JULY 29 2010

# Perceptual assimilation of American English vowels by inexperienced Russian listeners

Yana D. Gilichinskaya; Winifred Strange



*J. Acoust. Soc. Am.* 128, EL80–EL85 (2010)
https://doi.org/10.1121/1.3462988

# Perceptual assimilation of American English vowels by inexperienced Russian listeners

**Yana D. Gilichinskaya**[a] **and Winifred Strange**
*Speech Acoustics and Perception Laboratory, The City University of New York, Graduate School and University Center, 365 Fifth Avenue, New York, New York 10016*
*ygilichinskaya@gc.cuny.edu, strangepin@aol.com*

**Abstract:** In a perceptual assimilation task, 16 Russian speakers categorized American English (AE) vowels [iː, ɪ, ɛ, aeː, ɑː, ʌ, ʊ, uː] in /Vpə/ disyllables into Russian (RU) response categories and rated their perceived goodness on a 9-point Likert scale. Cross-language discriminant analysis established acoustic similarities to Russian vowels. For all but AE [ɛ], acoustic similarity predicted modal assimilation responses. Russian listeners consistently assimilated AE point vowels to their Russian counterparts, whereas assimilation of the remaining vowels was less consistent. These acoustic and perceptual similarity patterns provide a baseline for future studies of AE vowel discrimination by Russian learners of English.


**PACS numbers:** 43.71.Hw, 43.71.−k, 43.71.Es [JH]
**Date Received:** March 4, 2010 **Date Accepted:** June 1, 2010

## 1. Introduction

The goal of the current study was to evaluate cross-language perceptual similarity of American English (AE) vowels to Russian (RU) vowel categories by monolingual Russian listeners, before English immersion factors exerted their influence. Past studies of cross-language identification and discrimination have demonstrated the effect of L1 phonology on L2 perception: some phonetic differences that underlie L2 vowel contrasts while being semantically irrelevant in the native language remain difficult to perceptually differentiate even by advanced L2 learners (e.g., Levy and Strange, 2008; Levy, 2009). Filtered through the "funnel" of L1 phonology, perception of L2 phonetic segments by L2 learners differs from that of native speakers, resulting in accented speech production (Rochet, 1995).

Best's Perceptual Assimilation Model (PAM), originally developed to predict perceptual assimilation by naïve listeners (Best, 1995), was later adapted to predict perceptual patterns by experienced L2 learners (Best and Tyler, 2007; Guion *et al.*, 2000; Levy and Strange, 2008). According to PAM, perception of a non-native L2 segment is shaped by the phonological structure of the L1. Similarities and discrepancies in the gestural realizations of the non-native and native phonetic segments set conditions for the perception of non-native phones and their perceptual assimilation to native categories. For speech sounds assimilated into the native phonological space, PAM identifies two possible categorization scenarios (Best, 1995): (1) the non-native segment is clearly assimilated to a native category (Categorized) and rated as a good/acceptable/deviant exemplar of that category; (2) the non-native segment is assimilated into the native phonological space but does not fall within any native category (Uncategorized). In the first case, PAM makes predictions regarding contrastive L2 phonetic segments: a) if two categorized L2 phonetic segments are assimilated to two different native (L1) categories, then the contrast will be easy for L2 listeners to differentiate (Two-Category [TC] pattern); b) if two L2 segments are assimilated as equally good tokens of the same L1 category, the differentiation of the L2 contrast will be most challenging (Single-category [SC] pattern); and c) if both L2 segments are assimilated to the same L1 category but with a different perceived goodness (Cat-

[a]Author to whom correspondence should be addressed.

egory Goodness [CG] pattern), the contrast will be easier or harder to discriminate depending on the difference in perceived goodness. Finally, when one member of an L2 contrast is Uncategorized and the other Categorized, discrimination is predicted to be relatively easy.

