

UDC 81.342.1

DOI <https://doi.org/10.24919/2308-4863.1/32.214486>

Nataliia ZADORIZHNA,
orcid.org/0000-0001-7727-8447

Ph. D. in Philology,
Assistant Professor at the Department of English Philology and Intercultural Communication
The Institute of Philology of the Taras Shevchenko National University of Kyiv
(Kyiv, Ukraine) zadorizhna@gmail.com

PROSODIC VARIATION OF SPEECH IN A VICTORIAN NOVEL

This paper is devoted to characterization of prosodic features of the narrator's speech and direct speech fragments in a Victorian novel by Jane Austen «Pride and Prejudice» represented by a professional actor in audio version. The aim of the research is to conduct a contrastive analysis of the narrator's speech prosody and personified speech prosody and reveal the way they change due to communicative intent. This phonetic research is based on auditory and acoustic analysis and investigates speech melody (different configurations of pitch curves and types of terminal tones), FO range and intensity variability, tempo (mean syllable duration), pausation, rhythm and types of voice quality. The research was conducted on the basis of different types of utterances (exclamations, descriptions, evaluative and emotional statements, questions and answers) which differ according to the degree of prosodic prominence and implicitness or explicitness of emotion expression.

The results of the research showed that the narrator's speech is less emotional than direct speech fragments of Austen's characters. The latter demonstrate a wide variety of prosodic means of expressiveness to manifest different emotions. In the present paper the expression of positive and negative emotions in personified speech by means of prosody were studied. To express different types of binary emotions in stylized speech of characters the actor uses a particular combination of prosodic components which help to create voice portraits of characters to be easily distinguished by listeners.

Prosodic means perform a number of important functions. They divide the text into sense groups according to the demands of communicative situation. Moreover, the emotive aspect of utterances can be defined and expressed only by means of prosody. In the process of comparing and contrasting prosodic characteristics of narrator's speech and direct speech realizations of characters we arrived at the conclusion that each type of speech has some inherent prosodic features which seem to be highly idiosyncratic.

Key words: *prosodic components, prosodic variation, prosodic functions, narrator's speech, personified speech.*

Наталія ЗАДОРІЖНА,
orcid.org/0000-0001-7727-8447

кандидат філологічних наук,
асистент кафедри англійської філології та міжкультурної комунікації
Інституту філології
Київського національного університету імені Тараса Шевченка
(Київ, Україна) zadorizhna@gmail.com

ПРОСОДИЧНА ВАРІАТИВНІСТЬ МОВЛЕННЯ У ВІКТОРІАНСЬКОМУ РОМАНІ

Стаття присвячена характеристиці просодичних особливостей мовлення оповідача та фрагментів прямої мови у вікторіанському романі Джейн Остін «Гордість та упередження», що репрезентовані професійною акторкою в аудіоверсії. Метою дослідження є проведення контрастивного аналізу мовлення оповідача та просодії персоніфікованого мовлення, а також визначення того, яким чином вони змінюються відповідно до комунікативної інтенції. Фонетичне дослідження базується на аудитивному й акустичному аналізі, охоплює мелодику (різні конфігурації мелодичних рисунків і типи термінальних тонів), варіативність діапазону ЧОТ та інтенсивності, темп (середньоскладова тривалість мовлення), паузацію, ритм та типи тембру. Дослідження здійснювалось на основі різних типів висловлювань (вигуки, описи, оціночні й емоційні твердження, запитання та відповіді), що відрізняються ступенем просодичного виділення й імпліцитністю або експліцитністю вираження емоцій.

Результати дослідження показали, що мовлення оповідача менш емоційне, ніж фрагменти прямої мови героїв Джейн Остін. Останні демонструють широкую варіативність просодичних засобів вираження для демонстрації різних емоцій.

У статті досліджується вираження позитивних та негативних емоцій за допомогою просодії в персоніфікованому мовленні. Для вираження різних типів бінарних емоцій у стилізованому мовленні персонажів актор послуговується певною комбінаторикою просодичних компонентів, що допомагають створити голосові портрети героїв, які легко розрізняються слухачами.

