Записки научных
семинаров ПОМИ
Том 529, 2023 г.

U. Kochetkova, P. Skrelin, V. Evdokimova, D. Novoselova

USING AUDIO SIGNAL MODIFICATION TO VALIDATE THE ACOUSTIC FEATURES RELEVANCE

Abstract. In this work, we focus on experiments using modifications of prosodic parameters separately and in various combinations. For this purpose, original homonymous neutral and ironic utterances were modified to achieve the opposite effect, i.e. to get ironic utterances from neutral ones and vice versa. These original homonymous utterances were read by Russian native speakers in mini-texts and dialogues implying neutral or ironic meaning, then they were extracted from the context, any markers of ironic or non-ironic meaning were eliminated, then they were suggested to native listeners during the auditory perceptual experiments. The acoustic analysis was done for pairs of homonymous ironic and non-ironic utterances whose meaning was correctly recognized by the majority of listeners. The most salient acoustic features of ironic meaning were calculated. Then, modifications were done in the pairs of ironic and non-ironic utterances containing these features. In the three series of experiments, temporal, dynamic, and melodic parameters were modified using Wave Assistant and Praat software. The results of perceptual experiments with modified stimuli not only demonstrated the prevalence of the melodic pattern, but also revealed the complex of acoustic features of irony and indicated the challenges with modified ironic stimuli due to their spectral density.

§1. Introduction

Audio signal modification or resynthesis is an artificial change of the acoustic characteristics of the original speech signal in order to reproduce the characteristics of another (target) signal as closely as possible. This method of experimental research allows to determine the influence of the value of each acoustic parameter and its contribution to achieving the maximum perceptual similarity of the modified source signal with the target one [3, 13, 22]. Therefore, this method becomes a very subtle instrument that helps identifying the role of different parameters within the complex acoustic characteristic of a concrete meaning, emotion or modality. One

Key words and phrases: Audio Signal Modification and Acoustic Analysis and Auditory Perceptual Experiments and Prosodic Features and Ironic Speech.

of such modalities, which remains a challenging point for the dialogue systems using AI technologies in different languages, is irony, notably ironic negation or antiphrasis. This phenomenon consists in negating the direct word meaning by intonation or paralinguistic means. Its correct interpretation and production is very important for conversational agents because of its pragmatic role. If the utterance with antiphrasis is understood in its direct sense, it can mislead or disturb the communication process. That is why nowadays this topic attracts such a high attention of the researchers [1, 2, 4–7, 11, 12, 15, 18–21].

At the first stage of our research, we built a corpus of ironic speech containing ironic and non-ironic homonymous stimuli. In total, it gave more than 4.5 thousand pairs of ironic and non-ironic target sentences. The methodology of the corpus construction, as well as the results of the acoustic analysis are given in our previous works [16, 17, 23].

The current study aims at verifying the perceptual relevance and relationship of the acoustic parameters obtained at the previous stage of our work. We are interested in modifications of intonation characteristics. Since our material consists of stimuli, which represent identical sequences of sounds, which differ only by its connotation (neutral or ironic), read by the same speaker, we have the opportunity to fully use the technique of resynthesis. Alternately changing prosodic characteristics (pitch frequency, duration of sounds or syllables, their intensity or amplitude) will determine to what extent each of the studied parameters affects the achievement of perceptual similarity of the initially neutral utterance with the ironic and vice versa. All iterations of changes in prosodic characteristics were subjected to perceptual evaluation based on auditory perceptual experiments. A similar approach (apart from intensity level changes) can be found in the study of French ironic speech [14], an analysis of ironic speech through synthesis is also given in the work based on the material of the German language [10]. Furthermore, this method has been used in emotional speech analysis, see, e.g., [8, 9].

§2. The Pilot Experiment

2.1. Material and Method. The material was selected from the main corpus of ironic speech mentioned above. For the pilot experiment we chose the reading of four Russian native speakers (two male and two female speakers, aged 18–40). They read a set of short texts and dialogues containing homonymous utterances. The texts were composed in a way to

suggest to the speakers neutral or ironic implementation. Each speaker read 86 such texts. The recording was accomplished in a soundproof cabin at the Phonetic Department of the Saint Petersburg State University. At the second stage, ironic and non-ironic target utterances were extracted from the recording and presented in a series of auditory perceptual experiments. The snippets suggested to the native listeners did not contain any context or lexical marker of ironic or neutral (non-ironic) meaning. The participants associated each of these snippets, given in a randomized order, with one of the contexts written on the screen: ironic or non-ironic. A third option was "I am not sure".

