MANNER ASYMMETRIES IN THE PERCEPTION OF LARYNGEAL CONTRAST: A NOISE-MASKING EXPERIMENT IN RUSSIAN

Mayuki Matsui

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

ABSTRACT

The issue of the effect of consonant manner on the perceptibility of laryngeal contrast was addressed. Previous studies of perceptual confusion suggest that, in noise, the laryngeal contrast in fricatives is more difficult to perceive than stops for listeners in English, which is typologically classified as an "aspirating" language. The present study examined whether a similar pattern is observed in Russian, which is typologically classified as a "true-voice" language. Native listeners of Russian identified voiced and voiceless obstruents in Gaussian white noise conditions. The results demonstrate that voicing in stops is more confused than that in fricatives in noise, indicating a trend that is the opposite of that observed for English. The results suggest that the pattern of manner asymmetries in the perception of laryngeal contrast is not universal.

Keywords: Speech perception in noise, Laryngeal contrast, Manner of articulation, Russian

1. INTRODUCTION

Over the past two decades, data from speech perception have played a crucial role in the development of phonological theories [e.g., 7, 13, 14] and in turn phonology has clarified speech perception [e.g., 3, 6]. The focus of the present study is the effect of consonant manner on the perceptibility of laryngeal contrast.

Although the classic study of perceptual confusion [9] showed that the voiced-voiceless contrast in English was relatively less confused in white noise, recent studies replicating Miller and Nicely [9] suggest that, for native English listeners, the laryngeal contrast in fricatives is more difficult to perceive than stops [11]. Such manner asymmetries are observed not only in white noise [11] but also in speech-shaped noise [5, 10], suggesting that the pattern is, to some extent, consistent across different masker types.

An intriguing but unsolved question is whether or not such asymmetries are consistent across languages. Additional data from languages other than English would help address this question.

Here, the effect of consonant manner on the perceptibility of laryngeal contrast in Russian was investigated. Such as English, Russian has two-way laryngeal contrast, and the contrast is traditionally described as voiced versus voiceless. However, according to recent developments in typological studies of laryngeal contrast [e.g., 1], 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. Russian is classified as one of the "true-voice" languages, which are typically characterized by prevoicing and short-lag VOTs in word-initial stops [1, 12]. Since the acoustic manifestation of laryngeal contrast is different between aspirating and true-voice languages, the perceived confusability (or similarity) of contrasts might be different between the aspirating language English and the true-voice language Russian.

Two perception experiments are described in the following sections. The results of Experiment 1 confirmed that the laryngeal contrast in stops and fricatives is equally audible for listeners in a quiet condition. The results of Experiment 2 demonstrated that the voicing in stops is more confused than in fricatives in noise.

2. EXPERIMENT 1

2.1. Participants

The participants were 17 native speakers of Russian (mean age, 21 yrs; range 18–28 yrs). Of the 17, five were male and 12 were female. The data of one participant were excluded from the analysis because he did not complete the experiment.

At the time of the experiment, the participants were undergraduate or graduate students in Orenburg, Russia. None reported using any language other than Russian on a daily basis. None reported any known history of speech or hearing disorders.

2.2. Auditory stimuli

Coronal stops and fricatives (/t, d, s, z/) were embedded between open vowels /a/, and thus the stimuli consisted of four types of vowel-consonant-vowel (VCV) sequences: /ata/, /ada/, /asa/, and /aza/.

These materials were pronounced by two native speakers of Russian (a female and a male). Two possible stress patterns were recorded, and two repetitions of each condition from each speaker were used as stimuli. Tokens were recorded onto a portable recorder (PCM-M10, Sony) with a stereo microphone (ECM-MS907, Sony) using a 44.1-kHz sampling rate at a 16-bit quantization level. The first channel of the stereo recordings was extracted for creating the stimuli.

2.3. Procedures

An identification experiment was conducted in a quiet room at Orenburg State University (Russia). Stimuli were presented to the participants via headphones (MDR-Z700, Sony), and the participants were instructed to choose what they heard from two choices: voiced or voiceless. The possible choices were orthographically presented on a screen, and the participant responded by clicking the button as quickly as possible. The possible place effect of orthographic representation was counterbalanced: half of the trials were presented 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. For each participant, the number of trials was 64 (4 consonant types × 2 stress patterns × 2 speakers × 2 tokens × 2 button positions). Before the experiment, each participant performed nine practice trials. The experiment was implemented by the speech analysis software Praat [2].

2.4. Results and discussion

Table 1a is the cross table for stops, and Table 1b is that for fricatives. Not surprisingly, both tables show that the listeners' performances were almost on the ceiling level, confirming that both the stop and fricative stimuli used were equally audible enough for the listeners in the quiet condition.

Table 1a: Cross table for stops in the quiet condition. Each number represents the number of responses.

Stop, quiet	Response		
Stimulus	Voiceless	Voiced	Row total
Voiceless	256 (100%)	0 (0%)	256 (100%)
Voiced	1 (0.4%)	255 (99.6%)	256 (100%)
Column total	257 (50.2%)	255 (49.8%)	512 (100%)

Table 1b: Cross table for fricatives in the quiet condition. Each number represents the number of responses.

