DOI 10.1007/s11055-024-01619-5
Neuroscience and Behavioral Physiology, Vol. 54, No. 3, March, 2024

Experimental Studies

Features of the Recognition of Speech Signals in Conditions of Vocal Competition in Health and in Impairments to Auditory Speech Function

A. A. Balyakova, O. V. Labutina, I. S. Medvedev, S. P. Pak, and E. A. Ogorodnikova

UDC 612.85+616.8

Translated from Sensornye Sistemy, Vol. 37, No. 4, pp. 342–347, October–December, 2023. Original article submitted September 5, 2023. Accepted September 25, 2023.

We report here studies of the characteristics of the perception of speech signals in conditions of vocal competition based on a sex-related characteristic (male/female voice) in subjects of different ages and states of auditory-verbal function. Psychophysical measurements were carried out during simulation of a "speech cocktail" situation by simultaneous pronunciation of different words by male and female speakers. The mean fundamental tone frequency (FTF) of the male speaker was 108 ± 5.92 Hz, while that of the female speaker was 185 ± 12.03 Hz. Both speakers were standard native speakers of the Russian language. Digital recordings were equalized in intensity and mixed to generate a total test stimulus consisting of a mixture of words spoken by a male (M) voice and a female (F) voice, with synchronization of sounding start times. Test speech signals were presented via headphones or through a loudspeaker positioned in front of the listener at a distance of 50 cm. Measures of reaction time and the number of correct recognitions of words from the target speaker (M or F) were compared in four groups of subjects of different ages and hearing-speech status: adult subjects with normal hearing and speech (n = 35), adult subjects with hearing impairment (n = 26); schoolchildren with normotypical development (n = 26) and schoolchildren with speech disorders (n = 25). The results of this comparative study demonstrate deterioration in the ability to isolate target speech streams in conditions of vocal competition in hearing- and speech-impaired subjects. Differences in the perception of male and female voices were identified in subjects with hearing loss and speech problems, and this may have a biological and social basis. These results have practical significance for the development of auditory-speech training systems and modern hearing aid technologies.

Keywords: voice competition, recognition of speech signals, speech cocktail, sex-related differences in voice, hearing loss, speech disorders, auditory-verbal training.

Introduction. Auditory speech perception is characterized by some degree of noise immunity and selectivity, allowing people to detect and identify a target speaker on the backgrounds of acoustic interference and the sounding of other sound and speech sources. These properties of hearing play an important role in speech communication in different communication conditions and provide people with the ability to focus on one speaker, ignoring other interlocutors and surrounding noise. The problem solved by the auditory system in these situations is usually termed the "party problem" or "the cocktail-party problem" [Cherry, 1953]. Decades of research has not only addressed the spatial selectivity of hearing speech [Bronkhost, 2015; Andreeva, 2018], but has also developed the direction of "auditory scene analysis" in psychoacoustics, this being focused on studies of the mechanisms by which sound and speech streams in complex acoustic environments are separated and combined (perceptual grouping) [Bregman, 1990].

Results from experimental studies have shown that the following parameters are significant for the implementation of auditory analysis processes in difficult conditions with a

Pavlov Institute of Physiology, Russian Academy of Sciences, St. Petersburg, Russia; e-mail: ogorodnikovaea@infran.ru.

spatial component: the proximity of the spectral-temporal characteristics of sound or speech sources, the synchronicity of their operation, and separation in space [Bregman, 1990; Shamma et al., 2011 ; Gutschalk and Dykstra, 2014; Bronkhost, 2015; Andreeva, 2018]. In addition to the effects of spatial release from masking mediated by binaural perception mechanisms, the properties of the speech signal itself, which contribute to solving the "party" problem, should also be noted. First, this refers to concentration of the energy of the speech signal in limited spectral regions and to its redundancy, which allows missing or masked elements of the speech stream to be "restored" at the perceptual level [Kalikow et al., 1977; Moore, 2012; Fogerty et al., 2015].

In scenes with no significant spatial component, the perceptual grouping of the speech stream is largely determined by the sex-related and individual characteristics of the speaker's voice (pitch, timbre). Some contribution is made by phonemic coherence, pronunciation characteristics, and the context of the speech message, as well as cognitive factors, particularly selective attention [Shamma et al., 2011; Moore, 2012; Gutschalk and Dykstra, 2014; Popper and Fay, 2015]. In practical terms, studies of the processes of perception in complex acoustic environments are important for increasing rehabilitation effectiveness in people with impaired hearing and speech functions and for the development of technical means of designing hearing prosthetics.

The aim of the present work was to carry out a comparative assessment of measures of the perception of a target speech signal in conditions of vocal competition by subjects of different ages with normal and impaired hearing and speech.

