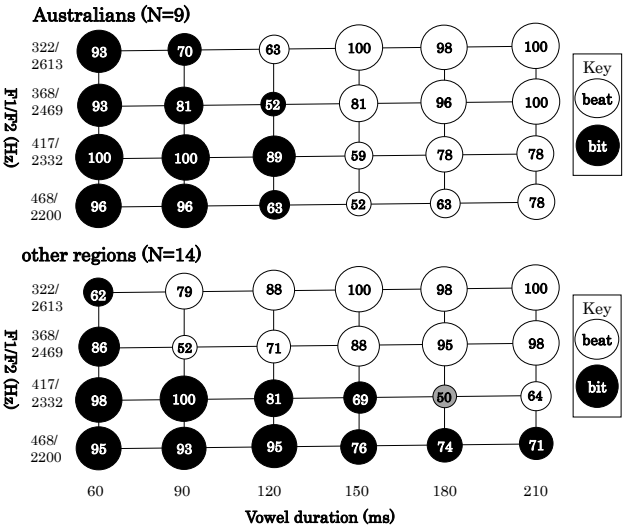



Table 2 reports the results of multiple regressions performed on the NA English data [5]. The effect of formants and vowel duration predicts 72% of the identification results in this model ($R^2 = .723$). North American listeners rely on changes in formants more ($\beta = .814$, $p < .001$) than on changes in vowel duration ($\beta = .247$, $p < .001$).

**Table 2**: Results of regression analyses on North American English data (N=24) [5].

| | B | SE B | B |
|---|---|---|---|
| Constant | -.521 | .032 | |
| Formants | .333 | .009 | .814* |
| Vowel Duration | .066 | .006 | .247* |

Note: Model $R^2$=723, *p<.001, B = regression coefficient, SE B = standard error of B, β= standardized regression coefficient.

Table 3 reports regression analyses performed on our English listeners with Australian listeners treated separately. For listeners from Australia, the effect of formants and vowel duration predicts 57% of the identification results ($R^2$ = .574). Changes in both vowel duration and formants played a significant role in the process of vowel identification. However, the effect of vowel duration (β = -.716, p < .001) was greater than the effect of formants (β = -.248, p < .001).

**Table 3**: Results of regression analyses on Australian English data (top) and English data from other regions (bottom).

| **Australian (N=9)** | B | SE B | β |
|---|---|---|---|
| Constant | 4.238 | .194 | |
| Formants | -.302 | .054 | -.248* |
| Vowel Duration | -.571 | .036 | -.716* |

Note: Model $R^2$=574, *p<.001,

| **Other regions (N=14)** | B | SE B | β |
|---|---|---|---|
| Constant | 4.568 | .149 | |
| Formants | -.737 | .042 | -.623* |
| Vowel Duration | -.344 | .027 | -.445* |

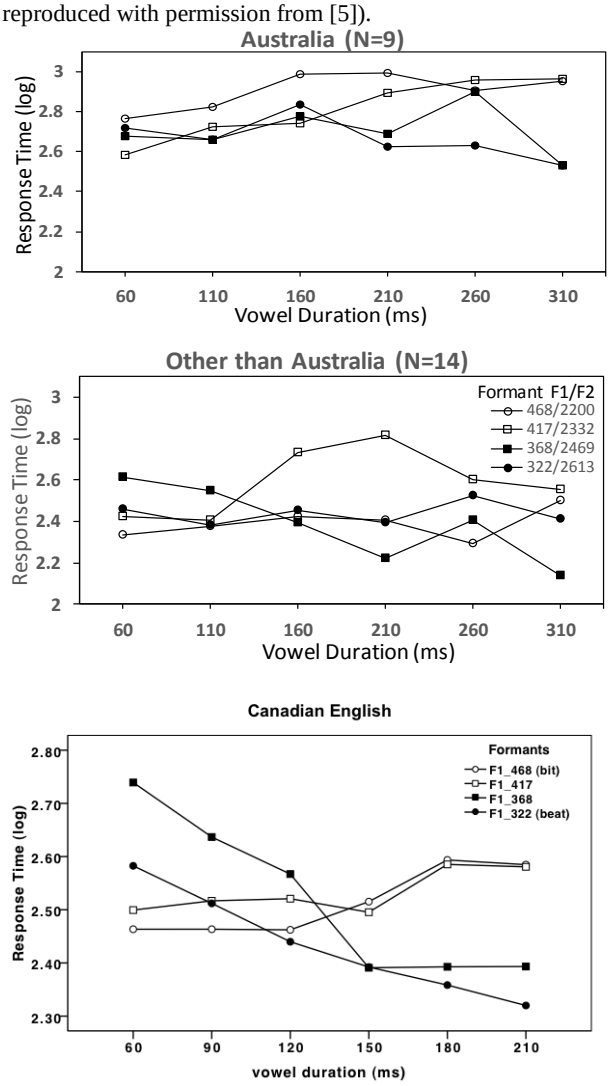
Note: Model $R^2$=586, *p<.001.

For listeners from other regions, the effect of formants and vowel duration predicts 59% of the identification results ($R^2$ = .586). Changes in both vowel duration and formants played a significant role in the identification process, but the effect of formants (β = -.623, p < .001) was greater than the effect of vowel duration (β = -.445, p < .001).

The regression results confirm that participants from other regions exhibited a pattern similar to English listeners from North America. That is, they used primarily changes in formants to classify the high front tense-lax vowel contrast, here manipulated from the speech of a Canadian English speaker. On the other hand, listeners form Australia used mainly vowel duration.

