

PRIDE AND PREJUDICE?

Judging non-native pronunciation of English

TON KOET* & HUUB VAN DEN BERGH**

* University of Amsterdam ** Utrecht University

Abstract

Can non-native judges of language proficiency be trusted? To answer this question, we asked Dutch (N = 70) and English (N = 57) listeners to judge English speech produced by Dutch speakers, and we asked Dutch listeners to judge Dutch speech produced by these speakers. Results showed that the Dutch listeners to the English speech produced by the Dutch speakers are stricter about pronunciation than the English listeners; they also arrive at rankings that are different from the rankings arrived at by the English listeners. We have found that the Dutch listeners, when judging the aesthetic quality and intonation of Dutch speakers' pronunciation of English, are influenced by aspects of the speakers' native language which interfere with the target language. Their judgements of English pronunciation may therefore be affected by their impressions of the speakers' Dutch.

Keywords: evaluation, L2, pronunciation, L1-influence

1. INTRODUCTION

There can be little doubt about the importance of a learner's native language in the process of becoming proficient in a foreign language. It is evident in the case of Dutch speakers of English. The scores of native speakers of Dutch on the 2016 International English Language Testing System (IELTS) tend to be in higher bands than those of native speakers of other languages (IELTS, 2018). The Netherlands occupied top rankings in the 2016 and 2015 EF English Proficiency Ranking, based on the results of more than one million test takers (EF, 2017). Although the 2016 report does not provide information on the native languages and does not mention the native language to account for the success of inhabitants of the four top ranking countries,

1

Koet, T. & van den Bergh, H. (2018). Pride and prejudice? Judging non-native pronunciation of English. Contribution to a special issue in honor of Gert Rijlaarsdam Making Connections: Studies of Language and Literature Education. L1-Educational Studies in Language and Literature, 18, 1-13. <https://doi.org/10.17239/L1ESLL-2018.18.03.01>

Corresponding author: Ton Koet, De Horst, Kennemerstraatweg 150, 1851 AV Heiloo, The Netherlands. Email: koetagm@hetnet.nl

© 2018 International Association for Research in L1-Education.

citing “key economic and social indicators” such as “service exports, internet access and [...] investment in research and development” instead, it seems hard to believe that the success is not in some way related to the fact that the major languages of those countries (the Netherlands, Sweden, Denmark and Norway) are closely related to English. This influence of the mother tongue, subtle as it at times may be, seems to be evident in all aspects of target language proficiency, although Hulstijn (2015) reviewing several studies on the interdependence of L1 and L2 literacy warned that it is “not a matter of knowledge and skill in L1 being transferred to knowledge and skill in L2” (Hulstijn, 2015, p. 132).

In the process of acquiring the pronunciation of a foreign language, learners are confronted with a number of obstacles, such as the inherent difficulty of the foreign language and the spelling system (Gussenhoven & Broeders, 1997). Another obstacle is the interference of the pronunciation of the native language. Thus beginning learners will be inclined to interpret the sounds of the target language as the sounds of their variety of the mother tongue and to use the sounds of this variant of the mother tongue in their attempt to produce the sounds of the target language (Escudero & Williams, 2012; Chládková & Escudero, 2012). It is a well-known fact that speakers of some languages have an advantage over speakers of some other languages in acquiring the sounds of certain target languages; the same applies to speakers of some varieties of languages.

Where the productive oral skills and in particular pronunciation are concerned, the problems caused by native language interference seem to be evident to lay listeners as well as experts. Dutch listeners, teachers as well as students, are highly critical of the English produced by other Dutch persons, in particular if these persons happen to be teachers (Hermans, 2017; Kleinjan, 2017; Boersma, 2017). Pronunciation mistakes made by non-native speakers are a source of mirth to native and non-native listeners alike. The merry listeners are not always aware of the unhappiness their laughter causes the unfortunate non-native speakers, who cannot be envied, suffering from economic and social deprivation as they were shown to be by Lippi-Green (1997) and even being “discriminated against because of their accent” as they claimed themselves to be according to Derwing (2003, p. 556).

