VINCENT J. VAN HEUVEN

Department of Applied Linguistics, University of Pannonia, Veszprém, Hungary
v.j.j.p.van.heuven@hum.leidenuniv.nl

Vincent J. van Heuven: An acoustic characterisation of English vowels produced by
American, Dutch, Chinese and Hungarian speakers
Alkalmazott Nyelvtudomány, XVI. évfolyam, 2016/2. szám
doi:http://dx.doi.org/10.18460/ANY.2016.2.004

An acoustic characterisation of English vowels produced by American, Dutch, Chinese and Hungarian speakers[1]

This is a comparative study of ten monophthongs of English pronounced in an /hVd/ environment by Chinese, Dutch, Hungarian, and American native speakers. Pronunciation problems were predicted by comparing traditional auditory vowel diagrams of source and target languages. Vowel duration and the resonance frequencies F1 (close-open) and F2 (front-back) were measured. Human vowel recognition was simulated by Linear Discriminant Analysis by training the algorithm with the vowel tokens produced by each of the four groups of speakers in turn and testing the model with the vowels of all four groups. The vowels of Chinese (66%) and Hungarian (59%) speakers are correctly identified less often by the American native model than the vowels of the Dutch speakers (77%). The native vowels were identified best (92%). Vowel identification was better overall when training and test languages were the same, which can be seen as a computer simulation of the interlanguage speech intelligibility benefit.

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* whenever professionals from different countries around the world have to communicate with one another (e.g. Rogerson-Revel 2007). The use of spoken English as a Lingua Franca (ELF) is not without problems, however. When an adult learns to speak a foreign language its pronunciation will differ substantially from that of native speakers 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. This has led to the use of somewhat disparaging names for such foreign accents as Chinglish (Chinese English), Dunglish (Dutch English) and by extension possibly also Hunglish (Hungarian English).

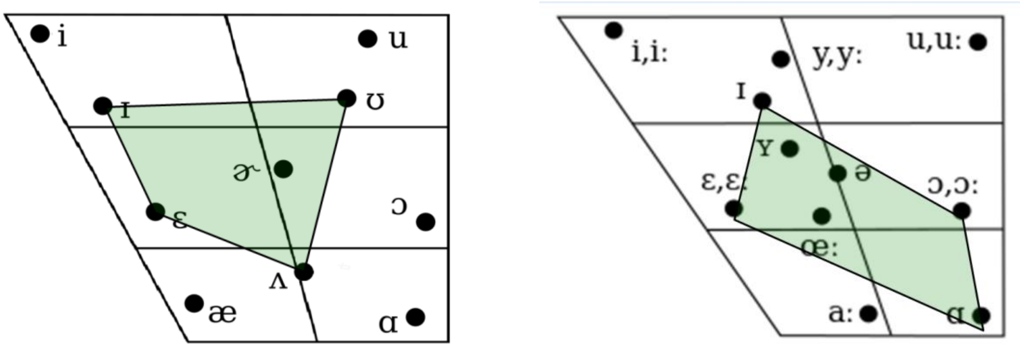
Communication is easiest when both interactants use the same native language (e.g. Munro & Derwing 1995). When at least one interactant is a non-native, communication generally suffers, not only due to improper use of words and flawed syntax but also, and especially, because the foreign-accented pronunciation compromises the recognition of words (Cutler 2012). Communication in ELF is somewhat more successful when both interactants share the same mother tongue. This so-called *interlanguage speech* *intelligibility benefit* (Bent & Bradlow 2003; Van Heuven 2015) may cause ELF users to overrate their pronunciation and listening skills.

This study is part of a research enterprise that aims to map out the difference in pronunciation of English vowels, single consonants, consonant clusters and short sentences by various speaker groups. Recordings of 20 native speakers of American English (10 males, 10 females) served as the baseline against which the performance of non-native speakers (20 Dutch and 20 Chinese speakers) was gauged. The intelligibility of the various types of native and non-native English was determined in listening tests. In a second stage, features of the pronunciation of vowels and consonants were established through acoustic analysis. Intelligibility of the produced tokens was additionally determined by software simulating the speech perception by human listeners (simulating all three listener groups) so that we were able to compare the performance by human listeners with that of the computer algorithm. The recognition of the English vowels by the three groups of human listeners (i.e. Chinese, Dutch and American) could be modelled accurately by the algorithm (Wang 2007).

The first aim of the present project was to include materials recorded from Hungarian speakers of English, and examine how the vowel sounds in Hungarian-accented English differ from those in native (American) English as well as from those in Chinese and Dutch-accented English. A second deliverable will be an algorithm that determines the native language underlying the non-native accent. This will tell us how a Chinese, Dutch and Hungarian accent can be efficiently discovered by a computer algorithm. The algorithm may also be applied to evaluate the strength of the foreign accent. This in turn can be used as feedback in language-learning curricula. The present extension of the project does not involve human listeners. Results will be entirely based on the measurement of relevant acoustic properties of the vowels and the simulation of human listeners by a computer algorithm.

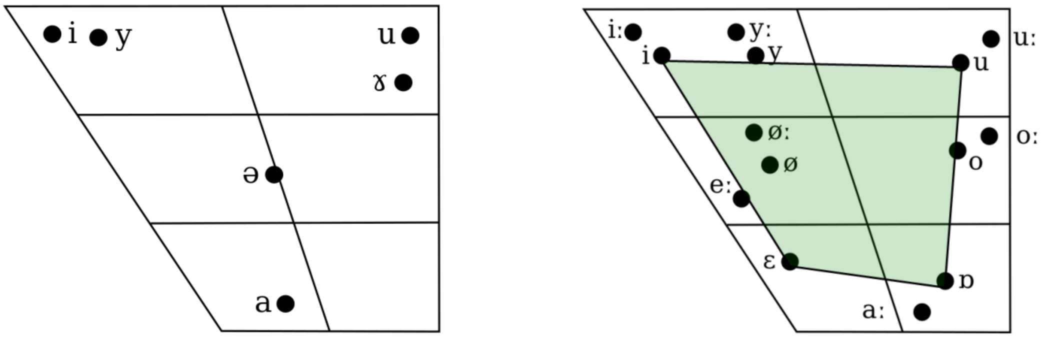
2. The target vowel systems

Figure 1 shows the vowel systems of the four target languages involved in the present study, i.e. Hungarian, Dutch, (Mandarin) Chinese and (American) English. The Mandarin inventory should probably be extended with the [ɛ] and [ɔ], as frequently occurring allophones of /a/. Dutch and English are relatively similar since both have tense and lax subsystems. Phonetically, tense vowels are produced with more extreme positions of the articulatory organs, which require more effort and time – so that the tense vowels are longer than lax vowels. Tense vowels form a natural class in English and Dutch on phonotactic grounds: they may occur in open syllables. Lax vowels are articulated with less effort, assume less extreme positions in the articulatory space, and have relatively short durations. Phonemically, they cannot occur in open syllables in English and Dutch but have to be followed by a coda consonant.[2] Since the contrast between the members of tense-lax pairs is coded along both color and duration dimensions, the difference in each dimension is expected to be smaller than in a language that makes the contrast exclusively in terms of duration (such as Hungarian). Mandarin differs from the European languages in that all its vowels are described as tense while length plays no role (Wang 2007).



