Determinants of English accents

Martijn Wieling*, Jelke Bloem†, R. Harald Baayen‡§ and John Nerbonne*¶
*University of Groningen, The Netherlands
†University of Amsterdam, The Netherlands
‡University of Tübingen, Germany
§University of Alberta, Canada
¶Freiburg Institute for Advanced Studies, Germany

***Abstract*—In this study we investigate which factors affect the degree of non-native accent of L2 speakers of English who learned English in school and mostly lived for some time in an anglophone setting. We use data from the Speech Accent Archive containing over 700 speakers speaking almost 160 different native languages. We show that besides several important predictors, including the age of English onset and length of anglophone residence, the linguistic distance between the speaker's native language and English is a significant predictor of the degree of non-native accent in pronunciation. This study extends an earlier study [1] which only focused on Indo-European L2 learners of Dutch and used a general speaking proficiency measure.**

I. Introduction

There is much research investigating the factors affecting the success someone has in mastering a second language (L2). Piske et al. [2] review a great deal of literature and note that the age of onset (the age at which one begins learning), length of residence (the time spent in a country where the language is dominant), and the amount of second language use may affect how nativelike pronunciations ultimately become. Furthermore, the difference between the speaker's L1 (first language) and L2 has been found to be a relevant factor [3].

Chiswick and Miller [3] developed a measure of linguistic distance by assessing how difficult it is for the speaker to learn the language. However, as Schepens et al. [1] mention, their measure is problematic due to motivational differences for speakers of different languages. Consequently, both Van der Slik [4] and Schepens et al. [1] used more objective quantitative measures of linguistic distance. Van der Slik [4] used a linguistic distance measure on the basis of the proportion of shared cognates (adapted from [5]) and he showed that this lexical measure significantly predicted the scores of L2 speakers of Dutch on the State Examination of Dutch as a Second Language (the STEX II test). While he only included 11 West European languages, Schepens and colleagues [1] extended this set to a total of 35 Indo-European languages. In addition, Schepens et al. [1] used two other lexical distance measures. One measure was based on the branch length of the Indo-European language tree of Gray and Atkinson [6]. The other distance measure they used was based on the ASJP database (Automated Similarity Judgment Program; [7]). The ASJP distances were calculated using LDND [8], a variant of the Levenshtein distance [9]. In contrast to the approach of Gray and Atkinson [6] in constructing the tree, the ASJP distances distinguish between more and less similar cognates. Schepens et al. [1] found that both methods were significant predictors of the scores of L2 speakers of Dutch on the STEX II test, with the Gray and Atkinson [6] method being best.

This study aims to investigate factors influencing the pronunciation performance of L2 speakers of English having learned English in school. In line with Schepens et al. [1] we also assess the influence of linguistic distance between the L1 and L2 (i.e. English) for each speaker. In contrast to their study, however, we additionally include speakers with a non-Indo-European background. Furthermore, rather than using a general measure of speaking proficiency assessed via a lengthy test (evaluating several aspects of speaking, including content, choice of words, vocabulary, coherence, syntax, speed, register and pronunciation), we use *automatic* judgments of how nativelike the speech is of the various L2 speakers. A final difference of our study compared to that of Schepens et al. [1] is that we focus on the pronunciation performance of English as an L2, in contrast to Dutch.

II. Material

Our dataset consists of data from the Speech Accent Archive (SAA [10]). The SAA is available at http://accent.gmu.edu and contains a large set of speech samples in English from people with various language backgrounds. Each speaker reads the same paragraph containing 69 words in English:

Please call Stella. Ask her to bring these things with her from the store: six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Speaker-specific information was collected and includes native language (people who are balanced bilinguals are excluded), other languages spoken, country of birth, age, gender, age of English onset (AEO; i.e. defined as being the age when first exposed to sustained English language input), cumulative residence length in an English-speaking country (LR), and learning style (i.e. naturalistic or academic). All speech samples are transcribed using the International Phonetic Alphabet. In 2010, we extracted all available transcribed samples from the Speech Accent Archive including speaker information. In this study, we focus on 712 adult non-native speakers who learned English in an educational setting (only a minority of speakers learned English in a naturalistic setting). There were slightly more men (383: 53.8%) than women (329: 46.2%) in this dataset. The average age of these speakers was 32.5 (SD: 12.1), and the average age of English onset was 12.2 (SD: 6.6). The mean residence length in an English-speaking country was

6.8 years (SD: 10.7) and the total number of unique native languages of these speakers was 159.