An aspect that has not received a lot of attention is the predicament of speakers of a non-standard variety of a language who are trying to learn a foreign language. It is well-known that such speakers are misunderstood and made fun of by speakers of the standard variety and also by speakers of other non-standard varieties of their native languages (see e.g. Omoniye & White, 2006). What is less well-known is that speakers of non-standard varieties also experience greater problems when learning a foreign language, as seems to be the case with learners of English in Mediterranean France (Ginesy & Hirst, 1972).

For these non-standard speakers it does not help that most textbooks for foreign language learners deal with two standard varieties: that of the native language and that of the target language. True, some textbooks may present several standard varieties of the target language (e.g. British and American English) but they usually

address speakers of a standard native language. Only few textbooks pay attention to the problems experienced by speakers of non-standard varieties (Jones, 1960; Collins & Mees, 1981; Gussenhoven & Broeders, 1976, 1997). There is evidence that these problems may be serious. For instance, the phoneme repertoire of some non-standard varieties may differ from that of the standard variety, which may make it more difficult for the speakers of that variety to learn the phonemes of the target language (see e.g. Luo, 2014).

A more important problem may be the behaviour of judges (teachers or external examiners) who are native speakers of the learner's first language, most probably of a standard variety. Admittedly, Yoshida (2004) found that non-native listeners were just as capable as native listeners of judging the intelligibility of Japanese speakers of English. But her listeners were highly trained and experienced professionals and her findings were exceptional. Most authors have found that non-native judges are stricter than native listeners. It may well be that these judges are biased against non-standard speakers, not only when they listen to their native language speech but also when listening to their target language speech, in which they recognize features of the non-standard language. Indeed, some non-native speaker judges are unable to identify native speakers (Wester & Mayo 2014; anecdotally, we have personally experienced situations where teachers of English who were native speakers of Dutch did not recognize native speakers of English amongst their students, often because these students were speakers of a non-standard variety of English, e.g. from the Windward Islands). If this is so, these non-standard speakers could experience a double disadvantage; not only could they have greater difficulty learning the sounds of the target language, but their efforts would also be judged more harshly than those of speakers of the standard variety of the native language (compare Van den Doel, 2006). This is the problem that was addressed in the study described below.

In the assessment of the pronunciation of the foreign language, assessors often refer to the extent to which learners have overcome the interference of the native language and approximated the pronunciation of the native speakers of the target language. It seems probable that in practice other factors play a part in the formation of the assessment. It is a well-known fact that listeners associate the varieties of a language with social and personal characteristics (Labov, 1966; Giles, 1970). There is no reason to assume that assessors of the pronunciation of a foreign language do not engage in such an association. As the pronunciation of a foreign language of most beginning and intermediate learners shows perceptible interference by the native language, it stands to reason that listeners who are also native speakers of that language can perceive what variety of the native language is responsible for the interference, a variety that may be associated with social and personal characteristics. As a result the assessment of the pronunciation of the target language is not only determined by the extent to which interference of the native language takes place but also by what particular variety of the native language causes the interference. It was the aim of the present study to indicate in what way the Dutch listeners' assessment

of the pronunciation of English by Dutch speakers was influenced by the interfering variety of Dutch.

Before this problem was addressed, however, the findings of earlier studies about Dutch listeners' strictness needed to be confirmed. In the present study an attempt was therefore made to answer two questions.

- 1) Are Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than English listeners?
- 2) Are Dutch listeners biased in their judgements of English speech produced by native speakers of Dutch?

2. METHOD

2.1 *The listening material*

Twenty upper intermediate to advanced learners of English (10 female, 10 male) had been asked to perform two types of task. Nearly all of them were 1st-year students of English in higher education, who performed these tasks voluntarily. First they were asked to read out two short texts, one Dutch text and one English text. Then they were asked to tell the same story, first in Dutch and then in English, on the basis of a number of pictures. The students had been chosen in such a way that the variety of Dutch of half had been judged "standard" and the variety of the other half "non-standard" by six trained Dutch listeners (3 female, 3 male).

