UvA-DARE (Digital Academic Repository)

Dialectal effects in the perception of vowels produced by first and second language speakers : North Carolinian versus Southern Welsh listeners

Hoffmann Bion, R.A.; Escudero, P.; Morrison, G.S.

DOI

[10.1121/1.3033938](https://doi.org/10.1121/1.3033938)

Publication date

2008

Published in

Proceedings of Meetings on Acoustics

[Link to publication](#)

Citation for published version (APA):

Hoffmann Bion, R. A., Escudero, P., & Morrison, G. S. (2008). Dialectal effects in the perception of vowels produced by first and second language speakers : North Carolinian versus Southern Welsh listeners. *Proceedings of Meetings on Acoustics*, 4(1), [060005]. <https://doi.org/10.1121/1.3033938>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Proceedings of Meetings on Acoustics

Volume 4, 2008

<http://asa.aip.org>

**155th Meeting
Acoustical Society of America
Paris, France
29 June - 4 July 2008
Session 2pSCa: Speech Communication**

2pSCa2. Dialectal effects in the perception of vowels produced by first and second language speakers: North Carolinian versus Southern Welsh listeners

Ricardo A. Hoffmann Bion*, Paola Escudero and Geoffrey S. Morrison

***Corresponding author's address: Psychology Department, Stanford University, Stanford, California 94305, ricardoh@stanford.edu**

This paper investigates the effect of listeners' dialect on the perception of vowels. Listeners from North Carolina and South Wales categorized natural tokens of the four English vowels /i/, /I/, /E/ and /ae/ which were produced by speakers of Californian-, Dutch-, Spanish-, and Portuguese-accented English. Randomization tests revealed a significant difference between the listener groups' confusion matrices. Territorial maps were constructed on the basis of logistic regression models that were fitted to each listener group's responses, and they revealed that the differences in categorization were due to differences between the boundaries of the vowel categories of the two listener groups.

Published by the Acoustical Society of America through the American Institute of Physics

1. Introduction

Listeners often have difficulty categorizing speech sounds produced by speakers of different dialects of their first language.^{1,2} They also have problems perceiving foreign-accented speech.³ The present study examines whether listeners of different dialects of English have differences in their perception of English vowels spoken with various accents.

The two groups of L1-English listeners were from South Wales and North Carolina respectively. Previous studies have shown that speakers of these two dialects differ in the way they produce English vowels, especially the low and mid-low front vowels.^{4,5} These differences in their production lead to the general hypothesis that they will also differ in their perception of these vowels.⁶

There were four groups of speakers: Californian English, Dutch-accented English, Spanish-accented English, and Portuguese-accented English. The speech sounds tested were the English vowels /i/, /ɪ/, /ɛ/, and /æ/. These four dialects were chosen because of the different ways they produce the front vowels: Californian speakers of English distinguish all the vowel contrasts,⁵ Spanish speakers have problems distinguishing the /i/-/ɪ/ contrast,³ Dutch speakers the /ɛ/-/æ/ contrast,⁷ and Brazilian speakers have problems with both these contrasts.⁸ This variation in vowel production should result in a range of values that form a natural continuum, which would not be possible if vowels from a single dialect were to be used.

2. Method

Participants: Eleven monolingual-English listeners from Raleigh, North Carolina and eleven monolingual-English listeners from Cardiff and Swansea, South Wales participated in the perception experiment. All listeners were female, and their ages ranged from 20 to 29 years. The listeners had grown up and had spent most of their lives in North Carolina or South Wales respectively.

Stimuli: The stimuli for the perception experiment were natural tokens produced by five first-language speakers of English from Sacramento, California, and by 15 second-language (L2) speakers of English, who were five from North Holland, five from Spanish Lima and Mexico City,⁹ and five from Florianópolis, Brazil. The speakers were all females and their ages ranged from 22 to 29 years. The L2 speakers had studied English for at least seven years, had not lived in an English speaking country, and reported that their target dialect of English was American English. The Californian speakers had grown up in Sacramento and reported no knowledge of any language other than English.