A. American English (Mannell et al. 2009)

B. Dutch (Gussenhoven 1992)



C. Mandarin Chinese Lee & Zee (2003)

D. Hungarian (Szende 1994)

Figure 1. The monophthongs of (A) American English, (B) Netherlandic Dutch, (C) Mandarin Chinese and (D) Hungarian. The shaded areas in the English and Dutch charts enclose the lax subsystems; in the Hungarian system they connect the short vowels. With the exception of /aː/ the long vowels in Dutch only occur in (typically French) loans.

We expect Dutch ELF speakers to be reasonably successful in approximating the English tense versus lax vowel contrast. Since Mandarin has no length contrast, Mandarin ELF speakers should differentiate relatively poorly between the long (tense) and short (lax) vowels in English. Hungarian ELF speakers are expected to exaggerate the length contrast in English. Since duration is the dominant cue marking the difference between the short and long vowels in Hungarian the stronger length difference will transfer to English.

The trapezoids in Figure 1 are a stylized version of the articulatory vowel space. The horizontal dimension indicates the location where the supralaryngeal tract (i.e. throat and mouth) is most narrowly constricted, whereas the vertical dimension represents the distance between the tongue at the place of narrowest constriction and the roof of the mouth (hard palate, for front vowels) or the back of the throat (pharynx, for back vowels). The outer perimeter of the trapezoid defines the ultimate possibilities for the human speech organs to articulate a vowel sound.

The center frequencies of the lowest two resonances (‘formants’) produced by the supralaryngeal tract, i.e. by the throat cavity and the mouth cavity, are a good indication of how a vowel sound is articulated. The first formant (F1) captures the degree of openness of a vowel (vertical axis in Figure 1), where low F1 values (ca. 200 Hz for a male voice) are characteristic for close vowels such as [i, u] and high F1 values (ca. 800 Hz) represent open vowels such as [a]. The second formant (F2) reflects the front-back dimension (horizontal axis in Figure 1), where low F2 values (ca. 600 Hz) are found for back vowels [u] and high F2 values define front vowels. The highest F2 found for a male speaker is at 2400 Hz, which would define an [i].

The charts in Figure 1 were drawn by experienced phoneticians who located the vowels in the chart by listening and imagining what articulatory gesture would be needed to produce a vowel with a particular color. Although expert listening is often quite accurate and reproducible, most phoneticians prefer to describe the articulation of vowels more objectively using formant measurements. This is the method that is used in our project.

3. Methods

In this article we will only consider the pronunciation of vowels. A list of words (Table 1) was compiled containing the 19 full vowels and diphthongs of English (excluding schwa) in identical /hVd/ contexts. This consonant frame is fully productive in English, allowing all the vowels of English to appear in a meaningful utterance, either a word, a name or a short phrase (Peterson and Barney, 1952).

When pronounced by American speakers, the so-called centring diphthongs (ending in a schwa-like element) will often be monophthongs followed by an approximant /r/ sound. Also, the contrast between vowels 9, 10 and 11 (the latter as in *father*) may be neutralized in American English. We decided to elicit the full set of potential contrasts, but kept post-hoc pooling of vowels (as stimulus and as response categories) as an option.

Table 1. The 19 vowel sounds of English in /hVd/ context, with transcription and key words.

Vowel		Trans.	Key words	Vowel		Trans.	Key words
1.	h**ee**d	/hid/	f**ee**d, n**ee**d	11.	h**ar**d	/hɑd/	c**ar**d, b**ar**red
2.	h**i**d	/hɪd/	m**i**d, k**i**d	12.	h**u**d	/hʌd/	m**u**d, bl**oo**d
3.	h**ay**ed	/hed/	pl**ay**ed, st**ay**ed	13.	h**ear**d	/hɜd/	b**ir**d, w**or**d
4.	h**ea**d	/hɛd/	r**e**d, b**e**d	14.	h**i**de	/haɪd/	sl**i**de, r**i**de
5.	h**a**d	/hæd/	b**a**d, s**a**d	15.	h**oy**ed	/hɔɪd/	t**oy**ed, empl**oy**ed
6.	wh**o’**d	/hud/	gl**ue**d, r**u**de	16.	h**ow**’d	/hɑʊd/	l**ou**d, all**ow**ed
7.	h**oo**d	/hʊd/	g**oo**d, w**oo**d	17.	h**ere**’d	/hɪəd/	b**ear**d, sn**eer**ed
8.	h**oe**d	/hod/	r**oa**d, sh**ow**ed	18.	h**oore**d	/hʊəd/	t**our**ed, m**oor**ed
9.	h**aw**ed	/hɔd/	s**aw**ed, fr**au**d	19.	h**aire**d	/hɛəd/	sh**are**d, c**are**d
10.	h**o**d	/hɒd/	g**o**d, n**o**d				

Participants in this study were 20 native speakers of Hungarian, either students at the University of Pannonia in Veszprém or fellows and staff at the ISES Research Centre in Kőszeg. Ten speakers were male, ten were female. None had received special training in English, had spent extended periods of time in English-speaking countries had had intensive contacts with native speakers of English. They are held to be representative of academically educated Hungarian ELF speakers. Speakers volunteered and were not paid.

Speakers were recorded in individual sessions. They received written instructions with the words and phrases listed in table 1, printed in the context of a carrier phrase *Now say ... again*. The list also contained the key words that contained the target vowel in common English words. Speakers were then asked to read out the list of materials twice with a short break in between.

Recordings were made in a quiet room on a noiseless hybrid notebook/tablet (44.1 KHz, 16 bit), using a good quality digital headset microphone (Sennheiser MM60).

4. Acoustic analysis

Given that our speakers, including the Dutch and Chinese speakers, used an American style of pronunciation, without centring diphthongs, there seems little point in measuring vowels followed by /r/. Therefore, we eliminated the tokens of *here’d, haired, hard*, *hoored* and *heard*. Next, we excluded full diphthongs as these would introduce the complication of having to trace the spectral change over the course of the vowels. This eliminated the types *hide, how’d* and *hoyed*. Finally we eliminated /ɔ/. Most of our Hungarian ELF speakers did not systematically differentiate between this vowel and /ɒ/. Moreover, some speakers incorrectly pronounced *hawed* as /haud/.

Onsets and offsets of target vowels were determined by ear and by eye, using the oscillogram and spectrogram displayed by the Praat speech analysis software (Boersma & Weenink 1996). Onsets were defined by the earliest absence of aspiration noise, offsets were at the point in time where formants disappeared from the spectrogram. Formants were estimated by the Burg LPC algorithm. The optimal LPC model order and frequency cut-off were found by trial and error, visually comparing formant tracks superposed on the spectrogram until the tracks matched the spectrogram. Vowel duration (in milliseconds, ms) and the centre frequencies of maximally five formants (in hertz, Hz) 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, using Traunmüller's (1990) formula.

