PhR 2019
Philological Readings

ASSOCIATIVE RECOGNITION AND CODING OF INTONATION UNITS JOINTS

Irina A. Gorbachyova (a)*
*Corresponding author

(a) Department of English Philology and Methods of Teaching English, Orenburg State University, 13, Pobedy Avenue, Orenburg, Russia, irgorby2017@gmail.com

Abstract

The properties of sounding speech do not lend themselves to speculative analysis. The perfection of methods for studying the structure of a speech signal contributes to the expansion of the possibilities of its study both in terms of speech production and in terms of perception. A simplified model of signal coding and decoding was replaced by a modern model of the speech flow, reflecting such elements of communication as phonological categorization, linguistic interpretation, the motivation of the utterance and communicative intention. The last aspect is of particular interest, since in the process of generating a speech signal it determines the choice of the neurobehavioral program. The mechanism of this program directly depends on the study of the prosodic organization of sounding speech. The analysis of the acoustic properties of intonation units would be much simpler if these properties did not depend on the phonetic context. The acoustic nature of the co-articulatory regularities at the intonation level sheds light on how the planning of the supra-segmental syntagmic model takes place.

1. Introduction

Issues of contextual variability of linguistic phenomena attracted the attention of scientists in all periods of linguistic development. In 1895, de Courtenay is presenting his theory of Phonetic Alterations claims the distorting influence of context on the intention of making a sound of a certain quality. Later Bergson (1998) in 1907 raises the question of attempts to discern the "tensive phase" and "the phase of the impulse" which occur at the moments immediately preceding the act of speech production. Introducing the definition of psycho-systematics as a new direction in science and methods of positional linguistics, Guillaume (1973) demonstrates the extreme abundance of combinatorics and brings us to understanding every linguistic phenomenon as being represented from the point of view of its horizontal deployment that is in the form in which our thinking comes about. One of the central issues in contemporary prosody investigations is that of coarticulation.

The huge array of articulatory adaptation that occurs as a result of the inter effecting of phonetic structures constitutes the coarticulation in its broadest sense (Gorbachyova & Filyasova, 2007; Volenec, 2015). An extraordinary breakthrough in the field of information technology in the field of constructing signal models and parameters of models of transmission systems and voice messaging (Kropotov & Paramonov, 2015; Ramasubmanian & Doddala, 2015) allows the linguist researcher to obtain objective data on the study of adaptive processes of prosodic units. The most successful heuristic approaches used in synthesis systems for foreign languages were tested; the accessibility of the technological base necessary for creating annotated speech corps and using statistical methods was evaluated (Kreychy, Krivnova, & Stupina, 2016). The process of studying the phonetic and phonemic correlates of prosodic borders also is going on (Knyazev, Krivnova, & Moiseyeva, 2016).

2. Problem Statement

In modern linguistics the contextual prosodic assimilation phenomena are becoming the issue of particular interest among linguists. Its study started with investigating intonation units (IU), such as question-reply unity, unison dialogues, intonation up-takings, intonation elements interactions in dialogical unity (Olshevskaya, 1982) and others.

Later, after recognizing the discrete nature of the intonation structures, determining the criteria for identifying intonation units (IU) (meaningful and formal) and developing fundamental approaches to their classifications, sufficiently logical intonation systems were created (O'Connor & Arnold, 1978). Information on the qualitative and quantitative features of intonation systems was summarized in the works of domestic and foreign linguists (Dubovskij, Dokuto, & Pereyashkina, 2018) with the aim of their further studying in applied researches. The latter made it possible to proceed to the study of not only the distinctive IU features, but also their integral properties.

With the invention of recognition systems in which each sequence of nodes (words) generated by the network was set as a sequence of allophonic spectrum patterns (Kachkovskaya & Skrelin, 2019) it became possible to transform the prosodic parameterized transcription into a digitized speech signal. The development of methods for analyzing the speech signal helped to significantly expand the possibilities of considering and studying the structure of the speech signal both in terms of speech production and in

terms of perception. (Tanenbaum & Bos, 2019). It is shown that linear prediction can be used as a tool for processing acoustic speech signals for almost all applications. A number of properties of the linear prediction method (Backstrom, 2017), its mathematical foundations, including parameter transformations, recursive methods for checking the stability conditions of the structure of speech synthesizers, estimation of the fundamental and formant frequency spectral analysis are considered (Kocharov, Kachkovskaya, & Skrelin, 2019). This simplified the task of studying combinatorial phenomena in intonation speech realization application area. Intonation, which is closely connected with the semantic level of an utterance, remains today one of the weakest aspects in understanding how the process of connecting the language with its mental system of concepts and intentions takes place.