A number of studies explored the acoustic and perceptual similarities among vowels in several languages. Until recently, most of these studies have focused on examining patterns of perceptual assimilation by speakers of languages with smaller native vowel inventories than the target language, i.e., L1 inventory < L2 inventory (Guion *et al.*, 2000; Strange *et al.*, 2004, 2005). Strange *et al.* (2004, 2005) investigated acoustic and perceptual similarities of American English (AE) and Northern German (NG) vowels. Spectral similarity of AE and NG vowels was established using cross-language discriminant analysis of syllabic midpoint formant frequencies (F1, F2, F3 in Barks). An AE vowel corpus served as a training set to establish centroids for each of the eleven categories, and then the NG vowel corpus was evaluated and cross-language classification matrices computed. Second, cross-language perceptual similarity was established using a Perceptual Assimilation task in which AE listeners were presented multiple tokens of the NG vowels, and were instructed to indicate to which AE category each token was most similar and then to rate its "category goodness" on a Likert scale from most native-sounding (7) to most foreign-sounding (1) [see Strange (2007) for a critique of methods of establishing perceived similarity]. Comparisons of discriminant analysis results with perceptual assimilation results showed that context-specific acoustic similarity between AE and NG vowels did not always predict perceptual similarity, especially for front rounded vowels that are allophonic variants of back rounded vowels in AE. In addition, perception of some NG vowels with similar counterparts in AE was not predictable from acoustic similarity patterns. The results also suggested that when cross-language duration and spectral similarity were in conflict, AE listeners tended to assimilate vowels based on spectral similarity.

The present study assessed perceptual similarity of AE and RU vowel categories as heard by inexperienced Russian listeners and related perceptual assimilation patterns to those predicted by acoustical similarities between the two languages. Russian has a small vowel inventory with no length contrasts between vowels (Jones and Ward, 1969). The Russian vowel system consists of 5 phonemes /i, ɛ, a, o, u/ with several allophonic variations. The so-called "hard" vowels have palatal counterparts $/^{j}$ɛ, $^{j}$a, $^{j}$o, $^{j}$u/. Furthermore, the phone [ɨ] is considered an allophonic variation of [i] by some phoneticians; however it does occur contrastively with [i] in some minimal pairs: *bila-bɨla*, *slil-slɨl*, *mila-mɨla*. In addition, there is a systematic correspondence between phonetics and orthography for stressed vowels in Russian.

Specific questions addressed in the present study were as follows: *Q1:* How accurately does acoustical similarity between RU and AE vowels predict their perceptual assimilation to RU vowels by RU listeners? *Q2:* What are the predicted difficult contrasts within the PAM framework?

## 2. Methods

Stimuli were produced by a monolingual male native speaker (27 years old) of AE who was born and raised in New York City. Nonsense disyllables /Vpə/ were read from a numbered list. The last item in the list was discarded to avoid any list-final prosodic effects. The speaker produced four randomized repetitions of the 11-vowel list, from which the first three tokens of eight vowels [iː, ɪ, ɛ, aeː, ɑː, ʌ, ʊ, uː] were used for this experiment: 8 vowels × 3 repetitions = 24 tokens. The recording was carried out in the sound-treated acoustic chamber. Digitized stimuli (22.05 kHz, 16 bit resolution) were analyzed using in-house customized Matlab-based software to obtain measures of the first three formants at syllable midpoint and the duration of the vocalic portion of the first syllable from voicing onset to stop consonant closure [see Fig. 1(a)].

Nineteen native speakers of Russian (19 to 68 years old, median = 36 years), were tested at the Soil Cryology Laboratory, Institute of Physicochemical and Biological Problems in Soil Science, Pushchino, Russia. All participants signed an informed consent form approved by the CUNY Graduate Center IRB and all but one had normal hearing by self-report. This participant was subsequently excluded from the data analysis. Most participants were researchers, employees and/or graduate students of the laboratory; others were current or former scientists or science students. Two participants reported having been exposed to other languages when growing up (Tatar and Ukranian,

