# Structural Organization and Features of Speech and Music Perception in Language Functions' Implementation with Timbre's Perceptive Assessment

I. G. Andreeva[1], M..Dymnikowa[2,3,4], E. A. Ogorodnikova[2]

[1] PhD in biology, Laboratory of comparative physiology of sensory systems, I. M. Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, Toreza pr.44,194223 Saint-Petersburg.
[2] PhD in biology Laboratory of Psychophysiology of Speech, I. P. Pavlov Institute of Physiology, Russian Academy of Sciences, Makarova nab. 6, 199034 Saint-Petersburg.
[3] Association of Musical Psychologists and Psychotherapists, Akademika Anohina st. 38/3, 119602, Moscow.
[4] PhD in psychology, Laboratory of Music-therapy, Institute of Psychiatry and Neurology, al. Sobieskiego 1/9, 02957 Warsaw.
*Authors' Email: ig-andreeva@mail.ru, dmwl@bk.ru, elena-ogo@mail.ru.*

_____

**ABSTRACT -** *In cognitive science, the psychological and biological foundations of music are often considered in the context of comparison with the functions of speech and language. The article describes some aspects of the similarities and differences between music and speech in the implementation of language functions, as well as a typological study of its features according to Buhler, Jakobson, Halliday and Kiklewicz classifications. A general comparison of structural organization and main perceptual characteristics of music and speech signals is carried out additionally, with separate consideration of neurophysiological foundations of the auditory analysis of timbre characteristics of speech sources (speaker's voice) and music (instrument sounds). Some aspects of the practical application of auditory and musical training in the context of training and rehabilitation measures for sensory-cognitive dysfunctions of various genesis are explained, with the prospects of further research aimed at studying the influence of the timbre on the indices of spatial selectivity of musical perception in comparison with the characteristics of the spatial selectivity of speech hearing in the perception of complex acoustic scenes.*

**Keywords -** music and speech similarities with differences, music and speech timbre characteristics, language's function.
_____

## 1. INTRODUCTION. SIMILARITY AND DIFFERENCE IN LANGUAGE FUNCTIONS' IMPLEMENTATION

The psychological and biological foundations of music are often considered in the context of comparison with the functions of speech and language [77, 99] in cognitive science. The biological basis for the similarity of speech and music is their auditory modality and phylogenetic relationship with communicative signals and singing in various living organisms, as well as with the signal function of the auditory system and with acoustic orientation in space. Speech and music have formed during human evolution process into relatively independent functions, the main content of which, despite a certain functional similarity, has become oriented either to the transmission and storage of information, communicative interaction in society (speech) or to figuratively emotional influence and self-expression (music). Their formation and further development led to the creation of a symbolic language, which in speech reached its maximum level - the alphabet, vocabulary, grammar, oral and written forms of speech. Linguistic elements in music are more limited and are mainly used in the professional field. However, there is also its own alphabet (notes) with prevailing rules for writing and reading. The similarity is manifested in the need for additional training to master them. In general, the presence of categorical perception and linguistic forms of their display allowed speech and music to become instruments of special forms of thinking, with the separation of the categories of verbal and musical intelligence and the formation of specific types of memory [37, 61, 62]. The general typological features of linguistic functions in speech and music can be distinguished according to Buhler, Jakobson, Halliday and Kiklewicz classifications.

K. Buhler [33] understood the function of language as a direct impact of language on the participants of the speech act. He distinguished three basic language functions of: 1. Symbol - which belonged only to natural languages.2. Expression of sender's psychological and physical attributes. 3. Impression - of influence by the speech on massage's receiver. Music language has all functional attributes in the frame of Buhler classification. The impression function exists in music theatre duration. R. Jakobson [44, 45] defined functions of language during developing the K. Buhler model. 1. Referential - as the context content with a description of the situation, object or mental state. 2. Emotive - directed to the sender with sound changes that do not change the word meaning, but inform about sender's internal state. 3. Conative (convention) - directed to receiver and concern command or order message. Added functions: 4. Poetic - concentrated on the 'message communication for its own-self'.

5. Phatic (physical) - as interaction matter of communication, observed in greetings and freedom conversations, especially with unknown people, it allows to open and close communication with the external social environment. 6. Meta-linguistic (meta-lingual or reflexive) - concerning the language using for itself description or discussion. Music language has the only meta-linguistic function of Jakobson added functions. The regular temporal organization of most music art pieces is due if its' order matter but do not concern the poetic role which exists in speech-language. In this context of expanding the classification, the musical language performs only a metalanguage function, because although the regular (rhythmic) temporal organization of most works of musical art is important, it does not correspond to the distinguished poetic role that is realized in speech.

M. Halliday [33] studied the functions of the child's language. He distinguished seven language functions: 1. Instrumental - as a tool for needs' expression. 2. Regulative - as a tool that commands and regulates interpersonal contacts. 3. Interactive - as a basic principle (way) of interpersonal contacts. 4. Representing - as a tool for knowledge processing. 5. Imagination - as a tool for thoughts' and imaginations' transmission. 6. Heuristic - as enabling the learning. 7. Personal - as a tool of emotions' and feelings' expression. Music language has preserved instrumental, imagination, heuristic and personal functional attributes in the frame of Halliday classification. Music matter is not the instrumental content for the function of contact between people. Its knowledge has a pitch-rhythmic background as 'art' production which is free from the knowledge that belongs to speech-language carrier processing.

A. Kiklewicz [50] studied the function of language in the direction of relation with the external environment. He distinguished six language functions: 1. Nominative (representational, meta-linguistic) - consists in the fact that units of language, words and sentences appear as signs of objects and states of affairs. 2. Cognitive (encyclopedic) - the role of language in the processes of understanding the world by a human. The source of all information is presented in literal signs, words, idioms. The cognitive function of language is most obvious when learning foreign languages. The non-verbal information is not included in this function. 3. Perceptive (decoding) - the role of language as a means of understanding texts. 4. Communicative (social) - enables social relations, implementation of joint activities, division of roles in society. It has three following manifestations of varieties: 4.1. An interactive (interpersonal) - where the language is the most important tool for communicating information and influencing partners through information. 4.2. Manifest - which consists in the fact that by language meaning a person expresses, manifests his belonging or non-affiliation to certain social groups. 4.3. Fatal function - based on the fact that human creates and sustain social relations by language meaning, in the form of speaking. 5. Active - consists in using the language to implement the mental, intellectual or emotional human activity. It is implemented as an emotive function during the: 5.1. Thought function (the realization of thinking by speaking). 5.2. Internal speaking (as speaking without words), egocentric speech. 5.3. Text function - using the language to write texts. In the case of written texts, the recipient is virtual and sometimes does not exist. 6. Magical (creative) - based on human's belief that signs and words are natural parts of named objects and things, so using them we can influence on things or on world. Music language has only perceptive (decoding) function in the frame of Klikewicz classification. Another components of this classification specify the speech-linguistic unique role that belongs only for speech processing.

