INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

The impact of manner of articulation on the intelligibility of voicing contrast in noise: Cross-linguistic implications

Mayuki Matsui

National Institute for Japanese Language and Linguistics, Japan
matsui-ma@ninjal.ac.jp

Abstract

The current study addresses the impact of manner of articulation on the intelligibility of voicing contrast in noise from a cross-linguistic perspective. Previous noise-masking studies have suggested that the impact of manner of articulation on the intelligibility of voicing contrast in noise is apparently different in Russian and English. In order to further assess the source of such a cross-linguistic inconsistency, the current study examines how Russian voicing contrast is perceived by English listeners. Native listeners of English performed a forced-choice identification task with Russian voiced and voiceless stimuli in quiet and noisy conditions. The results showed that the voicing contrast in stops were more confused than that in fricatives for English listeners, showing a pattern similar to Russian listeners. The results suggest that the source of the cross-linguistic difference identified in previous studies comes from the difference in the acoustic properties of the stimuli, reflecting the difference in phonetic implementation of voicing contrasts in each language. The results in turn suggest that perceptual cue weighting strategies for perceiving voicing contrast in different manners of articulation is similar among Russian and English listeners.

Index Terms: Cross-linguistic perception, Voicing contrast, Manner of articulation, Noise-masking, Russian, English

1. Introduction

The world is full of noises, where we every day perceive speech sounds. It might be easy for us to imagine that presence of background noise can more or less degrade intelligibility of phonological contrast compared with an ideal, quiet listening context. Perceptual characteristics in such *non-ideal* listening contexts have provided significant insights into the nature of speech perception and phonology, which cannot solely be derived from the examination in an ideal listening context. One of the important perceptual characteristics in non-ideal contexts is the emergence of asymmetries in the degradation pattern of phonological contrast, which is covert in an ideal listening context. As has been witnessed after a classic noise-masking study [1], the degree of degradation (or, confusion) depends on the type of the phonological contrast. The goal of the present study is to deepen our understanding of the degradation pattern emerged in noisy contexts, with special attention to cross-linguistic differences in the impact of consonantal manner on the perception of voicing contrasts in a noisy context.

Previous noise-masking studies suggest that the voicing contrast in fricatives is more confused than that in stops when English listeners hear English stimuli in white noise [2, 3, 4]. Such manner asymmetries are observed not only in white noise [2] but also in speech-shaped noise [3, 4].

On the contrary, a recent study revealed that the pattern in Russian is apparently opposite from that of English; the voicing contrast in stops is more confused than in fricatives when Russians heard Russian stimuli [5]. Taken together, it is suggested that the impact of manner of articulation on the voicing perception is apparently different between English and Russian. However, the source of such a cross-linguistic inconsistency is still not clear. Where does such a difference, if any, come from?

One possible source of the difference might lie in variability in speech perception itself. That is, the perceptual cue weighting strategies which are involved in the perception of voicing contrasts might be different among English and Russian. In this case, a prediction is that English and Russian listeners should perceive the same acoustic contrast in a different way.

Another possible source of the difference might be attributed to variability in the phonetic implementation of the voicing contrast, in other words, variability in speech production. In this case, a prediction is that English and Russian listeners should perceive the same acoustic contrast in a similar way.

In order to examine what contributes to the cross-linguistic difference, the current study offers new data on how Russian voicing contrast is perceived by English listeners, using the same stimuli in the previous study [5]. In the following sections, two identification experiments will be reported: In Experiment 1, English listeners are tested with no background noise, where symmetric perception with high accuracy is expected. In Experiment 2, the same listeners hear the same stimuli with background noise, where some type of asymmetric perception depending on the manner of articulation, triggered by the presence of a background noise, is expected.

2. Experiment 1: Quiet listening context

2.1. Participants

17 native speakers of English participated in the perceptual experiment (mean age, 20 yrs; range, 17–28 yrs; three males). All were native speakers of Canadian English with no reported history of speech or hearing disorders. At the time of the experiment, the participants were undergraduate or graduate students at the University of Toronto.

2.2. Auditory stimuli

The auditory stimuli were physically identical with what the Russian listeners heard in the previous study [5]; Russian

  http://dx.doi.org/10.21437/Interspeech.2016-697

voiced and voiceless obstruents (/t, d, s, z/) were embedded between open vowels /a/. Thus the stimuli consisted of four types of vowel-consonant-vowel (VCV) sequences: /ata/, /ada/, /asa/, and /aza/. These were pronounced by two native speakers of Russian (a male and a female) with two possible stress patterns and recorded onto a portable recorder (SONY PCM-M10) and a stereo microphone (SONY ECM-MS907) with a 44.1-kHz sampling rate at a 16-bit quantization level. Two repetitions of each condition from two speakers served as auditory stimuli.