Since formants for the same vowel differ when the vowels are produced by different individuals a simple vowel normalization procedure was applied, i.e. z-normalization of the F1 and F2 frequencies over the vowel set produced by each individual speaker. This is done by subtracting the individual speaker's mean F1 and mean F2 from the raw formant values, and then dividing the difference by the speaker's standard deviation. Z-transformed F1 values < 0 correspond to relatively close (high) vowels, values > 1 refer to rather open vowels. Similarly, negative z-scores for F2 refer to front vowels, whilst positive z-scores for F2 represent back vowels. Normalization was applied after Bark transformation.

Since some speakers speak faster than others, durations were also normalized within speakers. Here, negative z-values refer to relatively short vowel tokens, and positive values represent long vowel durations.

5. Results

I will first present the general characteristics of the English vowels as produced by the four groups of speakers. Specifically we will analyse the location of the various vowels in the acoustic vowel space, defined by the centre frequencies of the first formant (F1) as a correlate of vowel openness (or 'height') and of the second formant (F2) as a correlate of vowel backness. Vowel duration will be analysed as a third correlate. The locations of the vowels in the acoustic vowel space will be defined in Bark units (see above) so as to be auditorily realistic. Moreover, since male and female vocal tracts differ in length we will present the results separately for male and female speaker groups.

In a second part I will attempt automatic vowel recognition of individual vowel tokens. We will train an automatic classification algorithm (Linear Discriminant Analysis) with vowel tokens produced by American native speakers of English, and then see how well the tokens produced by the three non-native speaker groups are recognized by the native model. Since the classification algorithm applies to individual vowel tokens, the input data will be normalised within sexes and speakers before running the LDA. The automatic classification will be performed twice, once with and once without vowel duration as a third measure of vowel identity. We will also train the recognizer with tokens of Chinese-accented, Dutch-accented, and Hungarian-accented English, and determine how well the vowels produced by each of the speaker groups are then classified. This will allow us to quantify the magnitude of the so-called interlanguage speech intelligibility benefit (ISIB, see above) in our materials.

Finally, we will examine the possibilities of automatic identification of the speaker's native language background.

5.1. General characteristics

The mean F1 and F2 values are plotted (in Bark) in acoustical vowel diagrams in Figures 2A-H. Panels G and H contain the reference data collected for the American native speakers of English, for male and female speakers respectively. These data were collected in the predecessor project (Wang & van Heuven 2006, Wang 2007). Each plot contains the position of the ten monophthongs selected as explained at the beginning of this section. Panels E and F present the results obtained for the Hungarian ELF speakers.[3] For the sake of comparison panels A and B present the same information for male and female Chinese ELF speakers, while panels C and D contain the results obtained for Dutch speakers. The ten vowels have been subdivided into one group of six tense vowels and a second group of four lax vowels. The vowels in the lax group have been joined by a shaded polygon. It is easy to see in the American native speaker plot that the four lax vowels are the corner points of a much smaller subspace within the larger tense-vowel polygon (including ash, see above).

It is quite clear from Figures 2A-H that the configuration of vowels in the acoustic space is very much the same for male and female speakers. Starting with the American native speaker results, we see that the four lax vowels differ substantially in terms of their locations in the F1-F2 space from the six tense vowels on the outer perimeter of the space. It would seem attractive to analyse the native vowel system in terms of four pairwise oppositions, i.e. /i/~/ɪ/, /u/~/ʊ/, /æ/~/ɛ/, /ɒ/~/ʌ/, and two quasi-monophthongs (or semi-diphthongs) /e/ and /o/. The lax member in each pair is clearly more centralised than its tense counterpart; /e/ and /o/ will be mainly differentiated from their nearest spectral competitors by duration.

The Chinese ELF speakers hardly differentiate between the tense and lax members of the pairs /i/~/ɪ/, /u/~/ʊ/, /æ/~/ɛ/. They do, however, differentiate /ɒ/~/ʌ/ but the way this is done differs from the native system: Rather than being more centralised, the Chinese-accented /ʌ/ is more open than its tense counterpart (as well as being fronted). Note, finally, that the (half) close back vowels (/u/~/ʊ/ and /o/ have lower F2 values than open /ɒ/, which behaviour is clearly different than what we see in the native vowels, where the non-low back vowels are rather more centralised.

Dutch-English differentiates quite well between /i/ and /ɪ/, which contrast exists in the Dutch vowel system, and relies on the difference in vowel quality rather than on duration. The Dutch speakers of English do not differentiate between /u/-/ʊ/. They do, however, strongly differentiate between /ɒ/ and /ʌ/. Differentiation between /æ/ and /ɛ/ is modest, which probably means that some speakers make the difference and others do not (depending on their degree of familiarity with the English system).

We now come to the Hungarian-accented vowels of English. There is virtually no spectral difference between the members of the pairs /i/~/ɪ/, /u/~/ʊ/, /æ/~/ɛ/. There is a substantial spectral difference between the members of the /ɒ/~/ʌ/ pair, but the way the contrast is made is reminiscent of that produced by the Chinese speakers. In fact, the configurations of the Hungarian and Chinese ELF vowels, when expressed in terms of vowel quality, have a lot in common. The most salient difference between the two accents would seem to reside in the way the (half) close vowels are produced. In Chinese ELF there is a substantial distance along the F1 dimension between half close /e/ and /o/ and their close-vowel counterparts. In Hungarian ELF the half-close and close vowels have almost the same F1 values.

Table 2 summarizes the differences in the configurations of English vowels for the four speaker groups at issue.

Table 2. Distinguishing vowel quality properties of English monophthongs in four speaker groups. '✓': property present, '✗': property absent, '?': property unclear.

Pair	Property	L1			
		Chinese	Dutch	Hung.	Am. Eng.
1. /i/-/ɪ/	Centralization of /ɪ/	✗	✓	✗	✓
2. /u/-/ʊ/	Centralization of /ʊ/	✗	✗	✗	✓
3. /æ/-/ɛ/	Contrast along F1 (openness)	✗	?	✗	✓
4. /ɒ/-/ʌ/	/ɒ/ more open than /ʌ/	✗	✗	✗	✓
5. /e/-/i/	Contrast along F1 (openness)	✓	✓	✗	✓
6. /o/-/u/	Contrast along F1 (openness)	✓	✓	✗	✓
7. /ɒ/-/u/	Centralisation of /u/	✗	✓	✗	✓