Music language, in case of mathematical matter of music structure, has elements of mathematical language that is not manifested in speech-language. It has especially different named labels organization of musical alphabet which has repeated signs for different pitch frequencies in music system, where sounded pitch frequencies are noted equal of its naming (as it is presented in language phonemes pronunciation). In speech-language some analogy might be observed in the tonal language where the same tone has different meanings when is spoken on different frequency level, where different frequency level is not labeled in symbolic signs as it exists in music language. So the pitch - frequency level and timbre type do not specify the content meaning of structural units of speech-language, how it is presented in music language. The case belongs to the background of language matter – which in speech is the human vocal explanation, while in music it is mostly presented by music instruments description. Singing processing is the only one with a direct relation between music and speech material. An additional difficulty of music language is located in its formal presentation by kinesthetic activity on musical instruments on the active explanation of written linguistic code text. Speech-language engage vocal reproduction of written text, while music requires a kinesthetic instrumental presentation. Therefore learning process for speech and music matters differs from the beginning of its ontogenesis. Including the music learning as specialized education without population access to understand the musical notation system, as opposed to spoken language, the music language becomes as the specialized technical language required additional education for its acquisition. Musical language has its own peculiarity associated with the inclusion of mathematical language elements that are not in speech. It also involves a special organization of the named marks of the musical alphabet, which involves the repetition of characters for different tone frequencies in the musical system. In speech, a similar analogy is observed only in tonal languages. These differences are based on the characteristics of the perceptual characteristics of speech and music signals, as well as the processes of their processing in the human auditory system and memory systems.

## 2. FEATURES OF SPEECH AND MUSIC SIGNALS' PERCEPTIVE CHARACTERISTICS

Basic differences in acoustic features of music and speech signals [113] concern frequency, temporal and spectral cues. Fundamental frequency in music is pitch component of melody, as categorized and notated with its possible precision, while in speech it is a pitch component of prosody, as non-categorized and non-notated, with its common variability. Temporal regularity in music is a rhythmic component of melody, as categorized and notated with its possible precision, while in speech it is a rhythmic component of prosody, as non-categorized and non-notated, with its common variability. The silent break in music is the articulation that sometimes notated, in speech it is the part of plosive phonemes with its implicit notion. Components of instrumental timbre are non-notated and non-categorized in music, while the components of sustained and plosive phonemes are notated and categorized in speech.

In the context of the characteristics of perceptual units in speech, the main emphasis is on its spectral characteristics and the formant structure of speech units. The recognition of phonemes and the selection of the corresponding letter equivalents of the language alphabet is based on the selection and auditory analysis of precisely these features. Further, taking into account the language rules, the word is recognized, i.e. semantic content of the message. At the same time, prosodic characteristics are evaluated - the speaker's voice, his emotional state, phrase intonation, etc. It is important to note that the entire range of these parameters in speech is limited by the properties of a person's sound-producing apparatus. From this point of view, the closest conditions in music correspond to singing [35]. At the same time, in music, the content (semantic, informational) side of structural elements is not leading. To the forefront is an expressive function with a certain melodic and rhythmic organization of sound material, the display means of which are significantly expanded concerning speech and can rely on the entire range of characteristics reproduced by various musical instruments. In most cases, this implies the use of an additional means of generating music through kinesthetic activity - playing a musical instrument following the musical "text" of the work. At the same time, unlike the spoken language, the musical language acts as a special technical language that requires special education for its development and understanding of musical notation as applied to the performing skills of playing an instrument, which is formed during long training and mastering complex motor programs.

Music and speech are similar in the acoustic background with features used in different ways, especially for information encoding, although they both use sound and important class of music - like singing - is produced by the same apparatus that produces speech [35]. In the frame of modularity as a central theoretical construct in cognitive science for study possible relationships between language and music, both music and language (included written signed languages) are organized temporally with the relevant structures unfolding in time, they reach the same perceptual system as frequency spectra, i.e. pitches. All languages consist of phonemes and literal notes also, in music it concerns the pitches and musical notes. Musical and linguistic knowledge required to produce the systems, that are generated from a finite set of sounds (notes or phonemes), carried out a larger variability of sounds organized into discrete categories, which facilitating representation and memory. Here both materials for auditory perception are the matter of categorical perception and product of intelligence, as linguistic and music types. Thus even though speech and music rely on the common basic properties of hearing, they have significant differences concerning perceptual attributes used, sensorimotor programs, distinguished functional categories, and semantic content of language forms. In the direction of auditory analysis processes, this refers to the perception of both frequency (pitch, timbre), and temporal (duration, rhythm, tempo), and amplitude (volume, stress) characteristics of sounds [113]. In this context, the main frequency in the music is the step component of the melody, categorized through a musical scale and musical scale, requiring great accuracy in display and reproduction. In the speech, it is a step component of prosody with great variability. Its informal categories are conditional: for high-altitude separation - male, female, children's voice; when changing the contour of frequency of the main tone of the voice - a question, a statement, an exclamation. A similar situation is observed concerning the pace. Temporal regularity in music is a rhythmic component of a melody, classified, marked and performed with the necessary accuracy. In the speech, it is the rhythmic component of prosody, unclassified, without a special designation, and showing a high degree of variability. However, in the perception of the syllabic structure of a word, the implementation of verbal and phrasal stress, this component plays an important role.

The timbre characteristics of speech and music sources deserve special consideration. They are the perceptual basis for identifying the speaker (in speech) and musical instruments (in music). In this case, the first recognition associated with the mother's voice is observed in children at the very early stages of ontogenesis, which is manifested in the reactions of his body (sucking, breathing, heart rate) [8, 24]. On this basis, several methods have been developed for nursing premature babies and stabilizing the psychosomatic state of newborns in clinical settings [13, 28, 87].

Timbre characteristics are also an essential component of the processes of auditory grouping, analysis of complex acoustic scenes and spatial selectivity of speech hearing (for example, "party effect") [14, 15, 22, 65]. These characteristics are more included in music in the artistic fabric of a musical work (score) and perception of harmony. Despite the fact that they are distinguished elements of the language neither in speech nor in music, timbre characteristics have a significant impact on the perception of other parameters of sound signals, in particular, volume and pitch, and are also represented in long-term memory as a mental library of tones [29, 39, 69, 81, 94]. At the same time, the neurophysiological mechanisms of auditory analysis of timbre characteristics in the perception of speech and musical signals are not well understood.

