Science Discovery Psychology
2026, Vol. 1, No. 3, pp. 91–95
https://doi.org/10.11648/j.sdps.20260103.12

Research Article

Analysis of Research on the Factors of Speech Comprehension

Lobar Kambarova*

Uzbek Linguistics, Tashkent State University of Uzbek Language and Literature Named After Named Alisher Navoi, Tashkent, Uzbekistan

Abstract

Understanding speech requires a number of factors. Acoustic data goes through several stages from reception to processing in the brain. When studying international experience in neurolinguistics, it can be seen that the main attention is paid to the study of what processes occur in the brain during speech perception, what factors influence speech comprehension. Through the analysis of research, it is possible to increase information on neurolinguistics in the Uzbek language, to form an idea of how to conduct research in this direction. In the perception of speech, information is provided about the presence of such terms as speech envelope, fine structure, such phenomena as cocktail party, McGurk effect.

Keywords

Neurogram, Phonological Unit, Assumption, Speech Comprehension, Speech Shell, Precise Structure, Analysis Synthesis, Cocktail Party Effect, Mcgurk Effect

1. Introduction

Neurolinguistics is a new field of science that emerged at the intersection of psychology, neurology and linguistics, studying "the mechanisms of the speech process in the brain and changes in the speech process caused by brain injuries" [1]. The formation of neurolinguistics as a field of science is connected with the development of neuropsychology, on the one hand and linguistics and psychology on the other. Neurolinguistic research is conducted in the following areas:

1). in which part of the brain linguistic information is processed;
2). how information about language is processed as time changes;
3). which parts of the brain are involved in language learning and acquisition;
4). how neurolinguistics can contribute to working with speech pathology and other issues.

On a global scale, a number of studies have been conducted in the field of neurolinguistics on what factors are important in speech comprehension.

2. Literature Review

Speech perception is an extremely complex process. In this process, it became known by the 19th century that certain parts of the brain are responsible for speech. In 1861 French neurologist Paul Broca's speech at the Paris Anatomical Society was the first study in this direction. He conducted research with a

*Correspondence: Lobar Kambarova (qambarova.lobar@bk.ru)

Received: 4 November 2025; **Accepted:** 17 November 2025; **Published:** 28 July 2026

patient named Labrjin, who lost his ability to speak due to brain injury. Due to a brain injury, the patient could not speak for long and repeated the same phrases. After his death, an examination of his brain revealed damage to the left hemisphere of the frontal lobe (frontal lobe cortex). Broca soon conducted a similar experiment with a second patient. Both patients had speech difficulties due to damage to the left hemisphere of the brain. The greatest innovation introduced by Paul Broca to the 19th century medical world is that different parts of the human brain are responsible for performing different functions, including human speech in the left hemisphere.

In 1870, Karl Wernicke and his students put forward the hypothesis that a problem with understanding also arises in brain injury, and that another part of the brain is responsible for understanding speech [2]. Jonathan R. Brennan lists three main aspects. Firstly, there is a division between the two “language centers” mentioned above: the temporal lobe area is necessary for understanding speech, and the frontal lobe area is necessary for expressing speech. Secondly, these territories are interconnected. Thirdly, these parts are also connected to other parts of the brain for the full use of language. For example, words are associated with the motor area for articulation, with the memory area for storing conceptual knowledge [2].

To visualize the processes occurring in the brain during speech perception, it is advisable to know the language-related parts of brain anatomy. The aforementioned source provides detailed information about this [2].

3. Research Methodology

The description method was used in the study.

Since the emergence of humankind, speech has been valued as the most invaluable gift. The science of linguistics is noted as one of the most important sciences. In his work “On the Origin of Sciences” Abu Nasr Farabi, dividing sciences into types, separately distinguishes between linguistics and grammar: “When we come to the question of how to teach and receive education, how to express, explain, ask and answer thoughts, I confirm that the first of the sciences about this is the science of language, which gives names to bodies, that is, substances and accidents” [3].

In the East, one can observe a greater emphasis on the beauty of speech and the influence of words on a person. Yusuf Khas Hajib connects the emergence of life on earth with words:

From the green sky, words descended to this gray earth,
With his words, man illuminated the eyes [4].

Similar verses are also found in the works of Alisher Navoi:

Words separated humans from animals,
Know that there is no more noble pearl than it [5].

