Factors Influencing Flexibility in Vowel Categorisation

Stephanie Cooper, Brechtje Post

Phonetics Lab, University of Cambridge
sc2346@cam.ac.uk, bmbp2@cam.ac.uk

Abstract

Phonological category boundaries are flexible, and listeners are highly capable of interpreting variation in speech. This is particularly evident in vowel categorisation [1, 2]. Some factors that may facilitate or impede this flexibility are spectral distance from the listener's native vowel category, listener dialect, vowel frontness, and lexical bias [3, 4]. The current project explored the relative impact these factors may have on flexibility of category boundaries, to understand more about tolerance of variation. Continua of VC and CVC-structured sequences were used in a forced choice categorisation task. The stimuli varied in their lexicality; some continua contained only words or nonwords, and some contained both. The continua also varied in vowel frontness. Listeners from two different dialect groups (Standard Southern British English and New Zealand English) took part in the categorisation task. Results showed that lexical bias influences vowel categorisation, but only across an ambiguous boundary and categorising away from an underspecified vowel. This offers support for a range of theories such as Ideal Adapter [5] and Featural Underspecification [6, 7].

Index Terms: vowel categorisation, flexibility, lexical bias

1. Introduction

Listeners accommodate a range of genders, accents, and contexts in speech. They often process sounds from different speakers that may be similar in acoustic features, but in context be intended to target different phonological categories. Somehow, listeners sort variation into these discrete categories with few disruptions to communication. Phonological categories have prototypical members and boundaries, which may be "flexible" to include incoming sounds varying from the prototypes [8].

Vowel categorisation in particular appears to be more flexible than consonant categorisation. This may be because there are fewer discrete events in the spectrum of a vowel than in a consonant; it is therefore harder for listeners to hold a standard representation in their auditory short-term memory for as long [9]. Shaw and colleagues [2, 10] found that listeners were equivalent in their perception of isolated vowels in unfamiliar dialects and in their native dialect, and did not perform at ceiling in any dialect. In similar work [1], we found that New Zealand listeners successfully categorised Greek-accented English KIT[1] vowels (which are raised close to /iː/) whether or not they had prior exposure to the accent. The speech perception system is evidently capable of accepting a range of variant vowels, but it is currently unknown how far this acceptance can stretch.

Many factors may determine the limits of this flexibility. The distance in the vowel space an between an incoming vowel and the listener's native category influences how easily they will accommodate it [12]. However, in a recent experiment, we found that listeners were capable of accommodating variant vowels from a broader range of the vowel space than expected [13]. Listeners also have a strong lexical bias, in which they prefer to hear words over nonwords, and thus will preferentially categorise ambiguous sounds as phonemes leading to word comprehension [3, 4]. Listener dialect is also likely to play a role in determining category boundaries and prototypes. Other factors such as stimuli range and task structure have also been found to influence vowel categorisation [14], but they are beyond the scope of this study.

The present study set out to examine whether lexical bias, dialect spoken and vowel frontness interact with distance from native categories when listeners categorise vowels. Dialects examined were Standard Southern British English (SSBE) and New Zealand English (NZE), as these have the same phonological categories, but some differ in their placement in the vowel space [15, 16]. We also included measures from both front and back vowels, to see whether flexibility would be similar in different parts of the vowel space. We predicted that dialect groups would show peaks of categorisation where their native categories are, and that front and back vowels would show similar levels of flexibility. We also predicted that lexical bias would influence vowel selection, as no matter where on the vowel continuum, listeners would expand categories to include variant vowels when this would lead to perception of a word.

2. Method

2.1. Participants

50 monolingual participants aged 18-39 were recruited from Prolific and personal networks. They had been raised and were living in either the south of England (SE group; n = 25) or Aotearoa/New Zealand (NZ group; n = 25), without living outside of this region for more than 6 months in the past. They identified as speaking English with an accent representative of the region they grew up in and live in.

Participants were paid £5/$10NZD via Wise or Prolific upon completing the experiment.

2.2. Materials

2.2.1. Vowel continua

Eight continua consisting of a range of words and nonwords across front and back vowels were selected for use in this experiment (Table 1). These continua crossed from the highest points in the vowel chart (FLEECE and GOOSE) to lower points (TRAP and LOT). These continua varied based on lexicality; they consisted of either all words ("All"), all nonwords ("None"), words positioned medially ("Medial"), or peripherally ("Peripheral").