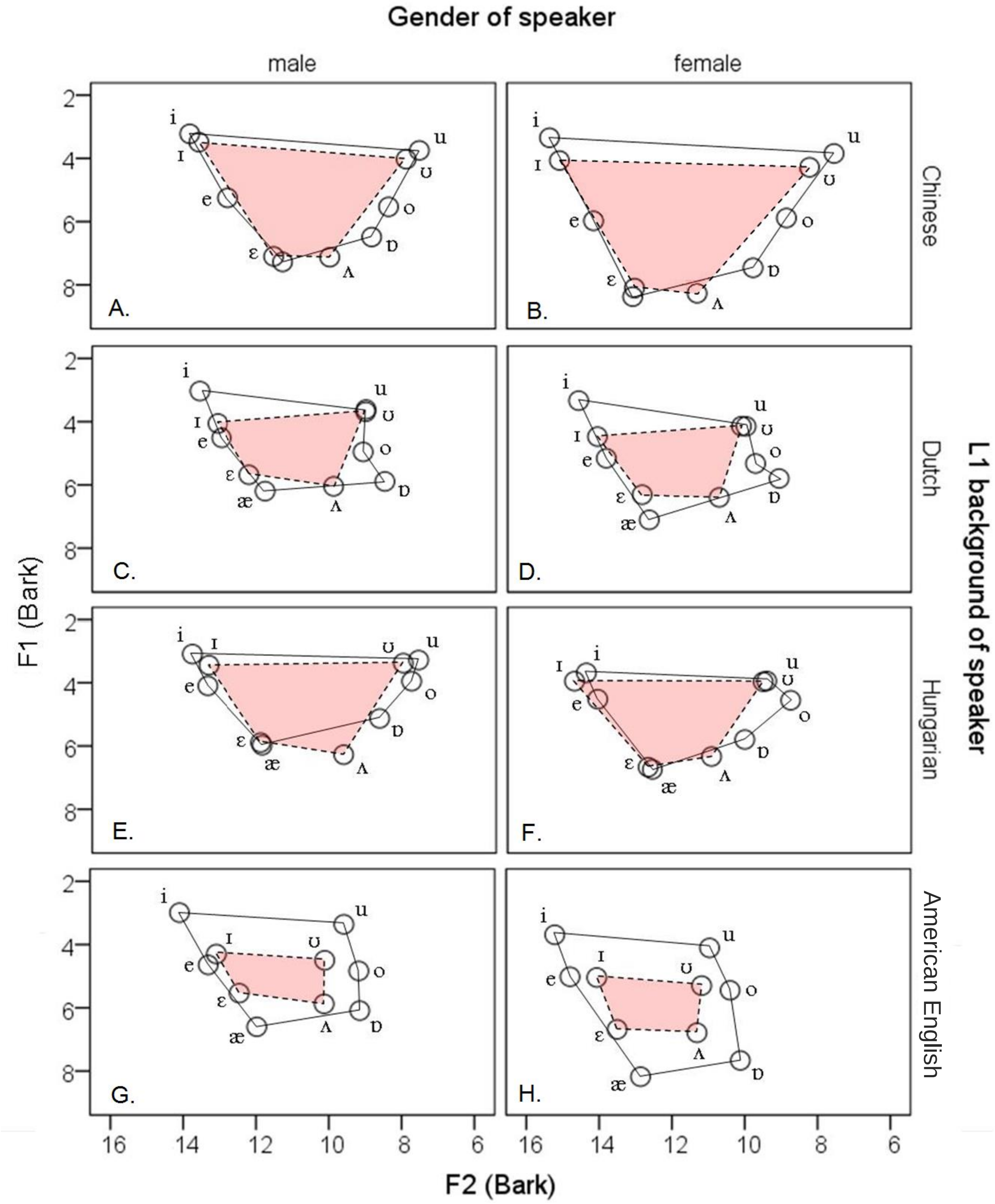


Figure 2A-H. Mean F1 and F2 (Bark) of the ten English monophthongs plotted separately for tense (solid polygons) and lax (dotted polygons) vowels for eight groups of speakers.

Hungarian ELF lacks seven properties that are all characteristic of (American) English. The crucial difference between Chinese and Hungarian English is in the greater distance between the close and half-close vowels (both front and back). On the basis of Table 2 we would predict Dutch ELF to be closest to American English (2 to 3 features differ), followed by Chinese ELF (5 features differ), whilst Hungarian ELF would differ most (all 7 features differ). These features pertain to vowel quality only. Let us therefore now consider vowel duration.

Figure 3 plots the durations measured for the ten vowels for each group of speakers. Vowels have been arranged in ascending order of length as observed in the reference language, i.e. American English.

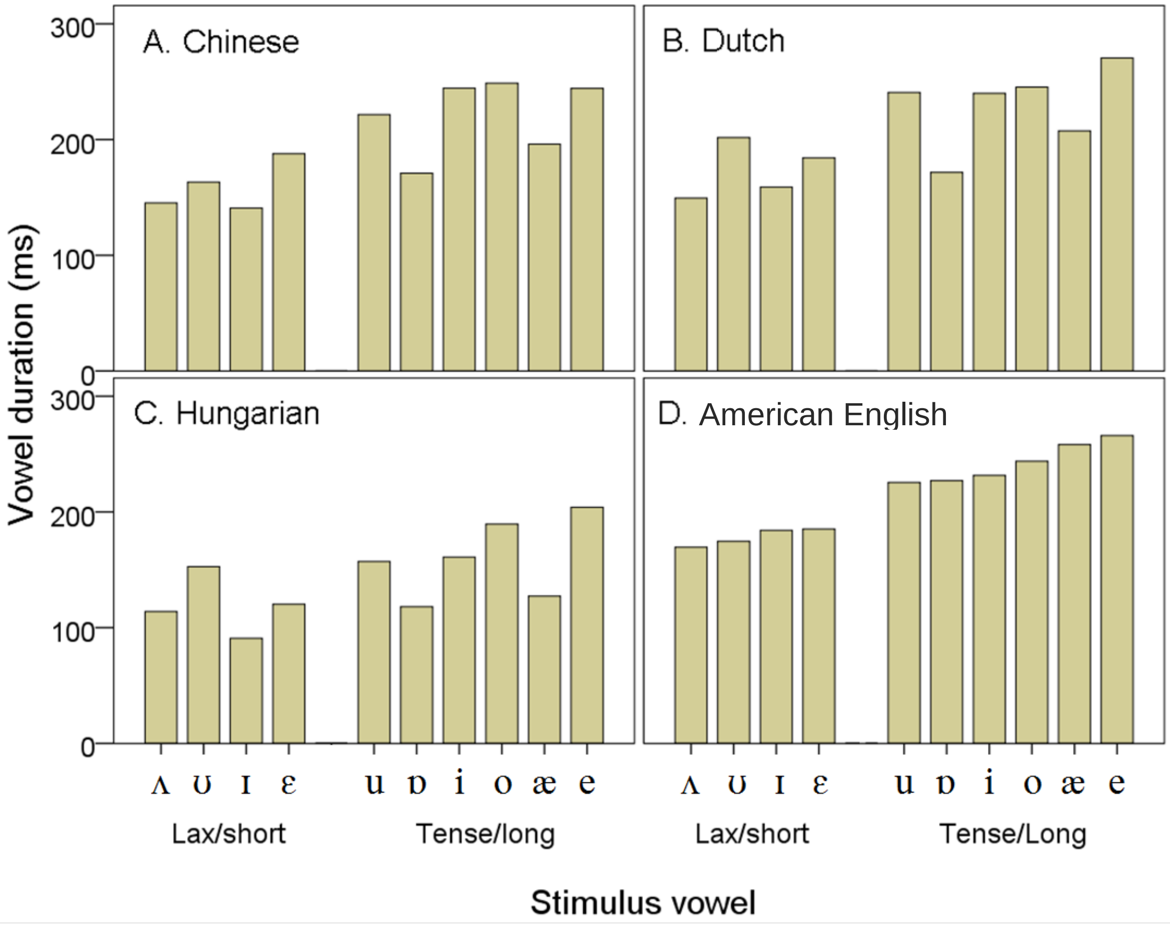


Figure 3. Duration (ms) of the four short/lax and six long/tense vowels of American English, spoken by Dutch (A), Chinese (B), Hungarian (C) and American (D) speakers of English.