Fricative, quiet	Response		
Stimulus	Voiceless	Voiced	Row total
Voiceless	255 (99.6%)	1 (0.4%)	256 (100%)
Voiced	1 (0.4%)	255 (99.6%)	256 (100%)
Column total	256 (50%)	256 (50%)	512 (100%)

3. EXPERIMENT 2

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 (for recent discussions on masker types, see Cooke *et al.* [4]). The Vocal Tool Kit in Praat was used to add noise to the stimuli.

Stimuli with three levels of signal-to-noise ratio (SNR) were prepared: −5 dB, −8 dB, and −10 dB. For the observation of confusion patterns, the levels of the SNR that induced each listener's perceptual confusion were carefully determined on the basis of the results of a pilot study.

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 (4 consonant types × 2 stress patterns × 2 speakers × 2 tokens × 2 button positions × 3 SNR conditions). Listeners performed the task in the following order: −5 dB (easiest), −8 dB (intermediate), and −10 dB (hardest).

3.4. Analysis

A preliminary analysis indicated that perceptual confusion was successfully induced at all three SNR levels. In the subsequent analyses, the responses were pooled across SNR conditions.

For the evaluation of each listener's sensitivity to the stimuli apart from response bias, *d'* values were calculated for each manner condition for each listener.[1] *d'* is a measure of sensitivity that was proposed in signal detection theory [8]. Higher d' values indicate higher sensitivity.

In addition to *d'*, criterion (*c*), which is a measure of the response bias, was also calculated for each manner condition for each listener.[2]

3.5. Results

Table 2a is the cross table for stops, and Table 2b is that for fricatives. No remarkable response bias is observed in Table 2b, whereas Table 2a shows a slight response bias toward voiceless. Table 3 summarizes the mean *d'* and *c* values for stops and for fricatives.

Table 2a: Cross table for stops in the noise condition (pooled across SNR conditions). Each number represents the number of responses.

Stop, noise	Response		
Stimulus	Voiceless	Voiced	Row total
Voiceless	561 (73.0%)	207 (27.0%)	768 (100%)
Voiced	298 (38.8%)	470 (61.2%)	768 (100%)
Column total	859 (55.9%)	677 (44.1%)	1536 (100%)

Table 2b: Cross table for fricatives in the noise condition (pooled across SNR conditions). Each number represents the number of responses.

Fricative, noise	Response		
Stimulus	Voiceless	Voiced	Row total
Voiceless	563 (73.3%)	205 (26.7%)	768 (100%)
Voiced	204 (26.6%)	564 (73.4%)	768 (100%)
Column total	767 (49.9%)	769 (50.1%)	1536 (100%)

Table 3: Mean *d'* and *c* values in the noise condition across 16 listeners and across SNR conditions. Standard deviations are represented in parentheses.

	Voicing in stop	Voicing in fricative
d'	1.79 (0.96)	2.31 (0.8)
c	−0.18 (0.31)	−0.004 (0.2)

For the investigation of the effect of manner of articulation on the perceptibility of voicing contrast, a linear mixed-effects model was constructed with 'Manner (stop vs. fricative)' as a fixed effect and 'listener' as a random effect. The dependent variable was the *d'* value. A likelihood-ratio test indicated that the fixed effect of Manner was significant [$X^2(1) = 7.5232$, $p<0.01$]. The *d'* values for the stop stimuli were significantly smaller than those for the fricative stimuli.

For the examination of whether any response bias was significant, one-sample *t*-tests were conducted on the *c* values. The results showed that, although the mean value of *c* for 'fricative' was not significantly different from zero [$t(15) = -0.0738$, $p = 0.9421$, n.s.], there was a significant difference between the mean *c* value for 'stop' and zero[$t(15) = -2.303$, $p < 0.05$].

Further, the effect of the stress position of the stimuli (stress on the first vs. second syllable) on the listeners' voiced-voiceless response was considered. The distributions of *d'* as a function of *c* with different stress positions are illustrated in Figure 1a (for stops) and Figure 1b (for fricatives).

A linear mixed-effects model was constructed with 'Stress Position (first vs. second syllable)' as a fixed effect and 'listener' as a random effect. The dependent variable was the *d'* value. No significant effect of Stress Position was obtained for stops [$X^2(1) = 1.4015$, $p = 0.2365$, n.s.] or for fricatives [$X^2(1) = 0.2021$, $p = 0.6531$, n.s.].

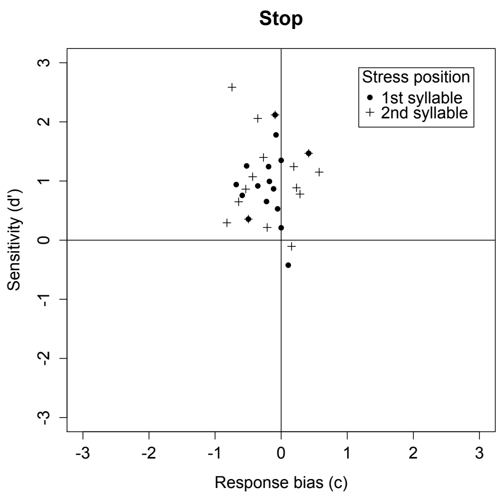

Figure 1a: Sensitivity (*d')* as a function of response bias (*c*) with different stress positions (for stops). Positive *c* values: the bias toward voiced response. Negative *c* values: the bias toward voiceless response.

