SPEECH SOUND PERCEPTION IN MONOLINGUALS, BILINGUALS AND LEARNERS – LANGUAGE BACKGROUND AFFECTS IDENTIFICATION AND DISCRIMINATION DIFFERENTLY

Henna Tamminen & Maija S. Peltola

Phonetics and Learning, Age & Bilingualism laboratory (LAB-lab), University of Turku, Finland
henna.tamminen@utu.fi, maija.peltola@utu.fi

ABSTRACT

The categorical nature of speech perception may result in learning difficulties when L2 speech sounds overlap with L1 sound areas. In addition to possible learning difficulties, it may create interesting overlapping phoneme category areas between the bilinguals' two native languages. Here, we tested how language learners, simultaneous bilinguals and monolinguals perceive the same rounded closed /y/-/ʉ/-/u/ vowel area. In learners and bilinguals this area is divided either into two or three categories in their two languages (Finnish and Swedish, respectively), whereas only the three category division is in use in the monolinguals (Swedish). It seems that second language learners' learning process is still ongoing since their /ʉ/–/u/ boundary was less sharp in vowel identification than the other two groups'. On the other hand, both learners and bilinguals seem to benefit from being language oriented and from using two languages in their daily lives as they were more sensitive discriminators than the monolinguals.

Keywords: perception, L2 learning, bilingualism, identification, discrimination

1. INTRODUCTION

Due to categorical perception, which is essential for speech perception, speech sound discrimination is easy at the immediate vicinity of category boundaries, whereas within category discrimination is difficult [16]. The native language speech sound categories are formed early in life [4, 15] and thereafter, speech is perceived through the phonological system of the mother tongue. According to the Native Language Magnet Model (NLM), the prototypical speech sound category representatives pull the nearby sounds towards the centre of the category impeding within category discrimination and creating a hierarchical native language speech sound system [13, 14].

The phoneme categories of two languages may overlap altogether or partially in varying degrees and different speech perception and learning models tackle these discrepancies, for example, through the assimilability or similarity of the foreign sounds to the native language phonemes. The Perceptual Assimilation Model (PAM/PAM-L2) [2, 3] offers four perceptual assimilation patterns according to which the non-native categories may assimilate into native categories. Three of these patterns predict perceptual difficulties. One pattern predicts problems when two non-native categories are assimilated equally well or equally poorly into one native category. Some problems are predicted, in two other patterns, when two non-native categories assimilate unequally into a single native category or when they are non-assimilable. The Speech Learning Model (SLM) [6, 7] describes non-native speech sounds according to their similarity with the native ones. The model predicts severe learning problems when a speech sound of the target language is *similar* to a sound in the mother tongue, whereas minor problems are predicted, in the beginning of the learning process, when the foreign sound is totally *new*, not representing any of the mother tongue sounds.

High proficiency bilinguals, whether simultaneous or sequential, are of special interest in speech perception since their two native languages usually have at least partially overlapping speech sound systems. Hence, a boundary of two phonemes in one of their native languages may locate in the middle of a phoneme category in the other language. This should be problematic since discrimination is challenging within category or when the sounds are *similar* [6, 7] or assimilate differently to the sounds of the other language [2, 3]. However, for example simultaneous bilinguals manage both their native languages without difficulties even though their native languages are intertwined and the other language cannot be switched off when the other one is in use [20]. Foreign language learners, even highly proficient, on the other hand, have two separate systems which can be switched off when only the other one is in use [17, 20]. Speech sound processing takes also more time, due to the intertwined phonological system, in the simultaneous bilinguals compared to the sequential bilinguals [20].

For the second language learners some of the target sounds are bound to be more difficult than others depending on the relationship of the native and foreign languages. By using the challenging speech sound differences in studies on foreign language learning it has, for example, been shown that native-like perception is not achieved through classroom learning [10, 18]. However, the achieved memory traces for non-native phonemes of highly proficient university level language learners can be strengthened through a short term listen-and-repeat training [22]. Hence, even though the subjects were highly proficient, the learning process was unfinished. Further, the continued use of the mother tongue, synchronous with the ongoing learning process, is shown to impede the learning of foreign language production [8, 9]. In contrast to learning in a classroom, learning in a natural environment seems to result in native-like perception [19, 23].

