Relative Contribution of Vowel Quality and Duration to Native Language Identification in Foreign-Accented English

Hongyan Wang
Shenzhen University
3688 Nanhai Ave, Nanshan, Shenzhen,
Guangdong 518060, P.R. China
+86 15013819676
wanghongyan0069@hotmail.com

Vincent J. van Heuven
University of Pannonia
Egyetem Ut. 10
8200 Veszprém, Hungary
+31 6 1514 3958
v.j.j.p.van.heuven@hum.leidenuniv.nl

ABSTRACT

Automatic identification of a speaker's native language background may have forensic applications. This paper explores the feasibility of automatic identification of the native language background of a foreign speaker of English, using phonetically interpretable measurements. The production of the ten monophthongs of (American) English by Dutch, Mandarin Chinese and American speakers was used as a test case. Vowel formants F1 (corresponding to articulatory vowel height), F2 (capturing vowel backness and lip rounding) and vowel duration were extracted. Clearly different duration and patterning of the vowels in the vowel space were seen. Automatic classification of the speaker's native language was 90 percent correct when all acoustic parameters were used as predictors. Language identification was slightly poorer when only formant data were used (85% correct) and substantially poorer – but much better than chance – when only vowel duration was used (60% correct). We conclude that vowel duration provides a weaker cue to foreign-accent identification in English than the spectral properties but that the combination of both information sources yields the best results.

CCS Concepts

• **Applied computing →Law, social and behavioral sciences; 300** • **Applied computing → computer forensics; 300** • **Applied computing → Arts and humanities; 500.**

Keywords

Foreign accent identification; vowel duration; vowels formants; English as a foreign language; American English; Dutch; Mandarin; Linear Discriminant Analysis (LDA); automatic classification; forensic application

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. [1]). The use of spoken English as a Lingua Franca (ELF) is not without problems, however. When a person learns to speak a foreign language, especially after the age of puberty, the pronunciation of the target language will differ markedly from that of native speakers of that language and will be reminiscent of the sound patterns of the learner's mother tongue (e.g. [2]). It is often easy for human listeners to recognize the native language background of an ELF speaker by his non-native accent.

In forensic applications it is often expedient to be able to identify the native-language background, and thereby the country of origin, of non-native speakers of English, and to ascertain that the speaker is a foreigner. In cases where a speech recording of an unknown speaker is evidence, it might be useful to identify the speech community that the speaker may belong to ([3]: 422).

The problem of dialect and foreign-accent detection has been studied by engineers ([4] and references therein), and with considerable success (e.g. [5] for English, French, German and Spanish speakers of French). The materials and techniques used in the engineering approach, however, do not yield a phonetically (or linguistically) interpretable characterization of the accents. Our goal is to investigate the relative contribution of vowel resonances ('formants') and duration to the automatic determination of the native language of a non-native speaker of English. The research is limited to two non-native speaker groups, whose vowel properties will be compared with those obtained from native speakers of American English. One of the non-native background languages, Dutch, is related to English. English and Dutch, are languages within the Germanic branch of the Indo-European language family. English and Dutch are not mutually intelligible [6] and differ substantially in their lexicon, phonology, morphology and syntax. The second non-native language is Standard Mandarin as spoken in the northern half of China. Mandarin is typologically very different from the two Germanic languages both in terms of its morpho-syntax and in terms of the phonology. The inventory of possible syllables is much smaller in Mandarin than in the Germanic languages. Mandarin allows just one consonant in the onset of syllables and none in the coda (with the exception of nasals /n, ŋ/, whereas Dutch and English allow complex consonant clusters with up to three onset and four coda consonants. The vowel inventories of Dutch and English, although quite different, are relative rich, with between 15 and 20 different vowel phonemes, which are divided into tense versus lax subsystems. Tense vowels are produced with more extreme positions of the articulators and require more effort and generally have longer duration than their lax counterparts, which are articulated

closer to the neutral position of the human vocal tract, require less effort and are pronounced short. The vowel inventory of Mandarin is smaller and has no subdivision into tense and lax types.