In addition to this data, we obtained the gross national income (per inhabitant), and the average number of years of eduction in 2011 ([11]: Statistical annex) for the country of birth of the speaker. Since the Gray and Atkinson [6] language tree only included Indo-European, we determined the linguistic distance between the native language of each speaker and English on the basis of the ASJP database [7]. (Note that tree-based distance measures are mostly unsuitable for our dataset, as our dataset contains languages from various major language families which are generally not combined in a single language tree.) The ASJP database contains the transcribed pronunciations (in a simplified phonetic alphabet) of 40 words from the Swadesh list for about half of the world's languages.

III. Methods

A. Determining the linguistic distance between languages

To evaluate the lexical distance between a pair of languages, the ASJP provides the implementation of a measure called LDND (Levenshtein Distance Normalized Divided), wich is an adaptation of the Levenshtein distance [9]. The Levenshtein distance (LD) measures the minimum number of insertions, deletions and substitutions to transform one string into the other and is a popular method in dialectometry [12] to assess the pronunciation similarity between different dialects. The difference between the LD and the LDND is (i) the normalization by length (frequently employed in dialectometry as well), and (ii) a normalization to correct for chance differences in two words with the same meaning, effectively dividing the raw distance by the mean distance between words with a different meaning. Wichmann et al. [13] compared the normalized LD (i.e. LDN) and LDND and concluded the latter was better able to assess relatedness between language pairs. More recently, Jäger ([14], [15]) developed and evaluated an alternative distance measure which employs weighted alignment (i.e. substitution costs may vary depending on the sound pairs involved; see next subsection) on the basis of Pointwise Mutual Information (PMI; [16]). Jäger [14] found that his PMI-based method resulted in a better language classification than the LDND method. In the following, we will evaluate the influence of all three measures of linguistic distance (on the basis of the ASJP dataset): LDN, LDND, and Jäger's PMI-based method.

Note that the word lists to which the three measures are applied consist of non-cognate words whenever the languages are unrelated, meaning that the measures are sensitive to lexical differences. Only in the case of related languages will each measure capture the pronunciation differences between the two languages.

B. Automatically determining nativelikeness of speakers

As we do not have nativelikeness ratings for all of the 712 pronunciations, we calculate these automatically. For this we use another modified version of the Levenshtein distance (LD) algorithm. The following example shows that the regular Levenshtein distance between two accented pronunciations of the word Wednesday, [wɛnzdeɪ] and [wɛnəsde] is 3:

w	ɛ	n		z	d	e	ɪ
w	ɛ	n	ə	s	d	e	
			1	1			1

The Levenshtein distance has been successfully used for comparing pronunciations in dialectometry ([17], [18]) and matches perceptual dialect distances well [19]. Unfortunately, and as can be seen above, the basic Levenshtein distance algorithm is quite crude and only distinguishes same from different sounds (i.e. substituting two completely different sounds, such as [u] and [ɛ] is not distinguished from substituting two more similar sounds such as [u] and [o]). To make the pronunciation comparison procedure more linguistically sensible, Wieling, Prokić and Nerbonne [20] proposed a method to incorporate (automatically obtained) sensitive sound distances in the Levenshtein distance algorithm and showed that this approach improved the alignment quality significantly. The procedure is based on calculating the Pointwise Mutual Information (PMI [16]) and works by counting how often two segments correspond in alignments and comparing this to how often they would correspond by chance. Segments which correspond more frequently than would be expected are assigned a low distance, while the distance is high for segments which correspond less frequently than expected. Wieling, Margaretha and Nerbonne [21] showed that the underlying sound (vowel) distances were linguistically sensible and corresponded well to acoustic vowel distances, with correlations ranging from $r = 0.63$ to $r = 0.76$ for several datasets. Applying this method to our example alignment yields the following associated costs (and a total pronunciation distance between the two pronunciations of 0.081):