Просодичні засоби виконують низку важливих функцій. Вони розділяють текст на смислові групи відповідно до потреб комунікативної ситуації. Крім цього, емотивний аспект висловлювань може бути визначеним та вираженим лише за допомогою просодії. У процесі порівняння та протиставлення просодичних характеристик мовлення оповідача та реалізацій прямої мови героїв ми дійшли висновку, що кожен вид мовлення має специфічні просодичні особливості, які є дуже своєрідними.

Ключові слова: просодичні компоненти, просодична варіативність, просодичні функції, мовлення оповідача, персоналізоване мовлення.

Recent years have seen a growing interest in prosodic variation detection on material of audiobooks of different literary genres. Scientists study speech variability in audiobooks from different perspectives. Some works have been done on expressive French audiobooks corpus read by a single amateur speaker covering various types of literary genres (Sini et al., 2018: 4294–4295.). Other works (Panayotov et al., 2015) are based on multi-speaker English corpus. Some corpora are multilingual (Stan et al., 2013) while others are based on the corpus recorded by a professional speaker. Long and coherent speech data is very interesting as it gives the possibility of studying voice expressiveness under different situations. Audiobooks are a good example of such data and are valuable for prosody modeling, especially in the field of storytelling. When reading the books, the speaker has an expressive speaking style and uses personification to make the characters distinguishable. These performances reflect the high diversity of vocal personifications during direct speech.

Corpus analysis of read speech is conducted by researchers to improve text-to-speech system able to read books of different literary genres. Encoding direct and indirect speech information and its emotional content on the level of prosody is important for developing expressive synthesized speech of good audio quality (Sini, 2018: 4295).

Prosodic features act as an alternative communicative channel that can be interpreted by listeners in addition to verbal means. In audio narration they perform a number of important functions. Crystal (Crystal, 2005: 249) distinguishes emotional, grammatical, informational, textual, psychological and indexical function. Emotional function expresses attitudinal meaning and other semantic nuances. Grammatical function helps to identify grammatical structure of speech, performing a role similar to punctuation. Informational function helps to draw attention to the new, important information in the utterance. It becomes obvious from the most prominent tone in a contour. Textual function helps larger units of meaning than the sentence to contrast and cohere. Changes in prosody reflect the progress of the action. Psychological function helps to organize speech into units that are easier to perceive. Indexical function is an important marker of personal or social identity.

The present paper is devoted to the comparative and contrastive analysis of narrator's speech and direct speech fragments in audiobooks. The narrator's speech is influenced by the attitude he wishes to convey (affective) and by the image he is trying to present (Nolan, 1983: 203). Prosodic variation hinges on the interaction between a multifaceted communicative intent and the mechanism of linguistic expression.

The database for the experiment was created out of the audiobook by Jane Austen "Pride and Prejudice". The analyzed novel contains two types of voice: the characters' voices and the narrator's voice. Utterances examined in the paper are exclamations, descriptions, evaluative and emotional statements, questions and answers. This phonetic research is based on auditory and acoustic analysis. It is confined to the dimension of speech melody, tempo, pausation, rhythm and voice quality. Auditory judgments were compared with acoustic correlates of the prosodic parameters. Acoustic analysis was carried out with the help of Speech Analyzer 1,5 speech analysis tool. The analyzed recordings were recorded by a single female professional actor. Mono-speaker corpus helps to study the strategy of one speaker over the entire book. The number of analyzed utterances is 290. Total duration of the whole corpus is 6 hours.