3. Research Questions

The study of IU contextual influences is complicated by the fact that not a single distinctive feature is insignificant in its composition. On the contrary, it finds its place in the holistic formation, correlating with the meaning of this or that level of abstraction and performs its basic functions in combination with its other features. For that reason it's necessary to know the ways of eliminating all redundant information in instrumental analysis of intonation variability. The idea of a language as of the theoretical construction of linguists whereas of speech as an observable real phenomenon arose in connection with the representation of a language as a model. A language is considered to be the area of constructs, and speech as the area of natural objects, observable and real. It implies the necessity of studying the following questions:

Is it possible that one and the same IU undergoes similar changes in all existing combinatorial positions? How can such changes influence the perception abilities?

Which syntagmatic intonation units can be selected for the analysis, and which type of diagrams is more illustrative in this regard?

However, in the analysis of the adaptive capabilities of the prosodic level units, one cannot do without the so-called "cross section" of the speech flow, i.e. of paradigmatic units, for example, Tone Groups.

Probably the pace, with which the choice of a language unit form corresponding to speech intention is made, depends on the quality of the whole system and on concrete positions held in it by various units.

What is the range of the unit variability necessary for providing the desired acoustic result?

In the process of finding the answers to these questions, a number of additional related problems are solved: the types of the acoustic correlates of IU adaptation data; the existence of a certain hierarchy by the importance of intonation techniques participating in the assimilation process.

4. Purpose of the Study

The concept of speech communication models is changing. Intonation, which is closely connected with the semantic level of utterance, remains today one of the weakest links in understanding how the phonetic form connects language with sensormotoric systems of perception and articulation

(Yevgraphova, Skrelin, & Shatalova, 2015). This simplified the task of studying. The ongoing instrumental analysis of the contextual influences of the congruent intonation units should bring us closer to deciding what prosodic characteristics (of an integral property) should be incorporated in speech producing program for pronouncing the intended intonation pattern. Since intonation is the closest of all other language components to the semantic level and carries a huge semantic load, it is necessary to foresee the nature of the change in the joint element, and, therefore, to determine the entire intonation pattern as a form that conveys a certain meaning.

5.   Research Methods

The material was processed in computer program WA (Wave Assistant) for processing speech signals. Its general characteristics are: a software package containing four analyser programs running through a sound card; oscilloscope is a simulator of a two-channel oscillograph; Analyzer - Spectrum analyser software; Audio Meter - a program that allows you to measure the input voltage levels on the sound card; Signal Generator - the program generates signals of various shapes in the frequency range 20Hz - 20 kHz, and also allows you to generate white noise.

The subject of the experiment is the acoustic description of the intonation model of a single-sense group utterance in dynamics.

The research material composition is aimed to eliminate unnecessary factors of variation, such as individual voice qualities, modifications associated with the use of various communicative types and excessive emotional colourings of statements.

The material is constituted by statements of one communicative type, containing identical lexical and grammatical composition.

For the analysis of contrasting adaptive conditions, the intonation units (IU) of O'Connor and Arnold (1978) Intonation system were used as the most developed from the point of view of practical application. The material is constructed in such a way that the main information sections of the contour, represented by terminal tones (nuclei), are located on the junction boundaries (e.g. see Table 01)

Table 01. Contrastive junction positions of terminal tones

ꝪHigh Drop + ³Low \Drop (HD+/LB)	ꝪHigh \Drop + ³High \Drop (HD+\HD)
╎Leaving for **Moscow** | /aren't you‖	╎Leaving for **Moscow** | \aren't you‖

6.   Findings

6.1. Brief survey of objective instrumental analysis data

The study of combinatorial interactions has been carried out over several years. It is complicated by the lack of the necessary number of English-speaking experiment participants and informants at our laboratories. However, for recording the material and conducting a perceptual experiment, the author of the study was given the opportunity to work with foreign students at the Department of Phonetics, St. Petersburg University. Before we start describing the results of the auditory experiment, and in order to understand which research problems should be solved, it is necessary to present the existing data of

instrumental research obtained and presented in a number of works on this topic. It certainly still requires some amendments.

