EXPLORING THE PROBLEM IN NATIVE DANISH LISTENERS' PERCEPTION OF THE ENGLISH *HOT* AND *HUT* VOWELS

Jonas Villumsen[1,a], Ocke-Schwen Bohn[2,a], Camilla S. Horslund[3,b]

[a)] Department of English, Aarhus University, Denmark
[b)] Department of Linguistics, Cognitive Science, and Semiotics
[1)] jv@cc.au.dk, [2)] ocke.bohn@cc.au.dk, [3)] horslund@cc.au.dk

ABSTRACT

Native Danish speakers tend to neutralise the vowels /ɒ, ʌ/ of Standard Southern British English in production, but research on the cause of this is sparse. This study examined the identification of these vowels by Native Danish and Native English listeners. As expected, Native Danish listeners identify /ɒ/ and /ʌ/ less accurately than Native English listeners. Exploration of by-token accuracy in F1–F2 space suggests that the location of the boundary between /ɒ/ and /ʌ/ differs for Native Danish and Native English listeners. Comparison with previous research indicates the contrast may span multiple Native Danish vowel categories. Taken together, these findings imply that inaccurate perception of the vowels by Native Danish listeners at least partially explains the vowels' neutralisation in production.

Keywords: second language speech learning, cross-language vowel perception, English, Danish

1. INTRODUCTION

Native Danish (ND) speakers have been observed to pronounce the Standard Southern British English vowels /ɒ, ʌ/ identically [5, 8, 9]. As identified by Native English (NE) listeners, both vowels sound like /ʌ/ when produced by ND speakers [5]. No explanation has yet been posited as to why ND speakers exhibit this exact pattern of mispronunciation. This paper explores the problem.

Perception of this vowel contrast likely plays an important role in its production problems. The /ɒ, ʌ/ vowels occupy an area of the vowel space which is sparsely populated in Danish (see Figure 1). This means that ND listeners will have encountered few vowel tokens in this area in native speech perception, which may have desensitised ND listeners to vowel contrasts in that area [4, 6]. It further adds to the problem that ND listeners are exposed to heterogenous varieties of English whose phonetic realisations of /ɒ, ʌ/ are diverse [5].

When ND listeners categorise /ɒ, ʌ/ in terms of Danish vowel categories, results suggest English /ɒ/ assimilates perfectly to Danish /ɔ/, whereas /ʌ/ assimilates inconsistently to Danish /ɔ/ as well as Danish /æ/ [7]. Viewed from the Perceptual Assimilation Model [2], /ɒ, ʌ/ are in a categorised–uncategorised relationship.

Few other tests of ND listeners' perception of /ɒ, ʌ/ have been conducted. Those that exist show great variation in performance. The vowels are discriminated in the range between 50–70% correct, depending on the order of vowel presentation [12]. Identification ranges between 65–85% [10], but those data were sampled from students of English at a university rather than from the general population.

The present paper reports on a study in which ND listeners and NE listeners identified /ɒ/ and /ʌ/ tokens presented in a /hVt/ environment, comparing the accuracy of the two groups. We also compare two types of ND listeners on the same task: high school students and first semester English students at a Danish university. ND students of English have been sampled previously in investigations into the /ɒ, ʌ/ problem [10]. Comparing their performance to that of a more general sample (high school students) tests the validity of that practice. Finally, as a step towards explaining the perceptual problem, this study explores whether misidentification of /ɒ, ʌ/ by ND

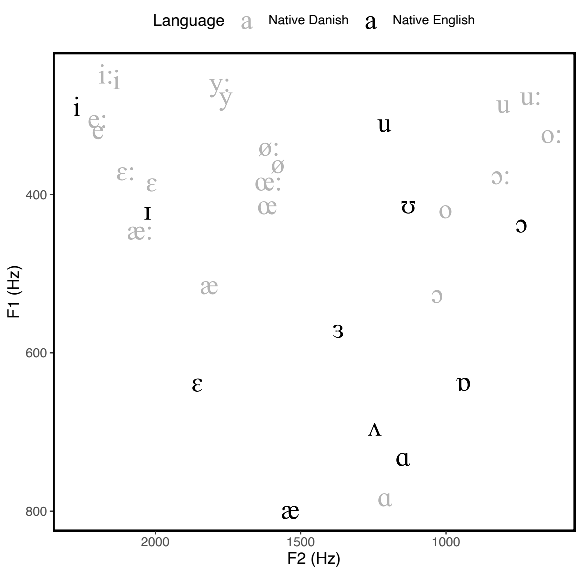


Figure 1. F1–F2 plot of L1 Danish vowels (grey) and L1 English vowels (black) based on data from [14].

listeners is related to the acoustic properties of individual vowel tokens.

