“Stability” of vowel categories is grounded in phonology: Evidence from English dialect comparison

Mathias Scharinger[a], William J. Idsardi[a]

[a]*Department of Linguistics, University of Maryland, USA*

<u>Corresponding author:</u>

Dr. Mathias Scharinger

Department of Linguistics

University of Maryland

College Park, MD 20742-7505

Tel.: +1-(301) 405-3127

Fax: +1-(301) 405-7104

Email: mts@umd.edu

ABSTRACT

Current models of speech perception emphasize either fine-grained acoustic properties or coarse-grained abstract characteristics of speech sounds. Here, we provide evidence for the view that vowel categories are grounded in abstract phonological representation and that these representations account for the successful access of the corresponding categories. In an auditory semantic priming experiment, American English listeners made lexical decision on targets (e.g. *pot*) preceded by semantically related primes (e.g. *pan*). A variation of the prime's vowel across categories (e.g. *pen*) was not tolerated, as assessed by a lack of priming, although the phonetic categories of the two different vowels considerably overlap. Compared to the outcome of the same experiment with New Zealand English listeners, where prime variations were tolerated, our experiment supports the view that phonological, but not phonetic representations guide the mapping process from the acoustic signal to an abstract mental representation.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grants XXX. We thank Yakov Konrod, Alan Mishler, and Alexis Wellwood for their help in the data acquisition.

KEYWORDS

Speech perception, phonetic vs. phonological categories, vowel representations, semantic priming, lexical decision

Introduction

English vowels show substantial variation in pronunciation across speakers. This can arise from several factors, most prominently gender, dialect, and social background (e.g. Hagiwara, 1997; Johnson, 1997; Lindblom, 1990; Thomas, 2001). Vowel categories considerably overlap based on their first (F1) and second (F2) formant values (cf. Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952), these measures being two salient acoustic cues for vowel identification and categorization across languages (Flege, Munro, & Fox, 1994; Ladefoged, 2001; Lindblom & Studdert-Kennedy, 1967; Pols, van der Kamp, & Plomp, 1969; Stevens, 1998). Although vowel categories thus seem to be rather fuzzy and vowel perception has shown to be less categorical than (stop) consonant perception (Pisoni, 1973; Schouten & van Hessen, 1992), subjects can nevertheless distinguish vowels in close vicinity (such as [ɛ] and [æ]) with high accuracy (Hillenbrand et al., 1995). Here we report further evidence that overlapping vowel categories in Standard American English (henceforth AE) do not cause perceptual ambiguities in a behavioral online task. The results are compared to previous findings from New Zealand English (NZE, Scharinger & Lahiri, 2010), where a substantial shift of vowel categories is observed. As illustrated in more detail below, we suggest that the underlying phonological system determines the stability of vowel categories.

Theoretical background: Phonetic vs. phonological variation

Some English varieties show a degree of dialectal variation which shifts certain vowels across categorical and phonological boundaries. An example is provided by the short front vowels of NZE. Here, original low [æ] moved to the mid position of [ɛ], while [ɛ] moved to the position of high [ɪ] (Gordon, Maclagan, Hay, Campbell, & Trudgill, 2004; Langstrof, 2006). Figure 1 illustrates the pronunciation consequences for the nouns *pan, pen,* and *pin* in comparison to AE. The column 'lexical set' refers to a dialect-independent label of the respective vowels (Wells, 1982). We conjectured that in order to correctly produce words with the TRAP vowel in NZE, the underlying phonological representation must have shifted, too (cf. Figure 1). Note that we consider the phonological representation as what may equal the distributional center in phonetic exemplar space. A change in phonological representation is not characterized by increasing overlap between phonetic categories, but rather in a shift of the category center. While changes in category overlap can be considered phonetic variation, shifts of category centers describe phonological variation (between dialects) or phonological change (within dialects). Naturally, these kinds of variation are not independent. Phonological variation implies phonetic variation, and phonetic variation can result in phonological variation as well.