A male speaker of SSBE produced all points of the continua. He was raised and currently lives in the south of England.

[1] Wells' [11] lexical set vowels are used here and throughout the text, as cross-dialectal vowel perception is discussed.

Table 1: *Continua across front and back vowels used in experiment. Words are in bold. "Canonical point" indicates the points on the continuum where these canonical productions lie.*

Frontness	Lexicality	Canonical point on continuum			
		1	9	17	25
Front	All	**heed**	**hid**	**head**	**had**
	None	heeb	hib	heb	hab
	Medial	eeg	igg	**egg**	agg
	Peripheral	**feet**	**fit**	fet	**fat**
Back	All	**shoot**	**short**	**shot**	
	None	foop	forp	fop	
	Medial	oob	**orb**	obb	
	Peripheral	**hoop**	horp	**hop**	

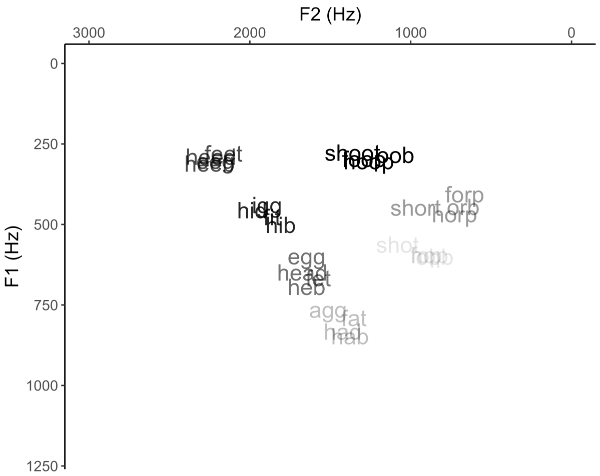

Figure 1: *F1 and F2 of words used to create continua.*

He speaks English as a native language, as well as beginner Russian and intermediate Italian.

Continua between adjacent points were created using Tandem-STRAIGHT [17] with a morphing percentage ranging from 0-100% in 9 equidistant intervals. These were combined to form larger continua between end points (e.g. between *feet* and *fat*). Due to their larger spectral range (Figure 1), front vowel continua contained 25 points between their FLEECE vowel endpoint and their TRAP vowel endpoint, while back vowel continua contained 17 points between their GOOSE vowel endpoint and their LOT vowel endpoint.

2.3. Procedure

The experiment was created using Gorilla Experiment Builder [18]. Participants were asked to use a desktop or laptop in a quiet place, with headphones to hear the auditory stimuli.

Participants read an information sheet and completed a consent form. They completed a short questionnaire to provide basic demographic information and answer questions about accents they hear in their everyday life.

The main task was forced choice, which examined how participants would classify sounds along continua between different words and nonwords. Forced choice was chosen over AXB or Go/No-Go in order to study categorisation, rather than discrimination, with multiple options available. In each trial, participants listened to a sound from a point on one of the continua, and then clicked on the screen which of the options they heard (Figure 2). While not present in the canonical productions, a PALM member of the front vowel continua was added to response options, in case participants perceived it. Response options otherwise reflected the words used to build each continuum, and varied depending on which continuum was presented.

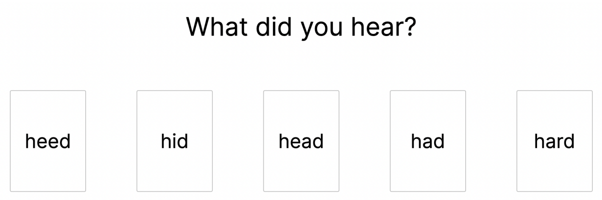

Figure 2: *Example of forced choice task screen. Participants hear a sound along a continuum (heed-had in this instance), and would click one of the buttons on the screen to classify.*

The number of response options provided can affect how participants categorise sounds [14]; we chose to give participants five for front vowel continua and three for back vowel continua to show where on the continua two or more categories may be activated. We did not offer more than these options to minimise distraction from options that would likely never be perceived.

There were 168 trials in total: 25 points per continuum x 4 types of lexicality for front vowels, and 17 points per continuum x 4 types of lexicality for back vowels.

