FACTORS DETERMINING PERCEPTUAL AND ACOUSTIC SIMILARITY BETWEEN NATIVE AND NON-NATIVE VOWELS

Please write XXX instead of the name(s) of the author(s)

Please write XXX instead of the affiliation(s)
please write XXX instead of the email address(es)

ABSTRACT

This study investigates perceptual assimilation of Norwegian vowels to Polish vowels, based on a task completed by 15 learners of Norwegian. The perceptual similarity is juxtaposed with the acoustic similarity operationalized as the Euclidean distance. The so-called marked lip rounding (as in the case of Norwegian front and central rounded vowels) and vowel length are examined as factors which may potentially influence perceptual assimilation. AIC comparison suggests that a model including Euclidean distances based on F1 and F2 only is better than a model including F3 as well. Neither vowel length nor marked lip rounding turned out to be significant in predicting assimilation count, but the interaction of lip rounding with Euclidean distance proved to be significant, meaning that for vowels with unmarked lip rounding, there is a stronger effect of Euclidean distance on assimilation. The conclusion for non-native vowel perception research is that marked lip rounding influences assimilation patterns.

Keywords: perceptual and acoustic similarity, non-native vowels, Euclidean distance, Norwegian.

1. INTRODUCTION

So far a number of studies have been conducted to test the perceptual assimilation of non-native vowels (for example, [1]), but relatively few studies have addressed the relationship between vowel perception and acoustic parameters. Strange et al. [2] examined acoustic and perceptual similarity of North German and American English vowels, to find that acoustic similarity did not always predict perceptual similarity, especially for front rounded vowels. Vowel duration differences did not affect perceptual assimilation patterns and spectral differences were more important than duration cues. [3] showed that detailed acoustic comparisons between native (Peruvian Spanish and Australian English) and non-native (Dutch) vowels predicted perception patterns more accurately than overall comparisons of inventory size. Although both [2] and [3] included front rounded vowels and observed peculiar assimilation patterns, they did not specifically investigate the role of lip rounding or F3 in the perceptuo-acoustic relationship and this is the point the present paper aims at addressing.

1.1. Perception of front rounded vowels

In perception studies it has been shown that speakers of languages that do not have front rounded vowels and are not experienced in contrasting them with front unrounded and back rounded vowels, find it challenging to perceive and produce front rounded vowels, which seem to be more marked (i.e. less frequent among world languages, and also more difficult to learn).

[4] showed relatively poor discrimination by inexperienced English listeners of the Norwegian high front unrounded versus outrounded /i–y/ in /bV/ syllables. Further studies [5] tested the perception of Norwegian vowels by English, French and Danish listeners and concluded that the results were largely in line with the phonologically contrastive and non-contrastive phonetic-articulatory properties of the listeners' L1s.

Most of previous research on this topic concentrated on perceptual differentiation of French or German front rounded vowels by English listeners. [6] pointed to the challenges that American English learners of French and American English speakers who had not learned French, had in comparison to native speakers of French when discriminating between French front rounded vowels and when identifying mid front rounded vowels.

In [7] Canadian English listeners had more difficulty discriminating German lax vowels /ʏ/ and /u/ than the tense contrast /yː/ and /uː/. A vowel identification and rating task using English keywords clearly showed that the four German vowels were distinguished from native English categories, but all the vowels were assimilated to English /uː/ and /ʊ/, pointing to the lip rounding being more important in assimilation decisions than the front position of the tongue.

1.2. Norwegian and Polish vowel systems

The two vocalic systems selected for scrutiny in the present study differ considerably. Polish has a simple six-vowel inventory: /i, ɨ, ɛ, a, ɔ, u/ with no distinctions in rounding, tenseness or duration (though some of the vowels have nasalized variants). Norwegian, on the other hand, has a rich a complex vocalic system with duration and rounding contrasts: /iː, i, yː, y, eː, e, øː, ø, ɑː, ɑ, oː, o, uː, u, ʉː, ʉ/, not to mention excessive dialectal variability.

2. THE PRESENT STUDY: RESEARCH QUESTIONS AND PREDICTIONS

L1 Polish listeners learning Norwegian, a language with rounding contrasts, but where rounding is not as strong as for example in German or French, seem to offer a good testing ground for investigating the relationship between perceptual and acoustic similarity and the role of lip rounding and F3. In the present experiment, perceptual assimilation of 16 Norwegian vowels to Polish vowel categories by Polish learners of Norwegian was investigated. Examining perceptual assimilation and juxtaposing it with the Euclidean distances between Norwegian vowel stimuli and reference values for Polish vowels (cf. [8]) permitted four research questions to be asked and predictions to be made.

RQ1: Is the Euclidean distance measure based on F1, F2 and F3 better suited to predict assimilation counts than the Euclidean distance based on F1 and F2 only?

