Cross-linguistic confusion of vowels produced and perceived by Chinese, Dutch and American speakers of English

Hongyan Wang and Vincent J. van Heuven
Jilin University / University of Leiden

1. Introduction

Foreign accent consists in a deviation from the generally accepted pronunciation norm of a language that is reminiscent of another language, i.e., the foreign speaker's native language. It is caused by established structures of language representation that have been shaped by the requirements of the native language (L1) that are confronted with speech data from the second language (L2). As a source of variability in speech, foreign accent can be detrimental to speech perception and may result in partial or complete misidentification when listeners are unable to recognize phonetic segments, words, or larger units.

Our research focusses on English as the target language and Dutch and Chinese as the source languages. We compare the intelligibility of Chinese-accented English, Dutch-accented English and native American English in an attempt to clarify how well these people understand each other and themselves when they are speaking English with their respective accents. As a part of a larger project, the present paper targets the intelligibility of English vowels. It may provide answers to questions such as: how well are English vowels identified by native American, Chinese and Dutch listeners? What is their confusion structure? Can we relate the confusions to specific interference patterns that reflect structural properties of the mother tongue of the non-native speaker and/or listener?

We hypothesize that foreign-accented English must be more difficult for English listeners as the source language deviates more from English, but native listeners still have strategies which non-native listeners lack for coping with all sorts of non-optimal speech, including foreign accents. Generally, then, native

Linguistics in the Netherlands 2004 21 (2004), **205–216.**
ISSN 0929–7332 / E-ISSN 1569–9919

English listeners will be at an advantage over foreigners when listening to non-native English. There may just be one exception to this rule: non-native listeners may understand their own accented English better than native English listeners do. Since the foreign listeners are acquainted with the interfering native language, they may be sensitive to cues in the source language that native English listeners fail to pick up (see Bent & Bradlow 2003, Wang & van Heuven 2003).

2. Method

For the present pilot experiment we recorded one male and one female speaker for each of three nationalities: Chinese, Dutch and American. All six speakers studied in the Netherlands at the time the recordings were made. Dutch and Chinese speakers had not studied English after secondary school. Speakers did not have, or had in the past, regular contact with English-speaking friends or relatives, nor had they ever lived in an English-speaking country. For a full description of the methods used in the pilot experiment see Wang & van Heuven (2003). For the present article we will just recapitulate the materials and procedures employed to assess the production and perception of the vowels.

A list of words containing 19 full vowels and diphthongs (so excluding schwa) in identical /hVd/ contexts was recorded: *heed*, *hid*, *hayed*, *head*, *had*, *who'd*, *hood*, *hoed*, *hawed*, *hod*, *hard*, *hud*, *heard*, *hide*, *hoyed*, *how'd*, *here'd*, *hoored*, *haired*. The /h_d/ consonant frame is fully productive in English, allowing al the vowels of English to appear in a word or short phrase.

Speakers were recorded on digital audio tape (DAT) in a sound-insulated recording booth through a Sennheiser MKH-416 microphone. Materials were downsampled (16 KHz, 16 bits), stored on computer disk, and then used to construct a listening experiment. The listening test contained the 19 /hVd/ items for all six speakers in random order across speakers, preceded by six practice items, yielding a total of 120 items.

Listeners were drafted from the same pool from which the speakers had been drawn. Twelve were Chinese, twelve Dutch, and twelve American. Across nationalities, subjects were roughly evenly divided over the sexes. Stimuli were presented over headphones. The listeners read standardized written instructions, and listened to the practice items in order to get familiar with their task, the layout of the answer sheets, and with the time constraints of the stimulus presentation.

3. Overall results

The overall results for vowel intelligibility are presented in Figure 1, broken down by nationality of the listeners and broken down further by nationality of the speakers.

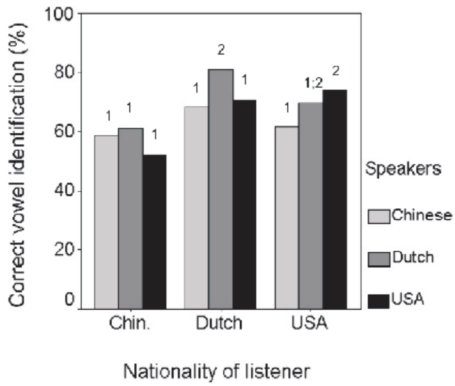

Figure 1. Percent correctly identified vowels broken down by listener group and by nationality of speaker.