Modifications of all components of the intonation contour were investigated, but the most significant among them were observed in the implementations of the most informative ones - fundamental frequency and pace. The instrumental analysis of IU boundaries explains the degree of adaptation, the characteristic features of the acoustic correlates of intonation adaptation, and the hierarchy of phonetic means involved in generating signal variability caused by a contrasting combinatorial context (Gorbachyova, 2014). The established rules and techniques of adaptation of IU can be safely attributed to: the predominance of the regressive type of adaptation of both smaller and larger prosodic units.

The most illustrative examples are given below as they demonstrate the most extreme fluctuations of tones at the junctions (Fig. 01). The capabilities of the analyzing electronic devices make it possible to cut out any part of the speech signal for its instrumental processing (as given below Fig. 02).
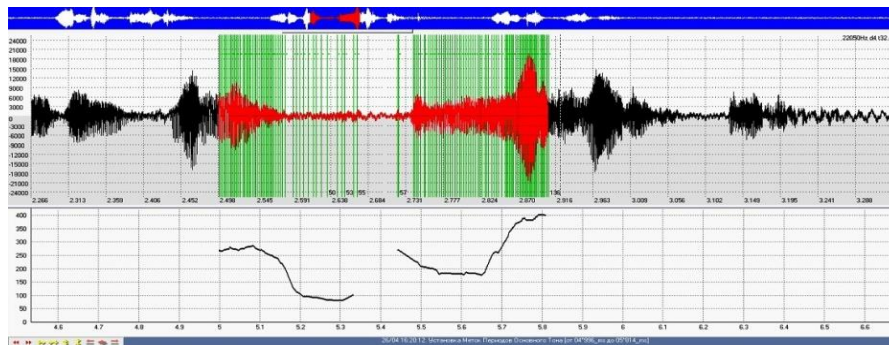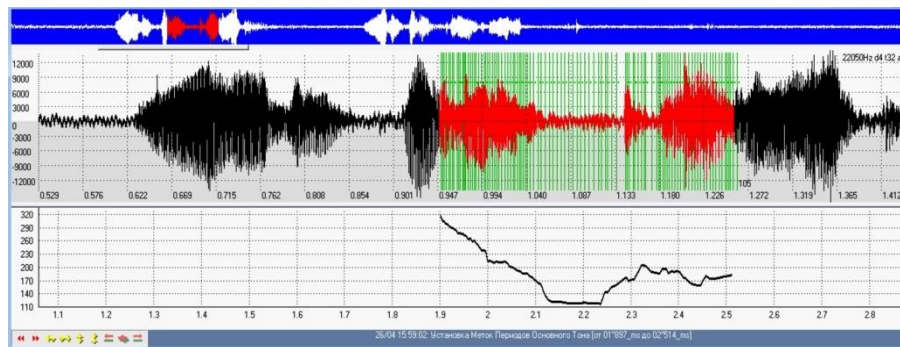

Figure 01. "Moscow" (HD + **HD**) contextual adaptation

Figure 02. "Moscow" (HD +**LD**) contextual adaptation.

Table 02. The variation of basic frequency parameters of the "interjacent syllable" /sk≅Y/ on the junction (Nucleus "Mo**scow**")

BF Parameters Value of / sk≅Y/	HD + /**LB**	HD + **HD**
Average BF	296 Hz	200 Hz
Maximum BF	375 Hz	219 Hz
Minimum BF	173 Hz	185 Hz
Range	202 Hz	34 Hz
Duration	0,22 sec.	0,21 sec.
Maximum range of fall	187 Hz	17 Hz
Maximum range of rise	14 Hz	34 Hz