2.3. Procedures

The participants were individually tested in a quiet room at the University of Toronto. The auditory stimuli were presented to the participants via headphones (SONY MDR-10RNC), and the participants selected what they heard from two choices: voiced or voiceless. The possible responses were orthographically presented using Latin alphabet, such as, '*t*' and '*d*'. The participants were instructed to respond by clicking the button displayed on the computer screen as quickly as possible. For each participant, the number of trials was 64 (4 consonant types × 2 stress patterns × 2 speakers × 2 tokens × 2 appearances). Half of the participants were tested with the 'voiced' button on the left and the 'voiceless' button on the right, and the other half with the 'voiceless' on the left and the 'voiced' on the right. A practice session with nine practice trials preceded the test session. The experiment was implemented by Praat [6].

2.4. Analysis

In order to evaluate listener's sensitivity to the stimuli apart from response bias, *d'* values were calculated for each manner condition for each listener. *d'* is a measure of sensitivity that was proposed in signal detection theory [7] [1]. In addition to *d'*, criterion (*c*), which is a measure of the response bias, was also considered [2].

2.5. Results and discussion

The total number of the obtained response was 1088 (64 × 17 listeners). Overall percentages of correct responses in the dataset were 98.2 for stops and 99.1 for fricatives.

Figure 1 describes individual listeners' *Hit* rates and *False-alarm* rates. Here, the Hit rates represent the rate of voiceless response to voiceless stimuli, thus reflecting correct responses. The False-alarm rates, on the other hand, represent the rate of voiceless response to voiced stimuli, thus reflecting incorrect responses (See [7] for further details on Hit and False-alarm rates in signal detection theory). Figure 1 shows that the listeners' performances were almost at ceiling level.

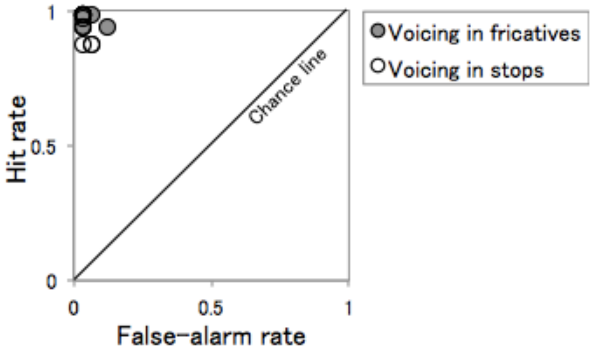

Figure 1: *Hit rate as a function of False-alarm rate in quiet listening context. Chance line corresponds to d' = 0, i.e., zero sensitivity to the stimuli.*

The sensitivity measure *d'*, which were calculated from Hit and False-alarm rates, were significantly greater than zero; The mean *d'* of the 17 listeners was 3.73 (SD = 0.49) for stops [one-sample *t*-test, $t(16) = 31.1017$, $p < 0.001$] and 3.87 (SD = 0.35) for fricatives [one-sample *t*-test, $t(16) = 45.5068$, $p < 0.001$][3].

Although the primary interest of the current study is intelligibility of the contrast, the presence/absence of response bias was also considered. The mean of the response bias measure (*c)* was not significantly greater than zero for stops [one-sample *t*-test, $t(16) = -0.9481$, $p = 0.3572$]. On the other hand, there was a significant response bias toward voiceless responses for fricatives (Mean $c = -0.13$, SD = 0.09) [one-sample *t*-test, $t(16) = -5.7241$, $p < 0.001$].

To summarize, Experiment 1 confirmed that both stop and fricative voicing contrasts in Russian were intelligible enough for English listeners as long as the listeners heard the stimuli in an ideal, quiet listening context. Additionally, significant response bias toward voiceless responses was observed for fricatives, but not for stops.

3. Experiment 2: Noisy listening context

3.1. Participants

The participants in Experiment 2 were identical to those in Experiment 1. The participants performed Experiment 2 after completing Experiment 1.

3.2. Auditory stimuli

A new set of stimuli was created by adding masking noise to the stimuli used in Experiment 1. The masker was Gaussian white noise, an energetic masker (see Cooke et al. [8] for recent discussions on masker types). The Vocal Tool Kit in Praat was used to add noise to the stimuli.