Across speaker groups, the Chinese listeners have the lowest vowel identification scores (50–60% correct). Dutch listeners perform best (65–80% correct), and the American listeners are intermediate (60–70% correct). Across listener groups, Dutch speakers were most intelligible, closely followed by American speakers, while Chinese speakers were poorest.[1] Crucially, the results also show that American listeners identified vowels produced by American speakers better than Chinese and Dutch listeners. Similarly, Dutch listeners identified Dutch-accented vowels better than Chinese and American listeners. Chinese listeners, however, identified Dutch-accented English vowels better than Chinese-accented or native American tokens. This small advantage of Dutch-accented English for Chinese listeners may have been caused by the circumstance that our Chinese listeners had lived in the Netherlands for some six months, and therefore had had more exposure to Dutch-accented English than to L1 American English.[2]

For the purpose of the present paper we are not so much concerned with differences in overall intelligibility. Rather we will examine the extent to which errors in the production and perception of English vowels by the non-native speakers and listeners can be traced to properties of the L2 languages of these individuals.

4. Confusion structure

The experimental literature on foreign-language interference typically addresses one specific contrast at a time. For instance, there is a vast literature on the acquisition of the English /r~l/ contrast by speakers of Asian backgrounds (where the contrast is no part of the phonology). In the area of vowels much effort has been made to study the details of the acquision of such 'new' contrasts as English /e~æ/ by Germans, or to give just one final example of the English /iː~ɪ/ contrast by hispanic learners (Flege, 1995). However, experimental studies targeting the confusion structure in an entire vowel inventory in a cross-linguistic setting are far and few between.

Before we present and analyse the confusion structure in Chinese, Dutch and American tokens of English vowels, let us consider the vowel inventories in the three languages in an attempt to derive specific predictions as to where confusions may arize in Chinese and Dutch-accented varieties of English.

English vowel inventory. General American English (GA, as exemplified in the American English pronouncing dictionary by Kenyon & Knott 1944), has four degrees of vowel height (high, high-mid, low-mid and low) and three degrees of backness (front, central, back). Also, GA has a series of tense vowels with peripheral locations and a more centralized series of lax vowels. The phonetic relationship between the tense and lax vowels is difficult to capture within the 4×3 height-backness grid. In Figure 2 we have therefore superimposed (more or less) concentric rings on the grid, as proposed by Delattre (1971). In the outer ring we find the tense vowels; the more centralised lax vowels are located in the middle ring. The inner circle, finally, is taken up by /r/-coloured vowels.[3] GA has two diphthongs, /ai, ɑu/, which start at an open position and glide towards a close position along the front and back side of the vowel space, respectively. The third diphthong is /ɔi/, which runs from back to front in the mid part of the space. The GA English set then contains six lax vowels, six tense vowels, five r-colored vowels, and three diphthongs (Table 1).[4]

The Dutch vowel system (Table 2) is in many respects similar to English. It also has tense and lax vowels, and distinguishes four degrees of height and three

Table 1. Structural representation of the General American vowel inventory.

	front		central		back	
	tense	lax	tense	lax	tense	lax
high	iː, ɪə^r				uː, ʊə^r	
hi-mid	eː, ɛə^r	ɪ			oː, (ɔə^r)	ʊ
lo-mid		ɛ	əː^r	ʌ	ɔː, ɔi	ɔ
low	ai	æ			ɑː, ɑu	

