THE CONTRIBUTION OF TEMPORAL CUES TO PERCEIVED NATIVENESS IN THE NATIVE SPEECH OF ENGLISH MIGRANTS TO AUSTRIA

Ineke Mennen[1], Ulrich Reubold[1], and Robert Mayr[2]

[1]University of Graz, [2]Cardiff Metropolitan University
ineke.mennen@uni-graz.at, johann.reubold@uni-graz.at, rmayr@cardiffmet.ac.uk

ABSTRACT

Learning a second language can affect pronunciation in the first language, and may result in the impression of non-nativeness. The role of segments and prosody in perceived non-nativeness is unknown. We investigated the role of temporal features to perceived non-nativeness in the L1 speech of English migrants to Austria, using a design where segmental and temporal cues are swapped. Results revealed that the relative importance of temporal features in listeners' impression of non-nativeness depends on whether the segmental string comes from monolingual (i.e. non-attrited) or bilingual (i.e. attrited) speakers. In the former situation, listeners attend to the transferred temporal information and judge the manipulated stimuli as less native. However, when segments originate from bilingual speakers, listeners ignore the transferred temporal information. This suggests that there needs to be some correspondence between segmental and prosodic information for listeners to attend to temporal information in their judgements of nativeness[i].

Keywords: L1 phonetic attrition, temporal cues, durationional cues, prosodic manipulation.

1. INTRODUCTION

The languages used by bilingual individuals are in constant interaction with each other (e.g. [1]). At the phonetic level, such interaction typically leads to foreign accented speech, where a speaker's second language (L2) accent retains perceivable traces of an individual's native language (L1) accent (see [2] for an overview). Far less is known about the opposite effect, i.e. that the L2 can exert a long-term influence on a bilingual's L1 pronunciation with traces of the L2 system occuring in a speaker's L1 accent. This phenomenon is usually referred to as L1 phonetic attrition (e.g. [3]). Listeners are very sensitive to such L2-induced changes to L1 pronunciation: accent rating studies show that listeners often detect a non-native accent when judging the L1 pronunciation of late-sequential bilinguals who have been long-term exposed to an L2 [4, 5, 6].

There is ample evidence for L2-induced changes to L1 pronunciation in a wide range of segmental and prosodic areas of speech production. Changes at the segmental level have been observed for among others voice onset time in plosives, formants of vowels and laterals, rhotics, and sibilants (see [7], for a recent overview). At the prosodic level, changes have been observed in the realisation of prosodic prominence, pitch range, the choice and frequency of use of intonation patterns, and how pitch accents are timed in relation to segments (see [8] for an overview).

Only few studies have tried to determine what cues listeners use when judging individuals as non-native in their L1. [9] asked monolingual Spanish listeners to rate short Spanish speech samples produced by Spanish English bilingual speakers living in the UK and indicate which features they associated with non-nativeness. Listeners predominantly listed segmental features; there were fewer comments on prosody. [10] examined the salience of different speech cues in the L1 of English migrants to Austria using a similar methodology. Contrary to [9], their results showed that listeners associated prosodic and segmental features approximately equally often with non-native speech. Given these contradictory findings, and the fact they were based on comments by listeners who may not have been able to verbalize what they based their judgments on, it remains unclear to what extent prosodic cues play a role in judgements of nativeness.

This study investigates the role of temporal features to listeners' impression of non-nativeness in the L1 speech of English migrants

to Austria. We use the prosody transplantation or morphing paradigm [11], where we swap segmental and temporal cues, so that the two levels are disentangled and their role in perceived non-nativeness can be examined. In our study, we created stimuli where the segments of monolingual native speakers of Standard Southern British English (SSBE) who live in England are transplanted ('morphed') onto the temporal (durational) features of late-sequential English-Austrian German bilinguals (i.e. English migrants to Austria), and vice versa. The resulting stimuli were then presented to monolingual English listeners and judged in an accent rating task. So far, no studies have used this method for assessing L1 attrited speech. However, transferring native temporal cues to L2-accented speech has been found to improve foreign accented ratings [12, 13]. If one assumes that listeners use similar cues in rating L1 naiveness and in rating L2 accentedness, then the prediction is that the accentedness ratings will improve when monolingual temporal cues are transferred to BIL speech but worsen in the reverse scenario.