w	ɛ	n		z	d	e	ɪ
w	ɛ	n	ə	s	d	e	
			0.031	0.020			0.030

Wieling et al. [22] showed that the PMI-based Levenshtein distance is a valid measure of how nativelike accented pronunciations are. Using audio samples from the Speech Accent Archive, they obtained human nativelikeness ratings for 286 speech samples. In their study, 1143 participants judged 41 speech samples on average, resulting in consistent judgments (Cronbach's alpha: 0.85). For each of the 286 distinct transcribed speech samples, the PMI-based Levenshtein distance was calculated with respect to the transcriptions of 115 speech samples of native American English speakers. Subsequently, these 115 distances were averaged and represented the distance from that speaker to the "average American English speaker". Wieling et al. [22] reported a correlation between the PMI-based Levenshtein distance and the human nativelikeness judgments of $r = -0.78$ ($p < 0.001$). When log-transforming the Levenshtein distances, this correlation increased to $r = -0.81$ ($p < 0.001$). The correlation is negative as higher nativelikeness implies a lower pronunciation distance. Given that this correlation was also very close to how well individual raters agreed with the average nativelikeness ratings ($r = 0.84$, $p < 0.001$ [22]), the PMI-based Levenshtein distance can be used as a valid measure of non-nativelikeness (i.e. the strength of foreign accent).

We use Levenshtein distance as a measure of pronunciation difference in accents because it is sensitive to all the segmental variations that accents are associated with, not merely typical ones such as [t]:[θ] or [s]:[θ], and because it is sensitive to the frequency with which segments are inserted, deleted or modified. It is thus a global measure of difference in segmental realization. In the following we will use this measure of foreign accent strength as our dependent variable.

It might appear that we use one variant of the Levenshtein distance (e.g., LDN or LDND) to predict another (i.e. the PMI-based Levenshtein distance). However, the predictor is based on the 40-word AJSP set and is sensitive to the lexical differences between English and the foreign language under scrutiny, while the dependent variable is the difference between the average American English pronunciation and the English pronunciation of the foreign speaker, where no lexical differences arise.

C. Mixed-effects regression

To evaluate the influence of the various predictors on the computed nativelikeness, we conduct a mixed-effects regression analysis, taking into account the potentially important predictor variables (outlined above) and the structural variability linked to the native language and country of birth of each speaker. For conducting the analysis, we use the `R` package `mgcv` (version 1.8.8) [23], as this also allows us to assess potential non-linear dependencies between the predictors and the dependent variable. The optimal model is determined via AIC (Akaike Information Criterion [24]) comparisons, offsetting the goodness of fit against the complexity of the model. A decrease in AIC indicates a better fitting model (given the increased complexity), and we only opt for a more complex model if it results in an AIC decrease of at least 2.

IV. RESULTS[1]

We used the log-transformed PMI-based Levenshtein distance between the average American English speaker and each of the 712 non-native English speakers as our dependent variable. While our results showed clear support for the inclusion of country of birth as a random-effect factor (i.e. having structural variability associated with it), this was not the case for native language (yielding no AIC decrease). This might seem strange at first sight, but several speaker-related and language-related predictors were already included in the model. Furthermore, there can be much variation between speakers of the same language in different countries. For example, Canadian French is markedly different from continental French [25] and their English accents may be as well.

Table I shows all significant factors and covariates of the best model for our data on the basis of 711 speakers. One speaker was excluded, as the residuals of the initial fitted model revealed a single extreme outlier during the model criticism phase [26]. Excluding this outlier increased the explained variance of our model from 35.3% to 36.8%. Furthermore, the residuals of our model followed the normal distribution. To compare the relative effect of each predictor fairly, we added a measure of effect size by specifying the increase or decrease of the dependent variable when the predictor increased from its minimum to its maximum value (following the approach of [26]).

