An acoustic analysis of English vowels produced by speakers of seven different native-language backgrounds

Heuven, V.J.J.P. van; Gooskens, C.; Wieling, M.; Kroon, M.; Noort, G. van; Bouma, G.

From Semantics to Dialectometry

Festschrift in honor of John Nerbonne

From Semantics to Dialectometry

Festschrift in honor of John Nerbonne

Martijn Wieling, Gosse Bouma and Gertjan van Noord (eds.). 2016. *From Semantics to Dialectometry* (Festschrift John Nerbonne). Groningen: University of Groningen.

Original template:
Language Science Press
Habelschwerdter Allee 45
14195 Berlin, Germany
langsci-press.org

Cover and concept of design: Ulrike Harbort
Fonts: Linux Libertine, Arimo, DejaVu Sans Mono
Typesetting software: XƎLaTeX

Chapter 13

An acoustic analysis of English vowels produced by speakers of seven different native-language backgrounds[1]

Vincent J. van Heuven
University of Pannonia; Rijksuniversiteit Groningen; Universiteit Leiden; Fryske Akademy

Charlotte S. Gooskens
Rijksuniversiteit Groningen; University of New England

We measured F1, F2 and duration of ten English monophthongs produced by American native speakers and by Danish, Norwegian, Swedish, Dutch, Hungarian and Chinese L2 speakers. We hypothesized that (i) L2 speakers would approximate the English vowels more closely as the phonological distance between the L2 and English is smaller, and (ii) English vowels of L2 speaker groups will be more similar as the L2s are closer to one another. Comparison of acoustic vowel diagrams and Linear Discriminant Analyses (LDA) confirm the hypotheses, with one exception: Dutch speakers deviate more from L1 English than the Scandinavian groups. The Interlanguage Speech Intelligibility Benefit was convincingly simulated by the LDA.

1 Introduction

In the past century English has evolved into the *Lingua Franca* of the world. It is now the language of commerce, international relationships and science *par excellence* (e.g. Rogerson-Revell 2007). The use of spoken English as a Lingua Franca (ELF) is not without problems, however. When a person learns to speak a foreign language, the pronunciation of the target language will differ from that of native speakers of

[1] This work was supported by The Netherlands Organisation for Research (NWO grant number 360-70-430) and by the European Union (TÁMOP-4.2.1.D-15/1/KONV-2015-0006).

Vincent J. van Heuven & Charlotte S. Gooskens. 2016. An acoustic analysis of English vowels produced by speakers of seven different native-language backgrounds. In Martijn Wieling, Gosse Bouma & Gertjan van Noord (eds.), *From semantics to dialectometry*, 113–123. Groningen: University of Groningen.

that language and will be reminiscent of the sound patterns of the learner's mother tongue (e.g. Flege 1995). It is often easy to recognize the native language background of an ELF speaker by his non-native accent.

The present study compares the pronunciation of the (monophthongal) vowels of English produced by American L1 speakers with the same vowels pronounced by speakers of English from six different non-native backgrounds, i.e. Chinese, Dutch, Hungarian, Danish, Norwegian and Swedish. Chinese and Hungarian do not belong to the Indo-European language family. The other four languages are rather closely related to each other and to English. They are all members of the Indo-European Germanic language family group, be it in different branches. English is in the Anglo-Frisian branch, whereas Dutch is a West-Germanic language. The three Scandinavian languages are more closely related to each other (as members of the North-Germanic branch) than they are to either Dutch or English (Hendriksen & Auwera 1994). We will test the hypothesis that native languages that resemble one another yield foreign accents that are also similar to each other. Specifically, we expect the Danish, Norwegian and Swedish foreign accents in English to be so similar that one might speak of a Scandinavian accent.

Swarte (2016) measured linguistic distances between English, Dutch, German, Danish and Swedish. The two Scandinavian languages {Danish, Swedish} had the shortest distance between them, followed by the pair {Dutch, German}. English was more closely associated with the West-Germanic pair than with the Scandinavian pair. These linguistic distances were matched by experimentally established intelligibility results obtained in spoken-word translation and comprehension tasks. We predict that the Englishes spoken with a Scandinavian accent will be more alike than each of these accents is like the Dutch accent. Moreover Swarte's (2016) results lead us to expect that Dutch-accented English is closer to native English than the Scandinavian accents are. Finally, we predict that the Chinese and Hungarian accents differ more strongly from native English than the Germanic accents but we have no way of predicting which one will be closer to English.