The following research questions are addressed:

- How do the English pure vowels ('monophthongs') produced by Mandarin, Dutch and American speakers of English differ, in terms of formant frequencies and duration?
- How well can we automatically determine the native-language background of a speaker of English as American, Dutch or Mandarin, from the acoustics of only the monophthongs?
- What are the relative contributions of formant frequencies and vowel duration to this automatic identification process?
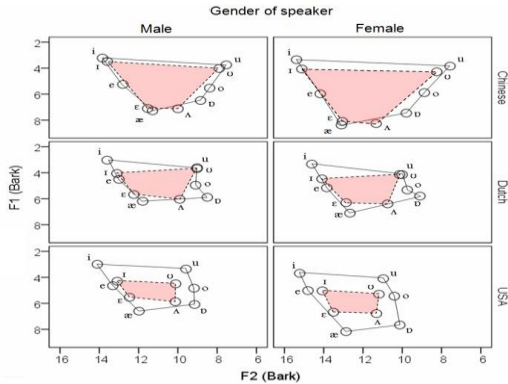

Figure 1. Mean F1 and F2 (Bark) of ten English monophthongs for tense (open polygons) and lax (shaded polygons) vowels for three groups of men and women.

We hypothesize that a Mandarin background will be correctly identified more often than a Dutch background due to the fact that Mandarin differs more from English and does Dutch. We also predict that the difference between the tense and lax vowels will be the most useful information source in the automatic determination the ELF-speakers language background. It is unknown at this juncture whether deviations in the duration of the ELF vowels will make a larger or smaller contribution to the automatic language identification than the differences in articulatory position (also called quality) of the non-native vowels.

2. METHODS

The data for English spoken with American, Dutch and Mandarin accents were described in detail by [7, 8]. 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. The American and Chinese participants were recent arrivals who had come to study at Leiden University.

Speakers produced all the 19 full vowels of English in an /hVd/ environment in a fixed carrier sentence *Now say h..d again* (following [9, 10, 11]). 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.

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 [12, 13]. 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 center 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 psycho-physically scaled in Bark units [14].[1]

3. RESULTS
3.1 Spectral properties
In American English ten vowels are recognized as monophthongs. The vowels in *caught*, *hot* and *father* are often not differentiated (and count as tense vowels in our analysis). The mid vowels in *pay* and *show* are commonly considered monophthongs (even though they are diphthongized to some extent in many varieties of English). The central tense vowel /ɜ/ in *bird* is not included since it only occurs immediately before /r/ – which makes it a positional allophone of the lax vowel /ʌ/ in *but*. For the same reason we excluded all other /r/-coloured allophones. This left ten vowels for acoustic analysis: /i/ (*heed*), /ɪ/ (*hid*), /e/ (*hayed*), /ɛ/ (*head*), /æ/ (*had*), /ɒ/ (*hod*), /ʊ/ (*hood*), /u/ (*who'd*), and /ʌ/ (*Hud*). Figure 1 displays the location of these ten English monophthongs in the acoustic vowel diagram separately for male and female speakers, 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.

The male and female vowel configurations are basically identical within each L1 group and yet differ systematically between groups. Using the American vowels (bottom row) 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 the vowel space, while the four lax/short monophthongs form the corner points of an inner quadrilateral. The members of the two pairs of higher-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 (e.g. [7, 8, 9, 10]).

The Dutch ELF speakers (middle row in Figure 1) deviate from English in several respects. Although the Dutch-accented vowels can also be divided into a tense and lax subsystem [7, 8, 16], 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 Mandarin speakers (top row) 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 [15]), 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.

3.2 Vowel duration
Figure 2 plots the durations of the ten English vowels as produced by the three speaker groups. Vowels are plotted from left to right in ascending order as determined for the native speaker group.