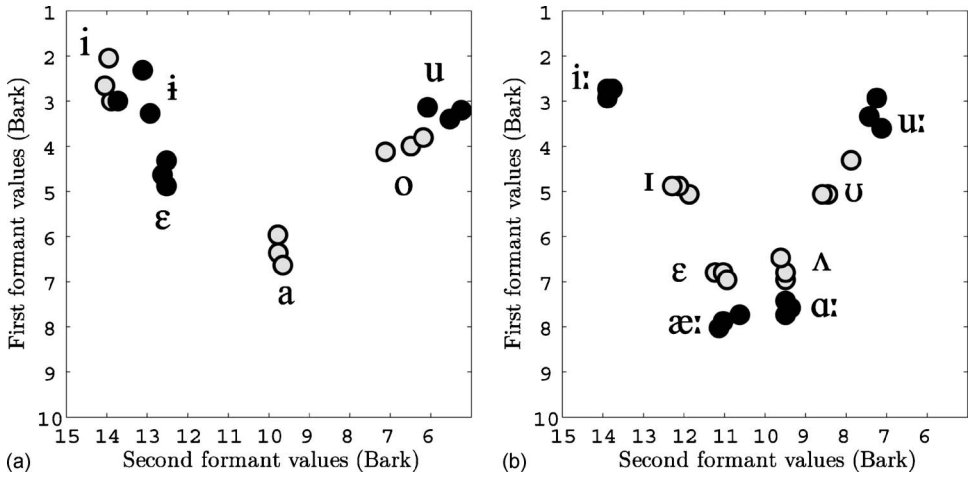


Fig. 1. Midpoint F1/F2 (bark) frequencies for Russian vowels in [Vbə] produced by Russian speakers (means from 3 speakers) (a) and American English [Vpə] stimuli from one speaker (b).

respectively). The majority of the participants had studied English in high school and some received formal instruction in English at the college level. Two participants reported having lived and attended high school in Canada for 6 months before the age of 18 years and were excluded from the data analysis; of the 16 remaining participants, 3 spent some time in the U.S. as visiting scholars (3 years, 2 years, 8 months, respectively) after 25 years of age. Despite having more experience with English, these subjects were retained in the analysis for the reasons covered in the Results section below.

Testing was carried out in a quiet room. Stimuli were presented using Paradigm 4.0 (Tagliaferri, 2005) and delivered through Telephonics TDH-50P headphones at a comfortable listening level from a HP Business Notebook Laptop Computer. The test consisted of one familiarization block of 5 trials using stimuli with AE point vowels, spoken by a female native speaker of AE, and four experimental blocks, 24 trials each (8 vowels × 3 repetitions) spoken by the male speaker. The participants were instructed to listen to nonsense English words such as "eepa," "oopa," "appa," attending to the first vowel, and choose which Russian vowel or combination of vowels it sounded most like. The list of responses was composed of 9 possible alternatives: [i, ɨ, ɛa, a, $^{j}$ɛ, ɛ, o, u, uo] written in standard Russian orthography. The list was compiled on the basis of a prior pilot experiment with 5 native Russian listeners who transcribed the AE vowels in Russian orthography. After a Russian vowel category was selected, the same stimulus was replayed, and participants were asked to judge its perceived goodness on a scale from 1 (not similar to Russian) to 9 (similar to Russian). Participants were repeatedly encouraged in the written instructions and verbally during the familiarization block to use the entire scale for their ratings. Four experimental blocks of 24 trials (8 vowels × 3 tokens) presented in random order were completed. The first block was used to familiarize the listeners with the speaker's vowels and to calibrate their use of the rating scale. Data from the last 3 blocks were used for data analysis; each AE vowel was categorized/rated 9 times (3 tokens × 3 repetitions) by each listener.