2. METHODS

41 ND speakers and 18 L1 British English speakers participated in the study. Of the ND participants, 18 were first semester students of English at Aarhus University and 23 were senior year high school students. The age range of the participants was 18 to 26 years.

The stimuli comprised 48 unique tokens of *hot* and *hut* produced four times each by six male native talkers of Standard Southern British English. *Praat* [3] was used to analyse the first and second formant frequency (F1 and F2) of the stimuli at the midpoint of each vowel token.

Participants performed a two-alternative forced choice identification task with a total of 96 trials. Each of the 48 unique stimuli was presented twice, in a pseudo-randomised order. In a trial, participants heard either *hot* or *hut*, respectively containing /ɒ/ and /ʌ/. They then responded by pressing one of two buttons labelled *hot* and *hut*, with an inter-trial-interval of 1200 ms. The position of the buttons remained constant throughout the task, which lasted 4 to 7 minutes.

3. RESULTS

The data were modelled using mixed effects logistic regression, implemented in *R* [13] using the package *lme4* [1]. Model predictions were derived using the *emmeans* package [11]. Model selection proceeded from the maximal structure with stepwise reduction in random-effects until convergence. The final model estimated the correctness of per-trial responses as a function of Group (High school students, University English students, and NE listeners). Group was dummy coded with University English students as the reference level. The selected method for obtaining *p*-values was likelihood ratio tests. The final *lme4* formula, including random effects, was as follows:

```
Response ~ 1 + Group +
    (0 + Vowel | Participant) +
    (1 + Vowel + Group | Talker)
```

3.1 Model effects

The fixed effects of the levels of Group were as follows: University English students (intercept) were estimated at 1.84 log odds accuracy. High school students performed worse than University English students, β = –0.90, *SE* = 0.30, $p < .01$. NE participants perform considerably better than University English students, β = 2.76, *SE* = 0.51, $p < .001$. Model predictions in probabilities are plotted in Figure 2.

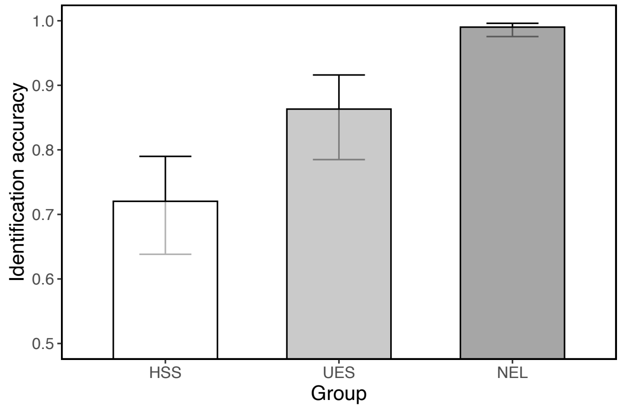


Figure 2: Identification accuracy by Native English listeners (NEL), university English students (UES), and high school students (HSS). Error bars: 95% confidence intervals.

In the random effects, there was variation between participants on their accuracy on each vowel, SD_{hot} = 0.78, SD_{hut} = 0.85. Accuracy on each vowel correlated highly within participants, *r* = 0.91. Within stimulus talkers, there was variation in accuracy on *hot*, *SD* = 0.78, with high variation in *hut*, *SD* = 1.26. Importantly, the intercept–slope correlation of Vowel within Talker was strongly negative, *r* = –0.87. This means that the difference between accuracy on *hot* and *hut* tokens from the same talker can be predicted from the accuracy of *hot* tokens: Higher accuracy on *hot* implies a negative difference—relatively lower accuracy—for *hut* from the same talker; lower accuracy on *hot* implies a positive difference—relatively higher accuracy—for *hut* from the same talker. This will be further explored in the next section.

3.2 Data exploration: accuracy by vowel token

The model's random effects suggested unexpected systematic variation in accuracy in vowel accuracy by talker. To explore what the model reflects, Figure 3 plots the averaged listener accuracies on both vowels within each talker. Note that the plotted values are not estimates from the model but manually computed averages. For ND listeners, the figure shows that most talkers' *hut* tokens are identified less accurately than their *hot* tokens; but tokens from talkers 4 and 5 exhibit the reverse relationship. This caused the model to estimate a high intercept–slope correlation in the random-effect term of vowel within talker; the lower the intercept (i.e., accuracy for *hot*), the higher the slope for vowel (i.e., accuracy for *hut*). The fact that NE listeners do not show the same pattern suggests that the reversal is not due to talker idiosyncrasy. Rather, the cause of the reversal is likely to be due to the acoustic properties of the vowels themselves, as well as ND listeners' use of these as perceptual cues.