We decided not to recruit human listeners but to simulate human listening through computer modeling. Linear Discriminant Analysis (LDA) has been used since the late 1990s to model and predict how human listeners with language background A identify the sounds of a foreign language B (e.g. Strange et al. 2005).

Intelligibility is greatest when interactants (speaker and listener) both use the same native language (e.g. Munro & Derwing 1995). When one interactant is a non-native, communication generally suffers (Cutler 2012). Communication in ELF may be as successful or even more successful when both interactants are non-native. This effect has been called the *interlanguage speech intelligibility benefit* (ISIB, Bent & Bradlow 2003). A realistic version of the ISIB hypothesis holds that the benefit will be found especially if speaker and listener have the same native language. If the LDA technique can be used to model and predict perceptual assimilation of non-native sounds to a native sound inventory (Strange et al. 2005), then the ISIB effect should also show up when we use the LDA technique.

2 Vowel systems of the seven languages

The languages targeted in this study have rich vowel inventories, as shown in Figure 1.

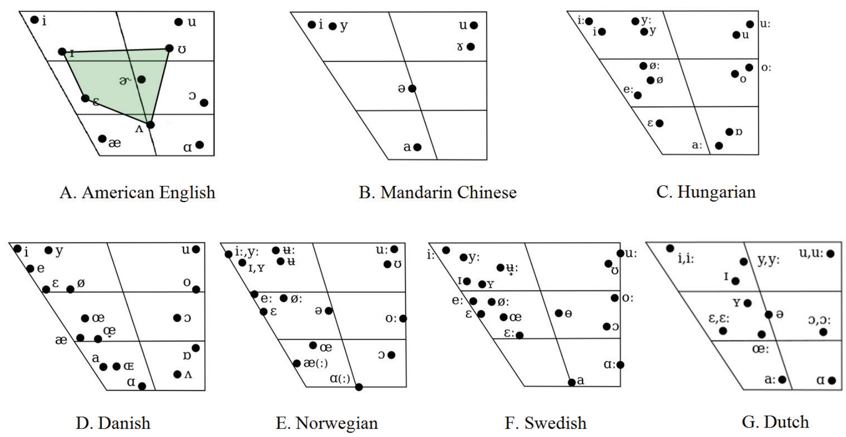

Figure 1: The monophthongs of (A) American English (Mannell, Cox & Harrington 2009), (B) Mandarin Chinese (Lee & Zee 2003), (C) Hungarian (Szende 1994), D. Danish (Grønnum 1998), (E) Norwegian (Vanvik 1979), (F) Swedish (Engstrand 1999) and (G) Netherlandic Dutch (Gussenhoven 1992). The shaded polygon defines the English lax vowel subsystem.

The Scandinavian inventories offer a plethora of vowel types that would qualify as viable substitutes in English ('similar sounds' in Flege 1995). Every English monophthong can be mapped onto a distinct vowel in Norwegian at approximately the same position in the vowel space. The Danish inventory has no vowel pair corresponding to the *fool~full* contrast, but there are one-to-one mappings for all other English monophthongs. Swedish does have vowels matching this latter contrast but lacks the contrast between the *bed* and *bad* vowels. Dutch, lacks both the *fool~full* and the *bed~bad* contrast.

3 Method

The data for English spoken with American, Dutch and Mandarin accents were described by Wang & Heuven (2006). For each language group ten male and ten female speakers were recorded. Non-native speakers were university students who had not specialized in English and had not spent time in an English-speaking environment. This type of speaker is representative of the typical ELF user in international settings. The 3 × 20 speakers lived in the Netherlands at the time the recordings were made, studying at Leiden University. The Hungarian ELF speakers (7 male, 10 female) were recorded in 2015 at the University of Pannonia (Hungary). The Scandivanian ELF speakers were recorded in 2016 at the University of Copenhagen (8 male and 12 female Danish speakers), the University of Oslo (10 male, 10 female Norwegian speakers) and the University of Stockholm (8 male, 12 female speakers). The inclusion criteria for the Hungarian and Scandinavian speakers were the same as for the Dutch and Chinese speakers.