---

[1] Bark = [(26.81 × *F*) / (1960 + *F*)] − 0.53,
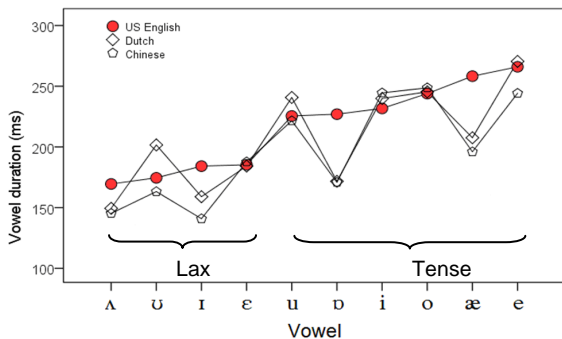where *F* represents the measured formant frequency in hertz.

Figure 2. Duration (ms) of ten monophthongal vowels of English produced by three speaker groups. Vowels are in ascending order of length as observed for the US English native speaker group.

Figure 2 shows that there is a good deal of similarity in the relative vowel durations across the L1 backgrounds. The four lax vowels (on the left in Figure 2) 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 (on the right of Figure 2). The lax vowels all have mean durations below 200 ms, whereas all the tense vowels are longer than 200 ms. Crucially, the vowels /ɒ/ and /æ/ are clearly long in American English: they are well over 200 ms.

The ELF vowel durations produced by the Dutch and Chinese speakers are remarkably similar to one another, and both sets approximate the American native vowel durations quite closely, with just a few exceptions. In the tense vowel set four out of six vowels have correct long durations above 200 ms. Two more tense vowels, i.e. /ɒ/ and /æ/ have durations below 200 ms and should therefore be considered too short by American native standards. It should be noted that these two vowels seem to constitute an in-between length category: they are shorter than 200 ms but still longer than the lax vowels. This is seen in the Dutch and in the Chinese results alike. All four lax vowels produced by the Chinese speakers are shorter than 200 ms, and also shorter than the intermediate /ɒ, æ/ category. therefore have the required length. This is also the case for three out of four lax vowels produced by the Dutch ELF speakers. However, the lax vowel /ɛ/ as produced by the Dutch speakers is longer than 200 ms, and falls into the intermediate length category. This makes the vowel duration behavior in Dutch-accented English different than that in Mandarin-accented English. It is also remarkable that the Mandarin speakers of English obey the temporal regularities of English vowels better than the Dutch speakers do, even though Dutch has a phonological contrast between tense/long versus lax/ short vowels and Mandarin does not.[2]

3.3 Native language identification

Now that we have observed characteristic patterns of deviations from the native English norms, both in terms of the spectral properties (F1 and F2) and in the duration of the monophthongs produced by Chinese and Dutch speakers of English, how well

can the native language background of the three speaker groups be predicted from these deviations, and do the spectral and the temporal properties contribute equally to the identification of the foreign accent? To answer these questions we performed automatic classification of language background using Linear Discriminant Analysis (LDA [20, 21]).[3] Before running the LDA we minimized individual differences between speakers (and between men and women) by applying within-speaker z-transformation to formants F1, F2 and vowel duration so that only the linguistically relevant differences would remain in the data.[4]

The total dataset for this part of the investigation comprised the vowel formants (F1 and F2, z-normalised within speakers after transformation to Barks), and (z-normalised) vowel durations of the ten monophthongs of English, including slightly diphthongal /e/ and /o/ as in *say* and *so*. This yields a set of 60 (20 Mandarin, 20 Dutch, and 20 American) speakers of English, as objects to be classified and 30 predictors, i.e. the F1, F2 and duration of each of ten different monophthongs. Data reduction was obtained by running the LDA in stepwise mode, starting with the predictor that differentiates best between the three speaker groups and including additional predictors one by one, but only if they made a significant improvement (in terms of Wilk's Lambda) to the overall performance of the decision algorithm. We ran three LDAs. The first time we entered only the duration values as predictors, the second analysis used only the spectral parameters; in the last LDA all 30 predictors were entered. Table 1 presents the results of the classification.

Table 1. Classification (%) by LDA of native language (L1) of three groups of speakers, using duration parameters only, spectral parameters only, and both types combined. Correct classification in bold face. N= 20 speakers per L1.