Methods. The study used a technique based on simulating a complex communicative scene without a spatial component [Ogorodnikova et al., 2022]. Stimulation corresponded to a simplified speech cocktail scheme consisting of simultaneous pronunciation of different speech signals (isolated words) by two speakers, a man and a woman. Both speakers were standard native speakers of the Russian language. The mean fundamental frequency (FTF) for the male voice was 108 ± 5.92 Hz and that for the female voice was 185 ± 12.03 Hz. Digital recordings were equalized in intensity and mixed in such a way that the total test stimulus was a mixture of words spoken by a male voice (M) and a female (F) voice, with synchronization of sounding start times. In total, six pairs of words were used: /*Yagoda* + Army/, /*Boroda* + Baraban/, /Vorobei + *Bereza*/, /Bumaga + *Derevo*, /Divany + *Limony*/, /Yagoda + *Vygoda*/,[1] in which the words of the male speaker are shown in roman font and the words of the female speaker are shown in italics.

Test speech signals were presented via headphones or through a loudspeaker positioned in front of the auditor at a distance of 50 cm. These conditions provided vocal competition and a procedure for selectively highlighting target words by sex-related, i.e., female/male voice. Subjects' numbers of correct recognitions and reaction times were assessed. The experiments were carried out at a comfortable level of stimulation using the "Learn to Listen" system developed by specialists from the Pavlov Institute of Physiology, Russian Academy of Sciences, and the St. Petersburg Institute of Ear, Throat, Nose, and Speech, Ministry of Health of the Russian Federation [Koroleva et al., 2013].

Four groups of subjects took part in the experiments: adult subjects aged 18–25 years with normal hearing and speech functions (n = 35), adults with hearing impairment (grade III sensorineural hearing loss, rehabilitation after cochlear implantation, n = 26), schoolchildren aged 10–14 years with normotypical development (n = 26) and schoolchildren with speech disorders: general speech underdevelopment (GSU) or signs of dyslexia or dysgraphia (n = 25). All procedures performed in the study involving subjects complied with the requirements of the Ethics Committee of the Pavlov Institute of Physiology, Russian Academy of Sciences, and the Declaration of Helsinki 1964 with its subsequent amendments.

Statistical processing of results used Student's t test for unrelated samples with distributions checked for normality using the Anderson–Darling test; the nonparametric Wilcoxon test was used for dependent samples.

Results and Discussion. The data obtained here showed that adult auditors and schoolchildren with normal hearing and speech successfully selected and recognized target words in conditions of vocal competition, with mean levels of correct recognition of target words pronounced in a male or female voice which were greater than the significant recognition level of 75% of the signals presented. These groups also showed the shortest reaction times, on average no greater than 3 sec (Table 1).

People with hearing impairments experienced the greatest difficulties in completing the task. This primarily applied to prelingually deaf patients after cochlear implantation (n = 10) at the first stage of auditory-speech rehabilitation. Performance in most of these cases was below 50% of correct answers and averaged 44.9 ± 3.4% (recognition) and 5.4 ± 0.2 sec (reaction time). Schoolchildren with speech impairments also failed to reach the level of reliable recognition. These subjects displayed scatter in the individual data, while mean reaction times were greater than the corresponding values in both adults and schoolchildren with normal auditory-speech function. Thus, significant differences were found in the perception of target speech signals in subjects of different ages with hearing or speech impairments in conditions of vocal competition as compared with normal values of indicators in the corresponding reference groups (normal hearing and normotypical development). In addition to hearing loss, this result, especially in subjects with cochlear implants and speech impairments, may be a

[1] Translator's note: words are given as transcripts of the Russian words as phonology may be a factor influencing results. In English: /*Berry* + Army/, /*Beard* + Drum/, /Sparrow + *Birch*/, /Paper + *Tree*/; /Sofas + *Lemons*/, /Berry + *Benefit*/.

TABLE 1. Mean Recognition Rates and Reaction Times on Perception of Competing Speech Signals in Reference Groups ($M \pm m$)

Reference groups	Adult subjects		Reference groups	Schoolchildren	
	N (%)	Reaction time (sec)		*N* (%)	Reaction time (sec)
Normal hearing	93.4 ± 1.1	2.5 ± 0.1	Normotypical development	86.9 ± 1.7	2.9 ± 0.2
Haring impairment	52.2 ± 2.9***	4.9 ± 0.2***	Speech impairment	69.1 ± 2.4***	3.3 ± 0.3**

N is the number of correct target word recognitions (%); significant differences compared with data from the corresponding normal groups: ***$p < 0.001$, **$p < 0.05$ (Student's *t* test for unrelated samples).

result of insufficient development of the central auditory analysis processes responsible for the perceptual grouping of speech streams based on the characteristics of the talker's voice [Koroleva et al., 2017]. This may be caused by an initial deficit of sensory experience (prelingual deafness, the initial stage of rehabilitation after implantation surgery, severity of hearing loss) and manifestations of central auditory disorders [Koroleva et al., 2017; Boboshko et al., 2014; Musiek and Chermak, 2014; Koroleva, 2022].