Participants also completed a perceptual acuity task to ensure that they could distinguish differences between sounds without having to classify them. Individual performances in this task and in the categorisation task were compared, and data was excluded from three participants who did not show both strong discrimination and categorisation.

2.4. Analysis

For an overview of all vowels, lexicalities were collapsed and classification curves for each vowel option were visually examined. For the purpose of this paper, only classification of DRESS and THOUGHT variants across different lexical continua were statistically analysed, as these vowels sit at interesting points between the two dialects; the SSBE DRESS vowel is near the NZE TRAP vowel, while NZE and SSBE THOUGHT vowels are more similar [15, 16]. Their placement in the centre of the continua that were used also allow the possibility of analysing the boundaries on either side of the category.

DRESS and THOUGHT data were analysed using a binomial generalised additive mixed model (GAMM) with the *mgcv* package in R [19]. GAMMs are appropriate for non-linear data and can account for both fixed and random effects [20].

The model was formed in a stepdown manner, in which a full model with all possible parametric and smooth terms was built, and then compared to models with one term removed at a time. AIC scores and the compareML function from the *itsadug* package in R [21] were used to compare models.

3. Results

3.1. All vowels

Figure 3 shows categorisation of points on the continua as different vowels (collapsed across all levels of lexicality) represented by members of Wells' lexical set.

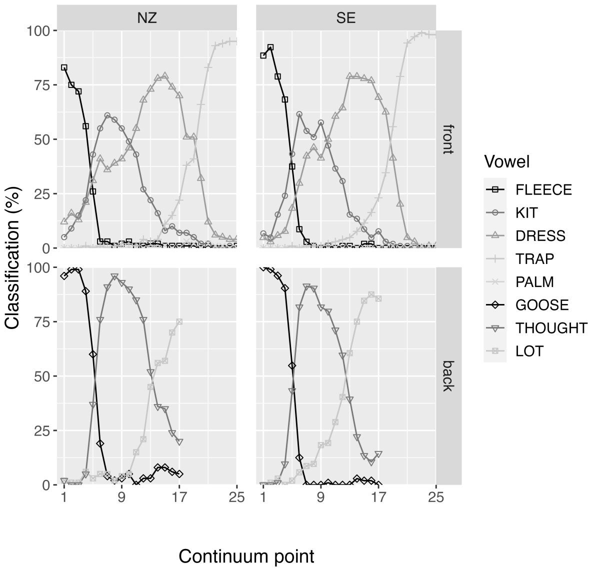

Figure 3: *Categorisation of points along front and back vowel continua as different vowel variants by NZ and SE listeners. Variants are represented by lexical sets. X axis marks indicate canonical productions by male SSBE speaker.*

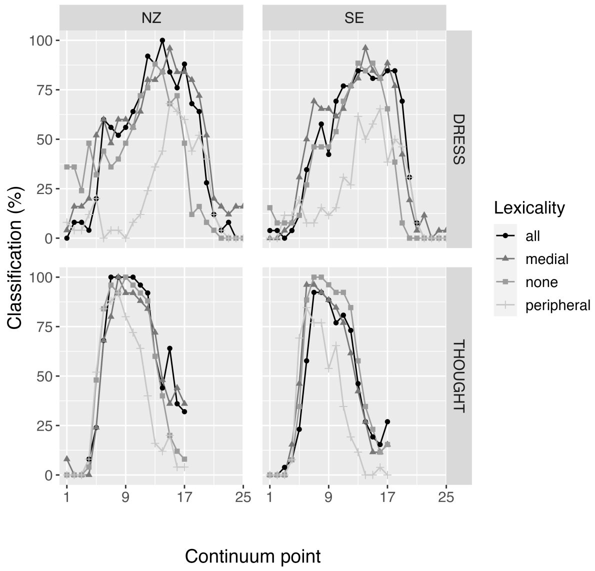

Figure 4: *Acceptance of points on the respective front or back vowel continua as containing* DRESS *or* THOUGHT *vowels by NZ and SE listeners. Responses to different continua with different levels of lexicality are indicated.*

The Dialect groups do not appear to differ greatly in their position and sizes of the categories. In both groups, the vowel categories are not very distinct and boundaries are blurred, both for front and back vowels. The KIT category in particular has overlap with the FLEECE and DRESS categories, and even at its peak of categorisation it is only chosen around 60% of the time.