	L1 of speaker	Predicted L1		
		Chinese	Dutch	USA
Duration only	Chinese	**55**	25	20
	Dutch	30	**50**	20
	USA	15	10	**75**
Mean correct	60			
F1, F2	Chinese	**80**	20	0
only	Dutch	10	**90**	0
	USA	0	15	**85**
Mean correct	85			
Duration +	Chinese	**95**	5	0
F1, F2	Dutch	10	**85**	5
	USA	0	10	**90**
Mean correct	90			

Overall, using leave-one-out cross-validation to prevent any cases being predicted that were also part of the training data, 60% of the 60 speakers were correctly classified in terms of their native language background when only vowel durations were used as predictors. Entering the two sets of spectral parameters (F1, F2) as predictors yielded an overall correct L1 classification of 85%, while the combination of spectral and temporal parameters

---

[2] It has been noted before that systematic differences in vowel duration are accurately reflected in foreign-accented speech produced by learners whose native language has no vowel length contrast, e.g. [15] for Dutch vowel length produced by Turkish immigrants, [16] for Catalan learners of English, and [17] for Turkish learners of German. It has been suggested that duration differences in a foreign language are easily perceived regardless of the native language of the learner [18].

[3] For plots of the spectral vowel spaces (as in Figure 1) for the three groups of speakers indicating individual scatter of the datapoints and spreading ellipses see [6, 7].

[4] Z-transformation (also called Lobanov normalisation) is generally the most successful procedure to eliminate linguistically irrelevant speaker-individual characteristics from formant and duration data [21, 22].

increased the overall performance of the classification to 90%. When only duration data were used, the confusion structure indicates that the Chinese and Dutch speakers resemble each other more than either of them resemble the American native speakers. In the remaining LDAs there are not enough error decisions for a meaningful interpretation of the confusion structure.

Table 2 lists, for each of the three LDAs, the acoustic parameters that were included in the model, in the order in which they were included in the analysis by the stepwise procedure. The parameters are identified as D, F1 and F2 for Duration, first formant and second formant, respectively. The vowel for which the parameter was computed is specified by its IPA symbol in parentheses. For instance, F1(ʊ) is the first formant frequency of the vowel /ʊ/ as in the English word *hood*.

Table 2. Order of inclusion of acoustic parameters in three LDAs (see text for explanation).

Duration only		F1, F2 only		Duration + F1, F2	
rank	parameter	rank	parameter	rank	parameter
		1.	F2(ɪ)	1.	F2(ɪ)
		2.	F1(ɪ)	2.	F1(ɪ)
		3.	F1(ʊ)	3.	F1(ʊ)
		4.	F2(ɒ)	4.	F2(ɒ)
		5.	F2(i)		
				5.	D(ɒ)
				6.	F1(u)
				7.	F2(ʌ)
1.	D(ʊ)			8.	D(ʊ)
2.	D(i)				
3.	D(e)				
4.	D(ɒ)				
5.	D(u)				
				9.	F1(o)

In the first LDA (with duration parameters only) the algorithm included five parameters, viz. (in descending order of importance) the durations of the vowels /ʊ/, /i/, /e/, /ɒ/, and /u/. In the spectral analysis the optimal model was again reached with five predictors, viz. the F2(ɪ), F1(ɪ), F1(ʊ)/, F2(ɒ), and F2(i), again in descending order. In the combined spectral and durational analysis nine acoustic parameters made a significant contribution to the pre-diction of the speaker's L1 background, viz. F2(ɪ), F1(ɪ), F1(ʊ), F2(ɒ), Dur(ɒ), F1(u), F2(ʌ), Dur(ʊ), and F1(o). The first four of these are the same, and in the same order, as was found in the spectral-only LDA. The first duration parameter, Dur(ɒ) was entered at step 5, and the second, Dur(ʊ), at step 8. This confirms that the contribution of differences in vowel duration is sub-stantially smaller than that of the spectral differences between the vowels. It is also noted that the order of the predictors is stable only for the first four parameters. From step 5 onwards the relative order of the spectral and the temporal parameters in the separate LDAs is not the same as those in the combined LDA. This is visualized in Table 2 by arranging the parameters such that the ranks obtained in the separate temporal and spectral LDAs can be optimally intercalated so as to yield the order obtained in the combined LDA. The switch in the rank of D(ɒ) is indicated by the arrow.