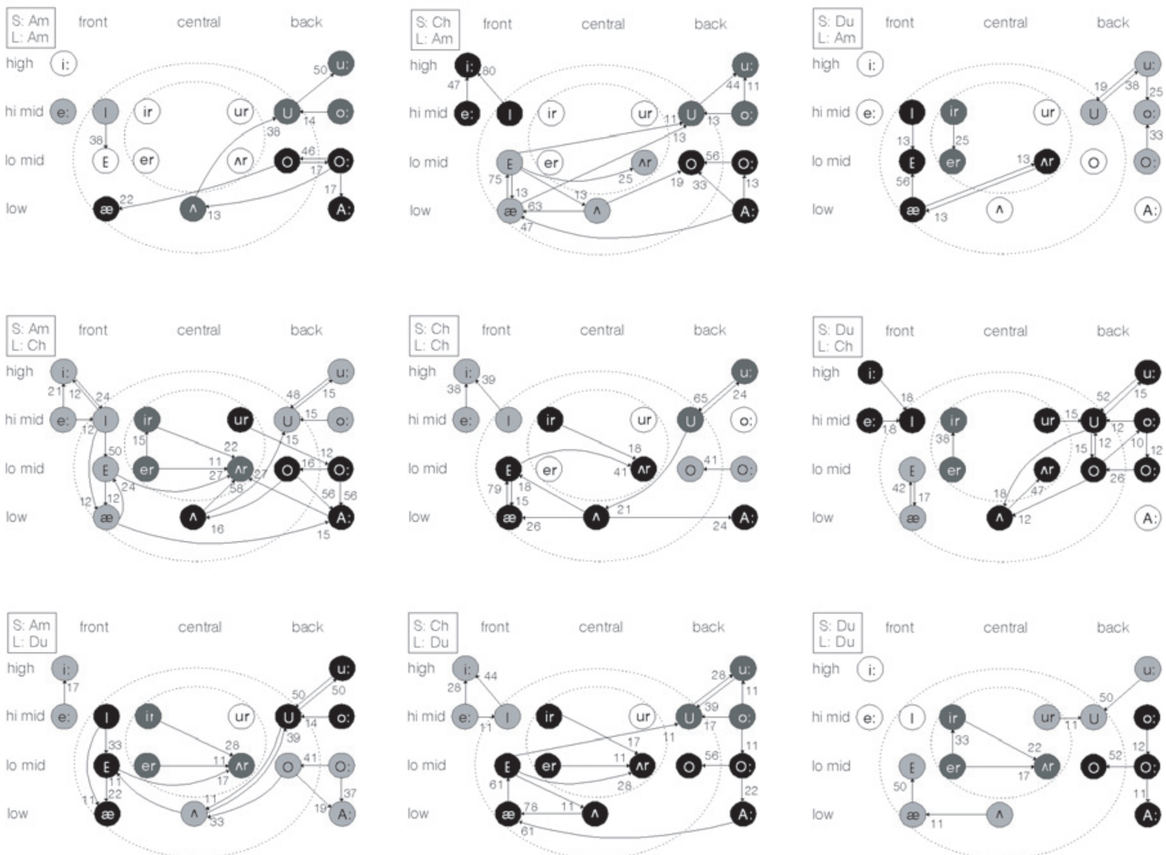

Figure 2. Vowel confusion graphs for nine combinations of speaker (S) and listener (L) language backgrounds.

Table 2. Structural representation of the Dutch vowel inventory.

	front		central		back	
	tense	lax	tense	lax	tense	lax
high	iː		yː		uː	
hi-mid	eː	ɪ	øː	ɵ	oː	
lo-mid	ɛi	ɛ	œy			ɔ
low			aː		ɑu	ɑ

degrees of backness. However, the central part of the vowel space is more densely filled as Dutch has (rounded) central high and hi-mid vowels. In the lax front vowels Dutch distinguishes two degrees of height for the /ɪ~ɛ/ contrast, where English has three: /ɪ~e~æ/. Dutch is underdifferentiated relative to English in the high back vowels, where English has the tense~lax opposition /uː~ʊ/ while Dutch only has /uː/. Confusions are expected in each of the vowels mentioned here. Dutch also has a number of vowels that are absent in the English system; such overdifferentiation is hardly ever a source of confusion.

Mandarin is basically a seven-vowel system. It has three high vowels /i, y, u/, three mid vowels /e, ɚ, o/ and one low vowel /ɑ/.[5] It has front, (rounded) central and back vowels, but no length (i.e. no tense~lax) contrast. We predict problems when Chinese learners try to pronounce or recognize English vowels. Chinese has no /iː~ɪ/, no /uː~ʊ/, no /ɔ~ɔː~oː/, and no /e~æ/ contrasts. We expect, then, that Chinese L2 learners of English vowels will experience all the problems of Dutch learners of English, plus a number of additional difficulties caused by the absence of the tense~lax parameter in Chinese.

Table 3. Structural representation of the Mandarin vowel inventory.

	front	central	back
high	i	y	u
mid	e	ɚ	o
low		ɑ	

Representation of vowel confusion patterns. Confusions in an identification task are customarily presented in a confusion matrix. Here the rows list the intended (stimulus) categories, while the columns represent the perceived categories. Correctly perceived stimuli appear in the cells along the main diagonal from top-left to bottom-right; errors are in the off-diagonal cells. As an illustration Table 4 presents the confusion matrix for the 19 English vowels as produced by Chinese speakers and identified by Dutch listeners.