## 3. NEUROPHYSIOLOGICAL BASIS FOR TIMBRE PERCEPTION

A timbre is defined as an attribute that allows us to distinguish sounds with the same duration, volume, pitch and location in space, for example, to distinguish musical instruments that play the same note [4]. The timbre in its physical basis is characterized by a fine spectral-temporal structure of sound, which can be described by several parameters, i.e. is multidimensional. The structure of the perceptual space of the timbre is still the subject of discussion. Three dimensions are considered generally [107]. Studies of Caclin [16, 17] and Kong [52] found that timbre differences in specific measurements are most effectively explained by attack time, spectral centroid and spectral irregularity (attack time, spectral centroid and spectral irregularity). The time of the attack is the transition state during which the amplitudes of the harmonics of the fundamental frequency increase from the level of the threshold of perception (audibility) to the maximum. The shorter the attack time, the sharper the beginning of the sound is perceived. It should be noted, that the response to a sound with a short attack time, a more synchronous excitation pattern should be formed than for sounds with gentle fronts [38]. The spectral center of gravity determines the relationship between low-frequency and high-frequency harmonics, taking into account their amplitudes. The higher the amplitudes of the high-frequency components concerning the amplitudes of the lower harmonics, the higher the spectral centroid, as a result, the sound is perceived as sharper and brighter.

Spectral irregularity is related to the intensity of even harmonics relative to the intensity of odd harmonics. At lower amplitudes of even harmonics concerning odd ones, the sound is perceived as more deaf. Thus, the timbre of a sound is determined by a fine spatial-temporal structure, which can be encoded in the auditory system by various characteristics and organized tonotopically differently. The perception of speech and music as special, different in natural languages, apparently, begins with a classification - assigning incoming acoustic signals to categories of speech sounds or musical sounds. Singing occupies a special place - musical sounds made by a person. How does this distinction occur? In the frequency range corresponding to the fundamental frequency of the tone of the voice, the height or level of intensity cannot sign on which such a distinction is based. Consequently, the primary attribution of sound to a particular category should arise concerning the fine temporal structure of sound or timbre. What is the fundamental difference between the timbres of a musical instrument and a human voice? The basis of this difference lies in the nature of the occurrence of sound. According to Euler's voice production theory, the vocal folds emit an impulse that causes a transient response in the vocal tract to produce a damped wave.

A series of such pulses produces a series of damped waves, and their superposition forms a loud sound. The step frequency in such an analysis is the pulse repetition rate or the pitch frequency of the voice. In contrast to the notes of musical instruments, the frequency of the main tone of the voice constantly varies relative to a certain characteristic value for this speaker. This possibility of varying the frequency of the fundamental tone of the voice forms intonation, emotional content, i.e. extra linguistic information of the speech. Apparently, similarly formed musical phrases create the emotional content of the music. However, we clearly distinguish between a person's voice and the sound of a musical instrument. What is the role of timbre as a key feature? The multidimensionality of the tone does not allow a clear answer to this question. The complexity of the analysis of the timbre consists in the higher variability of the frequency of the main tone of the voice and an additional set of factors (speech breathing, size and variability of the position of articulatory organs) compared to musical instruments. For this reason, singing is rarely used to assess the state of absolute hearing, i.e. due to the 'difficulty of objectively determining the pitch of a song' [100].

A comparative analysis of the variability of the fundamental tone for the piano, as well as the synthesized and live voices, showed that the latter has the greatest variability [109]. In the conventional analysis of a speech (voice) signal, fixed values of the time window and its shift are applied, which does not allow taking into account changes in the frequency of the fundamental tone in such a window. Thus, the evaluation of the overtones that characterize the timbre, and the frequency of the fundamental tone are poorly separated. This analysis method was proposed that does not have a fixed time window and is based on time intervals in which the pitch frequency does not change [19]. The authors suggest in this way to clearly distinguish between changes in the frequency of the fundamental tone and the timbre characteristics of the sound.

Such a method of analysis may allow a clearer quantitative assessment of the tonal characteristics of the voice. Despite the difficulties in an objective assessment of the parameters of the timbre, as already noted, it determines our perception and the ability to distinguish musical instruments, speech sounds and ambient sounds. In particular, the sound made when tapping on various objects often makes it possible to characterize the materials of which these objects consist, their structure and volume. Thus, the assessment of timbre by the auditory system is not a species-specific human task; it arose from the need to analyze natural sounds, including communicative ones. This allows us to involve in the study of the biological foundations of timbre perception, along with methods for studying human hearing, neurophysiological work on recording the activity of individual brain cells of animals [81, 105].

At the level of the auditory nerve, the temporal and spectral characteristics that determine the timbre of the sound source are represented in the population activity of the fibers of the auditory nerve. A detailed analysis of the fine spectral – temporal structure occurs in some areas of the auditory cortex [49, 53]. This determined the main experimental methods for studying the timbre, the results of which are presented in detail below. The correlates of musical timbre parameters proposed by multidimensional large-scale studies were observed using event-related potentials. In neurophysiological studies of timbre perception, components of long-latent evoked potentials P1, N1 and P2 were recorded using both EEG and MEG (P1m, N1m and P2m) [6, 71, 75, 96]. The N1m component depended on the type of instrument generating the sound, i.e. on its timbre [75]. Shahin [97] analyzed the components of the P1, N1, P2 DSWPs in the perception of piano sounds, violins, and pure tones. It was found that the P1 component has a greater amplitude for the sounds generated by musical instruments, compared with pure tones. This means that P1 is larger for complex spectrum sounds. The amplitude of the P2 component was highest for the sounds produced by the piano, medium for violin tones, and lowest for pure tones. These data indicate that P2 is most sensitive to changes in sound timbre. A similar Aramaki study [6], in which environmental sounds were used, confirms that the amplitude of the P2 component depends on the sound timbre. Component N1 in these timbre studies was uninformative. Later, Meyer [71] demonstrated that both components N1 and P2 have higher amplitudes in the perception of sounds generated by instruments than in response to pure tones. Different timbres were created using various musical instruments or materials in these studies. To evaluate the electrophysiological correlates of a timbre under three dimensions, it is desirable to use an analytical approach and methods for modeling or synthesizing sound signals with a specific spectral structure. Shahin [96] manipulated the spectral complexity of the piano sound by removing a certain amount of harmonics. The amplitudes of the P2 and P2m components in this work were larger in the case of sounds with a large number of harmonics, the response components N1 and N1m were still uninformative