4. CONCLUSIONS

We conclude from this pilot study that it is possible to identify the Chinese speakers of English with substantial (90%) precision. The input on which the determination is based is a set of 30 phonetically motivated acoustic parameters, i.e. the duration and the means of the center frequencies of the lowest two resonances of the vocal tract (F1, F2) for ten different English vowels, which were spoken just once by three groups of 10 male and 10 female speakers. The results show that the spectral parameters (F1 and F2) make by far the greatest contribution to the correct identification of the speaker's native language whereas the vowel durations make a significant but much smaller contribution. In fact, 85 percent correct L1 identification was obtained when only the first three parameters in Table 2 were used in the analysis, i.e. the F1 and F2 of the vowel /ɪ/ (as in the word *mid*) plus the F1 of /ʊ/ (as in *book*). The fourth and fifth spectral parameters, i.e., the F2 of /ɒ/ (as in *hot*) and the F2 of /i/ (as in *beat*), make a significant contribution in terms of Wilk's Lambda but do lead to a further reduction of classification errors over the model with only three spectral parameters. The determination of the native language background of our three groups of speakers, then, can be done with 85 percent accuracy on the basis of the way two specific vowels are pronounced, i.e. the two lax vowels /ɪ/ and /ʊ/. Why this should be so can be seen in Figure 1. Native speakers of American English produce strongly centralized tokens of these two lax vowels, by which they are clearly separated from their nearest competitors in the vowel space. The Dutch speakers of English make a similar distinction between /ɪ/ and its competitors but fail to do so for the English lax vowel /ʊ/ (which they pronounce identical to the tense back vowel /u/). The Mandarin speakers, who have no tense-lax subdivision in their native vowel system, do not distinguish between English /ɪ/ and /i/, nor between /ʊ/ and /u/.

5. DISCUSSION

The vowels in our study were produced in identical contexts, i.e. preceded by /h/ and followed by /d/. The spectral properties of vowels vary as a function of the preceding and following consonants. However, the influences of context are generally limited to the first and last quarter of the duration of the vowel, so that quite stable and reproducible formant values can be obtained if the measurements are based on the central portion of the vowel duration, especially if they are extracted from stressed syllables. Also, the process of locating the target vowels in the time domain and the formant measurements can be fully automated if a broad phonetic transcription is available for forced alignment [24] and if the formant measurements are made separately for back vowels. In order to obtain stable vowel duration measurements only vowels should be included which are followed by the same broad phonetic category of coda consonants, e.g. voiced plosives, as in our study, which have quite similar effects on the temporal organization of the syllable.

We have tested the feasibility of our approach on English vowels produced by speaker of two languages, Dutch and Mandarin, which are typologically very different from each other (see introduction). Given a basic set of 30 predictors, it would seem realistic to assume that the set of non-native languages (and thereby the number of different foreign accents in English) can be substantially extended. Given that every language has its own special arrangement of vowels in the articulatory space, there will generally be a unique combination of acoustic properties in the set of 30 that distinguishes the native language from its competitors.

There are also indications that the method is sensitive enough to afford classification of the regional background of a non-native speaker of English. It was shown, for instance, that Chinese speakers of English at the university level could still be classified in terms of their native regional dialect [25].

The method may fail for non-native speakers of English who have developed an exceptionally good approximation to the native English norm but this will only be the case for speakers who have learnt English during childhood (so-called early bilinguals ([26] and references therein) or who were trained to mimic the native pronunciation [27]. However, the method should work for the typical educated foreign speaker of English as a lingua franca found in international conferences and business meetings.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Rogerson-Revell, P. 2007. Using English for international business: A European case study. *Eng. Spec. Purp.*, 26, 103-120.