The aim of this study was to see how different language backgrounds, namely monolingual, bilingual and foreign language learner, affect the perception of the same vowel area. The particular vowel area has only one function for the monolinguals but in both bilinguals and language learners the area is divided into two categories in one language and into three in the other. The hypothesis was that the dual set of categories in one area might show some effects on perception in bilinguals and hinder perception especially in language learners.

2. METHODS

2.1. Subjects

The participants were divided into three groups. Group 1 (Learners) consisted of 9 Finnish university students majoring in Swedish, i.e., highly proficient language learners (aged 20–27 years, mean 24.3, 8 females). They had passed demanding entrance examinations to enter the Department of Swedish language at the University of Turku and had then studied Swedish for 1–6 years at the university level. Group 2 (Bilinguals) consisted of 12 Finnish-Swedish bilinguals (aged 16–31 years, mean 20.3, 7 females) who had acquired both languages from birth, one language from one parent, i.e., simultaneous bilinguals. Group 3 (Monolinguals) consisted of 9 monolingual Swedish speakers (Swedish as spoken in Sweden) (aged 20–36 years, mean 27.2, 6 females) who did not study any languages.

All subjects were tested for normal hearing with an audiometer with perceptually relevant frequencies. The study was approved by the Ethics Committee of the University of Turku, Finland.

2.2. Stimuli

The stimuli used in this study formed a continuum of isolated vowels from /y/ to /u/ varying in the second formant (F2). The stimulus continuum consisted of 18 vowels so that F2 ranged from 606Hz (703Mel) to 2077Hz (1553Mel) in 50Mel steps. This vowel area is divided into two vowels in Finnish, namely /y/ and /u/, and into three vowels in Finland-Swedish and in Swedish, namely /y/, /ʉ/, and /u/. The stimuli are described in detail in [20, 21]. According to NLM [13], PAM [2], and SLM [6, 7], this vowel area should be problematic to language learners and, on the other hand, it serves an interesting area for research on simultaneous bilinguals as well.

2.3. Procedure and analysis

First, in an identification task (ID), the subjects were asked to identify the 18 vowels as /y/, /ʉ/, or /u/ by pressing the appropriate symbol on a numpad. In order to ensure correct use of the symbols, the participants were given example words containing these sounds. The stimuli were presented ten times each in random order. The forced choice identification task was self-paced. On the basis of the individual results, we selected two stimulus pairs for an oddball discrimination task so that one pair was within the category /ʉ/ and one pair crossed the /ʉ/–/u/ category boundary so that one stimulus represented /ʉ/ and one represented /u/. The pairs were 100Mels apart from each other. The individual stimulus selection was done in order to make sure that the stimuli represented either one category or two different categories for each individual, since the exact location of the category boundary may vary between subjects and average stimuli may not represent intended categories for all subjects.

In the oddball discrimination task, one block was the between category block and the other was the within category block. The stimulus with higher F2 always functioned as the standard, whereas the one with lower F2 functioned as the deviant in both blocks. The subjects were asked to press a button as soon as they heard the deviating stimulus. There were 130 standards and 20 deviants in each block. All communication, including instructions for both tasks, were given in Swedish.

Category boundary location and boundary consistency were measured on the basis of the ID test data. The ID data was subjected to logit transformation analysis in SPSS to locate the point where the answers were distributed evenly to both categories and to obtain the steepness value for the boundary cross-over point. The category boundary and steepness values were then separately subjected to a Multivariate analysis (MANOVA, Group (3) × Boundary (2) or Steepness (2); post hoc Tukey tests when appropriate) (SPSS). The oddball discrimination task was carried out to measure discrimination sensitivity (d') and reaction time (RT). The hits, misses, correct rejections and false alarms were used in calculating the *d'* values and the RTs were measured from deviant stimulus onset to button press. Button presses within ±3 standard deviation were included in the analysis. Also the *d'* and RT values were separately subjected to a Multivariate analysis (MANOVA, Group (3) × *d'* (2) or RT (2); post hoc Tukey tests when appropriate).