Predictor	Estimate	Std. Error	p-value	Effect size
(Intercept)	-4.8160	.00869	$< .001$	
Length of residence	-.0055	.00103	$< .001$	
Age	.0021	.00068	.002	
Length of residence : Age	.0001	.00004	.008	.24
Age of English onset	.0099	.00102	$< .001$	.45
Nr. of language spoken	-.0194	.00576	$< .001$	-.10
Education (in years) per country	-.0197	.00301	$< .001$	-.22
LDND	.0037	.00091	$< .001$	.21

TABLE I. FIXED-EFFECT PREDICTORS OF THE OPTIMAL (LD) MODEL

Besides these fixed-effect predictors, the random-effects structure only consisted of a random intercept per country (no random slopes were supported by the data). Excluding this random intercept reduced the explained variance to 30.8%. We assessed if there were additional interactions, but none improved the fit of the model shown in Table I. In addition, no non-linearities were observed in the data.

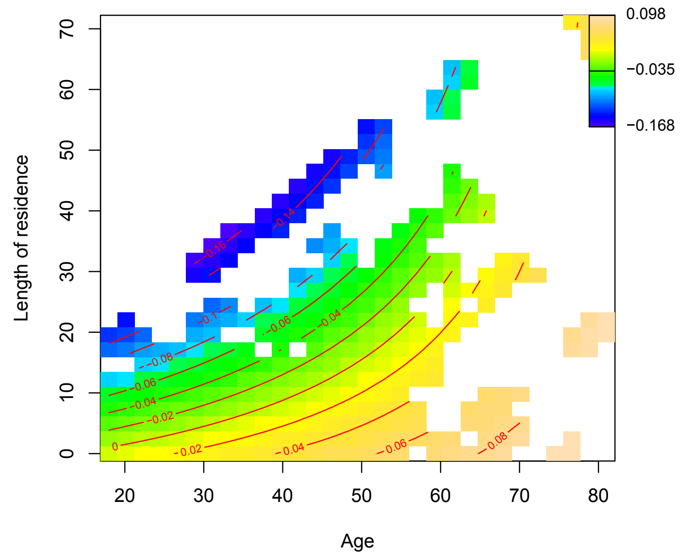

Fig. 1. Interaction between age and length of residence in their effect on non-nativelikeness in pronunciation.

Speaker age and length of residence interacted significantly as can be seen in Figure 1. Darker (i.e. blue and green) colors in this graph represent pronunciations which are more similar to average native American English, whereas a lighter (i.e. yellow and pink) color indicates the opposite. Clearly the beneficial effect of length of residence is dependent on age. For younger speakers a longer length of residence has the strongest effect (e.g., for a speaker aged 20, the contour lines are only a short distance apart), whereas the effect is smaller for older speakers (e.g., for a 50-year-old speaker the lines are further apart). The significance of the individual predictors in Table I indicates that the effect of length of residence is significant for the mean value of speaker age, and vice versa.

The strongest effect was found to be associated with age of English onset. Earlier learners had a more nativelike performance than later learners. Furthermore, we observed a significant effect of the number of languages spoken besides English. The more languages spoken, the more nativelike the pronunciation of the speaker. In addition, the average number of years of education per country was a significant predictor, with speakers from countries with longer average education having a more nativelike American English pronunciation. As this variable correlated highly, $r = 0.84$, with the average

[1] For reproducibility, the data, methods and results have been made available at the Mind Research Repository: http://openscience.uni-leipzig.de/index.php/mr2/article/view/124.

gross national income, this measure does not necessarily reflect education only, but also incorporates the wealth of a country. We did not observe a significant effect of gender.

The linguistic (i.e. lexical) distance between the speaker's native language and English played a significant role in predicting how nativelike a pronunciation was, and the LDND measure appeared to be the best predictor (i.e. using either LDN or Jäger's PMI-based measure resulted in a higher AIC value). Note, however, the high correlation between the three different measures (all r's > 0.96). By itself the LDND distance explained 9.0% of the variation in the data.