The analysis of speech samples, especially of stylized ones, illustrates the enormous plasticity of narrators' vocal tracts and a huge potential for variation. The actor's voice has its intrinsic and extrinsic characteristics. The degree of emotionality correlates with direct/indirect fragments and also with characters. In the given research we focused on positive and negative emotions expressed by characters. The most frequent positive emotions expressed by characters in direct speech segments are the following: playfulness, surprise, curiosity, amiability, love, adoration, admiration, passion, softness, hope, satisfaction, triumph, tenderness, praise, joy, gratitude. The most frequent negative emotions are: offence, sarcasm, reproach, mockery, rudeness, guile, disappointment, dissatisfaction, superiority, irritation, disgust, despair, envy, embarrassment, anger, fear, contempt, criticism. Binary emotion classification and speech segmentation of acted speech was done to conduct a contrastive analysis of expressive read

speech in terms of different speaking styles. Monospeaker read expressive speech corpus showed that personified speech is in stark contrast with the natural narrator's voice. The expressive strategy implemented by the actor is very specific, conveying different shades of meaning through pitch modulation and accentuation of focus words.

Prosodic pattern example (Figure 1) illustrating direct and indirect speech shows a striking difference between personified and natural speech of the narrator. Indirect speech fragment is marked by cursors. FO range is broad (21,4 semitones), while indirect speech pitch range is narrow (4,8 semitones). The same thing can be observed in terms of intensity range, which is broader in direct speech (16,9 dB) in comparison with reported speech (7,5 dB). Acting speech is

characterized by emphatic prosody which is achieved by elongation of focus words and simultaneous accentuation by means of pitch and intensity modulations together with spiky rhythm. The analyzed utterance expressing reproach voiced in a raised voice acquires more emotional coloring due to the use of rise-fall tone and contrasts with narrator's speech fragment produced with the help of unemphatic low fall tone. The temporal properties of direct and indirect speech also differ. Mean syllable duration, given in bold shows that the speech of the character is characterized by a slower tempo than the narrator's speech due to a higher degree of expressiveness. An emphatic pause indicates a transition from indirect speech to direct and adds emotional coloring to the utterance.

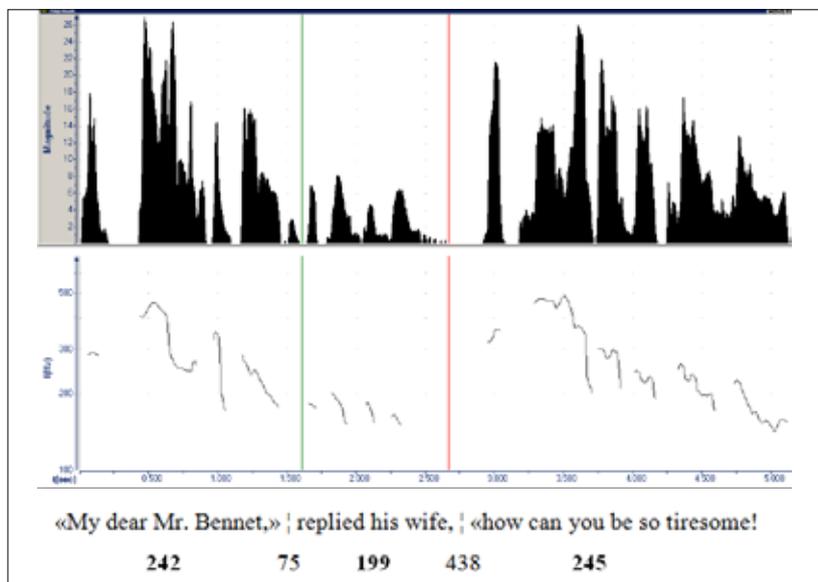


Fig. 1

The important prosodic effects are conveyed by the use and combination of different tones which express a wide range of meaning. The most frequent tones in the narrator's speech are low fall (65%) and high fall (25%). The first is used in neutral, unemotional utterances and the second one is emotionally colored. Direct speech differs from indirect speech and demonstrates great variability in terms of tones. Apart from low fall (8%) and high fall (21%), narrators use a wide inventory of simple and complex tones. Mid fall (6%) is used to express a routine, uncommitted comment. It sounds detached and unexcited. Level tone (8%) can express boredom, sarcasm or irony or a neutral emotion. Rise-fall (42%) implies a strong emotional involvement. Depending on the context, the attitude might be delighted, challenging, or complacent. This emphatic

tone can express an urgent request, a categorical denial, displeasure etc. It can also be used to sound persuasive or to add emphasis. The falling-rising tone (18%) is an emphatic variety of a rising tone and is used when something is implied but not expressed by words, for instance, caution, diffidence, indignation, surprise etc.