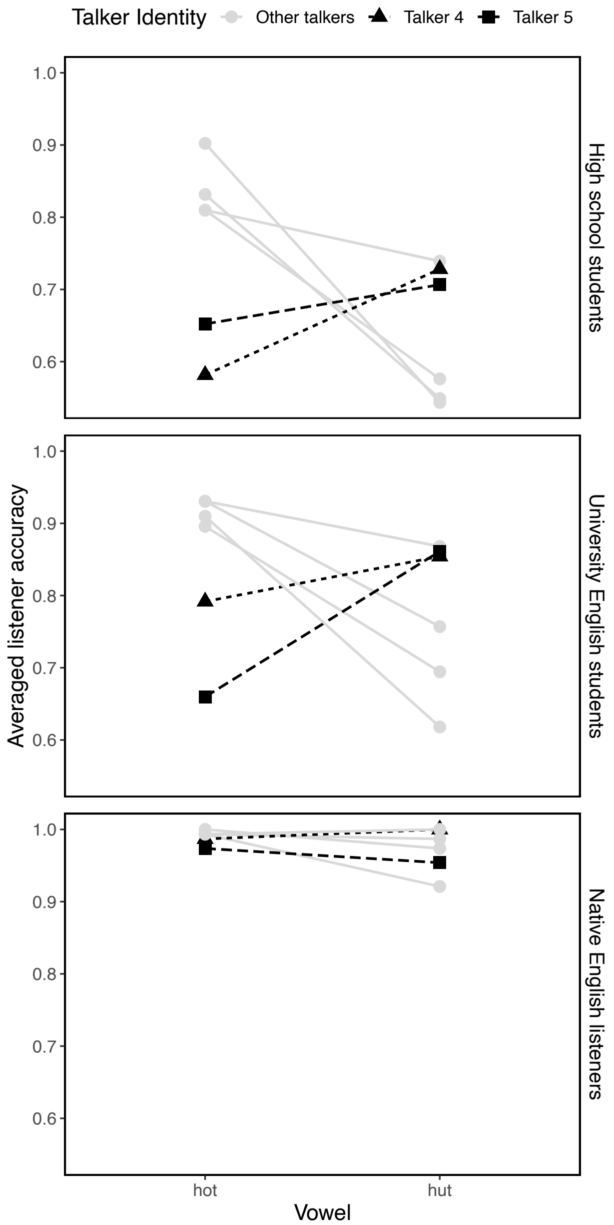


Figure 3. Group identification accuracy on vowels from each talker. To illustrate the reverse vowel relationship (see text), talkers 4 and 5 are highlighted.

Figure 4 plots all unique vowel tokens in the stimuli in F1–F2 space. The tokens are grey-scaled by average identification accuracy. Both ND groups exhibit the same pattern: With both vowels, some tokens are perceived highly accurately while others are perceived less accurately with a few being close to chance-level identification. Accuracy on each vowel appears to depend slightly differently on F1 and F2. Accuracy on /ɒ/-tokens appears to depend on F2, where most are identified accurately until F2 exceeds 945 Hz, at which point an abrupt decline in accuracy occurs. The /ɒ/-tokens where F2 > 945 Hz are exactly those produced by talkers 4 and 5 for whom the relative listener accuracies of /ɒ, ʌ/ were reversed (see Figure 3). Conversely, accuracy on /ʌ/-tokens does not appear to be sensitive to F2 variation. Instead, a small decrease in accuracy occurs when F1 surpasses about 675 Hz. Figure 5 depicts these relationships explicitly with smoothing lines (see especially the middle two panels). It appears that, in the case of /ɒ/-tokens with F2 > 945 Hz, ND listeners experienced an abrupt boundary shift at which point they became uncertain as to what they were perceiving. In the case of /ʌ/-tokens, these listeners experienced slightly more uncertainty as F1 increased past 675 Hz.