Well-recognized snippets, i.e., snippets for which ironic or non-ironic meaning was correctly identified by more than 60 percent of listeners, were analyzed using Wave Assistant and Praat software. To verify the significance of difference in acoustic parameters between neutral and ironic speech, a Student paired t-test analysis was made for the pairs of homonymous snippets read by the same speaker. Results of this analysis showed a complex multiparametric character of irony expression. The augmentation of the nucleus duration (i.e. the stressed syllable of the intonation centre of the utterance) and of its intensity were the most salient prosodic features, presented in 86 percent and 78 percent of utterances respectively. Another salient prosodic feature was the melodic range change. Its increase occurred almost as often as the decrease (in 48 percent and 46 percent of utterances respectively). At the same time, the melodic pattern is known to be the best recognized in differing various sentence types. Thus, we started by verifying the role of the melodic pattern in the perception of the ironic or neutral meaning.

For the analysis by synthesis we selected 14 pairs of ironic and non-ironic utterances, which were correctly evaluated by more than 80 percent of listeners at the previous stage of the study. Then the modification of the F0 was accomplished.

The experiment was based on the method of transposing the melodic contour. For this purpose the values of the envelope frequency of the main tone of the original utterance were modified, i.e. brought to the values of the envelope of the target utterance. In our experiments, the melodic pattern from the target fragments with ironies was transferred to the corresponding passages taken from non–ironic narrative, exclamatory and interrogative utterances, and vice versa – the contour from non-ironic fragments was superimposed on a similar passage with irony uttered by the

same speaker. Thus, ironic target fragments with melodic pattern from the corresponding non-ironic passages and non-ironic fragments with a melodic contour transposed from passages with irony were obtained.

2.1.1. *Resynthesis Procedure.* The procedure for transposing the melodic contour was as follows. In the phrases selected for resynthesis, the labels of the periods of the fundamental frequency were indicated, using the Wave Assistant sound processing program, then they were exported into a Praat file with the format .pitchtier. The resulting pattern in the Praat program was manually transferred to the corresponding segment based on the key points – the points of the local minimum and maximum of the pitch frequency. Thus, the original dynamic, temporal and spectral features were preserved and only the melodic design was changed. The fragments modified in this way were presented to the auditors for perceptual evaluation.

2.1.2. *Perceptual Experiments.* The organization of the perceptual experiment with resynthesized stimuli in general was very similar to the methodology of a previous survey mentioned in Section 2.1: participants were also asked to listen to the target passage and determine from which context (phrase or short dialogue) it was extracted. The answer options were compiled in such a way as to prevent direct indication of an ironic reading, while the participants were not aware of the purpose of the study. Participants could give comments at the end of the questionnaire. The survey was published on the SoSci Survey online platform. For this perceptual experiment, it was necessary to select reliably identified passages taken from ironic and non-ironic contexts. Some fragments had to be discarded due to changes in segment composition (for instance, two variants of the word "now" – "seichas" and "schas") and strong laryngalization. In total, 14 fragments were selected in this way. The informants also filled out a questionnaire with personal data. Passing the test took 5-7 minutes. 20 native speakers of the Russian language, including 17 women and 3 men aged from 19 to 40 years, took part in the perceptual experiment on the material of the resynthesized fragments.

Figure 1 represents a page from the survey. The following task was given: "Opredelite, iz kakoj frazy ili dialoga byl vzyat privedennyj otryvok" (Choose which sentence or dialogue the excerpt is from). Three options in the first paragraph are:

- "Drug, kak zhe! Ni razu ne pomog, kogda mne nado bylo!" (A friend, yeah! He's never helped me when I needed it!)