2. METHODS

2.1. Speakers

Two groups of speakers participated in this study: (1) English-Austrian German bilingual speakers (BIL, N = 7, 3 females, 4 males), who were raised as monolingual speakers of SSBE and moved to Austria in adulthood where they acquired Austrian German as an L2; (2) monolingual speakers of SSBE residing in England (MON, N = 7, 3 females, 4 males) who have never lived outside England and reported no more than high-school level knowledge of other languages.

Six of the bilingual participants were rated for their degree of perceived nativeness in an earlier global accent rating experiment [10] and were perceived as moderately accented in their L1, with an average score of 3 on a 6-point accentedness scale ("1"="certainly native" – "6"="certainly non-native") and a range of segmental and prosodic features were identified by listeners as deviant from the L1 norm. The remaining bilingual participant was recruited at a later stage, but was also perceived (by a native SSBE listener) as moderately accented with both segmental and prosodic features contributing to this perception.

2.2. Sentence material

Twelve sentences were used to create the stimuli for this experiment. The choice of these sentences was based on the fact that the sentences consisted of a variety of statements and questions, and provided plenty of scope for the speakers to produce a variety of segmental and prosodic cues. The audio recordings of the sentences were automatically segmented using text-to-phoneme conversion and forced-alignment algorithms [14] after which they were checked and manually corrected where necessary.

2.3. Stimuli

In order to create the stimuli, we paired the MON and BIL speakers in such a way that their versions of each sentence used for morphing contained the same number of syllables and that the speakers were similar in voice quality, median pitch, and speech tempo (in terms of the total durations of the paired utterances, cf. Fig.1). This was done to counteract a possible negative influence on the quality of the morphed stimuli. This resulted in 7 MON-BIL pairs.

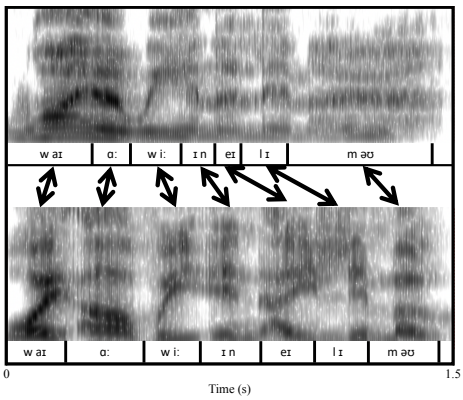

Figure 1: Duration morphing on syllable level of sentence "Why are we in a limo?" produced by a MON (top) and a BIL speaker (bottom).

We then created two sets of stimuli from the (sound level normalized) readings of the 12 sentences by the speakers: (1) unmorphed speech, which was, however, resynthesized to make sure that the listeners heard manipulated speech in both sets; (2) temporally morphed speech in which durations of syllables are

adjusted by means of the PSOLA algorithm [15][ii]. For the latter, we segmented the speech signals into syllables, measured their durations, calculated for each MON/BIL syllable pair the proportions dur_{MONO}/dur_{BIL} and vice versa, and used these as factors in duration tiers in the *To Manipulation* procedure in *Praat* [16]. This resulted in 336 stimuli, i.e. 168 stimuli per condition.

2.4. Listeners and procedure

We recruited 40 monolingual listeners (20 per condition) living in England through Prolific (https://www.prolific.co), an online research platform providing help with the recruitment of participants for online experiments (16 females, 4 males for the unmorphed condition; 18 females, 4 males for the duration morphed condition). Mean age for the unmorphed speech condition was 42 years (range 20-69), mean age for the duration morphed condition was 31 years (range 18-57). The experiment was presented to the listeners on Qualtrics [17]. Listeners were asked to use their desktop or laptop computers and a headphone for the experiment and given instructions to ensure adequate volume settings. They were informed that they would hear samples from fluent English speakers and would see an orthographic transcript of the sentence that was spoken. Samples were played in random order, and listeners were asked after hearing each sample whether it was spoken by a native or non-native speaker of English. They then had to indicate how confident they were of their choice (uncertain, semi-certain, certain). These two ratings were combined into a 6-point accentedness scale ranging from "1" = "certainly native" to "6" = "certainly non-native" – a method which is commonly used in studies on L1 attrition of speech [4, 5, 9, 10].