Differences associated with sex-related voice characteristics were also seen in the reference groups. These were more marked in subjects with hearing and speech impairments, but also appeared in schoolchildren with normotypical development (Fig. 1).

This shows that children with normal and impaired speech displayed significantly better ($p < 0.01$, Wilcoxon test) recognition of speech targets spoken in a female voice. At the same time, subjects with hearing problems, conversely, performed significantly better at identifying the words of a male speaker ($p < 0.01$, Wilcoxon test). It can be suggested that these differences have both biological and social bases. In subjects with reduced hearing or prelingual deafness preceding cochlear implantation, this may be a manifestation of the features of perceptual experience formed on the basis of residual hearing in the low-frequency region. This may contribute to more successful word recognition by a male speaker with lower FTF and lower voice pitch. In schoolchildren with normal hearing, differences in sex-related voice characteristics may be determined by the characteristics of communicative interaction during early childhood, which is realized mainly in "mother and child" dyads [Gaykova and Lyakso, 2011]. This is indirectly confirmed by reactions to mothers' voices as recorded in a study of psychosomatic processes and seen in clinical practice [Erkudov et al., 2019; Efendi et al., 2018]. In addition to biological connections, the perceptual strengthening of the female voice may be a result of the predominance of female educators and teachers in preschool and school educational institutions [Grinenko, 2014], especially among specialists involved in correctional work. The influence of these factors was also evident in a group of schoolchildren with normotypical development, where a definite predominance of female voices was also observed in recognition results ($p < 0.05$, Wilcoxon test). Data from adult subjects in the normal group show that this gradually disappears with age.

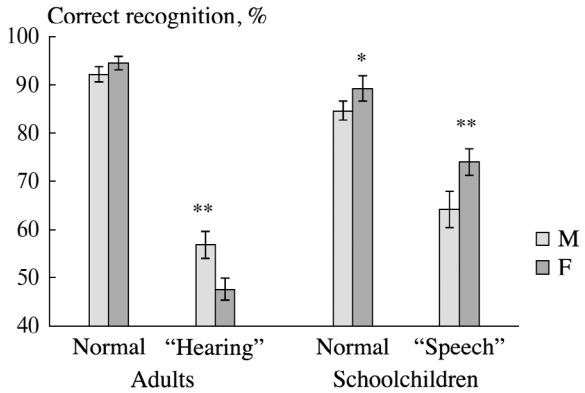

Fig. 1. Indicators of correct target word recognition as pronounced by a male (M) voice and a female (F) voice. The horizontal axis shows symbols for groups of subjects. The vertical axis shows numbers of correct recognitions (%). Significant differences, *$p < 0.05$ and **$p < 0.01$, Wilcoxon test (taking into account zero shifts).

Conclusions. In general, measurement results captured in conditions of vocal competition point to deterioration in the ability to isolate target speech streams and recognize speech in complex acoustic environments in people with impaired auditory-speech function, both from the point of view of auditory perception and in cases of problems in speech development. This deterioration is determined by insufficient formation of the central mechanisms for the auditory analysis of complex acoustic scenes due to lack of sensory experience. Taking a course of auditory-speech training can significantly improve this situation, primarily in patients in the initial stages of rehabilitation after cochlear implantation [Ogorodnikova et al., 2017; Koroleva, 2022]. In schoolchildren with speech impairments, low recognition rates in competitive conditions may also reflect problems with the central processing of acoustic information, which is in good agreement with data from an earlier study of auditory segmental analysis processes in children with speech, reading, and writing impairments [Ogorodnikova et al., 2012].

These results have practical significance for the development of auditory-speech training systems and modern hearing aid technologies. They confirm the advisability of including exercises addressing speech signal perception in conditions of vocal competition in perceptual training programs. When preparing such programs, new data on differences in the perception of sex-related voice characteristics, which are associated with age and the nature of hearing im-

pairment, should be taken into account. This applies, in particular, to the use of expanded databases of speakers' voices and recordings of speech material (words, syllables, short phrases, and much more).

This work was supported by state budget funds under a state assignment (topic No. AAAA-A18118050790159-4).

The authors would like to thank Head Researcher at St. Petersburg Science Research Institute of the Ear, Throat, Nose, and Speech Professor I. V. Koroleva and Defectology Lecturer N. Yu. Belova at School No. 10, Kalininskii District, St. Petersburg for assistance in conducting the study.

All authors took equal part in preparing and processing material for publication.

The authors declare that there are no obvious or potential conflicts of interests related to publication of this article.