The duration data show that the tense and lax vowel groups are quite systematically separated by the American native speakers. The longest lax vowel (/ɛ/) is clearly shorter than the shortest of the tense vowels (/u/). In the Chinese, Dutch and Hungarian results, however, the durations of the tense vowels /ɒ/ and /æ/ are appreciably shorter than those of the other tense vowels, and tend to fall within the range of the lax vowels. Within the lax subset the durations of /ʊ, ɛ/ are longer than those of /ʌ, ɪ/ in the three non-native Englishes.

Duration will be a helpful, if not necessary, cue to distinguish between vowels that are close to one another in the spectral vowel space. This may apply to the contrasts /i/~/ɪ/ and /u/~/ʊ/, and possibly even more so to the pairs /e~/ɪ// and /o/~/ʊ/.

5.2. Automatic vowel classification

Figure 4A-D plots the individual realisation of the vowels in the F1 by F2 plane as scatter clouds, enclosed by spreading ellipses (which include the central 46% tokens of a vowel type).

The Chinese speakers have more overlap between the ellipses of neighbouring vowels than is the case in the Dutch ELF realizations. Overlapping ellipses (i.e. poor separation between vowel categories) are observed in the Chinese results for the pairs /i, ɪ/, /æ, ɛ/ and /u, ʊ/. In the Dutch results there is complete overlap for /u, ʊ/, as well as partial overlap for /æ, ɛ/ and /o, ɒ/. The American native speakers have the smallest degree of overlap between neighbouring vowels. Only the ellipses of /ʊ/ and /o/ overlap somewhat but these two vowels differ substantially in their duration so that little confusion will arise between the two. The Hungarian ELF speakers resemble the Chinese: there is considerable overlap between the members of the pairs /i, ɪ/, /æ, ɛ/ and /u, ʊ/. There is also partial overlap between /ɒ/ and /ʌ/.

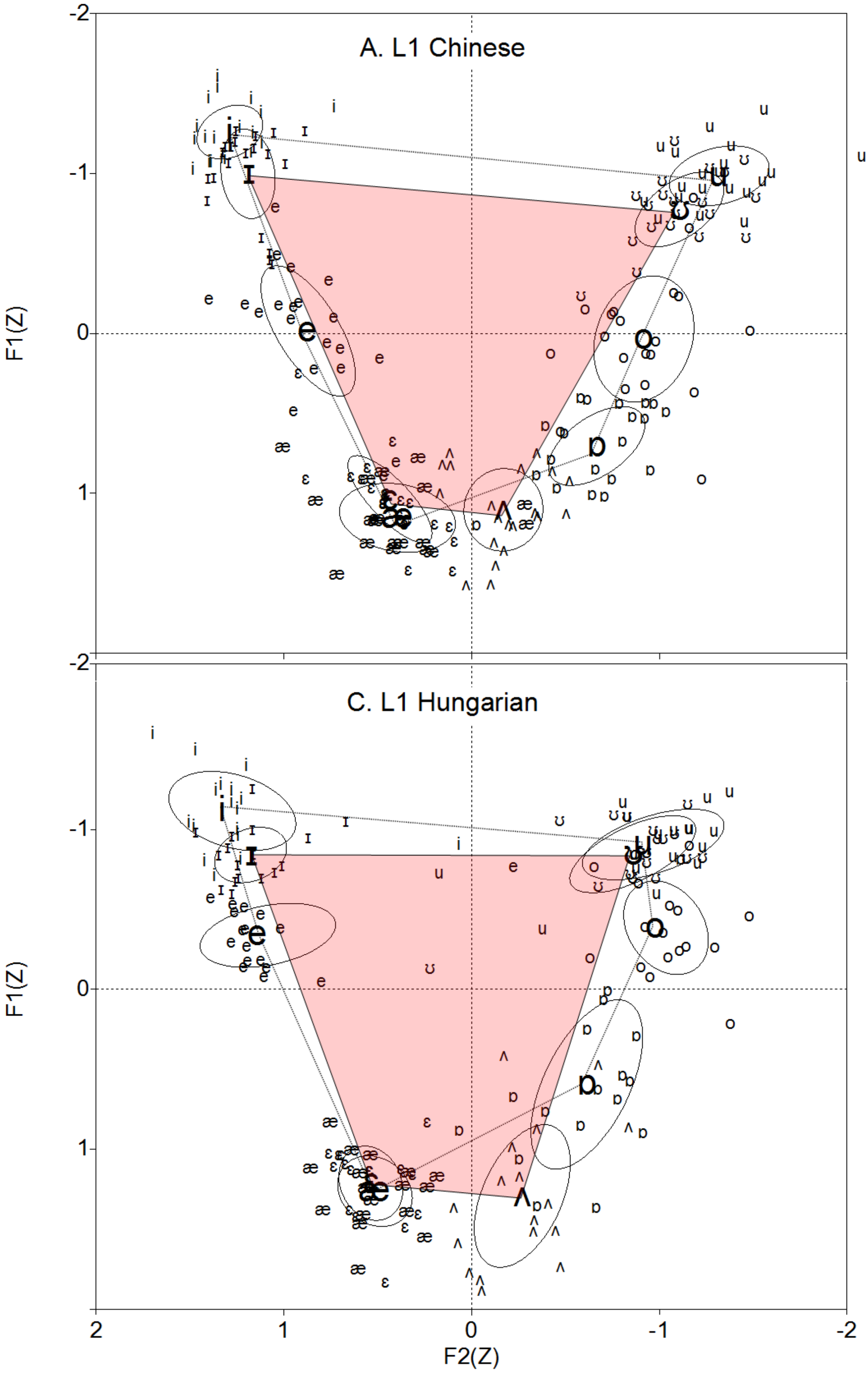


Figure 4. Individual vowel points for (A) Chinese and (C) Hungarian speakers of English (Bark transformed and then z-normalized within speakers) plotted in the F1 by F2 plane, with spreading ellipses drawn at +/−1 SD from the centroid along the first two principal component axes of the scatter clouds.