The steeper the gradient of the curve, the more distinct categories can be assumed to be. The boundary between GOOSE and THOUGHT, for example, appears to be more distinct than between THOUGHT and LOT, as the shift in classification between the vowels occurs over about 4 steps from GOOSE to THOUGHT compared to about 7 steps from THOUGHT to LOT.

3.2. DRESS vowels by lexicality

Figure 4 shows the classification of points as DRESS vowels, with the lexicalities of the vowel continua represented as different lines. A GAMM was fitted to represent this data, and the best fit contained an interaction between dialect spoken and continuum lexicality as a parametric term and as a smooth term, and participant included as a random smooth term.

Visual examination of the model using difference smooth plots for different Dialect*Lexicality conditions shows that for both Dialect groups (example in Figure 5), the Peripheral condition has the most distinctly different shape compared to other lexicality conditions. It differs significantly ($p < 0.05$) from All, None and Medial vowel continua conditions between points 5 and 20, outside of which classification as a DRESS vowel decreases for all lexicality conditions. The Peripheral condition produces a shallower curve with a lower peak of DRESS classification. The Dialect groups are similar to each other in the shape and position of their other lexicality conditions as well.

3.3. THOUGHT vowels by lexicality

Figure 4 also shows the classification of points along the vowel height spectrum as THOUGHT vowels by lexicality of continua. A GAMM was fitted to represent this data, and again the best fit contained an interaction between dialect spoken and continuum lexicality as a parametric term and a smooth term, and with participant included as a random smooth term.

Again, visual examination using difference smooth plots for Dialect*Lexicality conditions showed that the Peripheral condition curve is the most distinctly different for both Dialect groups. It differs significantly ($p < 0.05$ in all instances) from All, None and Medial vowel continua conditions between points 6 and 17, with the exception of the NZ graph where it differs significantly from the None condition only until point 16. Similarly to the DRESS curves, the difference between the Peripheral conditions and others is asymmetrical. In the THOUGHT curves, the Peripheral condition shows a steep increase in classification along the GOOSE-THOUGHT boundary, as the other conditions

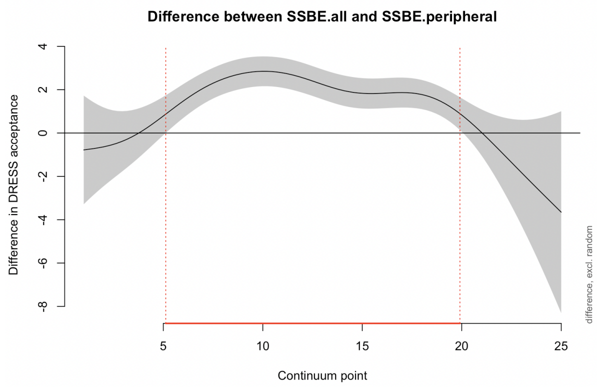

Figure 5: *Example difference smooth plot, here showing difference between SSBE.all and SSBE.peripheral smooths from fitted front vowel model.*

do. However, THOUGHT classification then drops off earlier along the THOUGHT-LOT boundary for the Peripheral condition than it does for the other lexicalities.

The NZ group's back vowel Medial condition does seem to show some difference from the other Medial conditions, as classification of the variants as "orb" stays high throughout the second half of the continuum. However, this also applies to the back vowel All condition, where classification as "short" also stays high. Apart from this NZ-specific pattern, the Dialect groups are again similar in their classification shapes and positioning along the continua.

4. Discussion

This study was an initial investigation into some of the factors that influence flexible vowel categorisation. Participants heard sounds across a range of vowel continua and classified what they heard as a word or nonword option. Similar to previous findings of vowel flexibility [1, 2, 13], we found that at least in English, with its large vowel inventory and wide range of dialects, vowel perceptual categories are not clearly defined and bleed into each other. One given vowel sound may be classified as a member of multiple categories depending on the context it is heard in. However, some category boundaries appear to be more distinct than others.