Speakers produced all the 19 full vowels of English in a /hVd/ environment in a fixed carrier sentence *Now say h..d again.* Stimuli were presented to the speaker printed in normal English orthography on a sheet of paper. The pronunciation of the target vowels was exemplified by everyday key words rhyming with the /hVd/ targets (e.g. the unfamiliar target word *hoed* was cued by the more familiar words *road*, *showed*). Only the /hVd/ target words were used for acoustic analysis. Each speaker produced one token of each vowel; only the Hungarians recorded two tokens of each vowel type.

The onsets and offsets of the target vowels were determined by ear and by eye, using oscillograms and spectrograms. Formants were estimated by the Burg LPC algorithm implemented in Praat (Boersma & Weenink 1996). The optimal LPC model order and upper frequency cut-off were determined by trial and error, visually comparing formant tracks with the spectrogram. Vowel duration and the centre frequencies of maximally five formants were extracted; for each vowel token each formant frequency was averaged over the duration of the vowel. Formant frequencies were then psychophysically scaled in Barks (Traunmüller 1990). In all, measurements of 1,540 vowel tokens were available for statistical analysis.

4 Results

In American English as spoken in Southern California, ten vowels are recognized as monophthongs. The vowels in *caught*, *hot* and *father* have merged in that variety (and count as tense vowels in our analysis). The mid vowels in *pay* and *show* (not in Figure 1A) are commonly considered monophthongs (even though they are diphthongized to some extent in many varieties of English, including Californian). The central vowel in *bird* is not included since it only occurs immediately before /r/ – which makes it a positional allophone of the vowel in *but*. For the same reason we excluded all other /r/-coloured allophones.

Figure 2 displays the location of the ten English monophthongs in the acoustic vowel diagram with the first formant frequency (F1, representing vowel height) plotted from top to bottom, and the second formant frequency (F2, representing vowel backness and rounding) from right to left. This type of plot affords immediate visual comparison with the vowel diagrams in Figure 1.

The male and female vowel configurations are basically identical within each L1 group and yet differ systematically between groups. Using the American vowels as the reference set, we observe that there is a strict separation between the tense and lax subsystems. The six tense vowels (including /æ/ and /ɒ/) are on the outer perimeter of

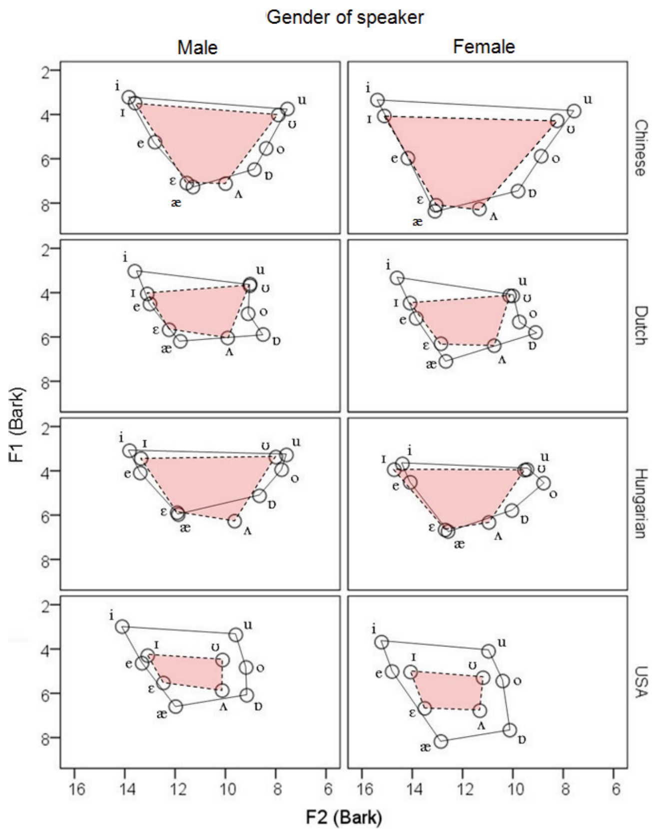

Figure 2: (Continued on page 134.)

the vowel space, while the four lax/short monophthongs form the corner points of an inner polygon. The members of the two pairs of mid vowels (/e, ɪ/ and /o, ʊ/ are rather close to one another in the spectral space but they are still distinct by a difference in duration – see below). This arrangement reproduces what has commonly been reported for (American) English and closely matches Figure 1A.