Today’s research in modern linguistics studies the speech process in connection with the human brain and makes new discoveries. In the human brain, information exchange is carried out with the help of neurons. There are several stages before the perception of speech. Speech reaches the listener’s ears as vibrations in the air and sets the eardrum in motion. Sound vibrations pass from the middle ear to the inner ear. In the inner ear, there is a shell-like organ (cochlea), the main function of which is to convert sound vibrations into electrical vibrations for the brain. This process is carried out by thousands of tiny hair cells. The hairs are excited by alternating vibrations, a stream of neurochemical processes is set in motion, and the resulting action potential is sent to the central nervous system. The shape of the shell is such that it begins from the greater base and gradually narrows. This means that different frequencies are trapped in different parts of the shell, i.e., high frequencies are trapped in the base part, while low frequencies are trapped in the narrow part. Hair cells are located in different parts of the shell and are excited at different frequencies. The reflection of different information received by the shell by different neurons is called spatial code. This code is considered one of the first principles explaining how the brain reflects sounds.

4. Analysis and Results

Brain signals produced in the shell pass through another part of the brain before reaching the part responsible for speech (superior temporal gyrus). This part is called the Heschl’s gyrus fold and is the primary auditory cortex. In hearing, as in the visual system of mammals, the hemispheres of the brain work in the opposite direction, that is, sounds coming from the right ear are processed in the auditory center located in the left hemisphere. Sound vibrations in the shell are converted into electrical signals for the brain and it takes approximately 50 milliseconds to reach the auditory area.

How do sounds manifest in the auditory part? Firstly, the codes generated in the shell are stored in the auditory part. As a result of studying the auditory process in humans and animals, it became clear that different parts of the auditory fold respond to different levels of frequencies. For a specific sound in the auditory part, the code is conventionally called *tonotopy*. To further clarify this issue, in the primary auditory fold, each neuron has a *receptive field (receptive field)* adjusted to different frequencies. Thus, tonotopy are neurons located in the auditory fold, possessing an area of stimulation and tuning to different frequencies. Neurons tuned to similar frequencies are located next to each other.

When a person receives speech signals, a unique adaptation is observed in them. The frequency of information in speech is not always important for understanding speech. Scientists studying speech divide the acoustic signal into two parts. The flow of speech has ascents and descents depending on the pitch of the sound. Such ascending and descending points are called the speech envelope *(speech envelope)* and represent the rhythmic form and stress of syllables in speech. Fast sound vibrations at different frequencies reduce speech energy forms and this specific structure is called *(fine structure).* In the above example, when analyzing the sentence *"Cats are soft and furry" ("Cats are soft and furry"),* it can be seen that the clear structure includes a narrow line representing such a feature as the differentiation of vowel sounds and a wide line representing noise associated with whispering and sliding. The presence of a speech shell and a clear structure is the main condition for understanding speech.

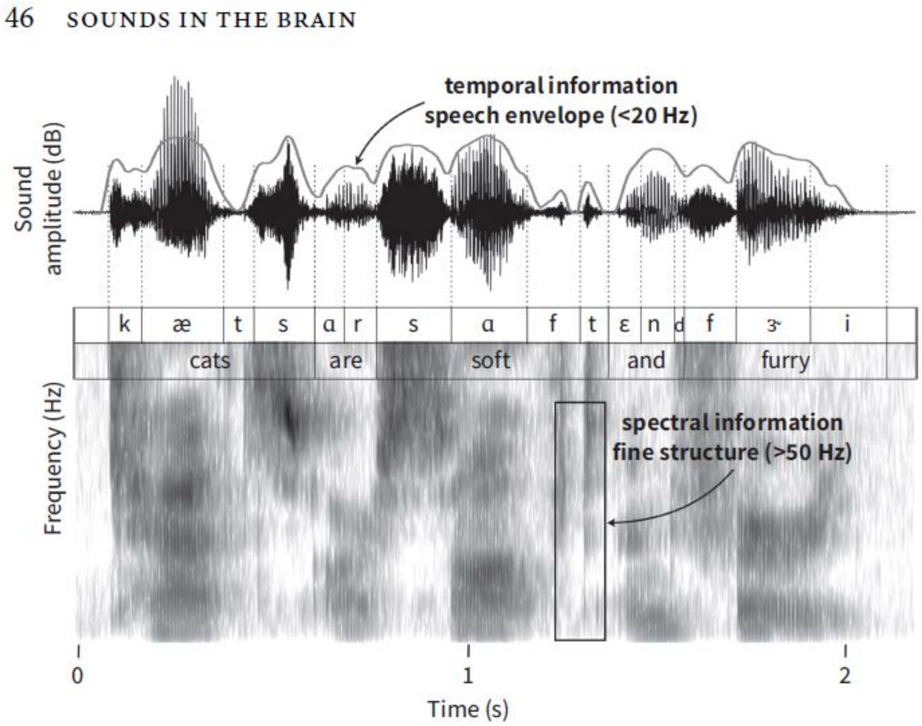

Figure 1. *Division of an acoustic signal into a speech envelope and a fine structure.*