In the Peripheral lexicality condition, listeners appear to classify points between the two categories as the one which will elicit perception of a word over a nonword, but only across indistinct boundaries. This occurs despite the mid vowels in the Peripheral condition having very similar acoustic qualities to those in the other conditions (Figure 1), thus reducing the likelihood that this can be attributed to other factors like the broad range of the stimuli [14]. Any shifts in classification can therefore be attributed to lexical effects. These top-down influences are most effective when listeners are uncertain about categorisation, as they are more evident across indistinct category boundaries. These influences are also transient and do not generalise to classification across word-word or nonword-nonword boundaries, as similar results for the same categories were not found in All or None lexicality conditions. Lexicality effects were expected, as they are quite pervasive [4], but the lack of effect in the Medial condition suggests an interaction between lexicality and blurry boundaries, as addressed further below.

It is interesting to note that despite documented differences in the vowel spaces of SSBE and NZE dialects, we found minimal difference between groups in the position of their vowel classifications. Additionally, for both groups, the peaks of KIT and DRESS vowels were at a slightly more raised point on the vowel continua (shifted to the left) than the canonical productions by the SSBE speaker. This was expected for the NZ group, who tend to have more raised front vowels [16], but is surprising for the SE group, who would be expected to classify vowels in a way consistent with the SSBE productions. We are currently replicating this experiment with a NZE speaker to further examine possible effects of speaker and listener dialect on these results. One explanation may be that the speaker's vowels are slightly influenced by other languages he speaks [22].

The NZ listeners reflecting SE classification patterns may also be due to the use of a SSBE speaker in stimuli creation. When asked in a concluding questionnaire what accent they thought the speaker had, NZ participants overwhelmingly responded with "British" or "English". Even though the continua extended across the vowel space, so any point could have been a canonical production from any accent, segments like the non-centralised KIT vowel and the low TRAP vowel may have led NZ listeners to correctly identify the accent as different to theirs and like SSBE, which they have a relatively high level of familiarity with. This is consistent with frameworks such as Ideal Adapter [5] which suggests that where possible, listeners select a model based on previous experience to help them to decode the incoming speech signal. NZ listeners in this experiment may have used the salient features above to select an SSBE model to aid in the task, thus shifting their vowel categorisation to be consistent with this model. Our replication with an NZE speaker will also reveal further information about this theory.

These results support previous findings that mid vowel categories are particularly capable of accommodating variation. In a comparison of NZ and American listeners, NZ listeners accepted more variation in their TRAP vowel than American listeners [6]. This was proposed to be due to TRAP being raised to a mid vowel in NZE, and thus lacking the feature specification of [+high] or [+low]. The authors advocate for a Featurally Underspecified Lexicon [7] in which some segments are defined by the absence of features, and thus may be more accepting of variation. The current study did find that FLEECE and TRAP seem to have more distinct vowel boundaries than DRESS, as indicated by their steeper classification curves. NZ listeners did not show as much range in their acceptance of TRAP vowels as found previously, which may again be due to participants identifying the speaker as British and shifting TRAP classification lower.

Featural underspecification may also provide an explanation for lexical effects in the Peripheral but not Medial condition. DRESS and THOUGHT categories are in the middle of the vowel space, so are likely to be the least specified for height. While in other circumstances this may lead to higher acceptance of variation, in the Peripheral condition when classification as a mid vowel led to nonword perception, the underspecification may have shifted classification in the direction of the adjacent category leading to perception of a word (i.e. KIT or LOT). The inverse may not be true; in the Medial condition where classification as KIT or LOT led to a nonword, their height specification may have been strong enough to resist a shift towards DRESS or THOUGHT, despite the lexical benefit of choosing a mid vowel in this condition. The exception to this is in the NZ Medial back vowel continuum, where some THOUGHT classification seems to extend further towards LOT than it does in other conditions (although the All condition shows a similar pattern). This may indicate underspecification of the LOT category in NZE, which is more of a mid vowel than it is in SSBE. It appears that lexical bias and category specifications interact to influence ambiguous vowel classification.

There is much further analysis of this data to be completed. We plan to look more at how lexicality affects other vowels with substantial overlap, like KIT, but also more specified vowels like FLEECE or GOOSE. We also hope to work with EEG to examine differences in neural responses when variants are accepted or rejected as members of a phonological category under different lexical conditions. Previous research has used differences in the N400 or other neural responses to examine how and when variation is accommodated in speech processing [23, 24].