Each speaker produced three tokens of each of the words “beat”, “bit”, “bet”, and “bat”: /bit/, /bɪt/, /bet/, and /bæt/. This resulted in a total of 240 stimuli. The words were recorded embedded in the carrier sentence *sVt and pVt sound like bVt*, however, the bVt words were presented to the listeners in isolation.

Fig. 1 provides the formant and duration values of all tokens produced by all four speaker groups (formant values were measured at 25% and 75% of the duration of the vowel, only the former are shown in Fig. 1). Note that the stimulus properties cover a range of values which form a natural continuum. This would not have been the case if only vowels from a single dialect had been used.

FIG. 1. a) First and second formant (F1 and F2) values measured at 25% of the duration of the vowel. b) Vowel duration and F2.

Procedure: The stimuli were presented via computer in random order, each stimulus was presented once. Listeners responded by clicking on one of four buttons labeled “beat”, “bit”, “bet”, and “bat”, which appeared on a computer screen.

3. Results and Discussion

Confusion matrices: Table I provides confusion matrices of the speakers’ intended vowel categories and the listeners’ response categories.

Stimuli		responses							
L1		American				Welsh			
		/i/	/ɪ/	/ε/	/æ/	/i/	/ɪ/	/ε/	/æ/
Californian	/i/	95	5			98			
	/ɪ/		90	8			85	12	3
	/ε/		9	75	15			30	68
	/æ/			13	87				98
Dutch	/i/	87	12			86	14		
	/ɪ/	4	92	4		3	96		
	/ε/		10	71	18			28	72
	/æ/		4	58	39			14	86
Spanish	/i/	73	25			65	28	5	
	/ɪ/	72	27			70	30		
	/ε/		70	28			30	61	9
	/æ/			9	91				99
Brazilian	/i/	78	21			78	22		
	/ɪ/	42	55			33	61	5	
	/ε/		3	46	51			17	82
	/æ/			44	55			16	84

TABLE 1. Confusion matrices for speakers’ intended and listeners’ perceived vowel categories. Cell values are perceived percentages for the intended vowel. Cells with a value of 2% or less are left blank.

The correlation between the Carolinian and the Welsh listener groups' confusion matrices were calculated.¹⁰ If there is a large difference between the perception of the two groups, then the correlation between their perception matrices will be relatively small. Randomization tests were conducted to determine whether the difference between the Carolinian and Welsh groups was significantly greater than the difference expected between two groups of listeners assigned at random: Half the listeners' data sets were randomly assigned to one group and half to a second group, irrespective of their dialects. The correlation between the two randomized groups was calculated. This procedure was repeated ten thousand times and the proportion of times that the correlation between the randomized groups was less than the correlation between the original Carolinian and Welsh groups was recorded. The results of the randomization tests indicated that there was a significant difference between the Carolinian and Welsh listeners' perception. That is, in all cases, the proportion of differences between randomized groups that had a correlation less than that obtained for the original groups of Carolinian and Welsh groups was less than 0.0001. Their perception was significantly different for the whole set of vowel stimuli and also for the sets of vowels produced by each L1 speaker group.

The confusion matrix for the Californian speakers' productions suggests that the North Carolinian and Southern Welsh listeners had similar perception of the Californian speakers' /i/ and /ɪ/ tokens. The major difference between the Carolinian and Welsh listeners is in their labeling of the vowel /ɛ/. The Carolinian listeners tended to correctly label the Californian speakers' /ɛ/ productions, whereas the Welsh listeners had a tendency to label them as /æ/. In addition, the Welsh listeners gave more correct responses to stimuli intended as /æ/.

Dutch speakers' English /i/ and /ɪ/ were generally correctly identified by both sets of L1-English listeners. With respect to Dutch speakers' English /ɛ/ and /æ/ productions, Carolinian listeners tended to identify both these vowels as /ɛ/, whereas Welsh listeners tended to identify both as /æ/. These results not only indicate that there is a difference between Carolinian and Welsh listeners' perception, but also that, from the perspective of both groups of listeners, the Dutch speakers had problems producing the /ɛ/–/æ/ contrast, as had been previously reported.⁷