Another Tuznik, Augustynowicz, and Francuz study [107] in this direction was performed using sound synthesis and changes in the spectral centroid and spectral irregularity. The P2 component was also informative. The amplitude of P2 was significantly greater in the case of sounds with an average spectral centroid compared with the other two sounds, the differences were most pronounced in the parietal and parietal-occipital leads. The spectral non-uniformity of the sound stimulus, depending on the degree of its severity, caused a complex redistribution of the amplitude P2 in both anteroposterior and left-right directions. So, based on the data of the electrophysiological approach, evidence of the topical organization of the neuronal structures involved in the analysis of the timbre was obtained. A more accurate topic of the cerebral cortex can be obtained by fMRI functional magnetic resonance imaging. When the fundamental tone or spectral shape of the envelope changed, overlapping bilateral activation regions in the temporal lobes were detected for such sounds, and additional activation was detected when the spectral shape of the envelope changed along with alternating harmonic and noise stimuli [111]. The authors suggested that the middle part of the right STS (superior temporal sulcus) contains a specific mechanism for processing spectral envelopes - the acoustic brightness correlates, which goes beyond areas that respond to height or a fine spectral-temporal structure. In this work, changes in tone and spectral form were not separated for perception, which made it difficult to interpret direct comparisons. In additional Allen and Oxenham study [3] for more accurate comparison of cortical processing of tone and timbre, sound sequences were formed that varied either in the frequency of the fundamental tone (pitch) or in the position of the harmonics of the spectrum (brightness). These changes in two perceptual dimensions were balanced in perception. One-Dimensional analysis of the results of the study did not reveal significant differences in terms of areas performing processing of stimuli that varied in these two dimensions. The exception was a small difference in the weighted center of mass of voxel clusters, the responses of which correlated with the step size in the direction parallel to HG (Heschl's gyrus) in the direction from the front-lateral to the rear-medial in the left hemisphere. These results did not provide evidence of modular and exclusive processing of two measurements in separate areas of the auditory cortex, but the results of using MVPA (Multivoxel pattern analysis) suggested that various schemes in these areas could be used to analyze height and brightness. In this regard, the results were consistent with the findings of Town and Bizley studies [105] on individual neurons in the auditory cortex of ferrets, which also proposed different codes for pitch and timbre within the same neuronal population. As it turned out in this work, many neurons were sensitive to changes in both dimensions.

In the Allen [2] study, large variations in tone or timbre in the sequences led to large changes in the 'bold' signal in both directions, which coincided with the conclusions of Zatorre and Belin [115] studies for spectral and temporal variations. These data support the hypothesis that regions that are selective in tone or timbre show an increase in activation with an increase in the size of the range covered in each sequence. Tone and timbre are treated as separate dimensions in studies of auditory perception [34, 55, 68]. Several studies have shown that these two independent characteristics of sound can interact [56, 67, 112]. Recent psychoacoustic studies have shown that variations in tone and brightness mixed with the perception of another dimension, and these interference effects were symmetrical [3]. In other words, tone fluctuations affected the perception of brightness as much as differences in brightness influenced the perception of pitch. The strong overlap of activation of the cortical regions when measured across two dimensions, found by Allen [2], may also reflect the perceptual difficulty of separating these two measurements. A more detailed idea of how in one population of neurons an organized topically analysis of various signs (cues) of sound can be performed can be obtained in studies of individual neurons of the auditory regions of the cortex. This methodological approach is reflected in attempts to identify neural substrates for sound recognition by searching for areas of the brain that are selective for certain categories of sounds, such as areas specific to the voice, in the secondary areas of the auditory cortex [9, 108] and to other sound categories: different instrumentation [63] or musical instruments [58].

Patil, Pressnitzer, Shamma and Elhilali study [81] raised the question of how the processing of spectral-temporal modulations in the cortex can provide recognition of musical instruments and perception of timbre. It revealed that cortical receptive fields and computational models obtained from them are suitable for classifying a sound source by evoked neural activity in a wide range of instruments, pitch and playing styles, as well as for accurate prediction of human judgments about timbre similarities. Reactions in the primary auditory cortex (A1) exhibit a variety of types of selectivity substantially greater than the tonotopy observed in the auditory nerve. Field neurons A1 are tuned not only to spectral energy at a given frequency but also to features of a certain spectral shape, such as its bandwidth [93], spectral symmetry [110] and temporal dynamics [92]. It was shown that such spectral-temporal receptive fields with various nonlinear corrections well capture and predict cortical reactions to various complex sounds, such as speech, music and modulated noise [20, 23, 31, 89, 102].

The resulting representation of sound in field A1 can be considered as a multidimensional map that spans at least three dimensions: (1) characteristic frequency; (2) the spectral shape of the tincture curve with a frequency bandwidth of from 2-3 octaves to 0.25 octaves and a different type of symmetry; (3) dynamics that range from very slow to relatively fast level changes, covering a frequency range of 1-30 Hz. This rich cortical representation can reflect the coding strategy of acoustic signals that provide the perception of various acoustic features (pitch, volume, location and timbre), as well as the recognition of complex sound objects, such as various musical instruments with their characteristic timbre profiles. This hypothesis was verified in [81], based on the use of a database of spectral-temporal receptive fields of 1110 neurons registered by the authors of the work in the primary auditory cortex of 15 ferrets awake non-behaving ferrets.

The studied receptive fields are linear characteristics of the selectivity of each neuron of the auditory cortex to spectral and temporal modulations, which is manifested in a cochlear 'spectrogram-like' representation of complex acoustic signals arising at the auditory periphery. Then, a neuro-computational model was created and applied based on the spectral-temporal characteristics of the receptive fields of the registered neurons and on the simulated neurons of the auditory cortex, supplemented by previously registered experimentally using a nonlinear classifier. The model was able to perform robust classification of a musical instrument regardless of the height and style of playing it with an accuracy of 98.7%. Using the same input, the model could reproduce perceptual judgments of the distance between the tones in the same way that these tones are evaluated by listeners. The study shows that the spectral-temporal features characteristic of mammals' primary auditory cortex neurons are crucial to providing a sufficiently rich representation necessary for the formation of perceptual judgments about the timbre, as well as human recognition of musical instruments.

In general, the wide involvement of timbre characteristics in the processes of auditory analysis, their influence on the perception of other parameters of musical and speech signals and on memory processes, makes these characteristics a basic component in relation to the organization of auditory training and the solution of problems of development of perception, both of music and speech. Such training is traditionally used in the training of musicians, as well as for the correction of auditory dysfunctions. In recent decades, they are widely used for the rehabilitation of patients after cochlear implantation [47, 54, 64]. At the same time, when developing training programs, the peculiarities of the systematic organization of perception and language functions in speech and music, as well as the conditions of their mutual stimulating influence, are taken into account.