The three Scandinavian groups of ELF speakers display roughly the same organization of the English vowel system. They, too, can best be described in terms of an outer hexagon of six peripheral, tense vowels and an inner tetragon with four lax vowels. Details of the configurations differ between Danish, Norwegian and Swedish groups but the basic organization is a very good match with the native system. The Danes differ from the other Scandinavian groups in their location of /u/, which is as far back as the mid vowel /o/. In the pronunciation of the Norwegians and Swedes, however, /u/ is centralized, which mimics the way this vowel is pronounced in present day English (both in England and in the United States). It seems practically impossible to

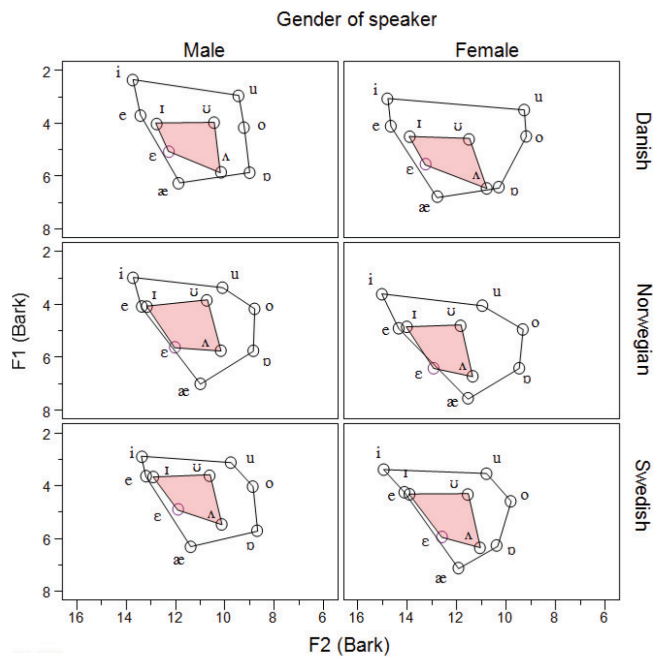

Figure 2: (Continued from page 133.) Mean F1 and F2 (Bark) of ten English monophthongs plotted for tense (open polygons) and lax (shaded polygons) vowels for 14 speaker groups.

tell the Norwegians and the Swedes apart on the basis of the spectral distribution of the English vowels. This is somewhat surprising, since we noted above that Swedes should be at a disadvantage when having to pronounce the contrast between /ɛ/ and /æ/, which does not exist in their L1 but does occur in Norwegian (and Danish). Either the Swedes have been so much exposed (through the media) to /æ/ in the neighboring languages Danish and Norwegian that this vowel is a familiar percept to them, or the pronunciation problem has been dealt with in the Swedish secondary school curriculum.

The Dutch ELF speakers deviate from English in several respects. Although the Dutch-accented vowels can also be divided into a tense and lax subsystem, the separation is poor in the bottom-left corner, where the contrast between /ɛ/ and /æ/ is weak (though not completely absent) and no difference is made at all between /u/ and /ʊ/.

The English vowel systems of the non-Indo-European groups do not seem to differentiate between the tense and lax subsystems – at least in terms of vowel quality. Although the Mandarin ELF speakers use a large vowel space – which confirms impressionistic claims (Zhao 1995) – they do not differentiate between /i/ and /ɪ/, nor between /u/ and /ʊ/. The vowel /æ/ is virtually the same as /ɛ/ while the distance between /ʌ/ and /æ, ɛ/ is so small that perceptual confusion can be expected.

The Hungarian speakers use a more contracted vowel space than the Mandarin speakers and show the same lack of contrast: poor separation between /i/ and /ɪ/, between /u/ and /ʊ/ and between /ɛ/ and /æ/. Moreover, the mid vowels /e/ and /o/ are spectrally quite close to /i, ɪ/ and /u, ʊ/, respectively – although they may still be distinct by duration.