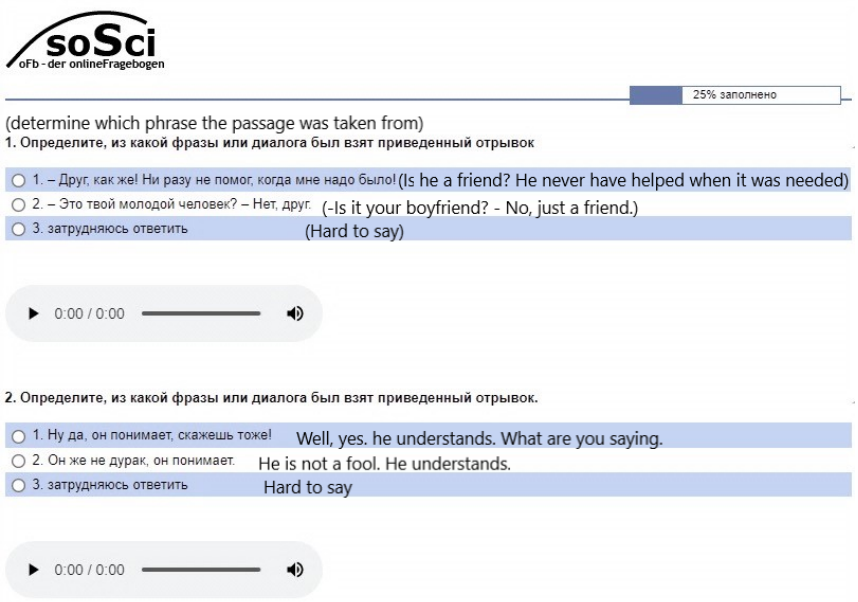

Figure 1. A page of the perceptual survey based on the material of the resynthesized excerpts (the translation is given in the text).

- "– Eto tvoj molodoj chelovek? – Net, drug" (– Is this your boyfriend? – No, <just> a friend.)
- "Zatrudnyayus' otvetit"' (Unsure).

The second paragraph contains the following options:

- "Nu da, on ponimaet, skazhesh' tozhe!" (Oh yeah, he understands, come on!)
- "On zhe ne durak, on ponimaet." (He's not a fool, he understands.)
- "Zatrudnyayus' otvetit"' (Unsure).

2.2. Results of the Pilot Experiment. In the course of a perceptual experiment with a resynthesized melodic contour, the following results were obtained:

- in 5 out of 14 stimuli the originally neutral utterances after modifications were perceived as ironic; at the same time the perception of

the sentence type changed, i.e. narrative sentences were perceived as questions, and vice versa;
- 2 originally ironic stimuli were perceived as non-ironic.

Changes in the rest of modified stimuli (50 percent of the excerpts) did not provide a significant change in neutral or ironic meaning perception, which may be explained by the influence of other parameters, such as duration and strong spectral changes.

2.3. Interim Conclusion. Based on the above, it can be concluded that melodic pattern may have an impact on the perception of ironic utterances in the Russian language, but the phonetic side of irony is far from being limited to the melodic figure, as spectral, dynamic and temporal characteristics seem also play an important role. To test the role of the remaining parameters, the following series of experiments were conducted.

§3. Experiments with a Series of Audio Signal Modifications

3.1. Material and Method.

3.1.1. *Material.* The methodology of material selection was similar to the method used in the Pilot experiment. Two native Russian speakers (one male and one female speaker, 19 and 22 years old) performed the same reading task, as described in Section 2.1. The snippets from there recordings were extracted with no context or remark, randomized and suggested to native listeners in the auditory perceptual experiments through the SoSciSurvey platform.

A set of reliably identified homonymous ironic and neutral stimuli was obtained. For this series of experiments the threshold of evaluation was higher than in the pilot experiment (correct evaluation by more than 85 percent of informants). It gave 13 pairs of the well-recognized homonymous ironic and neutral stimuli in male speech and 15 pairs in female speech. These snippets were subject to acoustic and statistical analysis (see Section 2.1). The same list of salient acoustic features was obtained: the increase of the stressed vowel (nucleus) duration, intensity level (see Table 1), as well as changes in melodic range and in melodic patterns on the whole.

As one can see from Table 1, the increase of average duration of the stressed syllable was almost the same in male and female speech, if we consider it as a percentage value (55 ms and 73 percent in male speech and 63 ms and 72 percent in female speech). The values of the intensity of the

Table 1. Acoustic characteristics of the stressed vowel in the original ironic and non-ironic stimuli, correctly identified by the listeners.

Type of Stimuli	Speakers (Male/Female)	Average Durations (ms)	Average Intensity (dB)
Ironic	M	138	70.2
Stimuli	M	153	70.2
Non-Ironic	M	83	65.3
Stimuli	F	90	64.8
Increase	M	55 (73 percent)	4.9
in Ironic Stimuli	F	63 (72 percent)	5.3

stressed vowel are even closer in male and female speech. These stimuli were selected for the modifications.