# 4. POSSIBILITIES OF SPEECH - MUSIC SYSTEMIC INTERACTION

The provisions discussed above allow us to conclude that speech and music have structural and perceptual similarities [12], which is manifested in the fact that they rely on close perceptual signs and are "rule-based" systems based on the basic elements of the language (phonemes - words; notes - chords) and language rules. According to these rules (grammar, harmony, syntax), the elements are combined into structures of a higher - order as musical phrases and sentences that ensure the transmission of semantic or emotional content. Language and music are combined into-higher-order structures (musical phrases and sentences, themes and topics) through the rules of harmony and syntax. From the direction of the systemic organization of speech and music, several researchers, first of all, highlight the differences and autonomy of the "modular" nature of their brain organization, as Peretz and Coltheart [83, 84] studies concluded. While Patel studies [76, 77, 78, 79, 80] stress their overlap, given the conflicting evidence regarding certain aspects of language and music (e.g. syntax) - emphasize the possibilities of "sharing brain resources". Patel has proposed a "resource sharing" framework, in which domain-specific representations for music and language are stored separately in long-term memory (LTM), but with deep connections in the cognitive processes that operate upon these. In Patel state common brainstem circuits are involved in pitch processing in speech and music matter, and musical training may drive subcortical pitch-encoding networks to function with higher precision than needed for ordinary speech processing, leading to benefits for speech processing because of the cognitive processes that music and speech share.

From modularity direction, as the central principle of organization in cognitive science, it is important to consider that both speech and music (their language and executive functions) are determined by structures that unfold in time, i.e. following the stages of ontogenetic development. Concerning speech function, at the initial stage of its development, the biological resource of the child provides the conditions for sensory processing of sound signals, inter-sensory and sensorimotor integration. Then, under the influence of internal and external factors, language forms of phonemic recognition and reproduction (articulation) are manifested, fixed in the speech environment following the perceptual rules of the native language [57], followed by the connection of letter designations, etc. The process of mastering the musical language, as a rule, proceeds at the later stages of ontogenesis, although biological preparedness for the perception of music (distinguishing high-altitude and rhythmic sequences) with the manifestation of hemispheric localization of functions is already recorded in newborns [82].

Differences in lateralization of speech and music functions are confirmed by clinical neuropsychological studies, where it has been observed general tendency with scientific evidence that injury of left temporal lobe often results in language impairment, while the existence of amusia deficit is related to damage to the right temporal lobe. Additional research with normal healthy adults has often indicated hemispheric lateralization of these functions [70]. The last scientific data of bilateral organization of music processing - i.e. spectral (pitch) and temporal (rhythm) attributes, with developing a diagnosis of the lateralization of music perception [30] allow to working out the new effective music-therapy method for speech disorders' correction and treatment, with studying possible relations and forms of synchronal laterality of speech and music features, especially for duration of hearing-loss and developmental aural perception deficits. Based on these data, a method for diagnosing musical perception [30] and an effective music-therapeutic method of musical training have been developed, which can be used to improve interhemispheric interaction with respect to music and speech, other bilateral cognitive functions (for example, teaching mathematics), as well as for correction hearing disorders. In general, the positive impact of music training, music education and musical experience on the formation, development and restoration of speech function is shown both in the conditions of early development of a child [116] and in the development of tonal languages [1, 21, 26, 27, 40, 59, 60, 88, 114]. The results of several recent studies of the musical deficit, pitch-amusia (i.e. tone-deafness) in both non-tone and tone language-speakers support the case for domain-general pitch perception across music and speech [48, 73, 103, 104]. That amusia type extends to speech and other pitch-related tasks-evidence for domain-general pitch perception across music and speech, also suggesting that tone long language experience does not compensate for this supposedly musical deficit.

Standley, Huges, and Overy studies [74, 98] examined the influence of musical training on language and reading skills using longitudinal training measure designs, as evidence of the strong relationship between musical ability (training) and language – literacy skills. It was revealed, that acoustic training might be beneficial for language developmental disorders. Music training during music processing has previously been shown in scientific evidences to improve auditory processing [86, 106], cognitive skills [36, 41, 42, 90], language and literacy skills [5, 7, 18, 43, 74, 98]. Here music processing concern melody, rhythm, metre, timbre, harmony, contour, with general auditory processing of pitch discrimination, pitch memory, rapid spectral-temporal processing, while language and literacy skills concern components such as syllables processing, reading, phonological awareness, pitch processing in speech, prosody perception and verbal memory with fluency [101].

These results are of practical importance for the organization of auditory training courses and rehabilitation of patients with cochlear implants. Based on them, special classes are held on the perception of musical sounds (various tones, pitch), the assessment of melodic intervals (pitch movement) and the contours of changes in the frequency of the fundamental tone (intonation), the distinction between the rhythmic patterns of sound sequences and the selection of the target voice of a speaker or musical instrument. The research results indicate the effectiveness of such classes concerning the formation and development of auditory analysis skills and accelerate the adaptation of patients to new hearing conditions [47, 54, 64]. Additionally phonological awareness is pivotal for reading and writing skills, and closely related to pitch awareness and musical expertise [25, 66], here preschoolers can benefit from a musical training program to increase their phonological awareness. Patel [78] proposed 'opera' hypothesis for explanation of music's beneficial role for language functions in case of speech-processing plasticity brain networks, overlap process acoustic features used in music and speech, precision with music's high demands on the shared network, emotion that elicit engage the network, repetition that engaged the network frequently often, attention associated with musical activity of that network engaged directly. With respecting these conditions the neural plasticity drives the network with higher precision than needed for standard speech communication, and the development of the verbal network is faster than in the situation of standard speech communication.

In general, the empirical research results showed the following relationships between music and speech [46]: - pitch imitation production mechanisms engaged in music and language material [10], - influence of musical aptitude on the linguistic skill acquisition in second language study [72], - implicitly acquired knowledge and implicit memory - used in linguistic and musical grammar acquisition [32], - musical expertise transfer from music to speech in musical actively humans [11], with facilitation the learning of linguistic and musical syntax form electrophysiological data [91], - musical training influence on speech perception at the sensory level, which might facilitate that process in humans with hearing loss [95], - Koelsch [51] neurocognitive model of music perception based on neuropsychological similarities and differences between the syntax of language and music.

Therefore in general conclusion musical training (perceptual, manual, cognitive, social, neuropsychological, plane aural listening) may aid in the prevention, rehabilitation and remediation of a wide range of language, listening and learning impairments, also in the complex treatment of neurological disorders. Here 'unique' function to music belong in therapy duration, with wide range of medical disciplines engaged and study the music's influence onto physiological and psychological health treatment, with evidence of scientific music therapy range in cognitive neuroscience of music direction [85] in medical industry, where some music becomes a pharmacological resource for human health stimulation and its dynamical changes. This function of music helps to increase the effectiveness of drug treatment and stabilize the physiological and psychological state of the patient. Thus the impact of music becomes a kind of non-pharmacological resource for stimulating human health and overcoming the negative effects.