Figure 3 plots the durations of the ten English vowels as produced by the seven speaker groups. Vowels are plotted from left to right in ascending order as determined for the native speaker group.

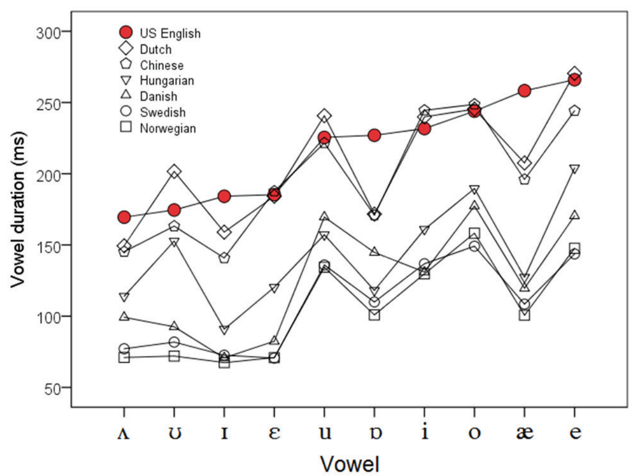

Figure 3: Duration (ms) of ten vowels of English produced by seven speaker groups. Vowels are in ascending order of length as observed for the US English native speaker group.

Although absolute durations differ from one language background to another (e.g. American, Dutch and Chinese speakers appear to produce longer vowels than Hungarian and Scandinavian speakers), there is a good deal of similarity in the relative vowel durations across the L1 backgrounds. The four lax vowels are indeed systematically shorter in the speech production of the native speakers, and these vowel durations do not overlap with any of the vowel durations of the six tense vowels. Crucially, the vowels /ɒ/ and /æ/ are clearly long. The same distribution of vowel duration is seen with the Scandinavian speakers: Danes, Swedes and Norwegians produce the four lax vowels with short durations, which are always shorter than any tense vowel duration. For the Scandinavian speakers, however, there is a tendency for the vowels /ɒ/ and /æ/, though longer than the lax vowels, to be shorter than the other tense vowels. The Hungarian speakers do not differentiate the duration of the tense versus lax members of the pairs /ɛ/~/æ/ and /u/~/ʊ/. The Dutch speakers have slightly longer durations for the tense members of these two pairs (suggesting that these contrasts exist for at least some Dutch speakers of English).

As a last exercise we performed a series of 98 Linear Discriminant Analyses (LDAs). Each of the seven languages provided the training data in turn, and the resulting models were then tested on the same set of seven languages, yielding a 7 × 7 matrix. Speaker normalization was performed prior to the LDA, by applying z-normalisation within individual speakers to vowel duration, F1 and F2. The results are shown in Figure 4, which plots the percentage of correctly identified vowels for each combination of training and test language. The left-hand panel presents the results when only F1 and F2 were entered as predictors; the right-hand panel shows the results of when the set of predictors was augmented with vowel duration. In both panels the data are plotted in ascending order of success of the training language with three predictors. Numerical values plotted in the right-hand panel are also given in Table 1.