Conveying emotions in speech is realized through complex prosody. The expression of negative emotions is characterized by a falling tone. Expressing positive emotions often requires a rising tone or accidental rise. The latter defines the logical center of the utterance, making it the most conspicuous (Levitsky, 2005: 41). Researchers point out that surprise is characterized by a sudden pitch shifting upwards and phonetic elongation, accentuation on the first syllable of a focus word (Sini, 2018: 4293).

It is interesting to observe the use of a rising tone in direct speech statements which belong to women characters. Since rising pitch patterns can also mark questions, there is a frequent perception that uptalk indicates a questioning nature, and that uptalkers are therefore insecure or lacking confidence (Warren, 2016). Linguistic studies (Crystal, 2005: 249) also point to the fact that the tone is preferred by the less powerful members of society. It may act as unconscious expression of uncertainty, even subservience and deference. Another phonetic study (Britain, Newman, 1992: 2) found that the tone is used in narratives in those parts of the story where the speaker wishes to heighten the interest of the listener.

Variability of FO range is a distinguishing feature of personified speech. A wide range is used to express joy, delight, surprise, some implication; it may also express indignation or anger. Intensity range varies from narrow to wide and depends on emotions expressed.

Pitch changes cannot be separated from loudness, tempo and other prosodic parameters in creating the effect of accentuation or certain vocal image. Tempo as well as loudness is a feature that can be varied from time to time. Differences in tempo are used to help the listener to differentiate the narrator's speech and the characters' speech. According to the obtained results narrator's speech is characterized by the prevalence of medium speed and a lower loudness. As for direct speech realizations of different characters, tempo and loudness varies from fast (*allegro*) to slow (*lento*) according to the emotional state of the speaker and the attitude conveyed, performing emotional and attitudinal functions. Fast tempo may be associated with anger or excitement etc.; slower than normal tempo may be associated with doubt, blame, accusation etc. Varying tempo is an important communicative feature that can convey different kinds of emotions and meaning. For example, speeding up can denote excitement or impatience, while slowing down can be connected with emphasis or thoughtfulness. The change of tempo may take place suddenly (*accelerando*) or gradually (*rallentando*).

The analyzed acted speech is characterized by an extensive use of pauses. They are used intentionally either to create an effect or to signal a major syntactic or semantic boundary. There may be silent pauses and filled ones, which are filled by interjections, coughing, prolonging sounds, etc. Dramatic, emphatic and hesitation pauses are the most conspicuous.

Another feature of narrators' speech is rhythmicality. Features of pitch, loudness, tempo

and pauses combine to produce a specific kind of rhythm. Narrator's speech is mostly perceived as rhythmical, while direct speech can be clipped or drawling. Shorter bursts of voicing contrast with the longer ones in alternating stretches of speech. Nolan (Nolan, 1983: 129) points out that clipped speech consists of relatively short (principally vocalic) stretches and relatively long intervening voiceless (principally consonantal) stretches, whilst drawled speech is perceived when the opposite relationship exists.

Speech rhythm can be described as rhythmic and arrhythmic, spiky and glissando, staccato and legato (Crystal, 1969: 163–164). Spiky rhythm utilizes sharp and rapid pitch jumps between syllables, glissando utilizes smooth and usually fairly slow glides. The contrast between strongly and weakly prominent syllables is heightened in staccato or lessened in legato by increasing or decreasing the loudness and duration differentials between them. Narrators make use of various speech rhythm models especially in reproducing characters' speech.

Together with verbal means of expressiveness different types of vocal timbre play an important role in the audio book. The author's text is mostly represented by a modal voice, while direct speech, which mirrors different characters, is tinged with emotions and various voice qualities. The voice can be cracked, dry, flat, hollow, husky, melodious, raucous, rough, raised, thin, tiny, resonant, velvety, breathy, whispery, rich, deep, flat etc. (Levitsky, 2005: 47). The actor personifies different characters of the book by changing her voice or her way of speaking. To personify a gloomy man, the speaker uses a low pitch, low energy and devoiced voice (Sini 4292).