Numerous studies of combinatorial conditions of smaller (e.g. Pre-heads, Heads, Nuclei) and larger intonation units (e.g. Tone Groups) have shown that contextual modifications take place in all the cases and affect all component parameters (Table 02). However, such parameters as the range, the slope of the fall, rise of the tone and the speed of its implementation can still be considered the most "vulnerable" (Fig. 03 and Fig. 04).
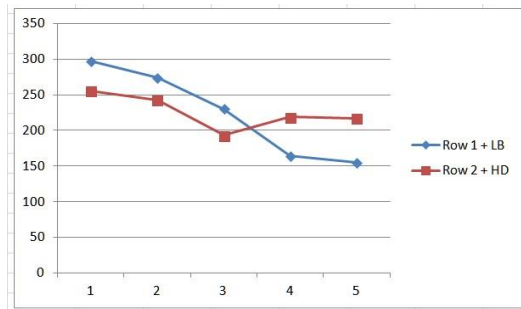
Figure 03. BF of / **sk≅Y** / ("Mo**scow**") in two combinatorial contexts. Informant 1
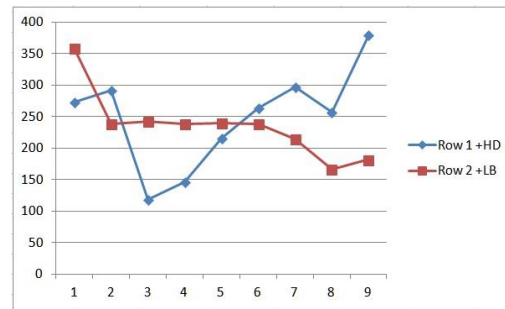
Figure 04. BF of / **sk≅Y/** ("Mo**scow**") in two combinatorial contexts. Informant 2

Examples of this kind were chosen intentionally, since they are directly related to the topic of further discussion, namely, the question of perceptual experiment.

6.2. The impact of IU contextual modifications on its perception character

The objective analysis of intonation units' junction phenomena demonstrates that they do not undergo the same modifications in any combinatorial position. The limits of variability of certain parameters of IU differ. Such observations allow the researcher to understand what way he should go in carrying out auditory experiment. It is necessary to answer the question of how contextual modifications of tones can affect their perception if they are cut out of the context. Will there be a sharp change in the meaning of the utterance in comparison with that implied in case the unit is reproduced in a contextual surrounding?

An auditory analysis, during which the informants were asked to correlate the cut-out initial segments of the intonation contour with the presented final sections of the IU, showed that in most cases the modifications at the junction occur without violating the sound and semantic identity. When the IU combinatorial conditions which constituted a certain utterance were changed, we could observe only some slight modifications of its emotional tinges while the basic emotive meaning preserved. Informants' answers expressed the indications pointing out that the phrase is incomplete: *sounds unfinished, going to finish the sentence, is about to go and explain the truth, sounds as if a sentence will follow on afterwards, sounds as if it would be followed by a normal question.*

The identification of the communicative type of truncated statements was also carried out (e.g. \HD+) (Table 03).

Table 03. High Drop + seven types of nuclei at the juncture

IU	Combinatorial position	Identification %
³**High \Drop +**	Low /Bounce	88
	Low \Drop	90
	High /Bounce	98
	High \Drop	**22**
	Switch ᵛ Back	31
	Jack ^Knife	26
	\High /Dive	28

The wide range of variation of terminal tones in most cases does not obscure the correct identification of the communicative type and the emotive meaning of the utterance. However, there are some exceptions, e.g. the junction of terminal tones (\HD + \HD) caused such modifications in the preceding tone, that most informants assigned it the emotive meanings of the complex tone Switch Back (Table 01.): *contradicting what someone has said, a little bit patronizing, as if she is correcting someone etc.*

It means that the range of variability of IU in certain combinatorial positions can cause its going beyond the planned value and lead to a semasiological alternation of a tone.

7.   Conclusion

The zone of intonation adaptation exists: the closer the contact between IUs is, the more intense the adaptation process reveals itself. Practically all the context-modified components of the intonation model can be considered as acoustic correlates of the above process. There is a hierarchy of phonetic means according to the degree of their significance in the formation of an adaptation section and the most important of them are such pitch parameters as: the range, the steepness of the fall or rise, the duration of the BF signal. The significant modifications of pitch occur in approximately 0.1 of a second and reach the greatest differentiation in 0.18 - 0.21 of a sec.

The perceptual approach used in investigating the contextual IU variations gives us certainty that the physical differences of the stimulus arising from its getting into different combinatorial conditions affect our perception. The instrumental research of the acoustic signal in its kinematics makes it possible to "observe" in what fraction of a second the contact between thought and language is set, that is, to be able to access mental operations, preceding the starting point of speech producing act. This again shows the ability of the brain in a fraction of a second to operate the degrees of freedom of the interaction of nerve cells when making this or that decision or mental operation. The objective methods for studying the contextual dependence of acoustic correlates of speech contribute to a more "correct" choice of properties for their use in heuristic algorithms. Intonology can be considered a field of fundamental and applied science, dealing with a set of theoretical and practical methods of research, analysis and synthesis of processes occurring in the human brain and responsible for the ability to express thoughts verbally.