The fragments were presented to the listeners as two recordings (one with Dutch speech and one with English speech) of some 30 minutes each, containing 40 fragments of 40 seconds; the 20 text fragments were presented first, the 20 fragments of free speech last. Some fragments were presented twice; the first fragments of each pair, intended to allow the listeners to familiarise themselves with the texts, were not considered in the analysis. The Dutch listeners had not been told that each speaker had produced Dutch and English speech; the listeners had not been told that some fragments occurred twice.

2.2 *The listeners*

The findings presented here are based on the judgements of 126 listeners, 76 female and 50 male, who judged some 4700 audio fragments. Among them were 70 Dutch listeners, 57 of whom were students in higher education (most of them 1st-year students of English) and 13 of whom were teachers of English, most of them in higher education. There were 56 English listeners, 54 of whom were students (most of them in the sixth form) and two of whom were teachers, none of them teachers of languages, in three secondary schools in county Suffolk in the UK.

The English listeners listened only to one recording containing the English fragments produced by each speaker; they did so at any time that was convenient for them. The Dutch students listened to the recordings in a Tandberg Educational IS 10

language laboratory; they first gave their judgements on the Dutch recordings (on the assumption that it would be easier for them to judge listening material in their native language); they were asked to listen to the English recordings at a later date. The Dutch teachers also listened to the Dutch recordings before they listened to the English recordings; they also listened at any time that was convenient for them.

The Dutch and the English students and the English teachers performed the task voluntarily and did not receive payment. The Dutch teachers received a small payment.

There were nearly as many judgements by Dutch listeners of Dutch as of English fragments (37% resp. 36 %), with slightly fewer judgements by English listeners (27%). 37 % of the judgements were made on the fragments of free speech, the remainder being on the texts read. There were slightly more judgements on fragments produced by female speakers (53%) than by male speakers.

2.3 *Listening situations*

The listening situations involving the two nationalities are presented in table 1. There were three listening situations: Dutch listeners judging Dutch speech (DD), Dutch listeners judging English speech (DE) and English listeners judging English speech (EE).

Table 1. Listening situations

Listener	Speech	
	Dutch	English
Dutch	x	x
English		x

2.4 *Instrumentation*

The listeners were invited to indicate what impression they were given by different ways of pronouncing English and, in the case of the Dutch listeners, Dutch on eleven 7-point scales (see Table 2). The English listeners used the scales in the left-hand column and the Dutch listeners used the scales in the right-hand column.

Table 2. Scales for judging each fragment in English and Dutch (Note, judgements were expressed on 7-point scales)

	English description	Dutch description
1.	unpleasant - pleasant	onaangenaam – aangenaam
2.	broad - cultured	plat – beschaafd
3.	ugly - beautiful	lelijk – mooi
4.	slovenly – polished	slordig – verzorgd
5.	strong accent - no accent	sterk accent – geen accent
6.	non-standard - standard	niet- standaard – standaard
7.	monotonous - melodious	monotoon – melodieus
8.	expressionless - expressive	expressieloos – expressief
9.	unintelligible - intelligible	onbegrijpelijk – begrijpelijk
10.	inaccurate - precise	onnauwkeurig – precies
11.	common - distinguished	ordinair – gedistingeerd

Some of the scales were “semantic differentials”, which had been studied by Osgood et al. (1957). The first five scales had been used by Koster and Koet (1993). In this experiment six more scales were used: the Dutch translations of three of which, *expressionless – expressive*, *slovenly – polished* and *inaccurate - precise*, had been used to study reactions to the pronunciation of Dutch by Fagel et al. (1983). Three new scales were introduced *unintelligible – intelligible*, *standard - non-standard* and *common – distinguished*. The *unintelligible – intelligible* scale was intended to allow listeners to indicate if they were able to easily understand what was said, whereas it was hoped that the *standard - non-standard* scale would allow the listeners to indicate in how far they felt the pronunciation approximated the standard set by educational institutions (the secondary school, the college, cf. Van der Haagen, 1998, p. 73). Scale 11 had been adapted from the *distinct – indistinct* scale, which was felt to be too similar to *inaccurate - precise*; it was hoped that the *common – distinguished* scale would allow the listeners to give another intuitive judgement on the extent to which the pronunciation was statusful, in addition to the *broad – cultured* scale.