The Carolinian and Welsh listeners had similar perception of the Spanish speakers' /i/ and /ɪ/, both vowels were predominantly identified as /i/, with a half to a third identified as /ɪ/. These results are consistent with earlier reports of Spanish speakers' difficulty with this English contrast.³ With respect to Spanish speakers' English /ɛ/ productions, both groups of listeners had a tendency to misidentify them as /ɪ/. Although the Welsh listeners correctly identified over half the Spanish speakers' /ɛ/ productions, the Carolinians identified the majority as /ɪ/. These results suggest that there is a difference between the two groups' perception, but also indicate that from the perspective of both groups of listeners, the Spanish speakers had problems producing English /ɛ/.¹¹ Neither group of listeners had difficulty correctly identifying the Spanish speakers' English /æ/ productions, as can be expected from the fact that Spanish speakers do not have problems differentiating between these two English sounds.³ Interestingly, correct identification rates for Spanish speakers' /æ/ tokens were higher than for Californian speakers' /æ/ tokens. This might

be expected if Spanish speakers substituted Spanish /a/-like vowels for English /æ/, these would have lower F2 than English speakers' /æ/, as can be seen in Fig. 1.

Consistent with earlier findings,⁸ it is apparent that from the perspective of both groups of listeners, the Brazilians had difficulty producing both the /ε/-/æ/ and the /i/-/ɪ/ contrast; although they had somewhat better performance on the /i/-/ɪ/ contrast. The Carolinian and Welsh listeners had similar perception of /i/ and /ɪ/, but for /ε/ and /æ/ the Welsh listeners had a greater probability of responding with /æ/.

Logistic regression analysis: The randomization tests indicated that there was a significant difference between the Carolinian and Welsh listeners' perception of the natural-speech tokens. Examination of the confusion matrices suggested that the two listener groups had similar perception of the /i/-/ɪ/ contrast, but differed in their perception of the /ε/-/æ/ contrast.

To investigate the relationship between the acoustic properties of the stimuli and the listeners' perception, a first order logistic regression model was fitted to each listener group's response data (for an introduction to the application of logistic regression to speech perception data see reference ¹²). For each response category, the models included a bias coefficient, and formant-tuned and duration-tuned coefficients. The formant-tuned coefficients consisted of a coefficient for F1 measured at 25% of the duration of the vowel (*F1*), and a coefficient for the change in F1 from 25% to 75% of the duration of the vowel ($\Delta F1$). Mutandis mutandi for F2 (*F2* and $\Delta F2$). Formant values were entered in log Hertz and duration values were entered in log milliseconds.

FIG. 2. Territorial maps showing the logistic regression models' modal predicted responses for vowel identification over the stimulus space. Top: Territorial maps based on the logistic regression model fitted to Carolinian listeners response data. Bottom: Territorial maps based on the logistic regression model fitted to Welsh listeners. Within each panel, the x axis represents $F1$ and the y axis represents $F2$. Panels in the middle column are plotted at the mean vowel duration of the perception stimuli (119 ms), panels in the left hand column are plotted at a vowel duration two standard deviations below the mean (79 ms), and panels in the right hand column are plotted at two standard deviations above the mean (180 ms). Panels in the middle row are plotted at zero formant movement ($\Delta F1 = 0$, $\Delta F2 = 0$), panels in the top row are plotted with diverging formant values two standard deviations above the mean $\Delta F1$ and below the mean $\Delta F2$ ($\Delta F1 = -66$ Hz, $\Delta F2 = +270$ Hz), and panels in the bottom row are plotted with converging formant values two standard deviations below the mean $\Delta F1$ and above the mean $\Delta F2$ ($\Delta F1 = +127$ Hz, $\Delta F2 = -171$ Hz). (Means and standard deviations calculated in log millisecond and log hertz).