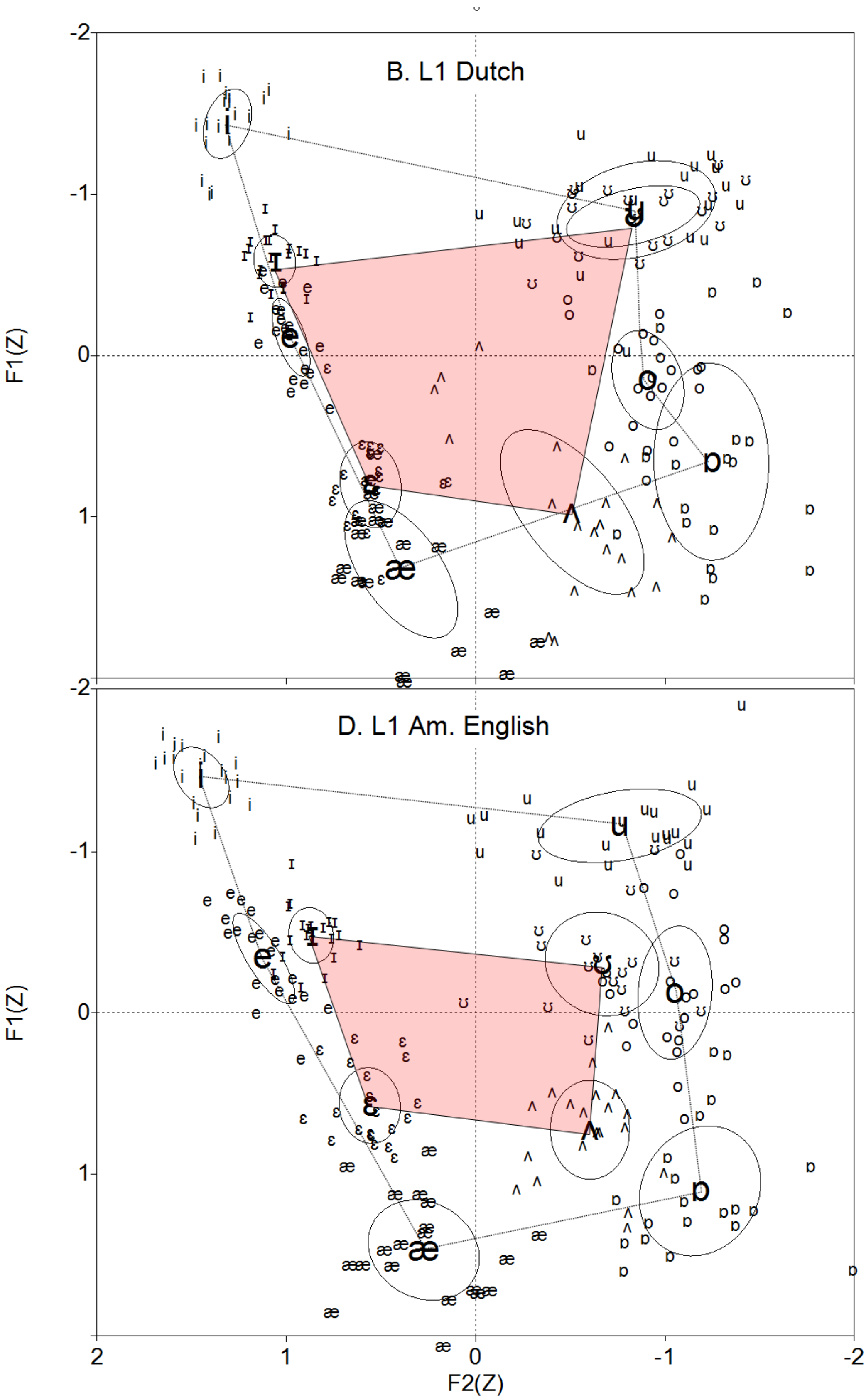


Figure 4 (continued). Individual vowel points for (C) Dutch and (D) American speakers of English. Tense vowels are joined by the non-shaded polygon; lax vowels are the corner points of the shaded polygons.

We will now quantify the difference between the four speaker groups in terms of the degree of success in keeping the ten vowels distinct. We have used Linear Discriminant Analysis (LDA) for this purpose. LDA is an algorithm that computes an optimal set of parameters (called discriminant functions) which automatically classifies objects in pre-established categories (e.g. Weenink 2006). The more distinct the categories are, the fewer the classification errors yielded by the algorithm. We ran the LDA twice. The first time we just included the two spectral parameters as predictors of vowel identity, i.e. F1 and F2 (converted to Bark and z-normalized within individual speakers). The second time we also included (z-normalised) vowel duration. Figure 5 presents these results.

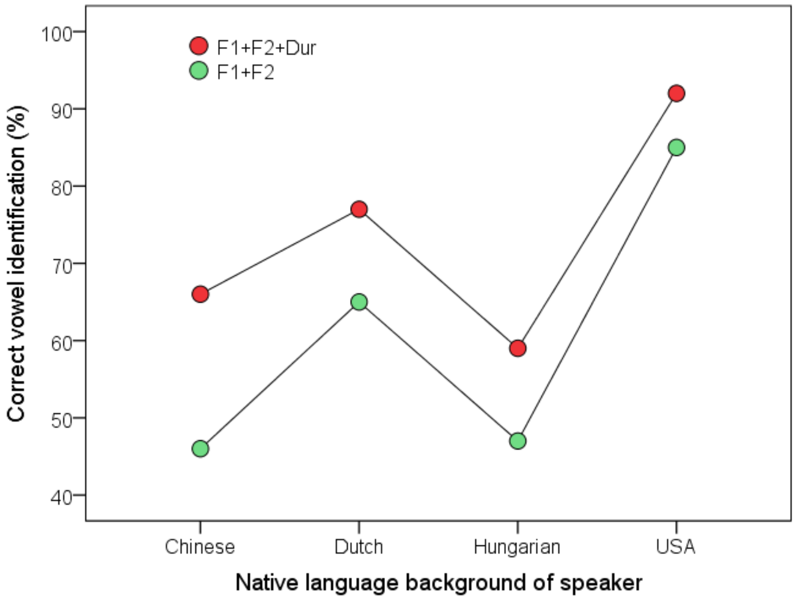


Figure 5. Correctly classified vowel tokens (%) by LDA with F1 and F2 (and duration) as predictors for eight groups of speakers.

Figure 5 shows that the vowels as spoken by the native speakers afford the best automatic identification, those spoken by the Dutch learners can be less successfully identified, and the Chinese and Hungarian ELF tokens are poorest. Adding duration to the set of predictors boosts the correct identification by 15 to 20 percentage points. The idea that Hungarian speakers would exploit duration more strongly, and Mandarin speakers least, on account of the native language phonology, finds no support in these results.

A more detailed view of the LDA results is presented in Table 3, where percent predicted vowel identity is cross-tabulated against the intended vowel identity for the four speaker groups.

Table 3. Crosstabulation of intended and recognized vowels (%) for four groups of speakers. Classification by LDA based on **American native tokens**. Predictors are F1, F2 (both Bark transformed) and duration (all z-normalised within speakers). Native speaker results based on leave-one-out cross-validation. Correct classifications are in the shaded cells.