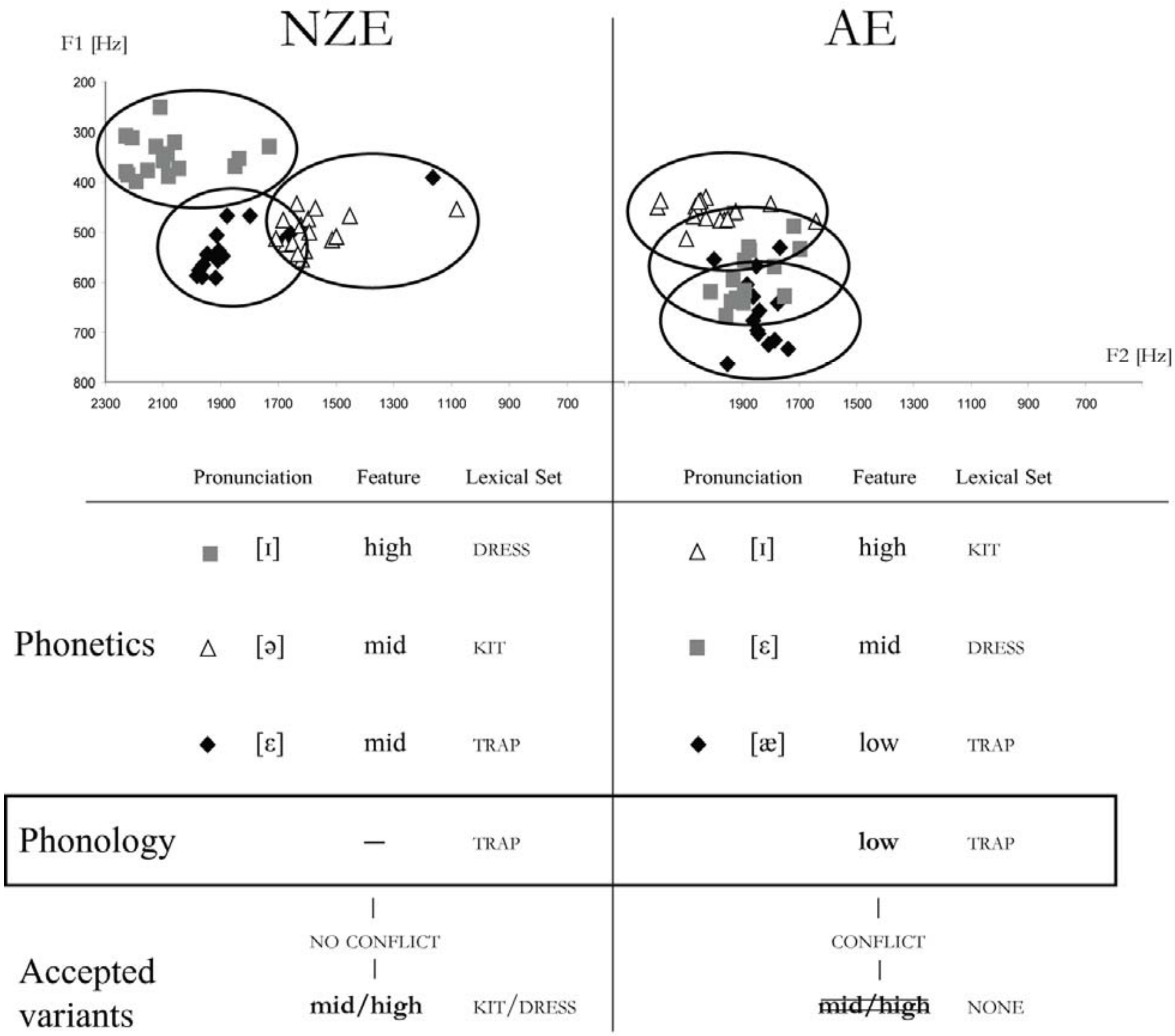

Figure 1: Phonetic realization of experimental stimuli in Scharinger & Lahiri (2010) [NZE] and this study [AE]. The top part shows the location of the vowels from monomorphemic English nouns in the F1/F2 space. The bottom part illustrates the assumed phonetic and phonological categorizations. Lexical set labels are given in order to compare the categories across dialects. Accepted variants are determined on the basis of feature conflicts as described in Lahiri & Reetz (2002) which are absent when the phonological representation has no specification for a given feature (here: low).