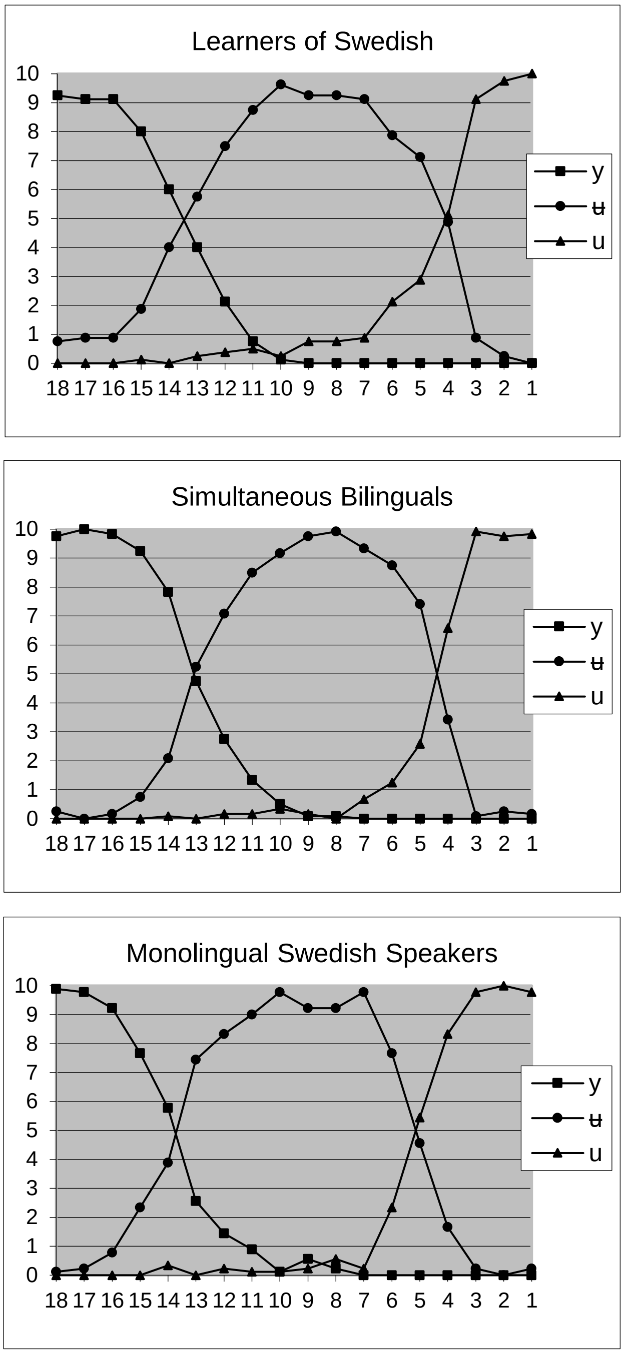
3. RESULTS

There was a significant difference between the groups in how they located the /ʉ/–/u/ boundary in the identification task and in the consistency of this boundary, as shown in the main effects of group $F(2,26)=3.920$, $p=0.033$ and $F(2,26)=5.821$, $p=0.008$, respectively. The post hoc (Tukey HSD) comparisons revealed that the consistency of the /ʉ/–/u/ boundary was statistically different between the students of Swedish and the bilinguals ($p= 0.007$). This is clearly seen in the mean steepness values as well, see Table 1 for comparison of the two groups. When we compared the steepness values of the category boundaries per group, we found that the only significant difference was within the simultaneous bilinguals ($t(11)=3.763$, $p=0.003$), indicating that one boundary was placed more systematically than the other. Table 1 shows that the /ʉ/–/u/ boundary was the more systematic one. The identification results in full are presented in Table 1 and Fig. 1.

Table 1: Mean, minimum, and maximum category boundary location and boundary steepness values for Learners (Group 1), Bilinguals (Group 2), and Monolinguals (Group 3). Standard deviations are in brackets.

Group	Boundary	Mean	Min/Max
1	/ʉ/–/u/	4.52 (0.80)	3.30/5.60
	/y/–/ʉ/	13.66 (1.41)	11.89/16.08
2	/ʉ/–/u/	4.59 (0.76)	3.51/5.94
	/y/–/ʉ/	12.86 (1.23)	10.41/14.70
3	/ʉ/–/u/	5.51 (1.00)	4.60/7.50
	/y/–/ʉ/	13.71 (0.96)	12.46/15.00
Group	**Steepness**	**Mean**	**Min/Max**
1	/ʉ/–/u/	1.56 (0.55)	0.95/2.62
	/y/–/ʉ/	1.55 (0.78)	0.55/2.78
2	/ʉ/–/u/	2.70 (0.93)	1.15/4.47
	/y/–/ʉ/	2.01 (0.54)	1.29/2.95
3	/ʉ/–/u/	2.07 (0.55)	1.00/2.82
	/y/–/ʉ/	1.69 (0.70)	0.59/2.44

Figure 1: ID scores for the Learners, Bilinguals, and Monolinguals. The *X*-axis shows the 18 vowel continuum where number 18 is the most /y/ like and number 1 the most /u/ like. The *Y*-axis shows the occasions the vowel was identified as a member of /y/, /ʉ/, or /u/ category (max 10 times).