3. RESULTS

3.1 Data reduction

In order to find if it was possible to cluster the scales, factor analyses were performed on the 11 scales for each listening situation. In order to obtain similar clustering for the three listening situations in each analysis four factors were allowed. Results of the (rotated) factor analyses are presented in Table 3.

Table 3. Factor loadings for Dutch listeners to Dutch speech (DD), Dutch listeners to English speech (DE) and English listeners to English speech (EE); small factor loadings (< .3) are not printed.

	Aesthetic quality			Intonation			Comprehensibility			Standardness		
	DD	DE	EE	DD	DE	EE	DD	DE	EE	DD	DE	EE
a1-pleasant	.7	.7	.8	.3				.4				
a2-cultured	.9	.8	.8									
a3-beautiful	.8	.8	.8	.3				.3				
a4-polished	.8	.5	.7		.6	.4	.4					
s5-no accent	.8	.7	.6	.4		.4						
s6-standard										1.	.9	.9
e7-melodious				.9	.8	.8						
e8-expressive				.9	.8	.8						
e9-intelligible							.9	.9	.8			
e10-precise	.4						.8	.9	.7			
s11-disting.	.7	.6	-		.4	.7						

Table 3 shows that there is a considerable convergence between the factor loadings in the three situations (DD, DE and EE). Not only do most scales load on the same factor, but the number of extra loadings on separate factors seems small.

The *pleasant*, *cultured*, *beautiful* and *no accent* scales always load on the same factor, which we have named *aesthetic quality*. The second factor, *intonation*, consists of judgements on the *melodiousness* and *expressiveness* scales. The third factor, *comprehensibility*, is the average of judgements on the *intelligible* and *precise* scales. The fourth and final factor consists of the judgments on the *standardness* scale.

Judgements on both the *polished* and *distinguished* scales appear to load on different factors depending the situation. For instance, *distinguished* loads on the first factor if it concerns the judgements of Dutch listeners, but not it if concerns the judgements of English listeners, in which case it seems to load on the second factor. The judgments on the *polished* scale load relatively high on the first factor, but also on the second factor for DE and EE. Therefore, these two scales were not taken into account in the analysis.

3.2 Reliability

The second step in the analysis concerns the interlistener agreement per (composite) scale. The reliability coefficients are presented in Table 4.

Table 4. Reliability coefficients for the three composite scales (intraclass correlation).

	Dutch listeners to Dutch speech	Dutch listeners to English speech	English listeners to English speech
Aesthetic quality	.85	.85	.84
Intonation	.89	.89	.89
Comprehensibility	.93	.93	.93
Standardness	.87	.87	.87

It appears from table 4 that for all four composite scales the reliability is relatively high: the proportion of variance that needs to be attributed to either listener or the interaction between listener and speaker (error variance) is low, and the reported proportion of variance associated with speakers is high. Hence, we conclude that the scores are trustworthy; there clearly is agreement between the listeners.

3.3 Average judgements

In Table 5 the mean judgements in each of the three listening situations are presented.

Table 5. Number of fragments judged (#Frag), Mean (M) and standard deviation (sd) per type of listener and language of the fragment.

	Dutch listeners judging Dutch			Dutch listeners judging English			English listeners judging English		
	#Frag	M	sd	#Frag	M	sd	#Frag	M	sd
Aesthetic quality	1714	4.0	1.4	1828	3.7	1.1	1207	3.9	1.1
Intonation	1753	3.5	1.6	1883	3.4	1.4	1217	4.0	1.4
Comprehensibility	1703	4.2	1.7	1892	4.1	1.4	1234	4.6	1.4
Standardness	1756	4.0	1.7	1868	3.9	1.2	1210	4.1	1.2

The average judgement of *aesthetic quality* of Dutch speech by Dutch listeners is 4.0, which is somewhat higher than the average judgements of English (3.7) by the same listeners. To test whether the difference between these averages is significant, a multilevel model was used (as judgements are nested both within listeners as well as within speakers; see appendix for variance estimates: variance between raters, variance between speakers and residual variance).