Table 4 shows, first of all, that the three English diphthongs were never a

Table 4. Sample confusion matrix for 19 Am. English vowels produced by Chinese speakers and identified by Dutch listeners. Correct responses (in percent) are in bold.

		Responded vowels																		
		iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɜː	ɔ	oː	ʌ	əː	ai	ɔi	ɑu	ɪə	ʊə	ɛə
Stimulus vowels	iː	**94**	6																	
	ɪ	44	**50**									6								
	eː	28	11	**50**										6						6
	ɛ				**44**		6	6	11				6	28						
	ɑː					**28**	61				6		6							
	æ				61		**39**													
	uː							**61**	39											
	ʊ							28	**72**											
	ɜː					22				**17**	56								6	
	ɔ					6					**94**									
	oː							11	17	11		**50**					11			
	ʌ						78						**22**							
	əː													**100**						
	ai														**100**					
	ɔi														3	**97**				
	ɑu																**100**			
	ɪə													17				**78**		6
	ʊə									6									**94**	
	ɛə													11				6		**83**

problem: /ai, ɔi, ɑu/ are the least confused vowel types. In order to reduce the complexity of the analytic problem, we will not be concerned with the diphthongs in the remainder of this article.

The next step in the analysis would normally be the generation of a hierarchical cluster scheme (HCS), a tree structure that visualizes which vowels constitute highly confusable subsets. Alternatively, data reduction can be attempted by Multidimensional Scaling (MDS). We feel, however, that neither HCS not MDS do justice to what actually goes on in the data. Both techniques presuppose a symmetrical confusion matrix, that is, the likelihood of vowel *x* to be confused with vowel *y* must be equal to that of *y* being confused with *x*. As

Table 4 shows, this is hardly ever the case. For example, /æ/ is perceived as /ɛ/ in no less than 61% but the reverse confusion, /ɛ>æ/, occurs in less than 10%. Such perceptual asymmetries cannot be expressed in HCS or MDS; the /æ>ɛ/ and the /ɛ>æ/ confusions would average to a symmetrical 34%. Therefore, we have taken recourse to an more informal representation by means of confusion graphs.

In Figure 2, we present the complete confusion structure in the English vowels as produced by American, Chinese and Dutch speakers (S, arranged column-wise) and as perceived by listeners of the same language backgrounds (L, arranged in rows). Vowels are arranged according to the 4 (height) × 3 (backness) vowel quality grid, with a finer distinction between tense, lax and r-colored vowels by means of superposed 'concentric' rings as explained above. Confusion between any two vowels is indicated by an arrow from the intended to the non-intended vowel. The confusion percentage is indicated at the tip of the arrow. Arrows were drawn only for 'problematic' vowel pairs, defined as pairs that were confused in at least 10 percent of the responses. Subsets of vowels that have strong mutual confusions are identified in the graphs by gray shades.

Let us begin by looking at the confusion structure when American listeners identify vowels produced by fellow American speakers (S: Am, L: Am). The literature shows that even in such situations vowel identification is far from perfect, with scores ranging between 54 and 88 percent correct (Peterson & Barney 1952, Strange, Verbrugge, Shankweiler & Edman 1976, and references therein). Our results are no exception. Confusions typically occur among lax vowels: /ɪ>ɛ/ (lax mid cluster), /ʌ>ʊ/ and /ɔ>æ/. Also, there are two prominent tense~lax confusions: /ʊ>uː/ (high-back cluster), and /ɔː>ɔ/ (low-back cluster). Eight out of the total of 362 possible confusion pairs (19×18, including diphthongs) were problematic, i.e., confused in more than 10 percent of the cases. Accordingly, there are eight arrows in the confusion graph.

When the American vowels are identified by Chinese listeners (S: Am, L: Ch), there is much more confusion. All front vowels are confused in height in one or even two degrees, as well as along the tenseness dimension. There is confusion among all three high-back vowels, and in the low-back vowels. There is also considerable confusion among the central vowels. There are 23 problematic vowel pairs in all.