		Mandarin Chinese ELF speakers: 66% correct (N = 20)									
		Vowel classified as									
		æ	**e**	**ɛ**	**i**	**ɪ**	**ɒ**	**o**	**ʊ**	**ʌ**	**u**
Intended vowel	**æ**	**60**		40							
	e	10	**75**			15					
	ɛ	40	5	**55**							
	i				**100**						
	ɪ				45	**55**					
	ɒ	5					**10**	5		80	
	o						5	**85**		5	5
	ʊ							5	**50**		45
	ʌ	5		20						**75**	
	u							10			**90**
		Dutch ELF speakers: 77% correct (N = 20)									
		æ	**e**	**ɛ**	**i**	**ɪ**	**ɒ**	**o**	**ʊ**	**ʌ**	**u**
Intended vowel	**æ**	**70**		30							
	e		**100**								
	ɛ	15		**85**							
	i				**100**						
	ɪ					**100**					
	ɒ						**60**	15	10	15	
	o						5	**85**	5	5	
	ʊ							5	**30**		65
	ʌ			15			15		5	**65**	
	u							15	10		**75**
		Hungarian ELF speakers: 59% correct (N = 17)									
		æ	**e**	**ɛ**	**i**	**ɪ**	**ɒ**	**o**	**ʊ**	**ʌ**	**u**
Intended vowel	**æ**	**42**		59							
	e		**91**			6					3
	ɛ	32		**68**							
	i		6		**82**	9					3
	ɪ		3		15	**82**					
	ɒ	9		3			**6**	3	21	56	4
	o			3			3	**71**	3		21
	ʊ			3	3	3		6	**21**		65
	ʌ	32		3					3	**59**	3
	u			3		6		12	6		**74**
		American native speakers: 92% correct (N = 20)									
		æ	**e**	**ɛ**	**i**	**ɪ**	**ɒ**	**o**	**ʊ**	**ʌ**	**u**
Intended vowel	**æ**	**100**									
	e		**95**			5					
	ɛ			**100**							
	i				**100**						
	ɪ					**100**					
	ɒ						**85**	10		5	
	o						5	**75**	5		15
	u			5				5	**75**		15
	ʌ						10		5	**85**	
	u										**100**

The results obtained for the Chinese-accented vowel tokens reveal two major problems, viz. the more or less symmetrical confusion of /ɛ/ and /æ/ and an asymmetrical confusion of lax /ʊ/ with tense /u/ (but not *vice versa*). These pronunciation errors follow from a traditional contrastive analysis, and were also noted in a pedagogical textbook (Zhao, 1995).

For the Dutch speakers we find two symmetrical error patterns, i.e. /ɛ/~/æ/ and /ʊ/~/u/, which were predicted by contrastive analyses (Table 3.4 in Wang 2007) and were noted in the pedagogical literature (Tables 3.5-6 in Wang 2007). The incorrect classification of intended vowel /ʌ/ as a front vowel /ɛ/ was not predicted.

Table 4 summarizes the results obtained in the LDA of vowel tokens produced by the four types of speaker of English and automatically identified by models trained by each of the four speaker groups in turn. A model trained on the tokens produced by speaker group X is considered a model of a listener with X as his/her language background.

Table 4. Identification of English monophthongs produced by four types of speaker and identified by listeners (simulated by LDA) with the same four languages as the L1. Conditions in which the L1 of speaker and (simulated) listener is the same ('matched interlanguage') are on the main diagonal (bold face in shaded cells). When either speaker or (simulated) listener is not a native speaker of English we have a situation of non-matched interlanguage.

Speaker/ test language	Simulated listener language					Simulated listener language				
	Chinese	Dutch	Hung.	USA	Mean	Chinese	Dutch	Hung.	USA	Mean
Mandarin	**76**	64	54	66	65	**12.1**	−2.4	−2.4	−7.2	.0
Dutch	63	**80**	54	77	69	−4.4	**10.1**	−5.9	.3	.0
Hungarian	58	51	**65**	59	58	.8	−8.7	**15.3**	−7.4	.0
USA	60	72	54	**92**	70	−8.4	1.1	−6.9	**14.3**	.0
Mean	64	67	57	74	**65**	.0	.0	.0	.0	.0
	Absolute scores					Relative ISIB				

Table 4 shows that vowel identification is better when the LDA was trained on the language data that it was tested with. The superiority of matched training and test data obtains even though the tests were never performed on identical tokens (using crossvalidation). The difference is 17.25 points in favor of the scores based on matched training and test data, $t(14) = 3.5$ ($p = .002$, one-tailed). The effect of the 'matched interlanguage' is easier to see if we express it in relative terms, using the computational procedure advocated by Van Heuven (2015). This is shown in the right-hand part of Table 7. The advantage of the matched training and test sets remains the same, but the variability in the scores is much reduced by factoring out the differences due to overall effects of speaker and listener native language background, $t(14) = 8.8$ ($p << .001$, one-tailed).

5.3 Automatic identification of language background

Finally, let us determine how well the native language background of the speaker can be established by examining the way s/he pronounces the (monophthongal) vowels of English. The data 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 diphthongized /e/ and /o/. There are 77 speakers (20 Mandarin, 20 Dutch, 17 Hungarian 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. This is a richness of predictors so that severe reduction of the number of predictors is called for. We did this by running the LDA in stepwise mode, starting with the predictor that differentiates best between the four speaker groups and including additional predictors one by one, and only if they made a significant improvement (in terms of Wilk's Lambda) to the overall performance of the decision algorithm. Table 5 presents the results of the classification. Overall, using cross-validation, 81% of the 77 speakers were correctly classified in terms of their native language background.

Table 5. Classification by LDA of L1 background of four groups of speakers. Results are based on cross-validation using the leave-one-out method.

	L1 of speaker	Predicted L1				Total
		Chinese	Dutch	Hungarian	USA	
Count	Chinese	**14**	3	3		20
	Dutch		**18**		2	20
	Hungarian	4		**13**		17
	USA		3		**17**	20
%	Chinese	**70**	15	15		100
	Dutch		**90**		10	100
	Hungarian	23		**77**		100
	USA		15		**85**	100

Interestingly, the Mandarin-Chinese and the Hungarian ELF speakers were less successfully classified for native language background (70 and 77% correct, respectively) than the Dutch speakers (90% correct) and the native speakers (85% correct). Moreover, it is never the case that a Mandarin or Hungarian speaker is mistaken for a native speaker of English.

There are three types of confusion: (i) Hungarians may be incorrectly classified as Chinese speakers (or *vice versa*), (ii) Dutch speakers are incorrectly classified as American native speakers (or *vice versa*), and (iii) Chinese speakers may be mistaken for Dutch speakers (but not *vice versa*. This confusion structure suggests that the overall vowel pronunciation of the Dutch speakers resembles that of native speakers of English more than that of either Chinese or Hungarian accented speakers, who would also have certain properties in common.

The performance obtained by the LDA is based on the contribution of just five predictors; Twenty-five other predictors were eliminated from the analysis as they did not make a sufficient contribution to the classification performance.