The log-transformed response time (RT) results, presented in Figure 4, suggests that in general Australian participants took longer to make their categorical decision, as compared with listeners from all the other regions (including North America). The response time results of Australian participants did not vary significantly as a function of vowel duration, suggesting that for Australian participants all tokens were similarly difficult to classify.

**Figure 4**: Average (log-transformed) response times for listeners form Australia (top) and other regions (middle) in this experiment and for listeners from North America (bottom) (the latter reproduced with permission from [5]).




The difference in perception of the target vowel contrast between listeners from Australia and other native English listeners (including North American) may be connected to the characteristics of vowels in Australian English. The tense-lax vowels in Australian English have a higher long-short duration ratio than North American English. Therefore, the length difference of vowels is more important and visible in Australian English, with a ratio of 1.89 (calculated based on data reported in [8]) as compared with Canadian English with a ratio of 1.62 [4] or American English with a ratio of 1.28 [6]. However, vowel length contrast is also prominent in UK English (ratio 1.91) [14], and yet, our UK listeners still relied on spectral differences to a larger extent than vowel duration. Thus, most likely, the more prominent vowel length contrast in Australian

English was not the factor responsible for the different performance of Australian participants.

It is possible that Australian listeners rely primarily on the duration cue simply because the formant values in our experiment are ambiguous for them. In American English, the average F1/F2 values for the tense and lax vowels are 338Hz/2571Hz and 437Hz/2192Hz respectively [15], roughly equivalent to the highest and lowest vowels in our experiment (highest 322Hz/2613, lowest 468Hz/2200Hz). The average F1/F2 values in Australian English, on the other hand, are 300Hz/2280Hz for the tense vowel and 370Hz/2210Hz for the lax vowel [2]. While the F1 values for the tense and lax vowels in Australian English are both within close range of the F1 value of the tense vowel in American English, their F2 values are closer to that of the lax vowel. That is, in the vowel space, the tense and lax vowels in Australian English have no direct correspondence with the American vowels. That may explain why one of the Australian participants reported that the vowels in our experiment sounded "heavily accented", and why they relied on vowel duration as an alternative cue and had longer response times.

The fact that Australian listeners appear unable to adjust their categorical boundary for vowel identification even after completing a practice block with all 24 test tokens, indicate that the perceptual system may not be as flexible to adjust to new vowel boundaries, as it supposedly is with consonantal boundaries [11, 9]. Hence, one implication of these results is that studies evaluating the perception of English vowels by second language learners, for instance, should be mindful of the dialect the learners have been exposed to, as this may have restricted their perceptual system to distinguish the vowel contrasts as they are produced in this particular dialect.

## 4. CONCLUSION

Except for native listeners from Australia, English listeners from all regions considered in this study used primarily changes in F1 and F2 to perceive the *bit/beat* contrast, rather than vowel duration, in a way similar to native North American English listeners as reported in [5]. Only speakers from Australia relied more on vowel duration. However, our response time results suggest that changes in formants are likely important for Australian English listeners as well, only that the vowels in our study were ambiguous for them in terms of spectral characteristics.

## 5. REFERENCES


[1] Bennett, D. C. 1968. Spectral form and duration as cues in the recognition of English and German vowels. *Language and speech*, *11*(2), 65-85.

[2] Bernard, J. R. L. B., Mannell, R. H. 1986. *A study of /h_d/ words in Australian English*. School of English and Linguistics, Macquarie University.

[3] Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009.

[4] Cebrian, J. 2007. Old sounds in new contrasts: L2 production of the English tense-tax vowel distinction. In J. Trouvain & W. J. Barry (Eds.), *Proc. 16th ICPhS.* Saarbrucken. 1637-1640.

[5] Grenon, I. 2010. *The bi-level input processing model of first and second language perception*. Ph.D. Dissertation. U. of Victoria, Canada.

[6] Hillenbrand, J.M., Getty, L.A., Clark, M.J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.*, *97*(5), 3099-3111.

[7] Hillenbrand, J. M., Clark, M. J., Houde, R. A. 2000. Some effects of duration on vowel recognition. *J. Acoust. Soc. Am.*, *108*(6), 3013-3022.

[8] Kato, A., Cox, F. 2006. Development of Japanese length contrast: A longitudinal study of L2 vowels produced by Australian learners of Japanese. *Proc. 11th Australasian International Conference on Speech Science and Technology*, Auckland, 170-175.

[9] McQueen, J. M., Norris, D., & Cutler, A. 2006. The dynamic nature of speech perception. *Language and Speech*, *49* (1), 101-112.

[10] Nagy, N., Zhang, X., Nagy, G., Schneider, E.W. 2005. A Quantitative categorization of phonemic dialect features in context. In A. Dey et al. (eds.) *Lecture Notes in Artificial Intelligence* 3554. Springer-Verlag. 326-338.

[11] Norris, D., McQueen, J. M., Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology*, *47*, 204-238.

[12] Qin, Y., Saeed, V. 2002. A comparative analysis of UK and US accents in recognition and synthesis. *Proceedings of Conference on Acoustics Speech and Signal Processing*, Florida, 413-417.

[13] Watson, C. I., Harrington, J., Evans, Z. 1998. An acoustic comparison between New Zealand and Australian English vowels∗. *Australian Journal of Linguistics*, *18*(2), 185-207.

[14] Wells, J. C. 1962. *A Study of the Formants of the Pure Vowels of British English* (MA thesis). University of London, London, UK.

[15] Yang, B. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24,* 245-261.