We had previously approached the issue of differences in processing TRAP vowels in NZE compared to AE by employing an auditory semantic repetition priming experiment (Scharinger & Lahiri, 2010). This behavioral experimental technique has shown to reflect lexical organization and processing (Forster, 1999). In our study, we measured the lexical decision times on targets semantically related to nouns with the TRAP vowel (e.g. *pot* in relation to *pan*, cf. Figure 2). Aside from conditions in which the TRAP vowel noun (e.g. *pan*) was used as prime for *pot*, we also selected DRESS and KIT vowel primes (e.g. *pen, pin*). Note that all primes were produced by a NZE speaker and therefore conformed to the pronunciation illustrated in Figure 1.

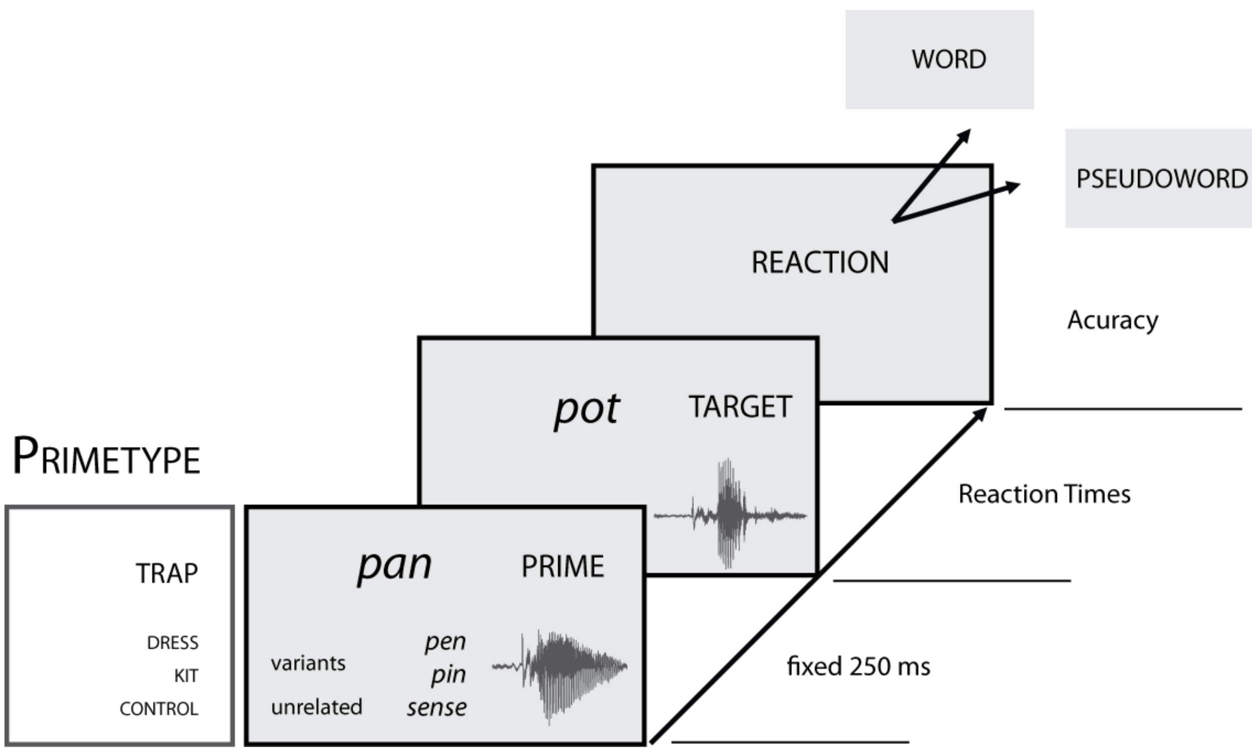

Figure 2: Experimental design of auditory semantic priming in Scharinger & Lahiri (2010) and this study. Prime variants are primes with different vowels than the semantic primes. They are labeled with lexical sets as explained in Figure 1. Semantic primes always contain the TRAP vowel.