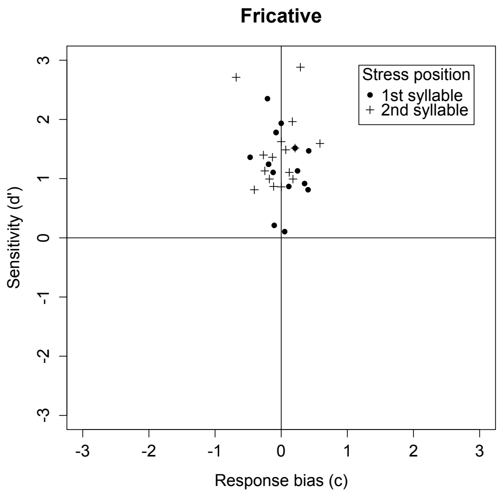

Figure 1b: Sensitivity (*d')* as a function of response bias (*c*) with different stress positions (for fricatives). Positive *c* values: the bias toward voiced response. Negative *c* values: the bias toward voiceless response.

To summarize, the results showed that the *d'* value was significantly smaller for the stops than for the fricatives. A significant response bias toward voiceless was obtained for the stops. The perceptibility of voicing contrast was not significantly affected by the stress position of the stimuli.

4. DISCUSSION AND CONCLUSIONS

The present study examined the effect of consonant manner on the perception of laryngeal contrast in Russian, which is typologically classified as a true-voice language.

The results of the noise-masking experiment showed that the sensitivity measure *d'* was significantly smaller for stops compared to fricatives. This suggests that voicing in stops is more confused than in fricatives in noise for Russian listeners, indicating a trend that is the opposite of that observed for English.

The results of this study suggest that the pattern of manner asymmetries in the perception of laryngeal contrast is not universal. If the perceived similarity of laryngeal contrast according to the manner of articulation is not universal, there should be at least three possible patterns:

(1) Fricative voicing is more difficult to perceive than stop voicing (e.g., English; [5, 10, 11]).

(2) Stop voicing is more difficult to perceive than fricative voicing (e.g., Russian; the present study).

(3) Stop and fricative voicing are symmetrically perceived (Language X, as yet unknown).

Future studies could extend these observations to the other languages, advancing our understanding of the source of such manner asymmetries.

ACKNOWLEDGEMENTS

I thank Liudmila Dokashenko and her students for their help carrying out the experiments at Orenburg State University. This research was partly supported by a Grant-in-Aid for JSPS Fellows (#246727).

REFERENCES

[1] Beckman, J., Jessen, M., Ringen, C. 2013. Empirical Evidence for Laryngeal Features: Aspirating vs. True Voice Languages. *Journal of Linguistics*, 49 (2), 259-284.

[2] Boersma, P., Weenink, D. 2010. Praat: doing phonetics by computer (Version 5.1.31) [Computer program]. Last retrieved on April 4, 2010 from http://www.praat.org/.

[3] Boomershine, A., Hall, K. C., Hume, E., Johnson, K. 2008. The impact of allophony vs. contrast on speech perception. In P. Avery, E. Dresher and K. Rice [Eds.], *Contrast in Phonology: Perception and Acquisition*, 143-172. New York: Mouton de Gruyter.

[4] Cooke, M. M., Lecumberri, L. G., Barker, J. 2008. 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.

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

[6] Huang, T., Johnson, K. 2010. Language specificity in speech perception: Perception of Mandarin tones by native and non-native listeners. *Phonetica*, 67, 243-267.

[7] Kawahara, S. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language*, 82 (3), 536-574.

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

[9] Miller, G. A., Nicely, P. E. 1955. An analysis of perceptual confusions among some English

consonants. *Journal of the Acoustical Society of America*, 27, 338-352.
[10] Phatak, S. A., Allen, J. B. 2007. Consonant and vowel confusions in speech-weighted noise. *Journal of the Acoustical Society of America*, 121(4), 2312-2126.
[11] Phatak, S. A., Lovitt, A., Allen, J. B. 2008. Consonant confusions in white noise. *Journal of the Acoustical Society of America* 124 (2), 1220-1233.
[12] Ringen, C., Kulikov, V. 2012. Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics* 20(2), 269-286.

[1]The formula used for the sensitivity measure is as follows: *d'* = z(H) − z(F). Z(H) is the *z*-transformed probability to 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"). See [8] for details.
[2]The formula used for the response bias measure is as follows: $c = -(z(H) + z(F))/ 2$.

[13] Stariade, Donca (1997) Phonetics in phonology: The case of laryngeal neutralization. Ms. from http://www.linguistics.ucla.edu/people/steriade/papers/PhoneticsInPhonology.pdf
[14] Steriade, Donca (2001) The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. Ms. from http://web.mit.edu/linguistics/people/faculty/steriade/publications.html