Dutch listeners to American-spoken vowels (S: Am, L: Du) have problems with the height distinction among the lax front vowels /ɪ~ɛ~æ/. They also predictably confuse /ʊ~uː/ (but symmetrically, unlike American listeners) and there is more confusion among the low-back vowels. There is also confusion among the central vowels, though not as much as with the Chinese listeners. There are 17 problem pairs.

When Americans listen to Chinese speakers (S: Ch, L: Am), there is confusion of height and tenseness in the high-front as well as in the high-back vowels. There is also heavy confusion in the lower front vowels /ɪ~ɛ~æ/ and in the low back vowels /ɔː~ɔ~ɑː/; there is even considerable confusion for /ɑː> æ/, so that the entire lower region proves massively confused. Fifteen vowel pairs are problematic.

When Chinese listeners respond to Chinese-accented vowels (S: Ch, L: Ch), the confusion pattern is very much the same as when listeners are American. It would seem, then, that the American listeners did not miss any subtle cues that Chinese speakers might have coded into their vowels. Sixteen vowel pairs are problematic.

Dutch listeners are less proficient in distinguishing among the Chinese-accented vowels (S: Ch, L: Du). By and large the same clusters of difficulty are found again: the high-front vowels /iː~ɪ~eː/, the high-back vowels /ʊ~uː/, the lower front vowels /ɛ~æ/, and the lower back vowels /ɔː~ɔ~ɑː/. The graph contains 19 problem pairs.

When Americans listen to Dutch speakers (S: Du, L: Am), confusion occurs in the lax front vowels /ɪ~ɛ~æ/ and the high-back pair /ʊ~uː/. These are the two English vowel contrasts that do not occur in Dutch. Unexpectedly, the three non-low tense back vowels are confused: /uː> oː/, /ɔː> oː/. There are eleven problematic vowel pairs.

Chinese listeners to Dutch-accented English (S: Du, L: Ch) have the same problems as the American listeners but add several more. Typically, there is widespread confusion among the central and back vowels (except /ɑː/).

When Dutch listeners are exposed to their own type of accent, a lot of the confusion disappears. Just three confusable clusters remain, predictably the /ɛ~æ/ and /ʊ~uː/ pairs, and the non-high back vowels /oː~ɔː~ɔ~ɑː/. The graph contains just ten problematic vowel pairs.

Table 5 summarizes the observations we made in the above on the incidence of problematic vowel pairs. The percentages are broken down for the nine combinations of speaker and listener language background.

Table 5 shows that, overall, native American listeners have fewer problems with the English vowels than L2 listeners. Dutch listeners are a good second, and Chinese listeners clearly have problems. More generally, the language background of the listener exerts a stronger influence on the number of

Table 5. Number of problematic vowel pairs (see text) broken down by nationality of speaker (down) and of listener (across)

Language background of speaker (down)	**Language background of listener (across)**			
	American	**Chinese**	**Dutch**	**Total**
American	**8**	23	17	48
Chinese	15	**16**	19	50
Dutch	11	21	**10**	42
Total number	34	60	46	

confused vowel pairs than the L1 of the speaker. The table also shows that listeners tend to identify those vowel tokens best that were produced by speakers of their own language background. In each column listeners who share the L1 of the speakers, have by far the smallest number of problematic vowel pairs. This, then, shows that knowledge of the phonemic codes of both target and source language is important in non-native communication.

5. Conclusion and discussion

Our first hypothesis was that English will be more difficult to understand as the foreign speaker's native language is more unlike English. We predict, then, that Chinese-accented English vowels will be more difficult to identify for native English listeners than, for instance, Dutch-accented vowels. Conversely, English vowels produced by native English speakers should then be more intelligible to Dutch listeners than to Chinese listeners. Both predictions were clearly borne out by the experimental results. Although these results can indeed be seen as experimental support for our typological distance hypothesis, it should be pointed out that cultural and educational differences between the People's Republic of China (with little exposure to English) and the Netherlands (with an abundance of English) will also have contributed to the difference in intelligibility.

The confusion structure in the foreign-accented Englishes can partly be accounted for by a contrastive analysis of the vowel inventories of the target and source languages involved. For Dutch-accented English, we predicted problems with the non-high lax front vowels /ɪ~e~æ/ and with the /uː~ʊ/ contrast. The results show that these were, indeed, the most frequent confusion types, not only when L1 English listeners identified Dutch-accented vowels, but also when Dutch L2 listeners identified native English vowel tokens. Moreover, our contrastive analysis predicted that Chinese-accented English would have all the problems of Dutch English but would additionally suffer from massive tense~lax vowel confusion, both in production and in perception. The experimental results show that this prediction is correct.