## 5. CONCLUSION

The topic discussed in the article is extremely complex and touches on issues related to the study of the similarities and differences in the functions of music and speech using several disciplines - psychology, physiology, neurology, linguistics, etc. An important aspect was the selection of typological features of speech and music from the duration of language competencies' description, as well as the possibility of practical use of the biological basis of the interaction of speech and musical signals. Scientific data testifies both to the modular organization of these functions, and to the systemic interaction of speech and music when sharing brain resources and community elements of perceptual organization. From this point of view, the possibility of joint consideration of the whole complex of mechanisms of auditory perception with the involvement of another important function of hearing - spatial orientation is of interest. She, like speech and music, is focused on processing the characteristics of pitch, timbre and sound intensity. However, bilateral connections play a much more significant role in its implementation, and linguistic elements are practically absent. The study of the spatial selectivity of musical signals seems to be a promising area of research, primarily in the context of comparison with the laws of spatial selectivity of speech hearing and modeling the mechanisms of auditory analysis of complex acoustic scenes with different localization of speech and music sources. The results of such a study can be used to develop artificial intelligence systems and the development of assistive technologies in medicine and pedagogy, as well as ways to prevent age-related sensory-cognitive dysfunctions.

# 6. REFERENCES

[1] Alexander, J. A., Wong, P. C. M., Bradlow, A. R. (2005). *Lexical tone perception in musicians and non-musicians.* Thesis of proceedings of interspeech', Eurospeech – 9th European Conference on Speech Communication and Technology, September 2005, Lisbon, Portugal.

[2] Allen, E. J., Burton, P. C., Olman, Ch. A., Oxenham, A. J. (2017). Representations of pitch and timbre variation in human auditory cortex. *Journal of neuroscience*, February, 37 (5), p.1284-1293.

[3] Allen, E. J., Oxenham, A. J. (2014) Symmetric interactions and interference between pitch and timbre. *Journal of the acoustical society of America*, March, 135 (3), p.1371-1379.

[4] ANSI (1999). *American National Psychoacoustic terminology*. S3.20. ed. New York, American National Standards Institute. 67 p.

[5] Anvari, S. H., Trainor, L. J., Woodside, J., Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of experimental child psychology*, October, 83 (2), p.111-130.

[6] Aramaki, M., Besson, M., Kronland-Martinet, R., Ystad, S. (2008). Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches. In: Ystad S., Kronland-Martinet R., Jensen K. (eds.) *Computer music modeling and retrieval. Genesis of meaning in sound and music. CMMR 2008. Lecture notes in computer science*, vol. 5493. ed. Springer, Berlin, Heidelberg. p.1-17.

[7] Barwick, J., Valentine, E., West, R., Wilding, J. (1989). Relations between reading and musical abilities. *British journal of educational psychology*, June, 59 (pt.2), p.253-257.

[8] Bates, E., Thal, D., Finlay, B. L., Clancy B. (2003). Early language development and its neural correlates. In S. J. Segalowitz, I. Rapin (ed.). *Handbook of neuropsychology*, vol.8, part 2, *Child neuropsychology*, chapter 5, ed. Amsterdam, Elsevier Science. p.525-592.

[9] Belin, P, Zatorre, R. J., Lafaille, P., Ahad, P., Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403 (6767), p.309-312.

[10] Bella, S. D., Berkowska, M., Sowinski, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in psychology*, July, vol. 2, art. 164, 11 p.

[11] Besson, M., Chobert, J., Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Frontiers in psychology,* May, vol. 2, art. 94, 11 p.

[12] Besson, M., Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, June, 930, p.232-258.

[13] Best, K., Bogossian, F., New, K. (2018). Language exposure of preterm infants in the neonatal unit: a systematic review. *Neonatology*, 114 (3), p.261-276.

[14] Bregman, A. S. (1994). Auditory scene analysis: the perceptual organization of sound. ed. Bradford book, MIT Press. 790 p.

[15] Bronkhorst A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attention, perception and psychophysics*, July, 77 (5), p.1465-1487.

[16] Caclin, A., Brattico, E., Tervaniemi, M., Naatanen, R., Morlet, D., Giard, M. H., McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of cognitive neuroscience*, December, 18 (12), p.1959-1972.

[17] Caclin, A., McAdams, S., Smith, B. K., Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *Journal of the acoustical society of America*, July, 118 (1), p.471-482.

[18] Chan, A. S., Ho, Y. C., Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, November, 396 (6707), p.128.

[19] Chen, C. J., Miller, D. A. (2019). Pitch-synchronous analysis of human voice. *Journal of voice*, February 7, pii: S0892-1997(18)30448-X

[20] Christianson, G. B., Sahani, M., Linden, J. F. (2008). The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *Journal of neuroscience*, January, 28 (2), p.446-455.

[21] Cooper, A., Wang, Y. (2010). *The role of musical experience in Cantonese lexical tone perception by native speakers of Thai.* Thesis of Proceedings of 5th International Conference on Speech Prosody. Chicago, IL. 4 p. http://www.isle.illinois.edu/speechprosody2010/papers/100184.pdf

[22] Darwin, C. J. (1997). Auditory grouping. *Trends in cognitive sciences*, December, 1 (9), p.327-333.

[23] David, S. V., Mesgarani, N., Fritz, J. B., Shamma, S. A. (2009). Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *Journal of neuroscience*, March, 29 (11), p.3374-3386.

[24] DeCasper, A. J., Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, June, 208 (4448), p.1174-1176.

[25] Dege, F., Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in psychology*, June, vol. 2, art. 124. 7 p.

[26] Delogu, F., Lampis, G., Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive processing*, 7 (3), p.203-207.

[27] Delogu, F., Lampis, G., Olivetti Belardinelli, M. (2010). From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. *European journal of cognitive psychology*, 22, p.46-61.

[28] Doheny, L., Hurwitz, S., Insoft, R., Ringer, S., Lahav, A. (2012). Exposure to biological maternal sounds improves cardiorespiratory regulation in extremely preterm infants. *Journal of maternal - fetal and neonatal medicine*, September, 25 (9), p.1591-1594.

[29] Donnadieu, S. (2007). Mental representation of the timbre of complex sounds. In J. W. Beauchamp (ed.). *Analysis, synthesis, and perception of musical sounds*. ed. New York, Springer. p.272-319.

[30] Dymnikowa, M. (2018). *Diagnostika muzykalnoj pamiati.* [Diagnosis of musical memory]. ed. Saint-Petersburg, Lema. 184 p.

[31] Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., Shamma, S. A. (2004). Dynamics of precise spike timing in primary auditory cortex. *Journal of neuroscience*, February, 24 (5), p.1159-1172.

[32] Ettlinger, M., Margulis, E. H., Wong, P. C. (2011). Implicit memory in music and language. *Frontiers in psychology*, September, vol. 2, art. 211, 10 p.