Intriguingly, all three prime types resulted in significantly faster lexical decision times on the respective targets compared to an unrelated control condition. We interpreted these findings as evidence for representational differences of the TRAP vowel in NZE: The absence of the feature "low" allows for the mapping of a phonetically high vowel (as in NZE *pen*) onto the phonological TRAP category (cf. Figure 1), that is, NZE *pen* ([pɪn]) is a no-mismatch to NZE *pan* (/pɛn/) in the terminology of Lahiri & Reetz (2002). The successful mapping of *pen* to *pan* leads to facilitated lexical access of *pan* and therefore, its semantic relatives (e.g. *pot*). The results of a further experiment with the same phonetic stimuli presented to AE listeners significantly differed regarding the prime variants. This allowed for the conclusion that the phonological representation of the TRAP vowel in AE differed with regard to NZE. Crucially, its low specification particularly excluded mappings from the high vowel variant, and consequently, priming was absent in this condition.

We conducted a similar experiment here in order to test whether the priming pattern was an artifact of the phonetic material (the NZE stimuli) or indeed the underlying representation of the TRAP vowel. For that purpose, we employed the same priming design as illustrated in Figure 2, but used stimuli recorded from an AE speaker and AE listeners as participants. We also plan to run the experiment with NZE listeners.

If it is in fact the phonological vowel representation that accounts for the priming pattern and not the dialect of the speaker, we expect that the prime variants are not accepted as exemplars of the TRAP category for AE listeners and thus should not prime. Based on previous studies (Allen & Miller, 2001; Miller, 1995), we further expect a relatively clear extraction of phonetic category labels for the vowel exemplars presented.

Methods

Experimental design

The same 64 mono-morphemic nouns (mean length 3.25 segments, SD=0.43) as used in Scharinger & Lahiri, 2010 were recorded from a New England speaker of AE and distributed over four experimental lists in a Latin Square design. This guaranteed that each subject heard a target only once while across subjects, each target could be paired with four different prime types: (1) TRAP vowel nouns (e.g. *pan*); (2) DRESS vowel nouns (e.g. *pen*); (3) KIT vowel nouns (e.g. *pin*) and (4) unrelated nouns (e.g. *sense*). There were 16 critical item pairs in each subject list together with 52 filler pairs, 34 of which had pseudo-words as their second (target) element. Pseudo-words were derived from existing English words by altering 1-3 segments. They were phonotactically legal and cross-checked by a native English speaker. All item pairs were pseudo-randomized.

Subjects & Procedure

68 students of the University of Maryland (52% females, mean age 20, SD=2.7) participated for class-credits and were randomly assigned one of the four experimental lists. They were tested individually and familiarized with the experimental task in a practice session with 10 prime-target pairs not occurring in the main experiment. In both tasks, experimental items were presented pair-wise, using the software DMDX (Forster & Forster, 2003), and subjects had to provide a word/pseudo-word decision on the second member of each pair (the target). In order to match the attribution of button presses to the study by Scharinger & Lahiri (2010), right-handed

subjects had to give word-responses by pressing J with their right index finger (pseudo-words: F), while left-handed subjects (N=2) were given the opposite instructions.

First (=prime) and second (=target) member of each experimental pair were separated by 250 ms, and subjects could respond within 2000 ms after target presentation. Reaction time measurement started at the onset of the target. The experiment, including a short briefing and the practice session, lasted for about 15 minutes.