---

1 The formula used for the sensitivity measure is as follows: $d' = z(H) - z(F)$. z(H) is the *z*-transformed probability in response X to X item, i.e., the probability of correct response ("Hit"). z(F) is the z-transformed probability to response X to Y item, i.e., the probability of incorrect response ("False alarm"). $d' = 0$ corresponds to zero sensitivity to the stimuli. Higher d' values indicate higher sensitivity to the stimuli (See [7] for further details).

2 The formula used for the response bias measure is as follows: $c = -(z(H) + z(F))/2$. $c = 0$ corresponds to zero response bias to the stimuli. Higher *c* values imply larger response bias toward a certain response category (See [7] for further details).

3 Some listeners scored 1 (all) or 0 (none) as a Hit rate and/ or a False-alarm rate, implying infinite *d'*. In order to avoid the infinite value, such rates were converted to finite values by applying the formula suggested in Macmillan and Creelman [7].

The magnitude of the signal-to-noise ratios (SNR) followed the values used in [5], with −5 dB, −8 dB, and −10 dB.

3.3. Procedures

The procedures in Experiment 2 were basically the same as those used in Experiment 1.

In Experiment 2, the number of trials was 192 per listener (4 consonant types × 2 stress patterns × 2 speakers × 2 tokens × 2 appearances × 3 SNR conditions). The participants completed the task in the following order: −5 dB (easiest), −8 dB (intermediate), and −10 dB (hardest).

3.4. Results and discussion

The total number of the obtained response was 3264 (192 × 17 listeners). Overall percentages of correct answers in the noisy listening context were 56.3 for stops and 70.0 for fricatives.

The properties in the noisy listening context vis-à-vis the quiet listening context described in Experiment 1 are shown in Figure 2. Comparison of Figure 2 with Figure 1 shows that the listeners' performance was degraded by the presence of the background noise.

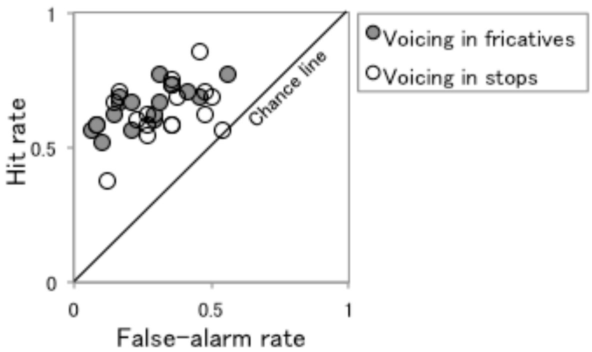

Figure 2: *Hit rate as a function of False-alarm rate in noisy listening context, pooled across SNR levels. Chance line corresponds to d' = 0, i.e., zero sensitivity to the stimuli.*

For closer inspection, Figure 3 illustrates *d'* scores broken down by SNR conditions. As illustrated in Figure 3, sensitivity to the stop stimuli is consistently lower than that to the fricative stimuli. Also, not surprisingly, English listeners' performance degrades as the signal-to-noise ratio increases.

In order to examine the effect of manner of articulation on the intelligibility of voicing contrast in the noisy context, a linear mixed-effects model was constructed by using lme4 package [9] in R [10]. *P* values were estimated by lmerTest package. In the model, 'Manner of Articulation (stop vs. fricative)' was specified as a fixed effect and listener and SNR level as random effects with random intercepts and slopes. The dependent variable was the *d'* score. The result indicated that the fixed effect of Manner was significant [Std. Error = 0.1006, DF = 16.4330, $t = -2.918$, $p < 0.01$]. The *d'* scores for the stop stimuli were significantly lower than those for the fricative stimuli; The mean *d'* of the 17 listeners in the noisy context was 0.90 (SD = 0.74) for stops and 1.19 (SD = 0.62) for fricatives.