The most influential parameter is the F1 (vowel openness) of the lax vowel /ɪ/. Predictably, this vowel embodies a pronunciation problem for Chinese and Hungarian speakers of English as their native language does not differentiate between lax /ɪ/ and its nearest competitor in English, i.e. the tense vowel /i/. The same remark can be made, *mutatis mutandis*, for the F1 (or openness) of the lax back vowel /ʊ/, which is not differentiated from tense /u/ in any of the three groups concerned. The F2 of /ɒ/ is typically too high in Mandarin and Hungarian ELF, suggesting the more frontish vowel quality resembling that of English /ʌ/, which is indeed the most frequent confusion seen in Table 3. The contribution of F1 of /o/ would typically relate to Hungarian ELF. As shown in Table 3, /o/ is often misclassified as /u/ in Hungarian ELF. As Table 3 also shows, there is massive confusion of /ɛ/ and /æ/ in each of the non-Englishes. The fact that the inclusion of the duration of /æ/ as the fifth (and last) acoustic parameter contributing to language background detection, rather than either F1 or F2, suggests that the difference in duration between /æ/, which is phonetically long and tense in American English, and the lax and short neighboring vowels /ɛ/ and /ʌ/ is the hallmark of nativeness.

6. Conclusion and discussion

We have compared the English monophthongal vowel produced by native speakers of (American) English and non-native approximations to these vowels by speakers from three different native language backgrounds, i.e. (Mandarin) Chinese, (Netherlandic) Dutch, and Hungarian. Rather than using human listening or measuring physiological differences between vowels, the results of this study are entirely based on acoustic measurements, i.e. the centre frequencies of the lowest two resonances of the vocal tract (F1 as a measure of vowel openness, and F2 as a measure of combined vowel backness and liprounding) and the duration of the vowel.

We predicted that Dutch speakers of English would have an advantage of the tense-lax similarity in their native language. Chinese learners would have to learn the difference between the English tense and lax vowels as a new phenomenon. Hungarian speakers of English were hypothesised to be at a disadvantage: we expected Hungarians to substitute their native length contrast for the English tense-lax difference, so that differences in vowel quality would be underestimated (i.e. smaller) between the members of tense-lax pairs while differences in vowel duration would be exaggerated (compensating for the absence of quality differences).

The results reveal that overall the Dutch-accented vowels were identified best by the vowel identification routine that was based on the production of the same vowels by native speakers of American English: 77% of the Dutch-accented vowels were correctly identified (against a baseline of 92% correct identification for vowels produced by the native speakers). The Chinese-accented vowels were identified at 66% correct while the Hungarian ELF vowels were correctly identified in 59% of the cases. The results contain no indication that vowel duration was used differently by any of the three non-native speaker groups (Figure 3). Typically lax /ʊ/ was too long while tense /æ/ was too short. All three speaker groups failed to discriminate between the tense and lax counterparts in the pairs /u/~/ʊ/ and /æ/~/ɛ/ (although the Dutch speakers made some difference in the latter case). The Chinese and Hungarian speakers additionally failed to differentiate spectrally between /i/~/ɪ/. On top of this, the Hungarian speakers insufficiently differentiate the vowel qualities (color) of the tense mid vowels /e, o/ from the close competitors /i/ and /u/, respectively. We should note here that these observations would not be expected from a contrastive analysis of Hungarian and English: Hungarian, like English, distinguishes between close and mid vowel qualities. Beside the tense-lax pairs Hungarians and Chinese speakers of English have a serious problem differentiating the vowel pair /ʌ/ ~ /ɒ/, at least when the recognizer is trained on American English vowels. It may well be the case that the Hungarian speakers would do better if the recognizer were trained with British English vowels, in which the rather open pronunciation of /ɒ/ is replaced by the less open vowel quality /ɔ/.

The results of the experiment show that the English vowels are better identified as intended by the speaker if the recognizer is trained and tested with vowels produced by the same speaker group. Correct vowel identification improves with 10, 3 and 6 points for Chinese, Dutch and Hungarian speakers of English, respectively. We take this as a convincing demonstration of the interlanguage speech intelligibility benefit (ISIB). The emergence of this effect in automatic vowel identification lends further credibility of the LDA technique as a realistic substitute for human vowel identification.

Our results can be of strategic use in the planning of the foreign-language curriculum. Communication in English as a foreign language is greatly facilitated if the sounds can be properly identified by the listener. Incorrect vocabulary and/or incorrect word order will only matter if words are recognized in the first place. It is imperative, therefore, to know which sounds in the target language are problematic for the foreign-language learner. Problems cannot be fully predicted from a contrastive analysis of the phonologies of the native and foreign languages involved (the present research is yet another example of the partial failure of contrastive analysis), so that the only option left is to establish the problem areas experimentally, which is what the present study aimed to contribute.

Notes

1. The Hungarian part of this research was part of TÁMOP 4.2.1.D-15/1KONV-2015-0006 “Development of the innovation centre in Kőszeg in the frame of the educational and research network at the University of Pannonia”, which is subsidized by the EU and Hungary and co-financed by the European Social Fund
2. The phonotactically lax open front vowel ‘ash’ (as in *had*) is generally considered phonetically tense in American English. It is pronounced at the edge of the articulatory space and its duration is clearly longer than that of lax vowels in English (e.g. Strange et al. 2004, Wang & Van Heuven 2006). In our own research we consistently treat ash as tense.
3. Three of the 20 Hungarian speakers had to be discarded from the dataset on account of the fact that they failed to produce a sufficiently complete set of vowel tokens.

References

Bent, T. & Bradlow, A. (2003) The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114. pp. 1600-1610.

Boersma, P. & Weenink, D. (1996) Praat, a System for Doing Phonetics by Computer. Report of the Institute of Phonetic Sciences Amsterdam, 132. www.praat.org

Cutler, A. (2012) Native listening: Language experience and the recognition of spoken words. Cambridge, MA: MIT Press.

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

Gussenhoven, C. (1992) Dutch. *Journal of the International Phonetic Association* 22, pp. 45-47.

Heuven, V. van (2015) A relative measure of the Interlanguage Speech Intelligibility Benefit: A meta-analytic exercise. In: Batyi, S. & Vígh-Szábo, M. (eds.) *Language – System, Usage, Application* (Studies in Psycholinguistics 5). Budapest: Tinta Könyvkiadó. 31-52.

Lee, W.-S. & Zee, E. (2003) Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33. pp. 109-112.

Mannell, R., Cox, F. & Harrington, J. (2009) *An introduction to Phonetics and Phonology*, Macquarie University. < http://clas.mq.edu.au/speech/phonetics/>

Munro, M. & Derwing, T. (1995) Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech*, 38. pp. 289-306.

Rogerson-Revell, P. (2007) Using English for international business: A European case study. *English for Specific Purposes* 26. pp. 103-120.

Szende, T. (1994) Illustrations of the IPA: Hungarian. *Journal of the International Phonetic Alphabet* 24, pp. 91-94.

Strange, W., Bohn, O.-S., Nishi, K. & Trent, S. (2004) Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 118, pp. 1751-1762.

Traunmüller, H. (1990) Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88. pp. 97-100.

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

Wang, H. & Heuven, V. van (2006) Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In: Weijer, J. van de & Los, B. (eds.) *Linguistics in the Netherlands 2006*. Amsterdam: John Benjamins. 237-248.

Weenink, D. (2006) *Speaker-adaptive vowel identification*. Doctoral dissertation, University of Amsterdam.

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