[33] Filipiak, E. (2015). Funkcje języka. [*Functions of language*] *Zeszyty Naukowe Wyższej Szkoły Pedagogicznej w Bydgoszczy. Studia Pedagogiczne 27*, Pedagogika Przedszkolna i Wczesnoszkolna 10, p.35-47.

[34] Fletcher, N. H. (1934). Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. *Journal of the acoustical society of America*, 6, p.59-69.

[35] Fletcher, N. H., Rossing, T. D. (2010). *The physics of musical instruments*. ed. New York, Springer-Verlag. 756 p.

[36] Gardiner, M. F., Fox, A., Knowles, F., Jeffrey, D. (1996). Learning improved by arts training. *Nature*, May, 381 (6580), p.284.

[37] Gardner, H. (1999). *Intelligence reframed: multiple intelligences for the 21st century*. ed. New York, Basic Books. 292 p.

[38] Greshuni, G. V. (1967). O znacenii wremennyh harakteristik w organizacji dejatelnosti sluhowoj sistemy. [*On the significance of temporal characteristics in the organization of the auditory system*]. In: M. N. Livanov, V. S. Rusinov (ed.). Sowremennyje problemy eketrofiziologii centralnoj nerwnoj sistemy. [*Current problems of the electrophysiology of the central nervous system*]. Ed. Moscow, Nauka. p.65-70.

[39] Golubock, J. L., Janata, P. (2013). Keeping timbre in mind: working memory for complex sounds that can't be verbalized. *Journal of experimental psychology: human perception and performance*, 39 (2), p.399-412.

[40] Gottfried, T. L., Xu, Y. (2008). *Effects of musical experience on Mandarin tone and vowel discrimination and imitation*. Thesis of Proceedings at the Acoustics 08, Paris, France. p.6161-6165.

[41] Hassler, M., Birbaumer, N., Feil, A. (1985). Musical talent and visual-spatial ability: a longitudinal study. *Psychology of music*, October, 13 (2), p.99-113.

[42] Hetland, L. (2000). Learning to make music enhances spatial reasoning. *Journal of aesthetic education*, September, 34 (3/4), p.179-238.

[43] Ho, Y. C., Cheung, M. C., Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology*, July, 17 (3), p.439-450.

[44] Jakobson, R. (1960). Linguistics and poetics. In T. Sebeok (ed.), *Style in Language*, Cambridge, MA, M.I.T. Press, p.350-377.

[45] Jakobson, R. (1995). *On Language*. ed. Cambridge, MA, Harvard University Press, 672 p.

[46] Jancke, L. (2012). The relationship between music and language. *Frontiers in psychology*, April, vol. 3, art.123. 2 p.

[47] Jiam, N. T., Deroche, M. L., Jiradejvong, P., Limb, C. J. (2019). A randomized controlled crossover study of the impact of online music training on pitch and timbre perception in cochlear implant users. *Journal of the association for research in otolaryngology*, June, 20 (3), p.247-262.

[48] Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, July, 48 (9), p.2630-2639.

[49] Kanwal, J., Ehret, G. (2006). *Behaviour and neurodynamics for auditory communication.* ed. Cambridge University Press. 384 p.

[50] Kiklewicz, A. (1999). *Lekcii po funcionalnoj lingvistike.* [Lectures on functional linguistics] ed. Minsk BGU, 214 p.

[51] Koelsch, S. (2011). Toward a neural basis of music perception – a review and updated model. *Frontiers in psychology*, June, vol. 2, art. 110, 20 p.

[52] Kong, Y. Y., Mullangi, A., Marozeau, J., Epstein, M. (2011). Temporal and spectral cues for musical timbre perception in electric hearing. *Journal of speech, language and hearing research*, June, 54 (3), p.981-994.

[53] Konig, R., Heil, P., Budinger, E., Scheich, H. (eds.). *The auditory cortex. A synthesis of human and non-human research.* ed. Mahwah, NJ. Lawrence Erlbaum Associates. 512 p.

[54] Korolewa, I. W., Orogodnikova, E. A. (2018). Ocenka progressa razwitia processow sluhorecewogo analiza u detej s narusheniem sluha, ispolzujusih kohlearnye implanty i sluhowye aparaty. [*Assessment of progress in the development of auditory-speech analysis processes in children with hearing impairment using cochlear implants and hearing aids*]. Westnik psichofiziologii [*Psychophysiology news*], No.2, p.106-113.

[55] Kraus, N., Skoe, E., Parbery-Clark, A., Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing: implications for language and music. *Annals of the New York academy of sciences*, July, 1169, p.543-557.

[56] Krumhansl, C. L., Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of experimental psychology. Human perception and performance.* August, 18 (3), p.739-751.

[57] Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, March, 9 (2), F13-F21.

[58] Leaver, A. M., Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of neuroscience*, June, 30 (22), p.7604-7612.

[59] Lee, C. Y., Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the acoustical society of America*, 124 (5), p.3235-3248.

[60] Lee, C. Y., Lee, Y. F. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *Journal of the acoustical society of America*, 127 (1), p.481-490.

[61] Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception and psychophysics*, 56 (4), p.414-423.

[62] Levitin, D. J., Cook, P. R. (1996). Memory for musical tempo: additional evidence that auditory memory is absolute. *Perception and psychophysics*, August, 58 (6), p.927-935.

[63] Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of neuroscience*, May, 25 (21), p.5148-5158.

[64] Looi, V., Gfeller, K., Driscoll, V. (2012). Music appreciation and training for cochlear implant recipients: a review. *Semin Hear.* NIH-PA Author manuscript, November, 33 (4), p.307-334.

[65] Lotto, A. J., Sullivan, S. C. (2008). Speech as a sound source. In W. A. Yost, A. N. Popper, R. R. Fay (ed.). *Auditory perception of sound sources*, ed. SHAR (Springer handbook of auditory research), vol. 29. Springer, Boston, MA, p.281-305.

[66] Loui, P., Kroog, K., Zuk, J., Winner, E., Schlaug, G. (2011). Relating pitch awareness to phonemic awareness in children: implications for tone-deaf-ness and dyslexia. *Frontiers in psychology*, May, vol. 2, art. 111, 5 p.

[67] Marozeau, J., de Cheveigne, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the acoustical society of America*, January, 121 (1), p.383-387.

[68] McDermott, J. H., Keebler, M. V., Micheyl, Ch., Oxenham, A. J. (2010). Musical intervals and relative pitch: frequency resolution, not interval resolution, is special. *Journal of the acoustical society of America*, October, 128 (4), p.1943-1951.

[69] McKeown, D., Wellsted, D. (2009). Auditory memory for timbre. *Journal of experimental psychology: human perception and performance*, 35 (3), p.855-875.