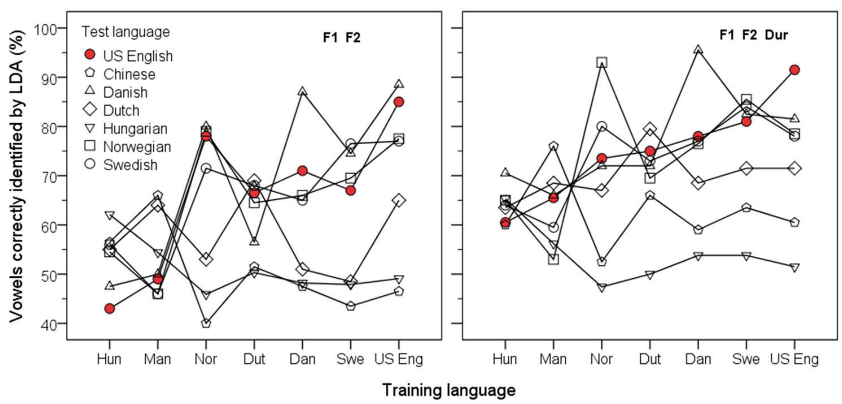

Figure 4: Correctly identified vowels by LDA (%) based on two (left-hand panel) or three (right-hand panel) acoustic predictors: F1, F2 and vowel duration as the optional third parameter. The models were trained in each of seven languages (x-axis), and tested in each of these languages (legend). There were 20 tokens (one token for each of 20 speakers) for each of ten English monophthongs (chance = 10%) per language, and 34 tokens (two tokens for each of 17 speakers with Hungarian L1). When training and test language were the same, the scores are based on cross-validation (leave-one-out method).

When including all three acoustic predictors, the vowels produced by the American native speakers are identified as intended most often when the test language is also US English (95.5% correct). The English vowels produced by Hungarian (51.5%) and Mandarin-Chinese (60.5%) ELF speakers are identified least successfully. The three Scandinavian groups are quite successful with scores around 80% correct. The Dutch ELF speakers are in the middle of the range (71.5%).

Table 1: Correct vowel identification (%) by LDA with F1, F2 and vowel duration as predictors (A). Also see Figure 4. Panel B lists the Relative ISIB values (see text).

		Training language						
		Chi	Dut	US	Hun	Dan	Nor	Swe
Test language	Chin	***76.0***	66.0	60.5	60.0	59.0	52.5	63.5
	Dutch	68.5	***79.5***	71.5	63.5	68.5	67.0	71.5
	USA	65.5	75.0	***91.5***	60.5	78.0	73.5	81.0
	Hung	56.2	50.0	51.5	***65.3***	53.8	47.4	53.8
	Dan	66.0	72.0	81.5	70.5	***95.5***	72.0	82.5
	Nor	53.0	69.5	78.5	65.0	76.5	***93.0***	85.5
	Swe	59.5	63.0	78.0	64.0	77.0	80.0	***84.0***

A. Raw scores

		Training language						
		Chi	Dut	US	Hun	Dan	Nor	Swe
Test language	Chin	***13.0***	0.1	-7.4	-3.3	-8.6	-13.4	-5.0
	Dutch	1.7	***9.9***	-0.1	-3.6	-2.8	-2.7	-0.8
	USA	-3.8	2.9	***17.4***	-9.1	4.2	1.3	6.2
	Hung	-2.6	-11.6	-12.1	***6.2***	-9.5	-14.3	-10.5
	Dan	-4.3	-1.2	6.3	-0.1	***20.6***	-1.2	6.7
	Nor	-16.0	-2.4	4.6	-4.3	3.0	***21.1***	11.0
	Swe	-9.1	1.5	4.5	-4.9	3.9	8.5	***9.9***

B. R-ISIB

The Interlanguage Speech Intelligibility Benefit (ISIB, see introduction) can readily be observed in panel B of table 1. The effect is seen most clearly if we convert the raw scores to relative ISIB scores (R-ISIB, Wang & Heuven 2015) by subtracting the row and column means from each cell in the matrix, and then divide the cell contents by 2. This removes the main effects of training language and of test language from the results, so that only the interaction term remains – which is our R-ISIB measure. When the test data are presented after the LDA was trained with data from the same speaker group (which simulates the shared interlanguage), vowel identification scores are much better than when test and training data are from different speaker groups, with R-ISIB values of 14.1 and -2.3, respectively, $t(47) = 6.2$ ($p < .001$).

When test and training languages are different Scandinavian languages (six pairs), vowel identification is better than when Scandinavian languages are paired with other non-native Englishes (eighteen pairs) with mean R-ISIB values of 5.3 versus -6.0, respectively, $t(22) = 4.8$ ($p < .001$). This confirms our hypothesis that the

ELF vowels of Danes, Norwegians and Swedes resemble one another more than they are like the vowels produced by the other ELF speakers.