The most common types of phonation used in acted speech are the following: whisper (for secrecy/conspiracy/confidentiality), breathiness (for deep emotion or desire/surprise/pleasure), huskiness (for unimportance or disparagement), nasality (for anxiety), extra lip-rounding (for intimacy) (Crystal, 2005: 249). Lipping is associated with effeminacy. The use of different timbres together with other prosodic means adds variability to acted speech and allows creating characters' voice portraits that can be easily recognized by listeners.

Thus, the manner of narration used in Victorian novels by Jane Austen is characterized by prosodic variation which can be observed in narrator's speech and especially in direct speech realizations. The latter act as voice portraits of Jane Austen's characters with their idiosyncrasies expressed by means of a specific complex of prosodic features.

BIBLIOGRAPHY

1. Britain D., Newman J. High rising terminals in New Zealand English. *Journal of the International Phonetic Association*. 1992. № 22. P. 1–11.
2. Crystal D. The Cambridge encyclopedia of the English language. Cambridge : Cambridge University Press, 2005. 499 p.
3. Crystal D. Prosodic Systems and Intonation in English. London : Cambridge University Press, 1969. 390 p.
4. Levitsky A., Haraschuk L. An advanced course of the theory of English Phonetics. Vinnytsia : Foliant, 2005. 71 p.
5. Nolan F. J. D. The Phonetic Bases of Speaker Recognition. Cambridge : Cambridge University Press, 1983. 221 p.
6. Librispeech: basn ASR corpus based on public domain audio books / V. Panayotov et al. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015. P. 5206–5210.
7. SynPaFlexCorpus: An Expressive French Audiobooks Corpus Dedicated to Expressive Speech Synthesis / A. Sini et al. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*. 2018. Miyazaki, Japan.
8. TUNDRA: a multilingual corpus of found data for TTS research created with light supervision / A. Stan et al. *14th Annual Conference of the International Speech Communication Association (Interspeech)*. 2013. Lyon, France. P. 2331–2335.
9. Warren P. Uptalk: The phenomenon of rising intonation. *Journal of International Phonetic Association*. 2016. Cambridge : Cambridge University Press. № 47 (3). P. 369–374,

REFERENCES

1. Britain D. and Newman J. High rising terminals in New Zealand English // *Journal of the International Phonetic Association* 22. 1992. P. 1–11.
2. Crystal D. The Cambridge encyclopedia of the English language. Cambridge: Cambridge University Press, 2005. 499 p.
3. Crystal D. Prosodic Systems and Intonation in English. London: Cambridge University Press, 1969. 390 p.
4. Levitsky A. E., Haraschuk L. A. An advanced course of the theory of English Phonetics. Vinnytsia: Foliant, 2005. 71 p.
5. Nolan F. J. D. The Phonetic Bases of Speaker Recognition. Cambridge: Cambridge University Press, 1983. 221 p.
6. Panayotov, V., Chen, G., Povey, D., and Khudanpur, S. (2015). Librispeech: basn ASR corpus based on public domain audio books. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015)*, pages 5206–5210.
7. Sini, A., Lolive, D., Vidal, G., Tahon, M., Delais-Roussarie, E. (2018) SynPaFlexCorpus: An Expressive French Audiobooks Corpus Dedicated to Expressive Speech Synthesis. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan.
8. Stan, A., Watts, O., Mamiya, Y., Giurgiu, M., Clark, R. A., Yamagishi, J., and King, S. (2013). TUNDRA: a multilingual corpus of found data for TTS research created with light supervision. In *14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pages 2331–2335, Lyon, France.
9. Warren, P., (2016). Uptalk: The phenomenon of rising intonation. In *Journal of International Phonetic Association*, 47 (3), pages 369–374, Cambridge: Cambridge University Press.
10. Levitsky A. E., Haraschuk L. A. An advanced course of the theory of English Phonetics. Vinnytsia: Foliant, 2005. 71 p.