3.1.2. *Resynthesis Procedure.* The first series of experiments was based on the imposition of acoustic characteristics of ironic target phrases (duration and intensity of the stressed vowel, as well as the pitch pattern) and their combinations on homonymous non-ironic stimuli. In the second series, the same experiments were carried out, but the characteristics of non-ironic statements were superimposed on ironic ones. In each series of experiments , seven types of modifications were carried out: (1) a change in the duration of the stressed vowel; (2) a change in the intensity of the stressed vowels; (3) a modification of the contour of the main tone of the target stimulus; (4) a combination of changes in the duration and intensity of the stressed vowel; (5) a combination of changes in the duration of the stressed vowel and the melody of the stimulus contour; (6) a combination of changes in the melody of the contour of the target phrase and the intensity of the stressed vowel; (7) a combination of changes in the duration of the stressed vowel, intensity and pitch contour of the stimulus.

Duration. Results of acoustic analysis showed that the duration of the stressed vowel of the intonation focus (nucleus) is an important acoustic parameter for listeners while detecting irony. Modification of this acoustic feature was done with the Wave Assistant sound processing software. First, segmentation of the stressed vowel in ironic and paired non-ironic stimuli was carried out manually. Then labels of the periods of the fundamental frequency were placed and corrected. Duration of the stressed vowel was

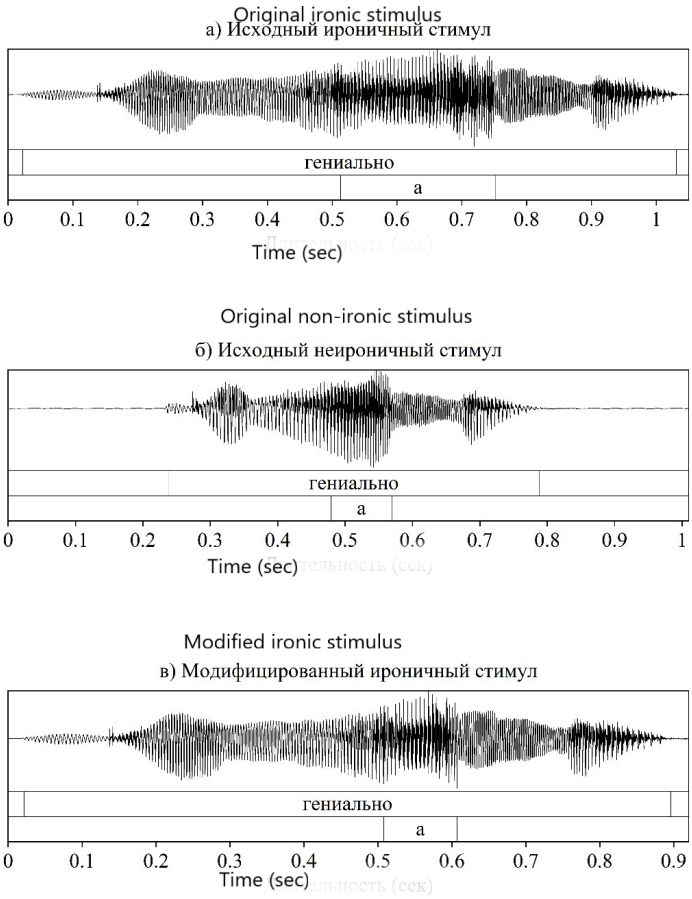

Figure 2. Transposition of the duration of a stressed vowel from a non-ironic target stimulus "Genialno" (Brilliant) (a) to an ironic homonymous source stimulus (b); the last example (B) is the resulting modified stimulus.

modified by adding or reducing the required number of periods of the fundamental frequency so that the resulting vowel corresponded in duration to the stressed vowel of the paired context. This was done in order to avoid clicks when adding or subtracting periods of the fundamental frequency.

Fig. 2 illustrates the modification of the duration of the stressed vowel for the ironic stimulus "brilliant" (“genialno”).
Intensity. Modifications of the intensity of the stressed vowel were carried out in the Wave Assistant program using the linear processing function, which allows you to change the amplitude of the sound by multiplying it by a constant. For each fragment, this number was selected individually so that when multiplying the amplitude by this constant, the stressed vowel had the same intensity as the stressed vowel from the paired context. Figure 3 illustrates the process of changing the intensity of the stressed vowel of the ironic stimulus “And this is the end” (“I eto konets”): the change in intensity is proportional to the change in the amplitude of the stressed vowel.
Melodic Pattern. Figure 4 shows the modification of melodic characteristics carried out using the intonation pattern transposition method, in which the melodic pattern of an ironic fragment is superimposed on a similar non-ironic passage, and vice versa, the melodic figure of a non-ironic stimulus is transferred to the corresponding stimulus with irony.