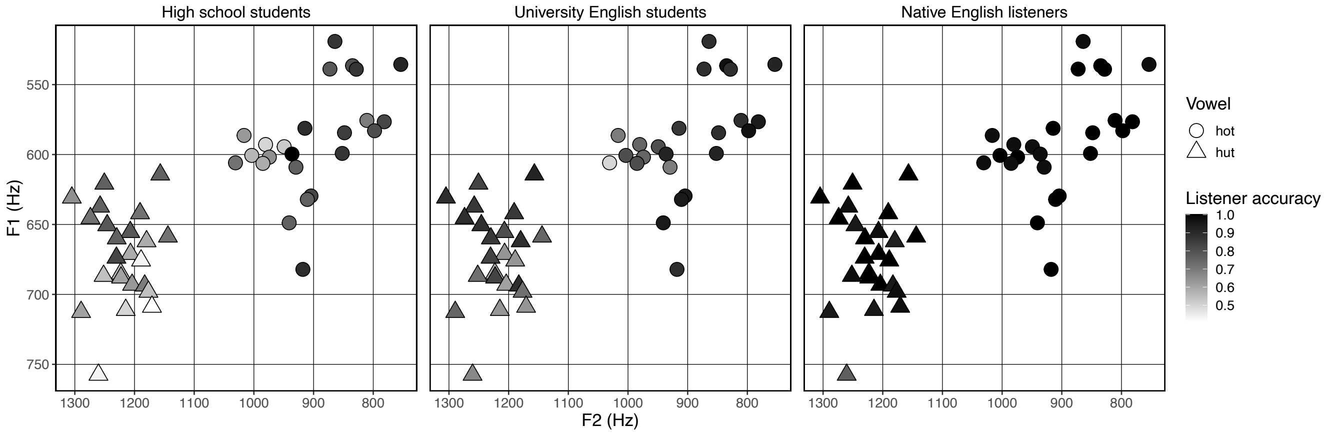
4. DISCUSSION AND CONCLUSION

The present study investigated ND and NE listeners' identification of Standard Southern British English /ɒ/ and /ʌ/. It furthermore compared university English students to high school students to test whether the former constitute an adequately random sample. A difference was found in both comparisons. ND university English students perform worse than NE listeners but better than ND high school students. Thus, University students of English constitute a poor stand-in for the general Danish population, at least when it comes to the present vowel contrast.

The study also examined whether the first and second formant frequency influenced ND listeners' responses. We found that (1) ND listeners identify some tokens of both vowels near ceiling, (2) they identify a few vowel tokens within each category with much lower accuracy, sometimes at chance-level, and (3) low-accuracy tokens tend to cluster in the same areas in F1–F2 space within each category.

When these results are compared to those of NE listeners, it is clear that ND listeners rely on F1 and F2 as cues for the two vowels in slightly different ways from NE listeners.

Foremost, the abrupt shift in accuracy on /ɒ/ at F2 > 945 Hz is only observed with ND listeners. This indicates that ND listeners perceive a category boundary at F2 ≈ 945 Hz. That category could represent an emerging second language vowel. Alternatively, it could represent a pre-existing native language vowel because ND listeners have previously been shown to categorise EN /ɒ/ highly consistently as DA /ɔ/ [7].



(Above) Figure 4. Vowel tokens in F1–F2 space. Dots are *hot* tokens and triangles are *hut* tokens. Each token is shaded according to its average proportion correct score (black is 1), split into the three groups