On the other hand, we found a number of problematic vowel contrasts that are not easily predicted from a contrastive analysis, e.g. the /uː> oː/ and /ɔː> oː/ confusions for Dutch speakers identified by American listeners. We did not encounter any cases where predicted problems did not arise. Our results, then, provide partial support for the transfer hypothesis in foreign language learning, which claims that L2 learners will not distinguish between contrasts in the target language that do not occur in their native tongue. At the same time, a weaker version of the transfer hypothesis seems called for in that, although it makes no false predictions, it predicts only a subset of the L2 vowel learning problems.

Notes

1. For an more elaborate presentation of these (and other) results and accompanying statistical analyses, see Wang & van Heuven (2003).

2. A larger-scale study is in preparation involving 20 Chinese, 20 Dutch and 20 American speakers of English. Optimal selections of recorded materials were submitted to identification tests with groups of 20 listeners. Within each speaker group the most typical (i.e. median) male and female speakers have now been selected on basis of vowel and consonant identification scores within their own language community, who will be representative for the populations of Chinese and Dutch L2 speakers of English and of L1 American speakers of English.

3. In non-rhotic varieties of English (such as Southern British English, but also American varieties spoken in the New England area) postvocalic /r/ is not pronounced but merges with the vowel into a so-called murmur diphthong or centring diphthong. In GA, which is a rhotic variety, the vowels are also realised as centring diphthongs (ending in a schwa-like vowel) but the coda /r/ is pronounced.

4. The r-colored vowels come in two types. The first type can always be analyzed as an allophonic variant of a tense non-low vowel with a phonetically lowered and centralized onset and an schwa-like offglide due to the presence of coda /r/. The normal realization, of course, applies in all other contexts. So, besides the normal variant of /iː, eː, uː, oː/ as in *steaming* /stiːmɪŋ/, *dating* /deːtɪŋ/, *tooling* /tuːlɪŋ/, *polling* /poːlɪŋ/, there are four allophones in complementary distribution with the former when followed by /r/ within the morpheme: *steering* /stiːr+ɪŋ/ > [stɪərɪŋ], *daring* /deːr+ɪŋ/ > [dɛərɪŋ], *touring* /tuːr+ɪŋ/ > [tʊərɪŋ] and pooring /poːr+ɪŋ/ > [pɔərɪŋ, pɔːrɪŋ]. Since [ɔə] generally monophthongizes to long [ɔː], only the latter vowel was elicited in our materials, yielding a vowel set of 19 rather than 20. The second set comprise /ɚː, ɑː/. The former, /ɚː/, exclusively occurs before tautomorphemic /r/; there is no other tense vowel in the inventory from which it can be derived as a centralized version due to r-coloring. Neither can it be analysed as an r-colored allophone of lax /ʌ/ given the existence of words such as *hurry* /hʌrɪ/, *worry* /wʌrɪ/, where /ʌr/ does not surface as [ɚː(r)]. The latter, /ɑː/, must be set up as a separate phoneme as it also surfaces in contexts where it is not followed by /r/, in no English present or past variety, e.g. *father* /fɑːðər/. However, occurrences of /ɑː/ not followed by (underlying) /r/ are extremely rare in GA.

5. The /ɚ/ is strongly r-colored and cannot be followed by a coda consonant. All other vowels can be followed by a coda nasal /m, n, ŋ/.

References

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

Delattre, P. (1971). *Comparing the phonetic features of English, German, French and Spanish.* Julius Gross Verlag, Heidelberg.

Flege, J. E. (1995): 'Second language speech learning: Theory, findings and problems'. In W. Strange (ed.) *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press, 233–277.

Kenyon J., and T. Knott (1944). *A pronouncing dictionary of American English.* Merriam, Springfield.

Peterson. G.E. and H.L. Barney (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.

Strange, W., R.R. Verbrugge, D.P. Shankweiler and T.R. Edman (1976). Consonant environment specifies vowel identification. *Journal of the Acoustical Society of America*, 60, 213–224.

Wang, H. and V.J. van Heuven (2003). Mutual intelligibility of Chinese, Dutch ad American speakers of English. In L. Cornips, P. Fikkert (eds.) *Linguistics in the Netherlands 2003*, John Benjamins, Amsterdam, 213–224.