Experiments have shown that the brain records the speech cortex (temporal information). Researchers conduct the following experiment. Several participants are given a typical speech. With the help of the ECoG device, signals coming from the auditory part of the left hemisphere are recorded. In this process, rapid changes in the brain coincide with rapid changes in the speech cortex. Such a correspondence between brain activity and perceptual input can be the key to the brain's method of "recording" important elements for subsequent processes in the sensory system. More precisely, this process involves the brain automatically recording the input perceptual information. Human attention is also of primary importance here. Because only part of the whispers heard is understood. There are several views on the separation of speech from various noises. One of them is the cocktail party effect: in a crowded room, a person hears their name being called from another corner. In this case, the name "jumped out" from among the voices in the noisy room.

The following conclusions are presented in the scientific source we are considering:

a) information input into the brain is called a neurogram;
b) processed data are considered phonemic features;
c) the transformation of a neurogram into a phonological feature is an extremely rapid process (occurring within 100-200 milliseconds after speech begins);
d) the above process takes place in the superior temporal lobe located adjacent to the primary auditory cortex.

The following experiment was conducted by Kourosh Saberi and David Perrott [6]. Participants are given an inverted speech. This stimulus lasts more or less a unit of time. For this, the sound stream is divided into syllables, which are received in 300, 200, 100, 50, and 20 milliseconds. Each syllable is reversed and rearranged according to the actual order. Participants hear a speech that is reversed in each case. They cannot understand words whose syllables last 300 or 200 milliseconds. When syllables were heard in less than 200 milliseconds, the meaning of words began to be understood. Importantly, they understood syllables with 100% accuracy in 30-50 milliseconds. For example, it took 20-30 milliseconds to distinguish between the syllables [ta] and [da]. Thus, a hypothesis has been put forward that even reversed speech can be understood if the distance between syllables is 1 in 20ths of a second. This study shows that the brain has *a sampling rate* for speech sounds, meaning the brain receives an "image" from acoustic information. It adds everything during the filming process to a part of the information, and this part serves to distinguish phonemes. This process is calledby J. Brennan *temporal windows of integration [7, 8]*. These windows are divided into two parts. The first window lasts 25-40 milliseconds, during which the spectral properties of different phonemes in a specific structure are distinguished. The length of the second window is relatively longer and lasts 200-300 milliseconds. This is the appropriate length for determining the characteristics of syllables and other structures that make up the speech shell. Such studies are based on the study of the periodic patterns of brain activity, manifested in neural oscillations - brain waves or special rhythms. Brain waves arise as a result of the coordination of the activity of a group of neurons, that is, joint excitation or opposite movement. Such coordination is important for the information exchange process; the entire group of neurons acts together in receiving and processing information.

Understanding speech also requires the co-operation of several sensory organs. For example, visual sensation. An experiment was conducted on combining the information seen and heard. As a result of the experiment, the concept of the McGurk effect [9, 10] arose. According to this study, the listener is shown a video of someone pronouncing a single syllable. One person pronounces the syllable [ba], but at the bottom of the video, it is shown that he pronounces the syllable [ga], that is, an artificial imbalance is created between sight and hearing. Some listeners say they hear the syllable as [ga], others as [da]. When the [ba] syllable is pronounced both in the video and in the broadcast, very few people say that they are hearing the [ba] syllable. Consequently, not only auditory but also visual perception plays an important role in understanding speech.

The next factor in understanding speech is the listener's assumption [11, 12] about what the speaker might say. This assumption is based on various data, such as the speaker's language being known to the listener (probable words, emerging speech sounds), the speaker's identity being known (what they might be talking about), and the direct speech context being known to the listener. Another idea in speech perception is the presence [13, 14] in the listener of knowledge about the internal synthesis of incoming speech. This internal process helps to analyze the speech signal upon arrival and is called analysis by synthesis (analysis by synthesis) [15].

The interconnected process from the introduction of acoustic information to the process of analysis through synthesis can be represented by the following diagram.