[70] McMullen, E., Saffran, J. R. (2004). Music and language: a developmental comparison. *Music perception: an interdisciplinary journal*, spring, 21 (3), p.289-311.

[71] Meyer, M., Baumann, S., Jancke, L., 2006. Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage*, 32, 1510-1523.

[72] Milovanov, R., Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: a review. *Frontiers in psychology*, November, vol. 2, art. 321, 7 p.

[73] Nan, Y., Sun, Y., Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain*, 133 (9), p.2635-2642.

[74] Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Annals of the New York academy of sciences*, November, 999, p.497-505.

[75] Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, January, 12 (1), p.169-174.

[76] Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, 6 (7), p.674-681.

[77] Patel, A. D. (2010). *Music, language and the brain*. ed. New York, Oxford University Press. 526 p.

[78] Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in psychology*, June, vol.2, art.142, 14 p.

[79] Patel, A. D. (2012). Language, music, and the brain: a resource sharing framework. In: P. Rebuschat, M. Rohrmeier, J. Hawkins, I. Cross (ed.). *Language and music as cognitive systems*. ed. New York, Oxford University Press. p.204-223.

[80] Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications. *Annals of the New York Academy of Sciences*, 1252, p.124-128.

[81] Patil, K., Pressnitzer, D., Shamma, Sh., Elhilali, M. (2012). Music in our ears: the biological bases of musical timbre perception. *PLoS Computational Biology*, November, 8 (11), art. e1002759, 16 p.

[82] Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the national academy of sciences of the United States of America*, March, 107 (10), p.4758-4763.

[83] Peretz, I. (2012). Music, language and modularity in action. In: P. Rebuschat, M. Rohrmeier, J. Hawkins, I. Cross (ed.). *Language and music as cognitive systems*. ed. New York, Oxford University Press. p.254-268.

[84] Peretz, I., Coltheart, M. (2003). Modularity of musical processing. *Nature neuroscience*, 6, p.688-691.

[85] Peretz, I., Zatorre, R. J. (2003). *The cognitive neuroscience of music*. ed. New York, Oxford University Press. 452 p.

[86] Peretz, I., Zatorre, R, J. (2005). Brain organization for music processing. *Annual review of psychology*, 56, p.89-114.

[87] Rand, K., Lahav, A. (2014). Maternal sounds elicit lower heart rate in preterm newborns in the first month of life. *Early human development,* NIH-PA Author manuscript, October, 90 (10), p.679-683.

[88] Ro, M. H., Behne, D., Wang, Y. (2006). The effects of musical experience on linguistic pitch perception: a comparison of Norwegian professional singers and instrumentalists. *Journal of the acoustical society of America*, 120, abstract 3168.

[89] Sadagopan, S., Wang, X. (2009). Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *Journal of neuroscience*, September, 29 (36), p.11192-11202.

[90] Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psych science*, 15, August, 15 (8), p.511-514.

[91] Schon, D., and Francois, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in psychology*, July, vol. 2, art. 167, 9 p.

[92] Schreiner, C. E., Mendelson, J., Raggio, M. W., Brosch, M., Krueger, K. (1997). Temporal processing in cat primary auditory cortex. *Acta oto-laryngologica. Supplementum*, 532, p.54-60.

[93] Schreiner, C. E., Sutter, M. L. (1992). Topography of excitatory bandwidth in cat primary auditory cortex: single-neuron versus multiple-neuron recordings. *Journal of neurophysiology*, November, 68 (5), p.1487-1502.

[94] Schulze, K., Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory*, April, 21 (3), p.377-395.

[95] Shahin, A. J. (2011). Neurophysiological influence of musical training on speech perception. *Frontiers in psycholology*, June, vol. 2, art. 126, 10 p.

[96] Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, November, 16 (16), p.1781-1785.

[97] Shahin, A., Roberts, L. E., Trainor, L. J. (2004). Enhancement of auditory cortical development by musical experience in children. *Neuroreport*, August, 15 (12), p.1917-1921.

[98] Standley, J. M., Huges, J. E. (1997). Evaluation of an early intervention music curriculum for enhancing prereading / writing skills. *Music therapy perspectives*, November, 15 (2), p.79-88.

[99] Stevens, C. J. (2012). Music perception and cognition: a review of recent cross-cultural research. *Topics in cognitive science*, October, 4 (4), p.653-667.

[100] Takeuchi, A. H., Hulse, S. H. (1993). Absolute pitch. *Psychological bulletin*, March, 113 (2), p.345-361.

[101] Tallal, P., Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in neuroscience*, July, 29 (7), p.382-390.

[102] Theunissen, F. E., Sen, K., Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of neuroscience*, March, 20 (6), p.2315-2331.

[103] Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., Peretz, I. (2011). Congenital amusia (or tone-deaf-ness) interferes with pitch processing in tone languages. *Frontiers in psychology*, June, vol.2, art.120. 15 p.

[104] Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umilta, C., Peretz, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *Journal of the acoustical society of America*, 130 (6), p.4089-4096.

[105] Town, S. M., Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Frontiers in systems neuroscience*, November, vol.7, art. 88. 14 p.

[106] Trainor, L. J., Shahin, A., Roberts, L. E. (2003). Effects of musical training on the auditory cortex in children. *Annals of the New York academy of sciences*, November, 999, p.506-513.

[107] Tuznik, P., Augustynowicz, P., Francuz, P. (2018). Electrophysiological correlates of timbre imagery and perception. *International journal of psychophysiology*, July, vol.129, p.9-17.

[108] Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., Patterson, R. D. (2006). Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage*, July, 31 (3), p.1284-1296.

[109] Vanzella, P., Schellenberg, E. G. (2010). Absolute pitch: effects of timbre on note-naming ability. *PloSOne*, November, 5 (11), e15449. 7 p

[110] Versnel, H., Kowalski, N., Shamma, Sh. A. (1995). Ripple analysis in the ferret primary auditory cortex. III. Topographic and columnar distribution of ripple response parameters. *Auditory neuroscience*, 1, p.271-286.

[111] Warren, J. D., Jennings, A. R., Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*, February, 24 (4), p.1052-1057.

[112] Warrier, C. M., Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and psychophysics*, February, 64 (2), p.198-207.

[113] Wolfe, J. (2002). Speech and music, acoustics and coding, and what music might be 'for'. In: C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Reniwick (eds.). *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney.* ed. Adelaide, Causal Productions. p.10-13.

[114] Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10 (4), p.420-422.

[115] Zatorre, R. J., Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, October, 11 (10), p.946-953.

[116] Zhao, T. Ch., Kuhl, P. K. (2016). Effects of enriched auditory experience on infants' speech perception during the first year of life. *Prospects*, June, 46 (2), p.235-247.