As F3 is related to lip rounding, and lip rounding plays an important role in many Norwegian vowels, including the front high and central rounded vowels, we expect the Euclidean distance measure including F3 to be able to predict assimilation counts more precisely than the Euclidean distance measure including F1 and F2 only.

RQ2: Does the smaller Euclidean distance between a Norwegian vowel stimulus and a Polish vowel category enhance the likelihood of assimilating a given Norwegian vowel to a given Polish vowel?

We hypothesize that acoustic similarity operationalized as the Euclidean distance will correspond to perceptual similarity operationalized as assimilation count.

RQ3: Is the impact of Euclidean distances stronger for vowels without marked lip rounding vs. vowels with marked lip rounding, i.e. /y(ː), ø(ː), ʉ(ː)/?

In the case of front and central rounded vowels, listeners with an L1 that has only less marked combinations of rounding in the case of high and mid back vowels, probably experience tension between choosing the assimilation target on the basis of the standard F1 and F2, or tongue height and advancement or F3 more related to lip rounding, when these features are incongruent in comparison to the standard unmarked configurations familiar from the L1. Therefore, in the case of Norwegian vowels with more marked lip rounding we expect the assimilation results to be influenced by the conflict between the position of the tongue typical for front vowels and lip rounding typical for back vowels in the L1, and less related to the Euclidean distance.

RQ4: Is the impact of Euclidean distances different for short and long vowels?

We hypothesize that vowel length may modulate the perceptuo-acoustic relationship, yet its relative influence remains to be established, as so the results of previous studies are inconclusive (cf. [2] where American English listeners relied on spectral cues when duration and spectral similarity were in conflict, and [9] where desensitization hypothesis emphasizes the role of duration in non-native vowel perception, but does not make claims regarding assimilation).

3. METHODS

3.1. Participants

15 participants, aged 21, 9 females, 6 males, who were 3rd year students in a Norwegian modern language BA programme, participated in the study. They filled in the Leap-Q Language Experience and Proficiency Questionnaire (Marian et al. 2007). Their language inventory included L1 Polish, L2 English and L3 Norwegian. Both L2 English and L3 Norwegian were acquired in a classroom setting. The exposure to Norwegian during the three years prior to the experiment was approximately 17 hours a week of speaking and listening and the subjects spent similar amount of time reading and writing in Norwegian. Their proficiency level in Norwegian was B2.

3.2. Stimuli

The stimuli were recorded by a Norwegian female native speaker and included 16 Norwegian vowels /iː, i, yː, y, eː, e, øː, ø, ɑː, ɑ, oː, o, uː, u, ʉː, ʉ/. They were embedded in nonce words in a /dVd/ framework, e.g. *dåd*, *dedd*, *did*.

3.3. Procedure

Due to the COVID-19 pandemic, the experiment was conducted online in June 2021. In addition to the perception experiment reported on here, the session included a battery of tasks in speech production and syntax. The part devoted to speech perception included assimilation and goodness of fit rating tasks presented to the participants in an online session using PsychoPy [10]. Prior to the audio stimulus, the participants were presented with a fixation point, at the onset of the audio stimulus (one of the 16 Norwegian vowels embedded in a /dVd/ nonce word), the participants saw a screen with a question in

Polish: *Which Polish vowel is the vowel you have just heard most similar to?* The participants were also presented with orthographic labels representing the six Polish vowels, as Polish vowel orthography is transparent, matched with numeric values to press on the keyboard (1, 2, 3 and 5, 6, 7 as these were deemed to be the most convenient to be used with both hands – the participants were instructed to keep the fingers on the keyboard). Once they made their choice, the participants saw another screen asking them to rate how well the vowel they heard fits the category they chose on a 7-point scale.

4. RESULTS, ANALYSIS AND DISCUSSION

4.1 Perceptual assimilation results

Table 1 presents the results of the behavioral tests.

Norwegian stimulus	Polish vowel targets					
	<i> /i/	<y> /ɨ/	<e> /e/	<a> /a/	<o> /ɔ/	<u> /u/
/i:/ TID /	95.6 (5.6)		2.2 (3)			
/i/ FIN	35.6 (4.3)	57.8 (5.1)	6.7 (4.3)			
/y:/ LYS	84.1 (4.6)	11.4 (3.6)			2.3 (1)	2.3 (3)
/y/ SYND	8.9 (3.3)	80 (4.7)			2.2 (1)	8.9 (4.5)
/e:/ STED	2.2 (4)		95.6 (5.1)	2.2 (7)		
/e/ BEST	4.4 (4)		93.3 (5.8)	2.2 (5)		
/ø:/ LØP		20 (4)	24.4 (2.9)		46.7 (3.7)	6.7 (4.7)
/ø/ SØNN		27.3 (4.4)	20.5 (3.2)		40.9 (4.2)	9.9 (4)
/ɑ:/ DAG				100 (5)		
/ɑ/ TAKK				100 (5.1)		
/o:/ RÅD	2.2 (7)				97.8 (5.1)	
/o/ NOK					100 (5.5)	
/u:/ BOK					22.2 (4.7)	77.8 (4.5)
/u/ ROM	2.2 (5)				64.4 (4.6)	33.3 (5.3)
/ʉ:/ GUD		16.7 (4)				83.3 (3.9)
/ʉ/ SLUTT		15.9 (4.3)	2.3 (2)		6.8 (4.3)	75 (4.2)