For all four response variables a significant difference between the three means was found (*aesthetic quality*: $F(2, 243.9) = 93.5$; $p < .001$; *intonation*: $F(2, 263.3) = 31.3$; $p < .001$; *comprehensibility*: $F(2, 248.2) = 26.3$; $p < .001$; *standardness*: $F(2,$

253.8) = 22.2; $p < .001$). Post Hoc analysis showed that in all cases the judgements of Dutch listeners to English speech are significantly lower than those of either Dutch listeners judging Dutch speech ($p < .001$), or English listeners judging English ($p < .001$).

We can conclude that the judgements of Dutch listeners to English speech are on average lower than the judgements of English listeners judging the same samples of English speech. This answers the question about strictness. However, it does not answer the question about bias completely, as we have to take the ranking of judgements into account as well; are the same speech samples judged as good (or poor) by both types of listeners? In Table 6 we therefore present the correlation between Dutch and English listeners both judging the same English fragments and Dutch listeners judging Dutch and English fragments by the same speakers.

Table 6. Correlations (r) between Dutch and English listeners judging English speech and Dutch listeners judging Dutch and English speech.

	r Dutch and English listeners judging English speech	r Dutch listeners judging Dutch and English speech
Aesthetic quality	.53	.45
Intonation	.91	.46
Comprehensibility	.90	.34
Standardness	.70	.21

The results presented in table 6 show that the correlation between judgements of the same fragments by English and Dutch listeners for *aesthetic quality* is relatively low ($r = .53$). For *aesthetic quality*, moreover, the correlation between Dutch listeners judging Dutch and English speech, although not strong, is certainly not negligible.

The results also show that the correlation between Dutch and English listeners, judging the same fragments, is strong for *intonation*, *comprehensibility* and, to a lesser degree, *standardness*. Moreover, the results show that the correlation between Dutch listeners judging Dutch and English speech of the same speakers is very low indeed where *comprehensibility* and *standardness* are concerned. There seems, therefore, little reason to assume that the Dutch listeners are biased by aspects of Dutch speech that are present in English speech for *comprehensibility* and *standardness*.

By contrast there seems, therefore, sufficient reason to assume that the Dutch listeners are biased by aspects of Dutch speech that are present in English speech for *aesthetic quality*.

Although the correlation between Dutch and English listeners for *intonation* is strong, the correlation between Dutch listeners rating Dutch and English speech is not negligible for *intonation* either, so that there is sufficient reason to assume that bias is involved here.

In Table 7 it is shown where greater strictness and bias were found in the judgements by the Dutch listeners.

Table 7. Presence of greater strictness and bias in Dutch listeners' judgements

	Greater strictness	Bias
Aesthetic quality	x	x
Intonation	x	x
Comprehensibility		
Standardness	x	

From Table 7 it appears that Dutch listeners' judgements show greater strictness and bias as far as *aesthetic quality* and *intonation* are concerned and that, moreover, these judgements show greater strictness as far as *standardness* is concerned. Therefore there seems to be every reason to treat these judgements with caution. It also appears that the Dutch listeners' judgements show no greater strictness and no bias as far as *comprehensibility* is concerned. Therefore there seems to be no reason to mistrust these judgements.

4. DISCUSSION

Limitations. Before considering the implications of our findings, we will recognize the limitations of our study.