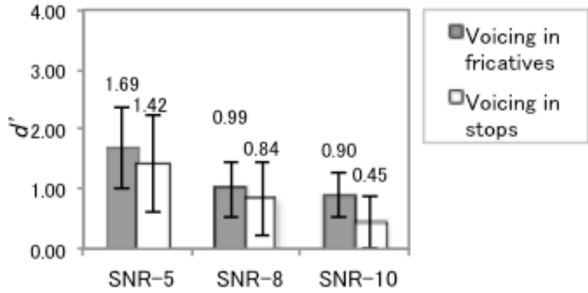

Figure 3: *Mean d' scores of 17 listeners broken down by SNR levels.*

Although the primary interest of the current study is intelligibility of the contrast, the presence/absence of response bias was again considered. For this purpose, additional linear mixed-effects models were constructed with *c* score as a dependent variable. In the model, 'Manner of Articulation (stop vs. fricative)' was specified as a fixed effect and listener and SNR levels as random effects with random intercepts and slopes. To test presence/absence of the significant difference between 0 and a reference level (i.e., intercept), two models with different intercept specifications were created. Namely, one is the model specifying stop as an intercept value. Another is the model specifying fricative as an intercept. The results showed that fricative intercept was significantly greater than zero [Intercept = 0.1864, Std. Error = 0.0771, DF = 16.9980, $t = 2.417$, $p < 0.05$]. This suggests that there is a significant response bias for fricatives; listeners were more likely to make "voiced" response for fricatives. Interestingly, this direction of the response bias was different in the quiet listening context reported in section 2.5. Unlike the case of fricative, stop intercept was not significantly greater than zero [Intercept = 0.04282, Std. Error = 0.07945, DF = 8.65900, $t = 0.539$, $p = 0.603$].

To summarize, Experiment 2 showed that the listeners degraded intelligibility of the voicing contrast when it was heard in noisy contexts, and the degradation patterns showed an asymmetry: voicing contrast in stops was less intelligible (or, more confused) than that in fricatives in the noisy context. Additionally, significant response bias toward voiced responses was observed for fricatives, but not for stops.

4. General discussion

The purpose of the current study was to examine what contributes to the cross-linguistic difference in the manner asymmetries in the intelligibility of voicing contrast in noise. Using physically the same stimuli in the previous study, the current study examined English listeners' perception of Russian voicing contrast both in a quiet listening context (Experiment 1) and in a noisy listening context (Experiment 2). One of the hypotheses was that English and Russian listeners should hear the same acoustic contrast differently if the variability in speech perception itself was the source of the difference. The alternative hypothesis was that English and Russian listeners should hear the same acoustic contrast similarly if variability in speech production was the source of the difference.

The results of Experiment 1 confirmed that English listeners were able to perceive Russian contrasts with high accuracy as long as the listeners heard the stimuli in a quiet listening context. The results of Experiment 2 showed that

English listeners were more likely to confuse the voicing contrast in stops, compared with that in fricatives, indicating a pattern similar to the previous study for Russian [5]. In [5], voicing contrast in stops was more confused than in fricatives when Russian listeners heard the stimuli. The comparison of the current study with the previous study [5] suggests that Russian and English listeners exhibited a consistent pattern as long as they heard the same stimuli. Thus, the results support that the apparent cross-linguistic difference identified in previous studies [2, 3, 4, 5] is attributed to the difference in the acoustic properties of the stimuli.

The findings of the current perceptual study are closely related to the overt cross-linguistic difference of the phonetic implementation of laryngeal contrast. In recent typological studies of laryngeal contrast [e.g., 11], English is classified as one of the "aspirating" languages, which are typically characterized by short-lag and long-lag voice-onset times (VOTs) in word-initial stops, and non-obligatory vocal fold vibration during constriction ([11], see also, [12]). On the other hand, Russian is classified as one of the "true-voice" languages, which are typically characterized by prevoicing and short-lag VOTs in word-initial stops, and robust vocal fold vibration during constriction [13, 14, 15]. Considering those overt differences in speech production, the acoustic properties of the stimuli used in the previous study for English [2, 3, 4] were presumably fairly different from those used in the Russian study, resulting in apparent cross-linguistic perceptual differences. If this interpretation was correct, it is predicted that Russian listeners should show a pattern different from the current study when they hear English stimuli. Testing this prediction is a task for future studies.

In addition to the primary issue on intelligibility of contrast, the current study revealed interesting patterns in response bias. The results of Experiment 2 showed that listeners tended to hear fricatives as voiced in the noisy listening context. Voiced response bias of English listeners in noise has also been observed in previous studies [2]. What is more intriguing is the opposite response bias pattern in quiet and noisy listening contexts; Listeners preferred voiceless choices when the stimuli were presented with no background noise (Experiment 1) whereas they were biased toward voiced responses when the stimuli were presented with background noise (Experiment 2). This suggests that background noise guides listener's response bias, and, insightfully, this effect was observable only in fricatives but not in stops. Although thorough examination of response bias is beyond the scope of the current study, this issue should also be a topic of future investigation.