Table 1: Confusion matrix with mean per cent categorization and goodness rating (in parentheses) of Norwegian vowel stimuli, presented here as phonetic symbols and keywords, in terms of Polish vowel categories, presented here as graphemes and phonetic symbols. The goodness of fit ratings are based on a scale from 1 *unlike* to 7 *identical*.

4.2 Analysis and discussion

Statistical analysis was run in R [11]. For significance testing, initially Poisson regression models with assimilation count (the number of times a given Norwegian vowel was categorized as a given Polish vowel) as the response variable were used, as they are restricted to non-negative values, and assimilation counts cannot be negative. Further, Poisson regression captures the association on a logarithmic scale, which seems appropriate for the data. Since the Poisson models suffered from overdispersion, negative binomial regression models were fitted using the 'glm.nb' function from the *MASS* R package [12].

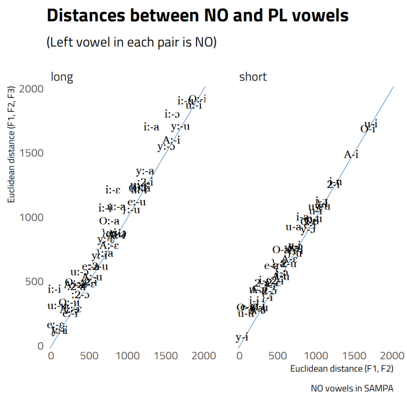

Figure 1: Distances between Norwegian and Polish vowels in terms of Euclidean distances based on F1, F2 and F3, and F1 and F2 only measures. Left vowel in each pair is Norwegian.

We fitted a model to examine assimilation count as a function of Euclidean distance, vowel length, unmarked vs. marked status with respect to lip rounding, and the interaction between the Euclidean distance and marked lip rounding. In order to answer the first research question, whether the Euclidean distance measure based on F1, F2 and F3 is better suited to account for assimilation counts than the Euclidean distance measure based on F1 and F2 only, the analysis was performed for the Euclidean distance defined in terms of F1 and F2 only and F1, F2, and F3. We then compared the models based on the Akaike Information Criterion. AIC comparison led to somewhat unexpected results, in the light of the fact that F3 was expected to be a proxy for lip rounding and enhance the model. Comparing the F1, F2, and F3 model (ac ~ euc_f1_f2_f3 * length) to an F1, F2 model (ac ~ euc_f1_f2 * length), we found that the F1 and F2 only model is lower by three AIC units (440 vs. 443, 7df) so it is a better model. It needs to be noted that both Euclidean distance measures (F1, F2 and F3 and F1 and F2) are highly correlated (r=0.98, p=0.001). For some vowel pairs (those above the diagonal in the plot in Fig. 1) there is a greater F1, F2 and F3 difference than F1 and F2 difference. For many vowel pairs (those on the diagonal in the plot in Fig. 1) there is no difference.

Term	Estimate	p-value
(Intercept)	2.778	<0.05***
euc_f1_f2	-0.001	1
lengthshort	0.139	0.758
non_marked_rounded	0.571	0.452
euc_f1_f2:lengthshort	-0.001	0.553
euc_f1_f2: marked_rounding	-0.002	0.029*

Table 2: Coefficient table of the negative binomial model of the assimilation count. *p*-values of fixed effects calculated with likelihood-ratio tests.

Let us now inspect the negative binomial model of assimilation count in Table 2. Answering the second research question, the negative estimate for euc_f1_f2 indicates that the higher the Euclidean distance between a given pair of vowels, the lower the number of times the Norwegian vowel is predicted to be assimilated to the Polish vowel, but the effect is not statistically significant (p=1).