Fig. 2 provides territorial maps based on the estimated coefficient values from each model. The territorial maps indicate the models' predicted modal responses in each part of the stimulus space. An examination of the territorial maps indicates that Carolinian and Welsh listeners had a similar location for the /i/-/ɪ/ boundary. However, their /e/-/æ/ boundaries clearly differed: The location of the /e/-/æ/ boundary for the Carolinian listeners was generally at higher $F1$ / lower $F2$ values than for the Welsh listeners. The Carolinian listeners' perception was more duration dependent: as duration increased, the location of the /e/-/æ/ boundary shifted towards lower $F1$ / higher $F2$ values, whereas duration had little effect on the Welsh listeners' /e/-/æ/ boundary. In contrast, formant movement had a greater effect on the Welsh listeners' perception: when $F1$ and $F2$ converged (negative $\Delta F1$ and positive $\Delta F2$), the /e/-/æ/ boundary shifted towards lower $F1$ /

higher $F2$ values, and when $F1$ and $F2$ diverged (positive $\Delta F1$ and negative $\Delta F2$), the / ϵ –/ $\ae$ / boundary shifted towards higher $F1$ / lower $F2$ values. This accounts for the greater use of / $\ae$ / responses by the Welsh listeners compared to the Carolinian listeners which was observed in the confusion matrices.

For both groups of listeners, the / i –/ ϵ / boundary appears to be based primarily on $F1$ properties; however, the boundary appeared to be at slightly higher $F1$ values for the Carolinian listeners than for the Welsh listeners. This may account for the Carolinian listeners giving more / i / responses to Spanish speakers' English / ϵ / productions than did the Welsh listeners. Also there appears to be a slight effect of formant movement which is greater for the Carolinian listeners: the boundary shifted towards lower $F1$ / higher $F2$ values when formants diverged (positive $\Delta F1$ and negative $\Delta F2$).

4. Conclusion

The perception of a natural vowel continuum by two groups of listeners of different dialects of English was tested. The most salient difference between North Carolinian and Southern Welsh listeners was in the boundary between English / ϵ / and / $\ae$ /. In particular, compared to listeners from South Wales, listeners for North Carolina had their / ϵ / and / $\ae$ / boundary at lower $F1$ and higher $F2$ values, and they made greater use of duration, classifying longer vowels as / $\ae$ / and shorter vowels as / ϵ /.

Acknowledgements

Many thanks to Dr. Robert Mayr for help with recruiting and testing of participants in Cardiff.

References and links

- ¹Cutler, A., Smits, R., and Cooper, N. (2005). "Vowel perception: Effects of non-native language versus non-native dialect," *Speech Communication*, **47**, 32-42.
- ²Evans, B.G., and Iverson, P. (2004). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *J. Acoust. Soc. Am.* **115**, 352-361.
- ³Flege, J.E., Bohn, and O-. S., and Jang, S. (1997) "Effects of experience on non-native speakers' production and perception of English vowels," *J. Phonetics* **25**, 437-470.
- ⁴Walters, J. R. (1999). "A study of the segmental and suprasegmental phonology of Rhondda Valleys English," PhD Thesis, University of Glamorgan, Cardiff.
- ⁵Clopper, C. G., Pisoni, D. B., and De Jong, K. (2005). "Acoustic characteristics of the vowel systems of six regional varieties of American English," *J. Acoust. Soc. Am.* **118**, 1661-1676.
- ⁶Kuhl, P. K. (2004). "Early language acquisition: Cracking the speech code," *Nature Reviews Neuroscience* **5**, 831-843.
- ⁷Flege, J. E. (1992). "The Intelligibility of English vowels spoken by British and Dutch talkers," in *Intelligibility in Speech Disorders: Theory, Measurement, and Management*, edited by R. Kent (John Benjamins, Amsterdam), pp. 157-232.

⁸Bion, R., Escudero, P., Rauber, A., and Baptista, B. (2006). "Category formation and the role of spectral quality in the perception and production of English front vowels," in Proceedings of Interspeech 2006, pp. 1363-1366.

⁹Two from Lima, Peru, and three from Mexico City, Mexico.

¹⁰Nearey, T, and Assmann, P. F. (1986). "Modeling the role of vowel inherent spectral change in vowel identity," J. Acoust. Soc. Am. **80**, 1297–1308.

¹¹Note that in Morrison (2007), Western Canadian English listeners tended to identify L1-Spanish speakers' Spanish /e/ as English /ɪ/.

¹²Morrison, G. S. (2007). "Logistic regression modelling for first- and second- language perception data," in *Segmental and prosodic issues in Romance phonology*, edited by M. J. Solé, P. Prieto, & J. Mascaró (John Benjamins, Amsterdam), pp. 219–236.