Discrimination sensitivity in the between category situation was statistically different between the groups as shown in the main effect of group $F(2,26)=4.641$, $p=0.019$. Further, the post hoc (Tukey HSD) comparisons revealed that the Swedish speakers' discrimination sensitivity differed statistically significantly from both students' *d'* ($p=0.049$) and from bilinguals *d'* ($p=0.026$). Further, as Table 2 shows, the native Swedish speakers were less sensitive in discriminating the /ʉ/–/u/ boundary stimuli than the other two groups. There were no statistically significant differences between the groups concerning the RTs. The *d'* and RT data are presented in Table 2.

Table 2: Mean, minimum, and maximum RT (ms) and *d'* values in the within category and between category situations for Learners (Group 1), Bilinguals (Group 2), and Monolinguals (Group 3). Standard deviations are in brackets.

Group	RT	Mean	Min/Max
1	within /ʉ/	556 (77)	462/686
	/ʉ/–/u/	546 (76)	439/634
2	within /ʉ/	484 (76)	379/671
	/ʉ/–/u/	471 (115)	337/768
3	within /ʉ/	523 (124)	382/694
	/ʉ/–/u/	532 (112)	412/711
Group	***d'***	**Mean**	**Min/Max**
1	within /ʉ/	4.54 (0.19)	4.06/4.61
	/ʉ/–/u/	4.44 (0.31)	3.72/4.61
2	within /ʉ/	4.34 (0.30)	3.68/4.61
	/ʉ/–/u/	4.45 (0.29)	3.70/4.61
3	within /ʉ/	4.27 (0.41)	3.44/4.61
	/ʉ/–/u/	3.87 (0.71)	2.28/4.61

4. DISCUSSION AND CONCLUSIONS

The present study was designed to see how Finnish university students majoring in Swedish, simultaneous Finnish-Swedish bilinguals and monolingual Swedish speakers perceive the closed rounded vowel area covering /y/, /ʉ/, and /u/ in Swedish. More importantly, the aim was to compare the three groups' identification and discrimination performance.

The identification task revealed that all three groups were able to categorise the vowel area into three categories, and that the Finland-Swedish and Sweden-Swedish closed rounded vowel areas are quite similarly identified as /y/, /ʉ/ and /u/ within our vowel continuum. The more fronted position of the Sweden-Swedish /ʉ/ [e.g., 1, 5] is not shown in the behavioural ID task which may very well be due to the forced-choice nature of the task and a fixed continuum of the 18 vowels. Despite the similar identification of the three groups, the language learners of Swedish had a less sharply defined /ʉ/–/u/ category boundary than the other two groups. This may imply that the students had not learned to perceive the target language sound system well enough to be consistent in identifying the stimuli of the foreign language. They also naturally still use their native language which may have affected the learning process [8, 9]. The simultaneous Finnish-Swedish bilinguals and Swedish monolinguals, on the other hand, were native speakers of the target language and hence, they were more consistent in identifying their native language vowels. The monolingual speakers positioned between the other two groups regarding the steepness of the vowel boundaries, simultaneous bilinguals were most stable in the identification whereas learners were least stable. This may further indicate that bilinguals need to be systematic in keeping their two languages separate, whereas monolinguals only have one language and hence there is room for hesitation, and the learning process is not yet finished for the language learners which shows as hesitation.

The Swedish /ʉ/ is *similar* [6, 7] to the Finnish /u/, or /y/, and Swedish /ʉ/ and /u/ assimilate unequally [2, 3] into the Finnish /u/ and hence, the Swedish /ʉ/ is problematic for the Finnish learners of Swedish. The bilinguals do not come across any learning problems due to these differences but naturally the vowels need to be kept apart from each other, which is possible due to the native language magnets [13]. The less sharp /ʉ/–/u/ boundary in the learners may indicate the predicted problems, whereas the bilinguals and monolinguals had sharp boundaries as they identified according to their native languages.