The transplant procedure was as follows. The tags of the main frequency of the target phrases were marked manually in the Wave Assistant program. The markup was then exported by Praat. After that, the resulting contour was manually super-imposed on the corresponding segment at the key points of the local minimum and maximum of the pitch frequency using the Praat script using the overlap-add method. This procedure allows you to change the melodic design of the utterance, while maintaining the original dynamic and temporal characteristics (see Figure 4).
Combinations of different modifications. In addition to the three types of modifications described above, four types of modifications consisting of their combinations were also carried out. These combinations include changing the duration and intensity of the stressed vowel, changing the duration of the stressed vowel followed by changing the pitch contour of the target phrase, changing the pitch contour followed by changing the intensity of the stressed vowel, and finally combining all three parameters together: controlling the duration of the stressed vowel, modifying the pitch contour and changing the intensity of the stressed vowel.

3.1.3. *Auditory Perceptual Experiments.* Seven types of modifications described above were accomplished for 10 utterances from the female speech and 8 utterances from the male speech. Other well-recognized utterances did not provide a satisfactory result during the resynthesis. Thus 252

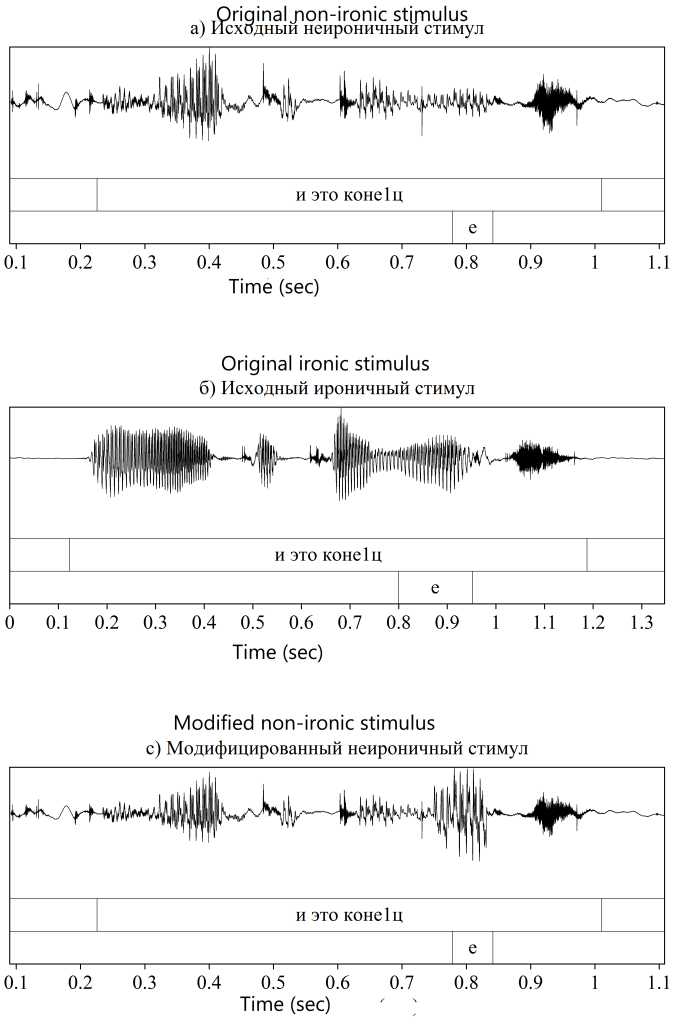

Figure 3. Transposing the intensity of a stressed vowel from a non-ironic target stimulus “I eto konets” (This is the end) (b) to an ironic homonymous source stimulus (a); the last example (c) is the resulting modified stimulus.