Results

Subjects showed acceptable performance on the targets (6.3% wrong responses, 0.8% time-outs), although 4 subjects and 2 items had error-rates above 25% and were excluded from further analyses.

The dependent measures *accuracy* (correct vs. incorrect lexical decision) and *reaction times* (as log values) were analyzed in Mixed-Effect Models (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000) with *subject* and *items* as random effects and *prime type* (TRAP, DRESS, KIT, control) as a fixed effect. For the accuracy analysis, we calculated a Mixed-Effect Logit Model (Agresti, 2002; Breslow & Clayton, 1993) and found a main effect of *prime type [ε]* (Wald-z = -2.20, $p < 0.05$), reflecting higher error rates for the DRESS primes compared to all other conditions.

For the reaction time analyses, we additionally removed outliers with more than 2.5 SD from the mean (8.5% of the data points). There was a main effect of *prime type* ($F(3,747) = 3.76$, $p <$

0.05), reflecting significant differences between the control and the TRAP vowel condition (cf. Table 1 & Figure 3).

Table 1: Random and fixed effects for the Mixed Model on reaction times. Fixed effects are given in relation to the control condition. The *p* values were calculated using 10000 Markov Chain Monte Carlo simulations (MCMC; Baayen, 2008).

Random	**Variance**	**Std.Dev.**	**MCMCmean**			
SJ (*Intercept*)	0.0177	0.1331	0.0970			
Item (*Intercept*)	0.0071	0.0845	0.0770			
Residual	0.0239	0.1545	0.1598			
Fixed	**Estimate**	**Std. Error**	**MCMCmean**	***p*MCMC**	***t* value**	***p*(>|t|)**
(Intercept)	6.8524	0.0301	6.8518	0.0001	227.8900	0.0000
Prime Type: TRAP	-0.0335	0.0152	-0.0340	0.0338	-2.2000	0.0280
Prime Type: DRESS	0.0146	0.0153	0.0145	0.3616	0.9500	0.3404
Prime Type: KIT	0.0042	0.0153	0.0036	0.8152	0.2700	0.7848

In order to assess potentially confounding factors, we paralleled multiple regression analyses by determining the best-fit mixed-model for reaction times (cf. Crawley, 2005) with the alternative fixed effects *phonemic length of prime*, *phonemic length of target*, *frequency of prime*, and *frequency of target* (Cobuilt *log* frequencies from CELEX, Baayen, Piepenbrock, & Gulikers, 1995). As a result of this procedure, the best-fit model included the random effects *subject* and *item* and the fixed effects *phonemic length of target* and *frequency of target* in a fully factorial design. Only the interaction *phonemic length of target* x *frequency of target* was significant ($t = 6.57$, $p < 0.05$). In particular, for low frequency targets, reaction times were faster with increasing target length. This is not surprising in an auditory priming design where decisions are made on the basis of the available acoustic material. It does not confound the reaction time main effect, since we distinguished between conditions on the basis of the primes, i.e. all conditions had the same targets.

In sum, reaction times robustly differed between the control and the semantic condition, but not between the control and the prime variant conditions. Hence, priming only occurred in the semantic condition and the prime variants were not accepted as variants to the TRAP nouns (cf. Figure 3).