5 Conclusion

We studied the vowels of English produced by six groups of non-native (ELF) speakers and compared these tokens with the vowels produced by native speakers of American English. Non-native speakers were representative of the academically trained professional with no specialisation in English. The vowel configurations of the six ELF groups differed substantially from those of native English as well as from each other, in ways that could often – but not always – be predicted from traditional impressionistic vowel diagrams of the first language of the speakers. Scandinavian, and especially Danish ELF speakers, approximate the English vowels most closely, better than Dutch ELF speakers and much better than Hungarian and Mandarin-Chinese ELF speakers do. Counter to what Swarte's (2016) results would suggest, it is not the case that Dutch ELF is closer to native English than the Scandinavian ELF varieties are. It is beyond the scope of the present paper, however, to examine whether the (monophthongal) vowel systems yield different linguistic distances than other aspects of the phonology, vocabulary, morphology and syntax.

These conclusions follow from visual comparison of the acoustic vowel diagrams of the seven varieties of English and are quantitatively corroborated by Linear Discriminant Analyses in which American native listeners are simulated. Moreover, the ELF vowels produced by the three Scandinavian speaker groups resemble each other more than they share properties with the ELF vowels of Dutch, Chinese and Mandarin speakers. This would suggest that the phonologies, specifically the vowel systems, of the three Scandinavian languages are rather similar and produce the same type of transfer from native to foreign language.

Finally, our results are in line with the idea that similarity of non-native accents in English (or any other language) may serve as an experimental means to quantify phonological distance between languages even if these languages are genealogically unrelated.

References

Bent, Tessa & Anne R. Bradlow. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114. 1600–1610.

Boersma, Paul & David Weenink. 1996. Praat, a system for doing phonetics by computer. *Report of the Institute of Phonetic Sciences Amsterdam* 132. www.praat.org.

Cutler, Anne. 2012. *Native listening: language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Engstrand, Olle. 1999. Swedish. In *Handbook of the International Phonetic Association: a guide to the usage of the International Phonetic Alphabet*, 140–142. Cambridge: Cambridge University Press.

Flege, James E. 1995. Second language speech learning: theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: theoretical and methodological issues in cross-language speech research*, 233–277. Timonium, MD: York Press.

Grønnum, Nina. 1998. Illustrations of the IPA: Danish. *Journal of the International Phonetic Association* 28. 99–105.

Gussenhoven, Carlos. 1992. Dutch. *Journal of the International Phonetic Association* 22. 45–47.

Hendriksen, Carol & Johan van der Auwera. 1994. The Germanic languages. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 1–18. New York: Routledge.

Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33. 109–112.

Mannell, Robert, Felicity Cox & Jonathan Harrington. 2009. *An introduction to Phonetics and Phonology,* Macquarie University. http://clas.mq.edu.au/speech/phonetics/.

Munro, Murrey J. & Tracey M. Derwing. 1995. Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech* 38. 289–306.

Rogerson-Revell, Pamela. 2007. Using English for international business: a european case study. *English for Specific Purposes* 26. 103–120.

Strange, Winifred, Ocke-Sven Bohn, Kanae Nishi & Sonja A. Trent. 2005. Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 118. 1751–1762.

Swarte, Femke. 2016. *Predicting the mutual intelligibility of Germanic languages from linguistic and extra-linguistic factors.* Groningen: Groningen Dissertations in Linguistics 150.

Szende, Tamás. 1994. Illustrations of the IPA: Hungarian. *Journal of the International Phonetic Alphabet* 24. 91–94.

Traunmüller, Hartmut. 1990. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88. 97–100.

Vanvik, Arne. 1979. *Norsk fonetikk.* Oslo: Universitetet i Oslo.

Wang, Hongyan & Vincent J. van Heuven. 2006. Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In Jeroen M. van de Weijer & Betteloe Los (eds.), *Linguistics in the Netherlands 2006*, 237–248. Amsterdam: John Benjamins. http://hdl.handle.net/1887/14899.

Wang, Hongyan & Vincent J. van Heuven. 2015. The Interlanguage Speech Intelligibility Benefit as bias toward native-language phonology. *i-Perception* 6. 1–13. http://ipe.sagepub.com/content/6/6/2041669515613661.full.

Zhao, Demei. 1995. *English phonetics and phonology: as compared with Chinese features.* Qingdao Shi: Qingdao hai yang da xue chu ban she.