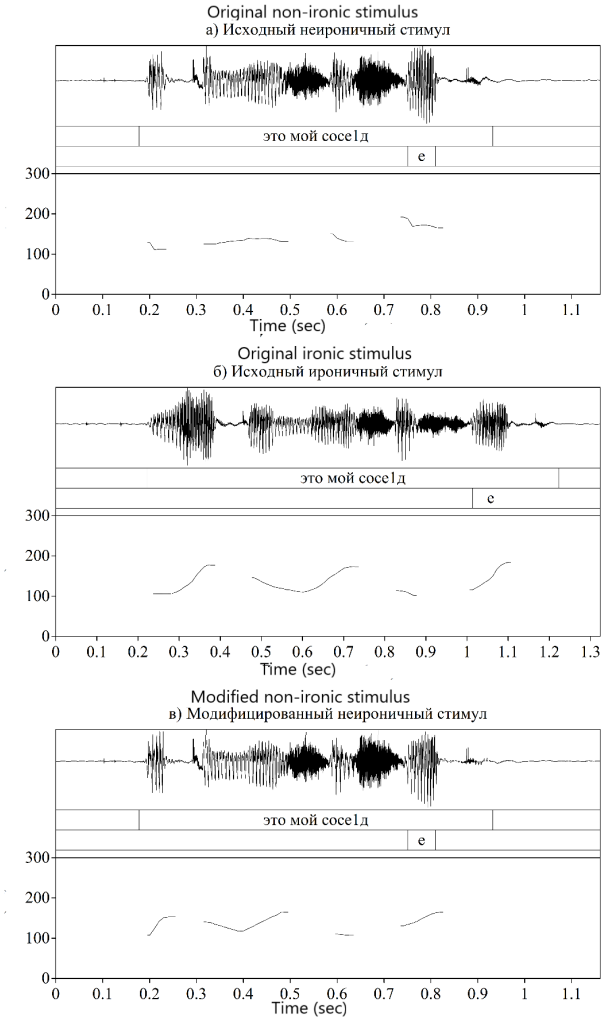

Figure 4. Transposing the melodic pattern from an ironic target stimulus “Eto moi sosed” (This is my neighbour) (b) to a non-ironic source stimulus (a); the last example (B) is the modified resulting stimulus.

modified stimuli (126 originally neutral stimuli and 126 originally ironic stimuli) were suggested to the listeners along with the source ironic or neutral stimuli. It resulted in a series of six perceptual experiments. The number of participants in each of the experiments was from 16 to 30; the average percentage of female native listeners was 56 percent, while male native listeners represented only 44 percent of the participants. Various age groups took part at the experiments, though the students formed the largest group (about 67 percent). The experimental task was the same, as in the previous auditory perceptual experiments (see Fig.1); all series of experiments were conducted using SoSciSurvey platform.

3.2. Results.

3.2.1. *Experiments with the neutral source stimuli.* Perceptual experiments showed that the combination of three parameters (duration, intensity and melodic contour), as well as the combination of the melodic pattern and duration changes were the most effective for the perception of irony in the stimuli modified from neutral source utterances using target ironic utterances (see Fig. 5).

The listeners' good recognition of irony in modified stimuli was also provided by a combination of intensity change with a change of pitch pattern, but modified stimuli in male speech were perceived as ironic less frequently, than in female speech. The pitch frequency contour plays an important role in the perception of irony, even if it is changed by itself. This observation was more typical for female speech in our study. The worst results gave separate modifications of the duration and intensity, as well as their combination. Nevertheless, we can suppose that a bigger increase of duration could give better result for the concrete intonation models and/ or sentence types (i.e. declarative, interrogative or exclamatory sentences). Such an analysis was not planned in the frame of the current work, but could be done in future.

3.2.2. *Experiments with ironic source stimuli.* The second series of experiments included modifications of ironic stimuli, on which, separately and in different combinations, the parameters of the original homonymous neutral stimuli read by the same speaker were superimposed. Statistical analysis of the perceptual experiments results showed that on average, the ironic passages that had undergone modifications were less often identified as neutral than the originally neutral ones. It can be due to the spectral characteristics that are typical for the ironic stimuli (higher spectral density was observed