Finally, the manner asymmetries observed in the current study have implications for phonological theories, where the basic assumption is that perceptual cue organization directly affects phonology [16, 17]. Steriade [16] claims that a phonological contrast is prone to be lost in the context where fewer perceptual cues are available. The current study demonstrated that the voicing contrast in stops is more prone to be missed than that in fricatives in noisy contexts, where perceptual cues to voicing were energetically reduced by white noise. Given Steriade's claim, one implication of the current results is that Russian stops might have fewer perceptual cues to voicing compared with fricatives. Oppositely, it is also hypothesized that English stops might have greater perceptual cues to voicing, compared with fricatives. Further cross-linguistic acoustic data are needed to support this hypothesis.

5. Conclusions

Through the examination of the degradation pattern of phonological contrasts in a non-ideal listening context, the current study revealed significant perceptual characteristics of phonological contrasts. Specifically, the current study offered new data on how Russian voicing contrasts are perceived by English listeners. The implication of the results is that the source of the apparent cross-linguistic difference identified in previous noise-masking studies is not attributed to the variability in perception itself, but rather, to the difference in phonetic implementation of the voicing contrast in each language. The results in turn suggest that perceptual cue weighting strategies for perceiving voicing contrasts in different manners of articulation are similar among Russian and English listeners.

6. Acknowledgements

This work was partly supported by NINJAL (Project name: Phonological Characteristics of the Japanese Lexicon). I also wish to thank Alexei Kochetov for cooperation in data collection, Haruo Kubozono and Tim Vance for valuable input on some phases of this paper, and Donna Erickson for English proofreading.

7. References

[1] G.A. Miller, P.E. Nicely. "An analysis of perceptual confusions among some English consonants". *Journal of the Acoustical Society of America*, 27, 338-352. 1955.

[2] S.A. Phatak, A. Lovitt, J.B. Allen. "Consonant confusions in white noise". *Journal of the Acoustical Society of America* 124 (2), 1220-1233. 2008.

[3] S.A. Phatak, J.B. Allen. "Consonant and vowel confusions in speech-weighted noise". *Journal of the Acoustical Society of America*, 121(4), 2312-2126. 2007.

[4] K.W. Grant, B.E. Walden. "Evaluating the articulation index for auditory-visual consonant recognition". *Journal of the Acoustical Society of America*, 100, 2415–2424. 1996.

[5] M. Matsui. "Manner asymmetries in the perception of laryngeal contrast: A noise-masking experiment in Russian." ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Paper number 1015. 2015.

[6] P. Boersma, D. Weenink. Praat: doing phonetics by computer (Version 5.1.31). 2010.

[7] N.A. Macmillan, D.C. Creelman. *Detection Theory: A User's Guide*. 2nd edition. Mahwah: Lawrence Erlbaum Associates Publishers. 2005.

[8] M. Cooke, L.G. Lecumberri, J. Barker. "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception". *Journal of the Acoustical Society of America*, 123, 414-427. 2008.

[9] D. Bates, M. Maechler, B. Bolker, S. Walker. Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. 2014.

[10] R Core Team. R: A language and environment for statistical computing (Version 3.1.1). R Foundation for Statistical Computing, Vienna, Austria. 2014.

[11] J. Beckman, M. Jessen, C. Ringen. "Empirical evidence for laryngeal features: Aspirating vs. true voice languages". *Journal of Linguistics*, 49 (2), 259-284. 2013.

[12] L. Davison. "Variability in the implementation of voicing in American English obstruents." *Journal of Phonetics* 54, 35−50. 2016.

[13] C. Ringen, V. Kulikov. "Voicing in Russian stops: Cross-linguistic implications." *Journal of Slavic Linguistics* 20 (2), 269-286. 2012.

[14] V. Kulikov. Voicing and voice assimilation in Russian stops. Doctoral dissertation, The University of Iowa. 2012.
[15] M. Matsui. Roshiago ni okeru yuuseisei no tairitsu to tairitsu no jakka: Onkyo to chikaku. [Voicing contrast and contrast reduction in Russian: Acoustics and perception]. Doctoral dissertation, Hiroshima University. 2015.
[16] D. Stariade. “Phonetics in phonology: The case of laryngeal neutralization.” Ms. 1997. Retrieved from: <http://www.linguistics.ucla.edu/people/steriade/papers/PhoneticsInPhonology.pdf>
[17] D. Steriade. “The phonology of perceptibility effects: The P-map and its consequences for constraint organization.” Ms. 2001. Retrieved from: <http://web.mit.edu/linguistics/people/faculty/steriade/publications.htm>