[2] Flege, J. E. 1995. Second language speech learning: theory, findings, and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, W. Strange, Ed. York Press, Timonium, MD, 233-277.

[3] Brown, G., and Wormald, J. 2017. Automatic socio-phonetics: Exploring corpora with a forensic accent recognition system. *J. Acoust. Soc. Am*., 142, 422-433. DOI = 10.1121/1.4991330

[4] Behravan, H. 2016. *Advances in automatic foreign accent recognition*. Doct. diss. University of Eastern Finland, Joensuu.

[5] Bartkova, K., and Jouvet, D. 2007. Automatic detection of foreign accent for automatic speech recognition *Proc. 16th Int. Con. Phon. Sc., Barcelona*, 2185-2188.

[6] Gooskens, C., and Heuven, V.J. van 2017. Measuring cross-linguistic intelligibility in the Germanic, Romance and Slavic language groups. *Speech Comm*., 89, 25-36. DOI = 10.1016/j.specom.2017.02.008.

[7] Wang, H., and Heuven, V. J. van 2006. Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In *Linguistics in the Netherlands 2006,* J. M. van de Weijer, and B. Los, Eds. John Benjamins, Amsterdam, 237-248.

[8] Wang, H. 2007. *English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English.* LOT dissertation series 147. LOT, Utrecht.

[9] Peterson. G. E., and Barney, H. L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am* 24, 175-184.

[10] Hillenbrand, J., Getty, L., Clark, M., and Wheeler, L. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am*., 97, 3099-3111.

[11] Chen, Y. Robb, M., Gilbert, H., and Lerman, J. 2001. Vowel production by Mandarin speakers of English. *Clin. Ling. & Phon.,* 15, 427-440.

[12] Boersma, P., and Weenink, D. 1996. Praat, a System for Doing Phonetics by Computer. *Report of the Institute of Phonetic Sciences Amsterdam,* 132.

[13] Boersma, P., and Heuven V.J. van. 2001. Speak and unSpeak with Praat. *Glot Int.*, 5, 341-347.

[14] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am*., 88, 97-100.

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

[16] Heuven, V. J. van 1986. Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In *Dutch Linguistics at Berkeley 1985*, J. van Oosten, and J. F. Snapper, Eds. The Dutch Studies Program, U.C. Berkeley, 67-84.

[17] Cebrian, J. 2006. Experience and the use of non-L1 duration in L2 vowel categorization. *J. Phon.*, 34, 372-387.

[18] Nimz, K. 2011. Vowel perception and production of late Turkish learners of L2 German. *Proc. 17th Int. Con. Phon. Sc.*, *Hong Kong*, 1494-1497.

[19] Bohn, O. S. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. York Press, Timonium, MD, 279-304.

[20] Klecka, W.R. 1980. *Discriminant Analysis*. Sage, Beverly Hills, CA, London.

[21] Weenink, D. 2006. *Speaker-adaptive vowel identification*. Doctoral dissertation, University of Amsterdam.

[22] Adank, P., Heuven, V. J. van, and Hout, R. van 1999. Speaker normalization preserving regional accent differences in vowel quality. *Proc. 14th Int. Con. Phon. Sc., San Francisco,* 1593-1596.

[23] Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.*, 49, 606-608.

[24] Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite, http://fave.ling.upenn.edu.

[25] Cui, R., and Heuven, V. J. van 2011. Mutual intelligibility of English vowels by Chinese dialect speakers. *Proc. 17th Int. Con. Phon. Sc., Hong Kong*, 544-547.

[26] Heuven, V.J. van. 2017. Perception of English and Dutch checked vowels by early and late bilinguals. Towards a new measure of language dominance. In *Future research directions for Applied Linguistics*, S. E. Pfenninger, and J. Navracsics, Eds.: Multilingual Matters, Bristol, Buffalo, Toronto, 73-98.

[27] Bongaerts, T., Summeren, C. van, Planken, B., and Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language, *Stud. Sec. Lang. Acquis.*, 19, 447-465.