To answer the third research question and check whether the effect of the Euclidean distance is stronger in the case of vowels with unmarked lip rounding vs. with marked lip rounding, we included marked lip rouding as a predictor in the model. It did not turn out to be significant on its own, but its interaction with Euclidean distance proved to be (b = -0.002, p = 0.029), meaning that for vowels with unmarked lip rounding, there is a stronger effect of Euclidean distance on assimilation count than for vowels with marked lip rounding. Looking at the empirical data in Fig. 2, we can see that usually a given Norwegian vowels is assimilated to the closest Polish category as determined by the Euclidean distance. On the plots in Fig. 2 such cases exhibit a standard decreasing pattern (as for vowels /iː, yː, e/ and /uː/). We can, however, also observe a few examples with rather low assimilation counts in the beginning, a rise afterwards, followed by a drop, as in the case of /ʉ(ː), ø(ː), oː/ and /u/.

As to the fourth research question, there are negative, but statistically insignificant estimates for short vowels (p=0.758) and the interaction between Euclidean distance and vowel length (p=0.553), which means that vowel length does not play a role in perceptual assimilation. This finding is in line with [2], but the duration variable was considered worth checking, because in L1 Polish tested in the present study does not use any duration cues, in contrast to English in [2].

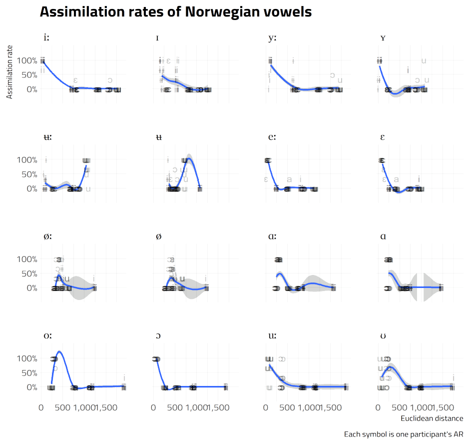

Figure 2: Plots of assimilation rates as a function of the Euclidean distance for each Norwegian vowel stimulus.

5. CONCLUSION

In this experiment, involving Polish advanced learners of Norwegian, a model with Euclidean distances based on the first two formants appeared to be better than a model with the first three formants. The association between Euclidean distances and assimilation counts was not statistically significant, but this finding will need to be further tested among larger pools of listeners and also the level of proficiency will need to be more closely looked at, as both [2, 3] referred to naïve listeners, whereas advanced learners may base their assimilation on other features in addition to simple formant similarity. There were no differences between assimilation predictions for long or short vowels. Finally, Euclidean distance was found to be related to vowels without the so-called marked lip rounding more directly than to front and central rounded vowels, absent from Polish, and more marked/less frequent among world languages. In conclusion, this study contributed to a better understanding of the role of Euclidean distance, marked lip rounding and vowel length in determining the perceptuo-acoustic similarity between foreign and native vowels, and motivates more research to elucidate this intricate phenomenon further.

6. ACKNOWLEDGEMENT

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

7. REFERENCES

[1] Tyler, M. D., Best, C.T., Faber, A., Levitt, A.G. 2014. Perceptual assimilation and discrimination of non-

native vowel contrasts. *Phonetica* 71, 4-21. DOI: 10.1159/000356237

[2] Strange, W., Bohn, O-S., Trent, S.A., Nishi, K. 2004. Acoustic and perceptual similarity of North German and American English vowels. *J. Acoust. Soc. Am*.115, 1791-1807. DOI: 10.1121/1.1687832

[3] Alispahic, S., Mulak, K.E., Escudero, P. 2017. Acoustic properties predict perception of unfamiliar Dutch vowels by adult Australian English and Peruvian Spanish listeners. *Frontiers in Psychology* 8, 52. DOI: 10.3389/fpsyg.2017.00052

[4] Best, C.T., Faber, A., Levitt, A. 1996. Assimilation of non-native vowel contrasts to the American English vowel system. *J. Acoust. Soc. Am*. 99, 2602.

[5] Best, C.T., Halle, P., Bohn, O-S., Faber, A. 2003. Cross-language perpcetion of non-native vowels: Phonological and phonetic effects of listeners' native languages. In: Solé, M.J, Recasens, D., Romero, J.(eds.), *15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3-9, 2003*. ICPhS Archive, 2889-2892.

[6] Gottfried, T.L. 1984. Effects of consonant context on the perception of French vowels. *Journal of Phonetics* 12, 91-114.

[7] Polka, L. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *J. Acoust. Soc. Am*. 97, 1286-1296.

[8] Weckwerth, J., Balas, A. 2020. Selected aspects of Polish vowel formants. In: Wrembel, M, Kiełkiewicz-Janowiak, A., Gąsiorowski, P. (eds), *Approaches to the study of sound structure and and speech: Interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk.* Routledge: 338-348.

[9] Bohn, O-S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In: Strange, W. (ed), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* York Press, 279-300.

[10] Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, Sogo, H., Kastman, E., Lindeløv, J.K. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51, 195-203.

[11] R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[12] Venables, W.N., Ripley, B.D. 2002. *Modern Applied Statistics with S.* Springer.