The experiment lasted approximately 35 minutes and listeners were paid a small fee for their participation. They could listen to each stimulus three times. Listeners were told that the samples they would hear may sound a bit artificial, that there was no right or wrong answer, and that they should respond intuitively. In order to familiarize listeners with the manipulated samples and experimental set-up, we presented three practice stimuli, randomly selected from the experimental stimuli.

2.5. Data analysis

Two measures were taken to examine whether the BIL and MON showed production differences in the temporal domain of the unmanipulated data: speech rate and VarcoSyll [18, 19], as speech rate and rhythm were expected to show cross-language differences [20, 21]. Speech rate was measured as the number of uttered syllables per second. VarcoSyll was computed as the standard deviation of a syllabic interval duration divided by the mean syllabic interval duration, multiplied by 100. It was computed for non-phrase final syllables only [22] to avoid pre-phrase boundary lengthening effects. We used paired t-tests to statistically examine the production differences in the temporal domain, unless otherwise mentioned.

To test whether potential production differences between the groups in speech rate or VarcoSyll could predict perceived nativeness in the unmanipulated stimuli, we applied linear models to the speaker-specific means of each measure and the ratings, both separately for MON and BIL, as for all speakers combined.

The accent ratings were tested for interrater reliability by means of Fleiss' Kappa [23], using the *R* package *irr* [24]. The 6-point accentedness scale was converted to an ordered factor, which then was the dependent variable in a Cumulative-Link Mixed Model using the function *clmm()* from the *R* package *ordinal* [25]. *Speaker Base* (*MON* vs. *BIL*) and *Duration* (*MON* vs. *BIL*) were treated as fixed factors, *Listeners* and *Utterance* as random factors (with random intercepts and slopes for each combination of random:fixed effects). Post-hoc comparisons were made with *emmeans* [26].

3. RESULTS

3.1. Production of temporal measures

As shown in Fig. 2a, BIL have a significantly lower speech rate in the unmorphed stimuli than MON ($t[6] = 3.0$, $p < 0.05$). No main effect was found for VarcoSyll ($t[6] = 0.3$, n.s.). However, as Fig. 2b suggests, variability of this measure is considerably higher in BIL as compared to the rather consistent MON. This was confirmed by an F-Test ($F[6,6] = 0.08$, $p < 0.01$).

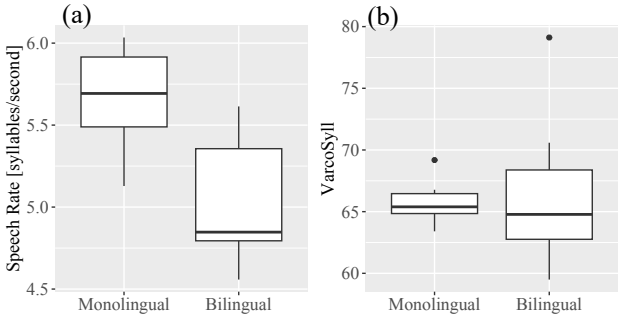

Figure 2: (a) Boxplots of speech rate and (b) VarcoSyll against Group.

3.2. Perceived non-nativeness in L1

We first tested the interrater reliability of the listeners' binary responses of (non-)nativeness. Fleiss' Kappa for the listener ratings of unmanipulated stimuli was 0.48 ($z = 85.7$, $p < 0.001$); for manipulated stimuli it was 0.43 ($z = 773$, $p < 0.001$). This constitutes 'moderate agreement' in both cases [27]. For the 6-point scale responses, Fleiss' Kappas were expectedly lower (unmanipulated: 0.26, i.e. 'fair agreement'; $z = 82.7$, $p < 0.001$); manipulated: 0.19, i.e. 'slight agreement'; $z = 64.2$, $p < 0.001$)).