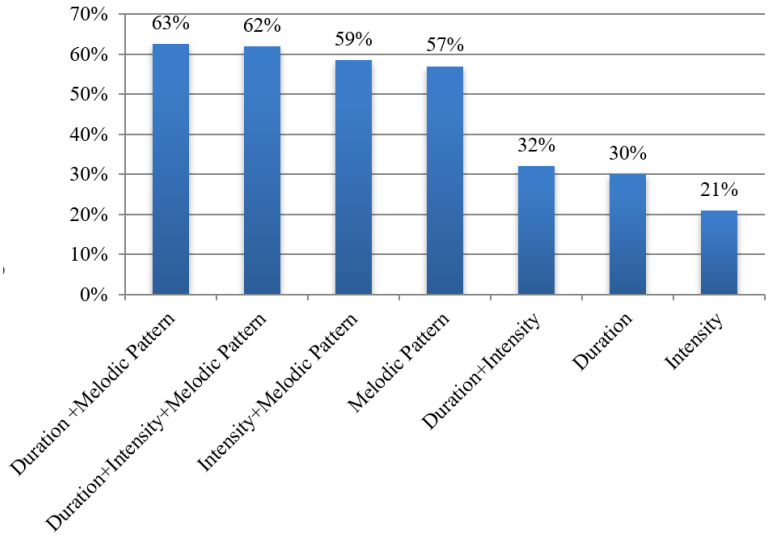

Figure 5. The average percentage of recognition of stimuli as ironic in accordance with various types of modifications based on the original neutral stimuli.

in more than 80 percent of ironic utterances). These characteristics can not be modified or extracted from the signal without a significant loss. Thus, these features of the source ironic stimuli remain in the signal even after the whole series of modifications. But, most often, it provides a new timbre. If the pitch pattern of the target neutral stimulus is placed lower or has less melodic peaks, the resulting quality produce the effect of a nasalized pronunciation.

As in the previous series of the experiments, the most significant were various kinds of pitch contour changes combined with changes of other acoustic characteristics (see Fig. 6). The importance of the intonation pattern for the transmission of ironic meaning is also evidenced by the fact that most of the fragments with a modified melodic contour (66.7 percent) were reliably identified by auditors as non-ironic (over 60 percent of auditors). At the same time, the change in duration and intensity does not have much effect on the perception of fragments as non-ironic: these

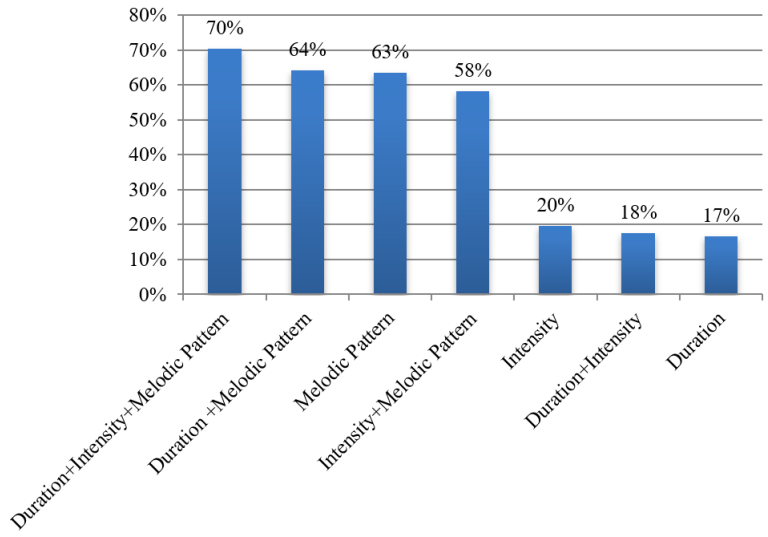

Figure 6. The average percentage of recognition of phrases as ironic in accordance with various types of modifications based on the original ironic stimuli.

modified fragments are generally recognized in the same way as the original ironic fragments.

§4. Conclusion

The results of the transfer of acoustic characteristics from ironic fragments to non-ironic and vice versa support the conclusions that the most important acoustic parameter responsible for the ironic meaning perception is the melodic pattern (including combinations of changes in the intonation contour with other acoustic characteristics). Simultaneous changes in duration, intensity and intonation contour have the greatest impact on the perception of a phrase as ironic or non-ironic. The same results were obtained in the study based on the French ironic speech [13], though the intensity of the stressed vowel was not considered in that study. At the same time, temporal and dynamic characteristics were the least important; changing these parameters individually or in combination with each other

does not significantly affect the interpretation of ironic and non-ironic passages.

At the same time, the melodic component seems to depend on the speaker individual strategies and habits while pronouncing utterances in an expressive way. If such an expressiveness exists, the melodic line by itself can reveal the ironic meaning. If not, the whole complex of acoustic parameters may be necessary to achieve the correct meaning: neutral or ironic. This consideration could be taken into account while preparing material for the machine learning models.