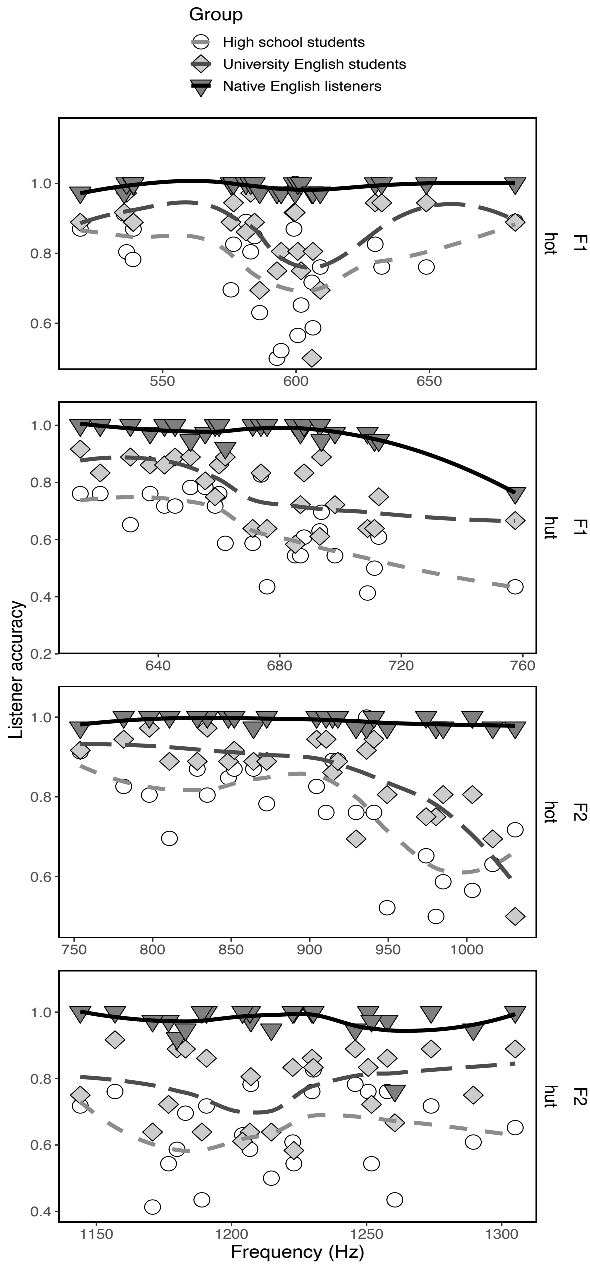


Figure 5. Listener accuracy by formant frequency for each vowel, fitted with separate local polynomial regression lines per group. Note that the label for each plot appears on the right.

It is possible that the acoustic properties of tokens in the /ɒ/–/ʌ/ contrast span multiple ND vowel categories, or even emerging L2 EN vowel categories, at various levels of goodness of fit. In the cross-language identification task reported in [7], ND listeners categorised EN /ʌ/ inconsistently as either DA /ɔ/ or DA /æ/. In the task employed in the present study, chance-level identification for any particular token could indicate one of two things: Either the token is truly intermediate between the two response categories, or it is a poor fit for either of the categories. In the present results, the /ʌ/ tokens that are identified at chance level may be a poor fit for either *hot* or *hut*. These tokens could be perceived as instances of DA /æ/. Future research should explore which categories ND listeners perceive in this area of F1–F2 space by presenting them with tokens from an array of orthogonally varied formant frequencies.

In conclusion, this study demonstrated that ND listeners perceive Standard Southern British English /ɒ, ʌ/ differently than NE listeners according to the acoustic properties of the vowel tokens. More research is needed to determine precisely why. With regard to ND speakers' propensity to neutralise the two vowels in production, the present results make it seem likely that at least part of problem is perceptual in nature.

6. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67 (1). https://doi.org/10.18637/jss.v067.i01.

[2] Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues*

in Cross-Language Research (pp. 171–204). York Press.

[3] Boersma, P., & Weenik, D. (2022). *Praat* (Version 6043) [Windows/MacOS]. http://www.fon.hum.uva.nl/praat/

[4] Bohn, O.-S. (2019). Cross-Language Phonetic Relationships Account for Most, But Not All L2 Speech Learning Problems: The Role of Universal Phonetic Biases and Generalized Sensitivities. In *Approaches to the Study of Sound Structure and Speech*. Routledge.

[5] Bohn, O.-S., & Bundgaard-Nielsen, R. L. (2008). Second language speech learning with diverse inputs. In T. Piske & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 207–218). Multilingual Matters. https://researchdirect.westernsydney.edu.au/islandora/object/uws%3A25770/

[6] Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303–328. https://doi.org/10.1017/S0142716400008912

[7] Bohn, O.-S., & Steinlen, A. K. (2003). Consonantal Context Affects Cross-Language Perception of Vowels. *ICPhS 15*, 4.

[8] Davidsen-Nielsen, N. (1970). *Engelsk Fonetik*. Gyldendal.

[9] Davidsen-Nielsen, N., Færch, C., & Harder, P. (1982). *The Danish Learner*. Antony Taylor.

[10] Kristiansen, A.-M. (2000). *The perception and production of the British English vowel contrast hot-hut by native Danish learners* [Unpublished MA thesis, English Department, Aarhus University].

[11] Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.5.4). https://CRAN.R-project.org/package=emmeans

[12] Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

[13] R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (4.0.4). R Foundation for Statistical Computing. https://www.R-project.org/

[14] Steinlen, A. K. (2005). The influence of consonants on native and non-native vowel production: A cross-linguistic study (Vol. 30). Gunter Narr Verlag.