## 3. Results

Frequencies of each selected RU category were computed for each AE vowel and summed across all listeners (3 repetition/vowel × 3 blocks × 16 listeners = 144). Table 1 presents these data, with each AE vowel (column 1) and each RU vowel response receiving at least 7 out of 9 responses by one or more listeners (column 2) listed. The third column gives the frequencies of each RU response receiving over 10% of total responses as percentages of total opportunities

Table 1. Perceptual assimilation and acoustic similarity results.

| AE-stimulus[a] | RU-response[b] | % chosen[c] | N subjects[d] | Rating[e] | Acoustic classification[f] |
|---|---|---|---|---|---|
| iː | i | 100 | 16 | 8.0 | i |
| ɪ | ɛ | 84 | 14 | 8.0 | |
| | i | | 1 | 9.0 | ɛ |
| | ɛa | | 1 | 4.0 | |
| ɛ | ɛ | 66 | 9 | 7.0 | |
| | a | 14 | | 4.5 | a |
| | ɛa | 19 | 1 | 5.0 | |
| aeː | a | 62 | 6 | 8.0 | a |
| | ɛa | 33 | 3 | 6.0 | |
| ɑː | a | 94 | 15 | 8.0 | a |
| ʌ | a | 69 | 9 | 8.0 | a |
| | o | 20 | 1 | 6.0 | |
| ʊ | o | 81 | 12 | 7.0 | a |
| | uo | 17 | 1 | 6.0 | |
| uː | u | 93 | 15 | 7.0 | u |

[a]AE vowel presented.
[b]RU vowel selected.
[c]Modal responses across subjects: only those >10% are shown.
[d]Number of subjects who chose that response at least 7 times out of 9 presentations.
[e]Median goodness rating.
[f]Discriminant analysis classification based on acoustic similarity.

(frequency/144 × 100), whereas the fourth column gives the number of individual listeners who consistently (at least 7 out of 9 responses) categorized each AE vowel as a particular RU vowel.

AE point vowels [iː, ɑː, uː] were assimilated most consistently overall to corresponding RU categories, whereas AE [ɛ, aeː, ʌ] yielded the least consistent modal responses. Finally, AE vowels [ɪ, ʊ] were assimilated to RU [ɛ, o], respectively, at a intermediate consistency level. These patterns of group consistency are reflected in individuals' responses as well. Perceptual assimilation patterns of the three participants with more English experience was similar to the rest of the group. Almost all listeners consistently categorized the point vowels as most similar to their RU transcriptional counterparts, whereas many fewer individuals were consistent in categorizing AE [ɛ, aeː, ʌ]. For AE [ɪ, ʊ], most listeners were consistent in their assimilation responses, but there were *inter*-listener differences across listeners as to which RU vowel these AE vowels were most similar.[1] The perceived goodness ratings associated with the modal responses were also scored, summed across all listeners (column 5). Median ratings for the group modal responses were quite high for all 8 vowels, whereas median ratings for most minority responses were lower.

To compare the perceptual assimilation patterns to those predicted by the acoustic similarities between AE and RU vowels, linear discriminant analysis was undertaken (Klecka, 1980). First, a classification matrix of Russian vowels based on F1/F2/F3 Bark values at the 50% point of the vocalic portion was established (training set). These data were obtained from previous research (Gilichinskaya *et al.*, 2007) and included 4 tokens of each of 6 RU (single)

vowels produced in citation /Vbə/ disyllables by three male native Russian speakers who had lived in U.S. less than one year at the time of recording [Fig. 1(a)]. Only the 6 vowel categories were used in the analysis. The 24 AE stimuli used in the experiment [Fig. 1(b)] were classified with regards to the centers of gravity of the RU training set: results are shown in Table 1 (column 6).

AE [iː, uː] were acoustically most similar to their RU counterparts, whereas, AE [aeː, aː, ɛ, ʌ] were all classified as spectrally most similar to RU [a]. AE [ɪ] was classified as most similar to RU [ɛ], whereas AE [ʊ] was acoustically most similar to RU [o]. With one exception, the results of discriminant analysis corresponded with the group modal responses obtained in the perceptual assimilation task: AE [ɛ] was perceptually assimilated by the RU listeners to RU [ɛ] or to an intermediate vowel [ɛa] by the majority of listeners.