The between category discrimination sensitivity data revealed that the native monolingual Swedish speakers were different from the other two groups: they were less sensitive in discriminating the vowels in comparison with the bilinguals and language learners. Even though individually selected stimuli were used, this may be an indication of the more fronted Sweden-Swedish /ʉ/, since the forced-choice ID task somewhat steers the identification and the more fronted /ʉ/ does not show in the ID result. Perhaps more importantly, it seems that the learners and the bilinguals might be more sensitive in discriminating these speech sounds as they are more language oriented than monolinguals and use at least the two languages more or less daily. They probably benefit from the fact that they are language oriented, as has also been shown by Immonen and colleagues [11] in children and by Jähi and colleagues [12] in elderly learners.

To conclude, it seems that the unfinished learning process affects the identification systematicity in the language learners so that they are more hesitant than simultaneous bilinguals and monolinguals. However, the continuous use of the two languages in both the language learners and the simultaneous bilinguals seems to be beneficial, since in comparison with monolinguals, these two types of bilinguals are sensitive discriminators.

5. ACKNOWLEDGEMENTS

We wish to thank all the subjects who participated in this study and both M.A. Heidi Toivonen and M.A. Minna Paavola for their assistance with data collection.

6. REFERENCES

[1] Asu, E. L., Schötz, S., Kügler, F. 2009. The acoustics of Estonian Swedish long close vowels as compared to Central Swedish and Finland Swedish. *Proc. FONETIK 2009* Stockholm, 54–59.

[2] Best, C. T., Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–330.

[3] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S., Munro, M. J. (eds), Language experience in second language speech learning: In honor of James Emil Flege. Amsterdam: John Benjamins, 13–34.

[4] Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Aljo, K., Näätänen, R. 1998. *Nature Neuroscience* 1, 351–353.

[5] Ewald, O., Asu, E. L., Schötz, S. 2017. The formant dynamics of long close vowels in three varieties of Swedish. *Proc. Interpeech 2017* Stockholm, 1412–1416.

[6] Flege, J. E. 1987. The production of ”new” and ”similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

[7] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 233–277.

[8] Flege, J. E., Frieda, E. M. Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25, 169–186.

[9] Flege, J. E., MacKay, I. R. A. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26, 1–34.

[10] Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., Brattico, E. 2014. Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioural and ERP study. *Frontiers in Human Neuroscience* 8, 1–14.

[11] Immonen, K. Peltola, M. S. 2018. Finnish children producing English vowels – Studying in an English immersion class affects vowel production. *Journal of Language Teaching and Research* 9, 27–33.

[12] Jähi, K., Alku, P., Petola, M. S. 2015. Does interest in language learning affect the non-native phoneme production in elderly learners? *Proc. 18th ICPhS* Glasgow. Paper number 0234.1–5.

[13] Kuhl, P. K. 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93–107.

[14] Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 979–1000.

[15] Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608.

[16] Liberman, A. M., Harris, K. S., Hoffman, H. S., Griffith, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368.

[17] Peltola, M. S., Aaltonen, O. 2005. Long-term memory trace activation for vowels depends on the mother tongue and the linguistic context. *Journal of Psychophysiology* 19, 159–164.

[18] Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R. 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters* 352, 25–28.

[19] Peltola, M. S., Kuntola, M., Tamminen, H., Hämäläinen, H., Aaltonen, O. 2005. Early exposure to non-native language alters preattentive vowel discrimination. *Neuroscience Letters* 388(3), 121-125.

[20] Peltola, M. S., Tamminen, H., Toivonen, H., Kujala, T., Näätänen, R. 2012. Different kinds of bilinguals – Different kinds of brains: The neural organisation of two languages in one brain. *Brain & Language* 121, 261–266.

[21] Tamminen, H. Peltola, M. S., Toivonen, H., Kujala, T., Näätänen, R. 2013. Phonological processing differences in bilinguals and monolinguals. *International Journal of Psychophysiology* 87, 8–12.

[22] Tamminen, H. Peltola, M. S. 2015. Non-native memory traces can be further strengthened by short term phonetic training. *Proc. 18th ICPhS* Glasgow. Paper number 0285.1–5.

[23] Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36(5), 638-642.