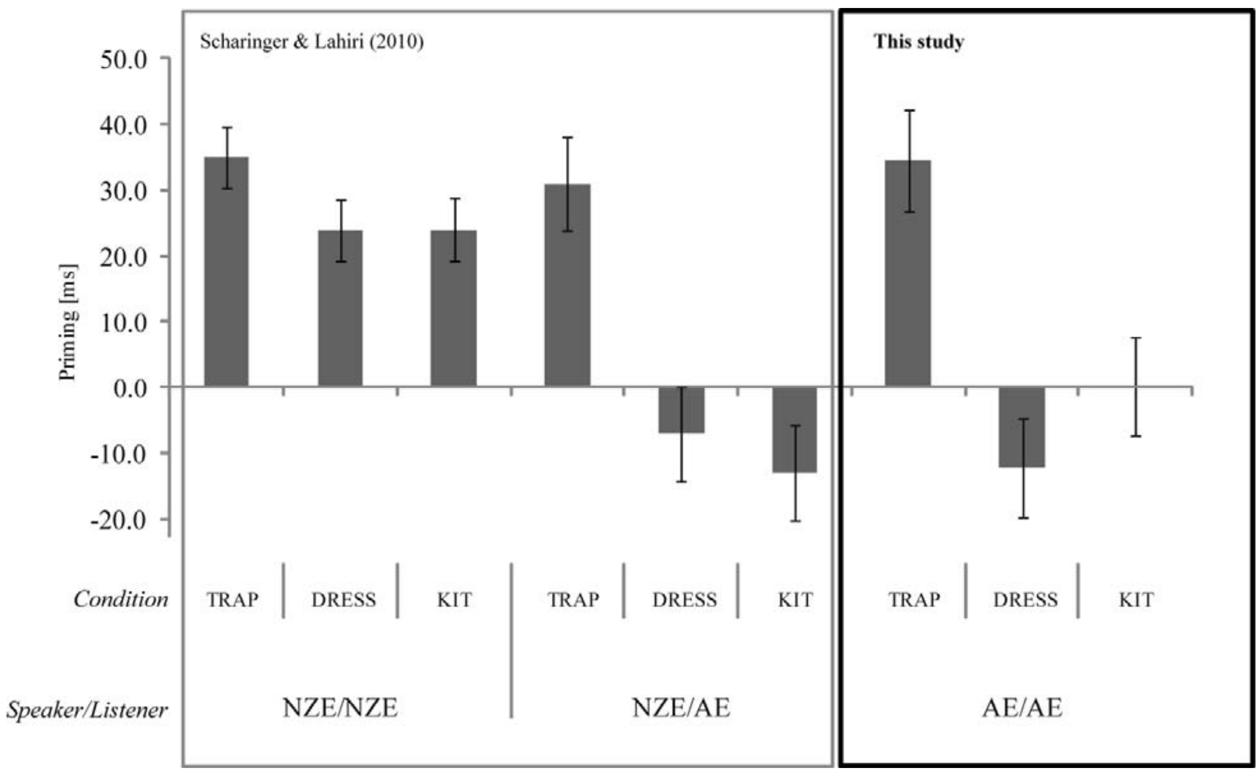

Figure 3: Amount of priming in the three vowel conditions (difference between reaction times in control and vowel conditions), compared to the Scharinger & Lahiri (2010) findings. Error bars indicate standard errors of the mean.

The patterns of results in the current study are thus parallel to those in the previous study when AE listeners heard NZE. Hence, the dialect of the speaker is not a major factor for the listener's interpretation of these vowels.

Discussion and Conclusion

In the study reported here, we modified a common auditory semantic priming experiment such that aside from the semantically related prime-target pairs (e.g. *pan-pot*) the design included primes which differed in their vowels from semantically related primes (e.g. *pin, pen*, cf. Figure 2). With this design, we wanted to test whether a deviance in the respective vowel would still yield significant target priming. Based on previous work (Scharinger & Lahiri, 2010), we expected that this is only possible if the differing prime is accepted as a variant of the semantic prime by virtue of its vowel. The acceptance should result from a comparison of the variant prime vowels (e.g. in *pin* and *pen*) with the phonological representation of the semantic prime vowel (e.g. *pan*). Based on the previous comparison between NZE and AE listeners, the phonological representation appears to be dialect-dependent and accounts for the observed priming pattern (cf. Figure 1, left). This study adds further evidence that the phonological, but not the phonetic representation determines the acceptance of vowel variants. In the experiment reported here, only the semantically related primes facilitated the lexical decision latencies for the respective targets. Primes with the DRESS and KIT vowel (e.g. *pen, pin*) did not elicit significant priming (Figure 1, right). In fact, the pattern paralleled the outcome of the Scharinger & Lahiri (2010) study for NZE stimuli and AE listeners (Figure 1, middle). Note that the phonetic realization of the DRESS vowel in NZE corresponds to the realization of the KIT vowel in