Linear models showed that nativeness ratings (6-point scale) of unmanipulated stimuli could not be predicted from VarcoSyll, neither for the groups separately, nor combined. However, speech rate (Fig. 3b) significantly correlated with nativeness ratings for BIL ($F[1,5] = 12.89$, $p < 0.05$; *Adj.* $R^2 = 0.66$), MON ($F[1,5] = 8.4$, $p < 0.05$; *Adj.* $R^2 = 0.55$), and all speakers combined ($F[1,12] = 45.7$, $p < 0.001$; *Adj.* $R^2 = 0.77$).

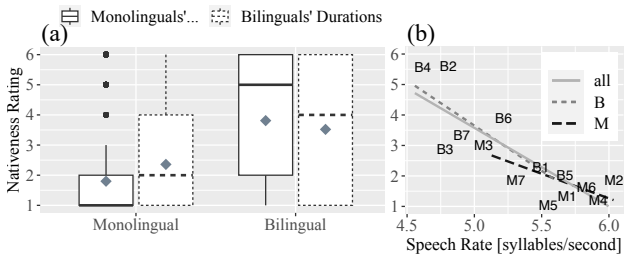

Figure 3: (a) Nativeness ratings and effect of duration on MON vs BIL speech. Diamonds show mean ratings; (b) nativeness ratings as a function of speech rate; for MON, BIL, and combined (all), with numbers indicating pairs.

Results of a Cumulative-Link Mixed Model (Fig. 3a) showed an effect of *Speaker Base*, with *BIL* rated as less native than the *MON* ($\chi^2[1] = 36.1$, $p < 0.001$). Also *Duration* did elicit significant changes in nativeness ratings ($\chi^2[1] = 6.0$, $p < 0.05$). There was a significant interaction between *Duration* and *Group* ($\chi^2[1] = 10.5$, $p < 0.01$). Post-hoc comparison showed a significant effect of *Duration* for the *MON Speaker Base* only (i.e., the manipulation led to more *non-native* responses: $p < 0.001$), but not for the *BIL Speaker Base* (n.s.). Under both conditions of *Duration*, MON and BIL *Speakerbases* differed significantly ($p < 0.001$), with the latter always being perceived as less native than the former.

4. DISCUSSION AND CONCLUSION

Results of the production data showed that BIL spoke the L1 more slowly and had a tendency of having a less stable rhythm compared to MON L1 speakers. This suggests an L2-induced influence, as Austrian German is reported to have a relatively slow speech rate [21] and research suggests that its rhythm may be situated more toward the syllable-timed end of the rhythm continuum than SSBE [20]. Moreover, the differences in speech rate were found to correlate with listeners' judgments of nativeness of the unmanipulated stimuli: slower speech sounds less native to listeners.

The influence of transferring durational cues, however, only partially confirmed our predictions. While transferring BIL durational cues to MON speech makes individuals sound less native, transferring MON duration cues to BIL speech did not have the expected positive effect on nativeness ratings. This suggests that the influence of durational cues depends on the degree to which other cues, such as segments and intonation, are native. When they conform to the native norm, as in the MON participants, durational cues influence perceptions of nativeness. However, when other cues deviate from L1 norms, as is the case in the BIL participants, the improved durational cues are simply ignored by listeners. This shows that segmental cues are more important than temporal ones, with the latter only considered when no other cues speak against them. This is in line with previous work on L1 perception of L2 speech showing that "there needs to be some degree of correspondence between segments and prosody for small deviances in prosody to be perceived" [28, p. 522]. Our results show that much the same holds true for nativeness perception in L1 attrited speech.

5. REFERENCES

[1] Green, D. W. 1998. Mental control of the bilingual lexico-semantic system. *Bil Lang Cogn* 1, 67–81.

[2] Edwards, H., Zampini, M. L. 2008. Phonology and Second Language Acquisition. Amsterdam: John Benjamins.

[3] Major, R.C. 2010. First language attrition in foreign accent perception. *Int J Bil* 14, 163–83.

[4] Bergmann, C, Nota, A., Sprenger, S.A., Schmid, M.S. 2016. L2 immersion causes non-native-like L1 pronunciation in German attriters. *J Phon* 58: 71–86.