The first limitation concerns the language background of the listeners. The greater strictness of non-native listeners is well attested as far as listeners with European native languages are concerned; there may be some doubt if the same is true for listeners with an Oriental native language; speakers of Chinese, for instance, may be well be more positive about the English of other speakers of Chinese and, indeed, about the English of other non-native speakers than native speakers of English. Also, the Dutch listeners shared the native language of the speakers; we did not address the situation where non-native listeners judged the English of native speakers of another language than that of the listeners, although we are aware of possible effects of accent familiarity (Huang, 2013).

The second limitation concerns the language background of the speakers. We only addressed English as a target language; we have no reason to assume that the situation with other European target languages would be radically different but we cannot rule this out.

The third limitation concerns the experience of the listeners. We tried to have inexperienced as well as experienced listeners; as far as the non-native listeners were concerned we managed to include a fair number of experienced listeners; as far as the native listeners were concerned we did not succeed. In order to estimate the seriousness of this omission we calculated the differences between experienced and inexperienced listeners and found no significant effect of experience. It may be argued that such a difference could have been found, had we included more experienced native listeners.

Another limitation concerns the scales. The fact that we used two versions, English and Dutch, may well have led to different interpretations by the listeners. Although we were careful to use what we believed to be the nearest equivalents in denotation, the terms may have carried differing connotations. We tried to address this problem through the factor analysis, the results of which prompted us to exclude two scales.

Finally there is a serious limitation as far as the speaking tasks are concerned; clearly, the tasks concerned spoken production, rather than spoken interaction; it may be objected that we did not address the pronunciation in a communicative situation. We were well aware of this limitation but decided to heed the warnings given about the difficulty of assessing pronunciation in communicative situations (Van der Haagen, 1998).

Recommendations. If our findings are correct, the consequences are serious. If Dutch, and by extension other non-native, listeners cannot be trusted to judge *aesthetic quality* and *intonation*, that will disqualify them as assessors and examiners in situations where these two aspects need to be taken into consideration. If they can be trusted to judge *comprehensibility* that would qualify them as assessors and examiners in situations where only this aspect needs to be taken into consideration. The Common European Framework may provide some guidelines. In Table 8 the levels of phonological control at five levels of proficiency are presented (Council of Europe, 2001).

Table 8. Levels of phonological control and language proficiency. Adapted from Council of Europe, 2001, p. 117.

C1	Effective Operational Proficiency	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2	Vantage	Has acquired a clear, natural pronunciation and intonation
B1	Threshold	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.

A2	Waystage	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time
A1	Breakthrough	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group

At the three lower levels A1, A2 and B1 it is a matter of pronunciation being “understood with some effort”, “clear enough to be understood despite a noticeable foreign accent” and “clearly intelligible even if a foreign accent is sometimes evident” respectively. At these levels intelligibility appears to be regarded to be of paramount importance. Only at levels B2 and C1 is it a matter of ‘clear, natural pronunciation and intonation’. Therefore it may be argued that Dutch listeners may well be allowed to act as judges where speakers whose English is at the three lower levels are concerned. Where speakers at the higher levels are concerned, however, it would seem advisable to have native listeners as judges.

REFERENCES

- Boersma M. (n. d.). *The horror and embarrassment: Dutch attempts at speaking English*. <http://www.martijnboersma.com/the-horror-dutch-accent-speaking-english>. (Retrieved 14 03 2018).
- Chládková, K. & Escudero, P. (2012). Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects. *The Journal of the Acoustical Society of America*, 131(2), 119-25. <https://doi.org/10.1121/1.3674991>
- Collins, B., & Mees, I. (1981). *The sounds of English and Dutch*. The Hague, The Netherlands: Leiden University Press.
- Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Derwing, T. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 59(4), 547-566. <https://doi.org/10.3138/cmlr.59.4.547>
- EF (2017). *English Proficiency Index*. <https://www.ef.nl/epi/>. (Retrieved 14 03 2018).
- Escudero, P. & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *The Journal of the Acoustical Society of America*, 131(5), 406-412. <https://doi.org/10.1121/1.3701708>
- Fagel, W., Van Herpt, L., & Boves, L. (1983). Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation. *Speech Communication*, 2(4), 315-326. [https://doi.org/10.1016/0167-6393\(83\)90048-1](https://doi.org/10.1016/0167-6393(83)90048-1)
- Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, 22(3), 211-227. <https://doi.org/10.1080/0013191700220301>
- Ginésy, M. & Hirst, D. (1972). The influence of regional accents (southern French) on the learning of English as a foreign language. *Etudes de Linguistique Appliquée*, 7, 28-37.
- Gussenhoven, C., & Broeders, A. (1976). *The pronunciation of English. A course for Dutch learners*. Groningen, The Netherlands: Wolters-Noordhoff-Longman.