AE, while the realization of the KIT vowel in NZE is close to the realization of the DRESS vowel in AE (although more centralized in the vowel space). Thus, it seems that the same listener groups processed different phonetic stimuli in a similar way, while different listener groups processed the same phonetic stimuli in different ways. This pattern is only explicable by referring to differences in underlying phonological representations in NZE versus AE listeners. These representations determine which phonetic variations are accepted and which are not (described in more detail in Lahiri & Reetz, 2002 and Scharinger & Lahiri, 2010). They ultimately enable the listener to distinguish between the rather fuzzy vowel categories. On the other hand, the observations of slightly less accuracy and a negative priming pattern in the DRESS condition may indicate that phonetic details are not entirely neglected in this task. Obviously, due to the acoustic proximity of the DRESS vowel to the TRAP vowel (i.e. *pen* to *pan*), lexical decision times were less accurate and longer when the target was preceded by DRESS vowel primes than when it was preceded by the originally semantic primes. In the same vein, we assume that the denser vowel space and the greater phonetic category overlap in the AE experimental stimuli co-varies with more detailed phonological representations (here: low TRAP vowel) in that dialect. Similarly we conjecture that the density of the vowel space correlates with the amount of phonetic information that needs to be extracted from the acoustic signal. In a denser space, more fine-grained distinctions are necessary. Previous research has shown that phonetic categories have internal structure with access to very detailed acoustic properties (cf. Allen & Miller, 2001; Miller, 1995).

References

Agresti, A. (2002). *Categorical data analysis*. New York: Wiley.

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception and Psychophysics, 63*, 798–810.

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory & Language, 59*, 390-412.

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* [CD-Rom]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*, 9–25.

Crawley, M. J. (2005). *Statistics. An Introduction using R*. Chichester: Wiley.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*, 3623-3641.

Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain and Language, 68*, 5-15.

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers, 35*, 116-124.

Gordon, E., Maclagan, M., Hay, J., Campbell, L., & Trudgill, P. (2004). *New Zealand English: Its Origins and Evolution*. Cambridge, UK: Cambridge University Press.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America, 102*, 655-658.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099–3111.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. W. Mullennix (Eds.), Talker Variability in Speech Processing (pp. 145-166). New York: Academic Press.

Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages*. Malden, MA: Blackwell.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), Laboratory Phonology VII (pp. 637-677). Berlin: Mouton de Gruyter.

Langstrof, C. (2006). Acoustic evidence for a push-chain shift in the Intermediate Period of New Zealand English. *Language Variation and Change, 18*, 141-164.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. *Speech production and speech modelling, 55*, 403–439.

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America, 42*, 830-830.

Miller, J. L. (1995). On the internal structure of phonetic categories: A progress report. *Cognition*, 333–347.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*: Springer Verlag.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*, 253–260.

Pols, L. C. W., van der Kamp, L. J. T., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *The Journal of the Acoustical Society of America, 46*, 458-458.

Scharinger, M., & Lahiri, A. (2010). Height differences in English dialects: Consequences for processing and representation. *Language and Speech, 53*, 1-28.

Schouten, M. E. H., & van Hessen, A. J. (1992). Modeling phoneme perception: Categorical perception. *The Journal of the Acoustical Society of America, 92*, 1841-1855.

Stevens, K. (1998). *Acoustic Phonetics*. Cambridge, MA: The MIT Press.

Thomas, E. R. (2001). An acoustic analysis of vowel variation in New World English. *American Speech, 85*, 1–14.

Wells, J. C. (1982). *Accents of English I: Introduction*. Cambridge, UK: Cambridge University Press.