Acknowledgements. This research was supported by the RFBR grant No 20-012-00552.

References

1. A. Abu-Shnein, *Acoustic prosodic features in sarcastic utterances.* — J. Kufa Studies Center **1**, No. 46 (2017).
2. L. Anolli, M. Infantino, M. Ciceri, *From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication.* — Intern. J. Psychology - INT J PSYCHOL **37** (2002), 266–276.
3. J. Barnes, S. Shattuck-Hufnagel, *Prosodic theory and practice*, MIT Press, 2022.
4. A. Braun, A. Schmiedel, *The phonetics of ambiguity: A study on verbal irony*, pp. 111–136, 10 2018.
5. G. Bryant, *Prosodic contrasts in ironic speech.* — Discourse Processes - DISCOURSE PROCESS **47** (2010), 545–566.
6. G. Bryant, J. Fox Tree, *Is there an ironic tone of voice?.* — Language and Speech **48** (2005), 257–77.
7. G. Bryant, J. Fox Tree, *Is there an ironic tone of voice?.* — Language and Speech **48** (2005), 257–77.
8. M. Bulut, C. Busso, S. Yildirim, A. Kazemzadeh, C. Lee, S. Lee, S. Narayanan, *Investigating the role of phoneme-level modifications in emotional speech resynthesis*, 09 2005, pp. 801–804.
9. M. Bulut, S. Lee, S. Narayanan, *Recognition for synthesis: Automatic parameter selection for resynthesis of emotional speech from neutral speech*, 05 2008, pp. 4629 – 4632.
10. F. Burkhardt, B. Weiss, F. Eyben, J. Deng, B. Schuller, *Detecting vocal irony*, pp. 11–22, 01 2018.
11. H. Cheang, M. Pell, *Acoustic markers of sarcasm in cantonese and english.* — The Journal of the Acoustical Society of America **126** (2009), 1394–405.
12. A. Del Ré, F. Hirsch, C. Dodane, *L'ironie dans le discours : des premiéres productions enfantines aux productions des adultes*, Cahiers de praxématique (2018).
13. J. L. Flanagan, *Speech analysis, synthesis and perception*, Springer Berlin, Heidelberg, 1965.

14. S. González-Fuente, P. Prieto, I. Noveck, *A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French*, 06 2016.
15. M. Igras-Cybulska, B. Ziolko, *Acoustic correlates of ironic intonation*, 01 2015.
16. U. Kochetkova, P. Skrelin, V. Evdokimova, D. Novoselova, *Perception of irony in speech.* — Proceedings of the 4th International Conference on Neurobiology of Speech and Language (Saint-Petersburg, Ruassia), 2020, pp. 72–73.
17. U. Kochetkova, P. Skrelin, V. Evdokimova, D. Novoselova, *The speech corpus for studying phonetic properties of irony*, Language, Music and Gesture: Informational Crossroads: LMGIC 2021 (Tatiana Chernigovskaya, Polina Eismont, and Tatiana Petrova, eds.), Springer Singapore, Singapore, 2021, pp. 203–214.
18. H. Leykum, *Acoustic characteristics of verbal irony in standard austrian german*, 2019.
19. S. Li, W. Gu, *Prosodic profiles of the mandarin speech conveying ironic compliment*, 01 2021.
20. H. Loevenbruck, M. Jannet, M. D'Imperio, M. Spini, M. Champagne-Lavau, *Prosodic cues of sarcastic speech in french: Slower, higher, wider*, 08 2013.
21. A. Nauke, A. Braun, *The production and perception of irony in short context-free utterances*, (2023).
22. S. V. Rybin, *Sintez rechi [speech synthesis]*, Saint Petersburg, Russia, 2014.
23. P. Skrelin, U. Kochetkova, V. Evdokimova, D. Novoselova, *Can we detect irony in speech using phonetic characteristics only? – looking for a methodology of analysis*, Speech and Computer (Cham) (Alexey Karpov and Rodmonga Potapova, eds.), Springer International Publishing, 2020, pp. 544–553.

St. Petersburg State University,
St.Petersburg, Russia

E-mail: u.kochetkova@spbu.ru

E-mail: p.skrelin@spbu.ru

E-mail: v.evdokimova@spbu.ru

E-mail: st065112@student.spbu.ru

Поступило 6 сентября 2023 г.