## 4. Discussion

The purpose of this study was to establish perceived similarity patterns of AE vowels to native categories by Russian listeners with very little experience with spoken English in order to provide a reference point for comparing perceptual patterns of Russian L2 English learners as they acquire L2 experience (Gilichinskaya *et al.*, 2007). Perceptual assimilation patterns from 16 nearly monolingual Russian listeners were assessed using 8 AE vowels [iː, ɪ, ɛ, aeː, ɑː, ʌ, ʊ, uː]. As summarized above, with the exception of AE [ɛ], acoustic similarity, defined by spectral patterns at syllable midpoint, corresponded well with the perceptual assimilation patterns for these inexperienced Russian listeners. As mentioned above, the pattern of the results exhibited by three more experienced with English listeners was not different from the rest of the group. Consequently, we treated their exposure to English as irrelevant to the task used in this study. These results are further supported by our perceptual assimilation data from Russian immigrants in the United States who, while assumingly having had more experience with English, assimilated AE vowels to Russian vowel categories in the same manner. These data were partially presented at the 154th Meeting of the Acoustical Society of America in New Orleans (Gilichinskaya *et al.*, 2007).

To address the second question regarding which AE contrasts are predicted to be difficult for RU L2 learners of AE, overlap in RU response categorization for vowel pairs was evaluated for individual subjects. AE [ɪ/ɛ] were consistently assimilated to RU [ɛ] by 9 out of 16 participants; similarly, AE [aː/ʌ] were consistently assimilated to RU [a] by 9 participants. In both cases, both members of the contrast received relatively high goodness ratings (7–9). Thus, these contrasts constitute Single Category assimilation patterns in terms of Best's PAM for the majority of participants. Consequently, we may predict that perception of AE vowel contrasts [ɪ/ɛ] and [ɑː/ʌ] will be difficult for RU learners of AE. Contrasts between other vowels show a Categorized-Uncategorized pattern for most listeners (e.g., [aː/aeː]), or a Two-Category pattern (e.g.,[iː/ɪ, uː/ʊ] and, thus, we would expect them to be quite easy for Russian L2 learners to perceive.

Research is underway that examines the extent to which perceptual assimilation patterns of AE vowels by Russian-speaking late L2 learners of English change from the patterns shown here, and how perceptual assimilation patterns predict relative difficulty in a speeded categorial discrimination task (Gilichinskaya *et al.*, 2007).

## Acknowledgments

We wish to thank Dr. Gilichinsky and the Soil Cryology Laboratory at the Institute of Physicochemical and Biological Problems in Soil Science, Pushchino, Russia for hosting this project.

## References and links

[1]No difference was observed in the response consistency of older (>55 years old) versus younger listeners.

Best, C. (**1995**). "A direct realist view of cross-language speech perception: New directions in research and theory," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-*

*Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.
Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning*, edited by O.-S. Bohn and M. Munro (John Benjamins, Amsterdam), pp. 13–34.
Gilichinskaya, Y. D., Law, F., and Strange, W. (**2007**). "Speeded discrimination of American vowels by experienced Russian learners of English," J. Acoust. Soc. Am. **122**, 3028.
Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–24.
Jones, D., and Ward, D. (**1969**). *The Phonetics of Russian* (University Press, Cambridge).
Klecka, W. R. (**1980**). *Discriminant Analysis*, Quantitative Applications in the Social Sciences (SAGE, Newbury Park, CA), Vol. **19**.
Levy, E. S. (**2009**). "Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French," J. Acoust. Soc. Am. **125**, 1138–52.
Levy, E. S., and Strange, W. (**2008**). "Perception of French vowels by American-English adults with and without French language experience," J. Phonetics **36**, 141–157.
Rochet, B. (**1995**). "Perception and production of L2 speech sounds by adults," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.
Strange, W. (**2007**). "Cross-language phonetic similarity of vowels: Theoretical and methodological issues," in *Language Experience in Second Language Speech Learning*, edited by O.-S. Bohn and M. Munro (John Benjamins, Amsterdam), pp. 35–55.
Strange, W., Bohn, O. S., Nishi, K., and Trent, S. A. (**2005**). "Contextual variation in the acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–62.
Strange, W., Bohn, O. S., Trent, S. A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **115**, 1791–807.
Tagliaferri, B. (**2005**). Perception Research Systems, http://www.perceptionresearchsystems.com (Last viewed 2/22/2010).