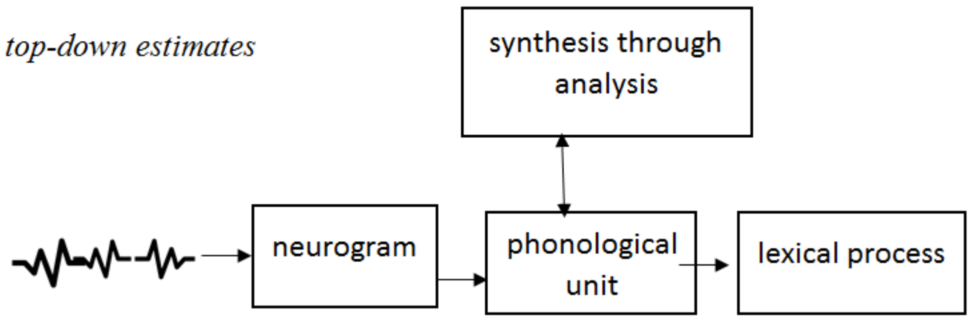

Figure 2. *Analysis by synthesis.*

5. Conclusions and Suggestions

Based on the above information, the following general conclusions are presented:

1). acoustic information is reflected in the cochlear part of the ear and the initial auditory part as spatial code. Various acoustic properties of speech are encoded by different neurons; Tonotopy is a code for data frequency, generated by adjacent neurons that respond to sounds of similar frequencies;

2). the transformation of the brain from a neurogram (a continuous neural image of sounds) into a phonological image - a phoneme, which is considered a categorical linguistic unit, occurs in 100-150 milliseconds in the superior temporal fold (superior temporal gyrus), located near the auditory fold;

3). the combination of acoustic information occurs in two temporal windows: a short window for a specific structure of speech; a long window for the speech shell;

4). the phonological unit is improved in a process called analysis through synthesis, which occurs as a result of the connection between acoustic information and linguistic knowledge.

Abbreviations

EcoG	Measures Neuronal Activity with Electrodes Implanted in the Cortex

Author Contributions

Lobar Kambarova is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Luria A. R. Fundamental Problems of Neurolinguistics. - Moscow: Moscow University Press, 1976. - 253.

[2] Language and the brain. Jonathan R. Brennan. – Oxford, 2022– Oxford, 2022. – P. 13.

[3] Abu Nasr al-Farabi. The City of Virtuous People. - Tashkent, 1993. – P. 179.

[4] Yusuf Khas Hajib. Kutadgu Bilig ("Knowledge Leading to Happiness") // Rendered in modern Uzbek and with a foreword by B. Tukhliev. - Tashkent, 2014. - p. 13. See also: Ulukov N., Abduhamidova L. Interpretations and Analyses of Speech Culture. – Tashkent, 2017. – P. 35.

[5] Garden of Thought. – Tashkent: Gafur Gulom Literature and Art Publishing House, 1989. – P. 165.

[6] Saberi K., Perrott D. R. (1999). Cognitive restoration of reversed speech. Nature, 398(6730): 760. 5.

[7] Jertberg RM, Soto-Faraco S, van Wassenhove V, Van der Burg E. Temporal Window of Integration XOR Temporal Window of Synchrony. Multisens Res. 2025 Oct 31; 38(6-8): 353-367. https://doi.org/10.1163/22134808-bja10171

[8] Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2006). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

[9] Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*, 725. https://doi.org/10.3389/fpsyg.2014.00725

[10] Eggermont, J. J. (2017). Multisensory processing. In *Elsevier eBooks* (pp. 71–90). https://doi.org/10.1016/b978-0-12-805398-0.00003-7

[11] Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, *121*(1), 22–40. https://doi.org/10.1016/j.cognition.2011.06.009

[12] Atal, B. S., & Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, *50*(2B), 637–655. https://doi.org/10.1121/1.1912679

[13] Alves, J. D. N., De Almeida, A. a. F., Yamasaki, R., & Lopes, L. W. (2024). The influence of listener experience, measurement scale and speech task on the reliability of auditory-perceptual evaluation of vocal quality. *CoDAS*, *36*(3), e20230175. https://doi.org/10.1590/2317-1782/20232023175

[14] Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. L., & Imada, T. (2014). Infants' brain responses to speech suggest Analysis by Synthesis. *Proceedings of the National Academy of Sciences*, *111*(31), 11238–11245. https://doi.org/10.1073/pnas.1410963111

[15] Bever, T. G., & Poeppel, D. (2010). Analysis by Synthesis: A (Re-)Emerging Program of Research for Language and Vision. *Biolinguistics*, *4*(2–3), 174–200. https://doi.org/10.5964/bioling.8783