- Gussenhoven, C., & Broeders, A. (1997). *English pronunciation for student teachers*. (2nd ed.). Groningen, The Netherlands: Wolters-Noordhoff.
- Hermans, F. (2017). *Studenten willen geen les meer in steenkolen-Engels van docenten* [Students do not want to be taught in broken English]. <https://www.gelderlander.nl/nijmegen-e-o/studenten-willen-geen-les-meer-in-steenkolen-engels-van-docenten~a5a997b1/> (retrieved 14 03 2018).
- Huang, B. (2013). The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech. *System*, 41(3), 770-785. <https://doi.org/10.1016/j.system.2013.07.009>
- Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam, The Netherlands: John Benjamins Publishing Company. <https://doi.org/10.1075/llt.41>
- IELTS (2018). *Demographic Data 2016*. <https://www.ielts.org/teaching-and-research/demographic-data>. (retrieved 14 03 2018).
- Jones, D. (1960). *An Outline of English phonetics* (9th ed.). Cambridge, UK: W. Heffer & Sons, Ltd.
- Kleinjan, G. (2017). *Steenkolen Engels bij Docenten* [Lecturers' broken English]. <https://www.trouw.nl/home/steenkolenengels-bij-docenten-experts-die-details-en-nuances-missen-zo-jammer-~ade0943b/> (retrieved 14 03 2018).
- Koster, C., & Koet, T. (1993). The evaluation of accent in the English of Dutchmen. *Language Learning*, 43(1), 69-92. <https://doi.org/10.1111/j.1467-1770.1993.tb00173.x>
- Labov, W. (1966). *The social stratification of English in New York City*. Arlington, DC: The Center for Applied Linguistics.
- Lippi-Green, R. (1997). *English with an accent*. London, UK: Routledge.
- Luo, J. (2014). A study of mother tongue interference in pronunciation of college English learning in China. *Theory and Practice in Language Studies*, 4(8), 1702-1706. <https://doi.org/10.4304/tpls.4.8.1702-1706>
- Omoniye, T. & White, G. (eds). *The sociolinguistics of identity*. London, UK: Bloomsbury.
- Van den Doel, R. (2006). *How friendly are the natives?* Dissertation. Utrecht, The Netherlands: LOT.
- Van der Haagen, M. (1998). *Caught between norms. The English pronunciation of Dutch learners*. The Hague, The Netherlands: Holland Academic Graphics.
- Wester, M. & Mayo, C. (2014). Accent rating by native and non-native listeners. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) in Florence, Italy* (pp. 7749-7753). New York, NY: IEEE. <https://doi.org/10.1109/ICASSP.2014.6855098>
- Yoshida, H. (2004). *An analytic instrument for assessing EFL pronunciation*. Unpublished doctoral dissertation. Temple University Philadelphia.

APPENDIX

In the models three variance components are estimated: variance due to listener, variance due to speaker and residual variance. In the table below the variance components are presented per response variable.

Table A1. Estimated variance components (est) and their standard errors (se) per dependent variable.

	S ² listeners		S ² speakers		S ² residual	
	est.	(se)	est.	(se)	est.	(se)
Aesthetic quality	.14	(.02)	.50	(.14)	.88	(.01)
Intonation	.21	(.03)	.52	(.14)	1.39	(.03)
Comprehensibility	.23	(.04)	.34	(.10)	1.72	(.04)
Standardness	.09	(.03)	.25	(.04)	1.63	(.03)