A. Validation

To validate that the results we observed were not contingent on the computational measure used as our dependent variable, we fitted a similar model using the average nativelikeness ratings from the study of Wieling et al. [22]. This reduced the number of speakers to 237, with 87 different languages. The same predictors were generally found to be significant as for the other model. There were only a few differences: (1) the random intercept for country of birth was not necessary for this model, (2) the LDN measure appeared to be a (slightly) better predictor than the LDND measure, and (3) the p-value for the effect of age (given the average length of residence) dropped below the significance threshold ($p = 0.06$). However, AIC comparisons revealed that the interaction between age and length of residence was still necessary, and showed a similar pattern as shown in Figure 1. The explained variance of this model was 39.3%, out of which 12.5% was explained by the linguistic distance (on the basis of LDN) between the native language of the speaker and English. In sum, the results using both dependent variables were very similar.

V. Discussion

In this study we have shown that nativelikeness of L2 speakers of English was significantly associated with various predictors, including the linguistic distance between their L1 and English. This result is in line with the result of Schepens et al. [1], who also found a signficant effect of linguistic distance on the variation in L2 speaking proficiency scores. In our study, however, the dependent variable (i.e. nativelikeness) was determined automatically, rather than requiring a lengthy test. Furthermore, our sample of speakers also included those with a non-Indo-European language background, and therefore our results appear to apply more generally.

Other predictors we found to affect the nativelikeness of the pronunciation included age of English onset, length of residence in an English-speaking country, age of the speaker, the number of additional languages spoken by the speaker (besides their native language and English) and the average number of years of education per inhabitant of the country of birth of the speaker. As mentioned by Muñoz [27], "[T]here is still a lack of empirical evidence to date confirming that, after the initial stages of foreign language learning, younger starters overtake older starters in school settings.". While our results suggest that younger learners do overtake older ones with respect to pronunciation, our dataset did not contain information about other relevant factors, such as the amount of exposure of English during education, and teacher proficiency [28]. Our finding that a longer time spent in an English-speaking country improved the nativelikeness of the speaker's pronunciation is in line with previous research (see, for example, [2]). Figure 1 shows that this effect was less beneficial for older speakers, approximately indicating that an earlier stay in an English-speaking country is more beneficial than moving to an English-speaking country at a later age (in line with the effect of age of onset in an immersion context [2]). The number of languages spoken had a positive influence on nativelikeness. If this predictor is interpreted as a measure of language aptitude, this finding is in line with DeKeyser [29], who proposed that language aptitude is relevant for explicit L2 learning. Finally, the average number of years of schooling in a country was positively related to nativelikeness. This is also in accordance with previous research, where the amount of input (for which this predictor is a very rough proxy) has been found to be important for L2 learning [28].

In sum, our results confirm that obtaining a nativelike pronunciation is dependent on various factors, including the similarity between the native language and the second language. Importantly, both this relationship and the pronunciation performance of an L2 speaker can be quantified by comparing the pronunciations of a small set of words in the different languages.

Acknowledgments

We thank Søren Wichmann and Eric Holman for supplying the ASJP LDND distances between English and the other languages, and Gerhard Jäger for supplying the corresponding distances on the basis of his PMI-based distance measure.

References

[1] J. Schepens, F. Van der Slik, and R. Van Hout, "The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language," in *Comparing approaches to measuring linguistic differences*, L. Borin and A. Saxena, Eds. Berlin: Mouton De Gruyter, 2013, pp. 199–230.

[2] T. Piske, I. R. MacKay, and J. E. Flege, "Factors affecting degree of foreign accent in an L2: A review," *Journal of Phonetics*, vol. 29, no. 2, pp. 191–215, 2001.

[3] B. R. Chiswick and P. W. Miller, "Linguistic distance: A quantitative measure of the distance between English and other languages," *Journal of Multilingual and Multicultural Development*, vol. 26, no. 1, pp. 1–11, 2005.