References

Backstrom, T. (2017). *Speech Coding with Code-Excited Linear Prediction*. Switzerland: Springer International Publishing AG. Retrieved from http: //www.springer.com/gp/book/9783319502021

Bergson, A. (1998). *Kreativnaja jevoljucija [Creative Evolution]*. Moscow: KANON-Press, Kuchkovo pole.

de Courtenay, J. N. B. (1895). Versuch einer Theorie phonetischer Alternationen. In *Ein Kapital aus der Psychophonetic*. Strassburg: Truebner.

Dubovskij. J. A., Dokuto, B. B., & Pereyashkina, L. N. (2018). *Osnovy anglijskoj fonetiki. [Basics of English Phonetics]*. Moscow: Flinta.

Gorbachyova, I. A. (2014). Prikladnye aspekty issledovanija adaptivnyh processov edinic prosodicheskogo urovnja. [Applied Research Aspects of Prosodic Units Adaptation]. *The World of Scientific Discoveries, 7*(55), 349-363.

Gorbachyova, I. A., & Filyasova, J. A. (2007). Lingvisticheskaja kombinatorika i intonacionnyj kontekst. [Linguistic Combinatorics and Intonation Context]. *The Proceedings of the XXXVI International Conference, 17,* 3-15.

Guillaume, G. (1973). *Principes de Linguistique Teorique de Gustave Guillaume. Recueil de textes inedits prepare en collaboration sous la direction de Roch Valin*. Paris–Quebec: Klincksieck. Les PUL.

Kachkovskaya, T. V., & Skrelin, P. A. (2019). Analiz razgovornoj russkoj rechi. [The Analysis of Colloquial Russian Speech]. In *Trudy vosmogo mezhdistsiplinarnogo seminara* (pp. 46-54). Saint-Petersburg: Politehnika-print.

Knyazev, S. V., Krivnova, O. F., & Moiseyeva, Y.V. (2016). Issledovanija prosodicheskogo chlenenija zvuchashhego teksta na materiale russkogo jazyka. [The Studies of Prosodic Division of Oral text on the material of Russian Language]. *Vestnik Moskovskogo universiteta [Bulletin of Moscow University], 9*(4), 17-44.

Kocharov, D. A., Kachkovskaya, T. V., & Skrelin, P. A. (2019). Prosodic boundary detection using syntactic and acoustic information. *Computer Speech and Language*, *53*, 231-241. https://doi.org/10.1016/j.csl.2018.07.001

Kreychy, C. A. Krivnova, O. F., & Stupina, Y. A. (2016). Problema identifikacii diktora v uslovijah shepotnoj rechi. [The problem of speaker identification in whispering speech systems]. *Speech technologies*, *2*, 34–44.

Kropotov, J. A., & Paramonov, A. A. (2015). *Metody proektirovanija algoritmov obrabotki informacii telekommunikacionnyh sistem audioobmena*. [Methods of designing information processing algorithms for telecommunication audio exchange]. Moscow-Berlin: Direct-Media.

O'Connor, J. D., & Arnold, G. F. (1978). *Intonation of Colloquial English*. London: Longman.

Olshevskaya, R. V. (1982). Rol' prosodii v sozdanii jemocional'nogo dialoga-unisona [The Role of Prosody in the Creation of an Emotional Unison-Dialogue], In *Scientific SMPIL reports, 201* (pp. 155-167). Moscow: Publishing House of Moscow State Pedagogical University.

Ramasubmanian, V., & Doddala, H. (2015). *Ultra Low Bit-Rate Speech Coding*. New York: Springer Heidelberg Dordrecht London.

Tanenbaum, A., & Bos, H. (2019). *Sovremennye operacionnye sistemy. [Modern Operation Systems]*. Saint-Petersburg: Piter.

Volenec, V. (2015). Coarticulation. In *Phonetics* (pp. 47-86). Croatia: Nova Science Publishers, Inc.

Yevgraphova, K., Skrelin, P., & Shatalova, D. (2015). Cross-language Perception of Emotional Children's Speech in German and Russian. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.