REFERENCES

Andreeva, I. G., "Spatial selectivity of hearing in speech recognition in speech-shaped noise environment," *Human Physiol*., **44**, No. 2, 226–236 (2018), https://doi.org/10.1134/.

Boboshko, M. Yu., Garbaruk, E. S., Zhilinskaya, E. V., and Salakhbekov, M. A., "Central auditory disorders (literature review)," *Ross. Otorinolaringol*., No. 5, 87 (2014).

Bregman, A. S., *Auditory Scene Analysis: the Perceptual Organization of Sound*, MIT Press, Cambridge (1990).

Bronkhorst, A. W., "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," *Atten. Percept. Psychophys*., **77**, No. 5, 1465–1487 (2015), https://doi.org/10.3758/s13414-015-0882-9.

Cherry, E. C., "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am*., **25**, No. 5, 975 (1953).

Efendi, D., Caswini, N., Rustina, Y., and Iskandar, A. D., "Combination of mother therapeutic touch (MTT) and maternal voice stimulus (MVS) therapies stabilize sleep and physiological function in preterm infants receiving minor invasive procedures," *J. Neonatal Nurs*., No. 6 (24), 318–324 (2018), https://doi.org/10.1016/j.jnn.2018.08.001.

Erkudov, V. O., Ogorodnikova, E. A., Pugovkin, A. P., et al., "Isolating the voice of the target speaker in conditions of speech competition among schoolchildren with different psycho-emotional status," *Pediatr.*, **10**, No. 4, 51–59 (2019), https://doi.org/10.17816/PED10451.

Fogerty, D., Ahlstrom, J. B., Bologna, W. J., and Dubno, J. R., "Sentence intelligibility during segmental interruption and masking by speech-modulated noise: Effects of age and hearing loss," *J. Acoust. Soc. Am*., **137**, No. 6, 3487–501 (2015), https://doi.org/10.1121/1.4921603.

Gaikova, Yu. S. and Lyakso, E. E., "Individual contribution of maternal speech characteristics to the speech development of a child in the first year of life," *Vestn. Sankt-Peterburg. Univ. Ser. 3 Biol*., No. 3, 66–74 (2011).

Grinenko, S. V., "Gender asymmetry in education," *Sovrem. Nauchn. Issled. Innov*., No. 12 (3) (2014), http://web.snauka.ru/issues/2014/12/41818.

Gutschalk, A. and Dykstra, A. R., "Functional imaging of auditory scene analysis," *Hear. Res*., **307**, 98 (2014).

Kalikow, D. N., Stevens, K. N., and Elliott, L. L., "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am*., **61**, No. 5, 1337–1351, PMID: 881487 (1977), https://doi.org/10.1121/1.381436.

Koroleva, I. V., *Basic Audiology and Hearing Aids*, KARO, St. Petersburg 2022.

Koroleva, I. V., Ogorodnikova, E. A., Pak, S. P., and Levin, S. V., "The importance of central hearing mechanisms in the restoration of speech perception in deaf patients after cochlear implantation," *Spets. Obrazov*., No. 3, 100 (2017).

Koroleva, I. V., Ogorodnikova, E. A., Pak, S. P., et al., "Methodological approaches to assessing the dynamics of development of auditory-speech perception processes in children with cochlear implants," *Ross. Otorinolaringol*., No. 3, 75–85 (2013).

Moore, B. C. J., *An Introduction to the Psychology of Hearing*, Brill, Leiden (2012).

Musiek, F. E. and Chermak, G. D., *Handbook of Central Auditory Processing Disorders*, Vol. 1, *Auditory Neuroscience and Diagnosis*, Plural Publishing, San Diego (2014).

Ogorodnikova, E. A., Balyakova, A. A., Zhilinskaya, E. V., et al., "Auditory training as a method of rehabilitation in patients with hearing and speech impairments," *Folia Otorhinolaryng. Pathol. Respir.*, **23**, 1, 33 (2017).

Ogorodnikova, E. A., Labutina, O. V., and Pak, S. P., "Simulation of complex acoustic scenes with stimulation through headphones," *Vestn. Psikhofiziol*., No. 2, 140–146 (2022), https://doi.org/10.34985/o0640-6924-4290-f.

Ogorodnikova, E. A., Stolyarova, E. I., and Balyakova, A. A., "Features of auditory-speech segmentation in school-age children with normal hearing and with hearing and speech impairments," *Sens. Sistemy*, **26**, No. 1, 20–31 (2012).

Popper, A. N. and Fay, R. R. "Perspectives on auditory research," in: *Springer Handbook of Auditory Research* (2014).

Shamma, S. A., Elhilali, M., and Micheyl, C., "Temporal coherence and attention in auditory scene analysis," *Trends Neurosci*., **34**, 114 (2011).