[5] de Leeuw, E., Schmid, M.S., Mennen, I. 2010. The effects of contact on native language pronunciation in an L2 migrant context. *Bil Lang Cogn* 13: 33–40.

[6] Hopp, H. Schmid, M.S. 2013. Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Appl Psycholinguist* 34, 361–94.

[7] Reubold, U., Ditewig, S., Endes, K., Mayr, R., Mennen, I. 2021. The effect of dual language activation on L2-Induced changes in L1 speech within a code-switched paradigm. *Languages* 6: 114.

[8] Mennen, I., Reubold, U., Endes, K. & Mayr, R. 2022. Plasticity of native intonation in the L1 of English migrants to Austria. *Languages*, 7(3), 241.

[9] Mayr, R., Sánchez, D., Mennen, I.. 2020. Does teaching your native language abroad increase L1 attrition of speech? The case of Spaniards in the United Kingdom. *Languages* 5: 41.

[10] Ditewig, S., Reubold, U., Mayr, R., Mennen, I. Under review. The relation between perceived non-native features and deviations from the L1 norm in adult sequential bilinguals.

[11] Boula de Mareüil, P., Vieru-Dimulescu, B. 2006. The contribution of prosody to the perception of foreign accent. *Phonetica 63*, 247–267.

[12] Polyanskaya, L., Ordin, M., Busa, M.G., 2016. Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Lang Speech* 60 (3), 333–355.

[13] Van Maastricht, L., Zee, T., Krahmer, E., Swerts, M. 2021. The interplay of prosodic cues in the L2: How intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Comm* 133, 81090.

[14] Kisler T., Reichel, U.D., Schiel, F. 2017. Multilingual processing of speech via web services. *Comput Speech Lang* 45, 326–47.

[15] Moulines, E. & Charpentier, F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Commun* 9, 453–467.

[16] Boersma, P., Weenink, D. 2006. Praat (Version 6.2.06). https://www.praat.org.

[17] Qualtrics. 2022. Qualtrics XM PlatformTM. Computer Program. Available online: http://www.qualtrics.com.

[18] Lai, C., Evanini, K., Zechner, K. 2013. Applying rhythm metrics to non-native spontaneous speech. *Proc. of SLaTE*, 159–163.

[19] Leemann, A. Kolly, M-J., Nolan, F., Li, Y. 2018. The role of segments and prosody in the identification of a speaker's dialect. *J Phon*. 68, 69–84.

[20] Kelly, N. Variability of rhythm across dialects: Austrian German. Unpublished Abstract. https://www.academia.edu/31492968/Variability_of_rhythm_across_dialects_Austrian_German.

[21] Kleber, F., Jochim, M., Klinger, N., Pucher, M., Schmid, S., Zihlmann, U. 2021. Sprechgeschwindigkeitsunterschiede zwischen den natonalen hochsprachlichen Varietäten Deutschlands, Österreichs und der Schweiz. Talk presented at the Österreichische Linguistik Tagung.

[22] Gibbon, D., Gut, U. 2001. Measuring speech rhythm. *Proc. of Eurospeech*, 95-98.

[23] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters, *Psychol. Bull.*, 76 (5), 378–382.

[24] Gamer, M., Lemon, J., Fellows, I., Singh, P. 2019. irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1.

[25] Christensen, R. H. B. (2019). Ordinal: Regression Models for Ordinal Data. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

[26] Lenth, R. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.1-1, <https://CRAN.R-project.org/package=emmeans>.

[27] Landis, J. R. & Koch, G.G. 1977. The measurement of observer agreement for categorical data, *Biometrics*. 33 (1) 159–174.

[28] Ulbrich, C., Mennen, I. 2016. When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *Int J Bil* 20: 522–49.

[i] We would like to thank the Austrian Science Fund (FWF) for their financial support under grant number P33007-G.

[ii] While the durational morphing transferred the proportional duration of syllables from MON to BIL speakers and vice versa, there is some overlap with speech rate (measured as the number of syllables per second), as we selected the MON/BIL pairs to have a similar speech tempo (in terms of the total durations of the paired utterances).