[4] F. W. Van der Slik, "Acquisition of Dutch as a second language," *Studies in Second Language Acquisition*, vol. 32, no. 03, pp. 401–432, 2010.

[5] A. McMahon and R. McMahon, *Language classification by numbers*. Oxford University Press, 2005.

[6] R. D. Gray and Q. D. Atkinson, "Language-tree divergence times support the Anatolian theory of Indo-European origin," *Nature*, vol. 426, no. 6965, pp. 435–439, 2003.

[7] S. Wichmann, A. Muller *et al.*, "The ASJP database (version 16)."

[8] E. W. Holman, "Programs for calculating ASJP distance matrices (version 2.1)."

[9] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.

[10] S. H. Weinberger and S. A. Kunath, "The Speech Accent Archive: towards a typology of English accents," *Language and Computers*, vol. 73, no. 1, pp. 265–281, 2011.

[11] UNDP, *Sustainability and Equity: A Better Future for All*. New York: Palgrave Macmillan, 2011.

[12] M. Wieling and J. Nerbonne, “Advances in dialectometry,” *Annual Review of Linguistics*, vol. 1, no. 1, pp. 243–264, 2015.

[13] S. Wichmann, E. W. Holman, D. Bakker, and C. H. Brown, “Evaluating linguistic distance measures,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 17, pp. 3632–3639, 2010.

[14] G. Jäger, “Phylogenetic inference from word lists using weighted alignment with empirically determined weights,” *Language Dynamics and Change*, vol. 3, no. 2, pp. 245–291, 2013.

[15] ——, “Support for linguistic macrofamilies from weighted sequence alignment,” *Proceedings of the National Academy of Sciences*, 2015. [Online]. Available: dx.doi.org/10.1073/pnas.1500331112

[16] K. W. Church and P. Hanks, “Word association norms, mutual information, and lexicography,” *Computational Linguistics*, vol. 16, no. 1, pp. 22–29, 1990.

[17] B. Kessler, “Computational dialectology in Irish Gaelic,” in *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*. Morgan Kaufmann Publishers Inc., 1995, pp. 60–66.

[18] W. J. Heeringa, “Measuring dialect pronunciation differences using Levenshtein distance,” Ph.D. dissertation, University of Groningen, 2004.

[19] C. Gooskens and W. Heeringa, “Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data,” *Language Variation and Change*, vol. 16, no. 3, pp. 189–207, 2004.

[20] M. Wieling, J. Prokić, and J. Nerbonne, “Evaluating the pairwise string alignment of pronunciations,” in *Proceedings of the EACL 2009 workshop on language technology and resources for cultural heritage, social sciences, humanities, and education*. Association for Computational Linguistics, 2009, pp. 26–34.

[21] M. Wieling, E. Margaretha, and J. Nerbonne, “Inducing a measure of phonetic similarity from pronunciation variation,” *Journal of Phonetics*, vol. 40, no. 2, pp. 307–314, 2012.

[22] M. Wieling, J. Bloem, K. Mignella, M. Timmermeister, and J. Nerbonne, “Measuring foreign accent strength in English,” *Language Dynamics and Change*, vol. 4, no. 2, pp. 253–269, 2014.

[23] S. Wood, *Generalized additive models: an introduction with R*. CRC press, 2006.

[24] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[25] D. C. Walker, *The pronunciation of Canadian French*. University of Ottawa Press Ottawa, ON, 1984.

[26] R. H. Baayen, *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press, 2008.

[27] C. Muñoz, “Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning,” *Applied Linguistics*, vol. 29, no. 4, pp. 578–596, 2008.

[28] S. Unsworth, L. Persson, T. Prins, and K. De Bot, “An investigation of factors affecting early foreign language learning in the Netherlands,” *Applied Linguistics*, 2014. [Online]. Available: dx.doi.org/10.1093/applin/amt052

[29] R. M. DeKeyser, “The robustness of critical period effects in second language acquisition,” *Studies in Second Language Acquisition*, vol. 22, no. 4, pp. 499–533, 2000.