The aim of this study was to explore the weighting of some factors that influence vowel categorisation flexibility. We found that across front and back vowels for two dialect groups, lexicality and specificity of vowel categories influence categorisation curves along a height continuum. These factors also interact in that underspecified categories are more susceptible to lexical bias. This provides a useful foundation for examining vowel flexibility further, and offers many avenues for future research.

5. References

[1] Cooper, S. and Cooper, S., "Exposure-independent comprehension of Greek-accented speech: evidence from New Zealand listeners," in Proceedings of the 20th International Congress of Phonetic Sciences, 6–10, 2023,

[2] Shaw, J. et al., "Resilience of English vowel perception across regional accent variation," Laboratory Phonology, 9(1), 1–36, 2018.

[3] Fox, R.A., "Effect of lexical status on phonetic categorization," Journal of Experimental Psychology: Human Perception and Performance, 10(4), 526–540, 1984.

[4] Gaskell, M.G. and Marslen-Wilson, W.D., " Mechanisms of phonological inference in speech perception", Journal of Experimental Psychology: Human Perception and Performance, 24(2), 380–396, 1998.

[5] Kleinschmidt, D.F. and Jaeger, T.F., "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychological Review, 122(2), 148–203, 2015.

[6] Scharinger, M. and Idsardi, W. J., "Sparseness of vowel category structure: Evidence from English dialect comparison," Lingua, 140, 35–51, 2014.

[7] Lahiri, A. and Reetz, H., "Distinctive features: Phonological underspecification in representation and processing," Journal of Phonetics, 38(1), 44–59, 2010.

[8] Xie, X., Theodore, R.M. and Myers, E.B. "More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories," 43(1), 206–217, 2017.

[9] Pisoni, D.B., "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Perception & Psychophysics, 13(2), 253–260, 1973.

[10] Shaw, J.A. et al., "Revealing perceptual structure through input variation: cross-accent categorization of vowels in five accents of English," Laboratory Phonology, 14(1), 2023.

[11] Wells, J.C., Accents of English: Volume 1. Cambridge University Press, 1982.

[12] Best, C.T., "The emergence of native-language phonological influences in infants: A perceptual assimilation model," in The development of speech perception: The transition from speech sounds to spoken words, 167–224, 1994.

[13] Cooper, S. and Post, B. "Flexibility of vowel categorisation in newly learned words", Colloquium of British Association of Academic Phoneticians, Cardiff, 2024.

[14] Benders, T., Escudero, P. and Sjerps, M.J. "The interrelation between acoustic context effects and available response categories in speech sound categorization", The Journal of the Acoustical Society of America, 131(4), 3079–3087, 2012.

[15] Hughes, A., Trudgill, P. and Watt, D. English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles, Fifth Edition, Routledge, 2012.

[16] Bauer, L., Warren, P., Bardsley, D., Kennedy, M. and Major, G. "New Zealand English", Journal of the International Phonetic Association, 37(1), 97–102, 2007.

[17] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H. "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation", in 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, 3933–3936, 2008.

[18] Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N. and Evershed, J.K. "Gorilla in our midst: An online behavioral experiment builder," Behavioural Research, 52(1), 388–407, 2020.

[19] Wood, S. "mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation", 2023, available at https://cran.r-project.org/web/packages/mgcv/index.html.

[20] Sóskuthy, M., "Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction," arXiv, 2017, http://arxiv.org/abs/1703.05339.

[21] van Rij, J., Wieling, M., Baayen, R.H. and van Rijn, H. "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.", 2022. https://cran.r-project.org/web/packages/itsadug/index.html

[22] Flege, J. "Second language speech learning: Theory, findings and problems," in Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 229—273, Strange, W. (Ed.), York Press, 1995.

[23] Sebastián-Gallés, N., Vera-Constán, F., Larsson, J.P., Costa, A. and Deco, G. "Lexical plasticity in early bilinguals does not alter phoneme categories: II. Experimental evidence," Journal of Cognitive Neuroscience, 21(12), 2343–2357, 2009.

[24] Politzer-Ahles, S., Lin, J., Pan, L. and Lee, K.K. "N400 Evidence That the Early Stages of Lexical Access Ignore Knowledge About Phonological Alternations", Language and Speech, 65(2), 354–376, 2022.