DOI: 10.61782/fa.2023.0156

forum acusticum 2023

Development of speech perception in noise: Effect of auditory scene analysis and musical abilities

Benocci Elena[1] **Axelle Calcus**[1]

[1] Center for Research in Cognitive Neuroscience (CRCN), Université libre de Bruxelles (ULB), Brussels, Belgium

ABSTRACT*

From classrooms to playgrounds, children communication occurs in noisy environments. Yet children and adolescents experience more difficulties than adults when perceiving speech in noise. Speech intelligibility in noise is likely influenced by selective auditory tracking, an aspect of auditory scene analysis that only starts improving at adolescence. Interestingly, musical abilities also seem to contribute to the development of speech perception in noise. The aim of this study was to investigate the respective contribution of auditory scene analysis (via stream segregation and selective auditory tracking) and perceptual musical abilities on the development of speech intelligibility in noise. Our results suggest a developmental improvement on the mechanisms of auditory scene analysis and speech intelligibility in noise. Furthermore, musical abilities predict auditory scene analysis, which in turn predicts speech-in-noise perception. Importantly, the mechanisms involved in auditory scene analysis and speech perception in noise likely continue to develop throughout adolescence, see Tiernye et al., 2019

Keywords: *Speech Perception In Noise - Development - Auditory Scene Analysis – Musical Abilities*

1. INTRODUCTION

Perceiving speech in noisy environments is a fundamental ability for human communication. However, this task can be challenging due to the interference of background noise, which degrades the quality of the speech signal and impedes its intelligibility. The development of speech perception in noise is a complex process that involves various perceptual, neural and cognitive mechanisms. Understanding how the perception of speech in noise develops is an important topic of research in the field of auditory cognition and speech perception.

In complex auditory environments, sound waves are mixed before reaching the ears. In the presence of interfering sounds, listeners must segregate concurrent auditory streams, to selectively track the stream of interest over time, hence performing what is called the auditory scene analysis[1]. Auditory tracking refers to the ability to follow a target stream over time, while ignoring concurrent streams. How auditory scene analysis develops in the first decades of human life remains poorly understood.

The peripheral auditory system is functionally mature soon after birth [2], but complex auditory processes such as speech perception in noise follow a protracted developmental trajectory [3]. Speech intelligibility in noise progressively improves between 9 and 17 years of age. The exact timing of maturation remains uncertain[4-6], and

* ***Corresponding author***: elena.benocci@ulb.be

likely depends on the nature of the background noise. The ability to perceive speech in the presence of interfering talkers may improve until at least 13 years[7-8], possibly even 16-17 years of age[9-11].Overall, immature attentional abilities, which could be associated with neural network development, may contribute to difficulties encountered by young listeners in noisy environments[12-13] .

To our knowledge, no study has specifically examined the ability of young individuals to selectively track an auditory in the presence of a similar interfering stream. Children aged between 9 and 11 years exhibit poorer stream segregation performance compared to adults. Children need larger frequency separations to distinguish two streams from each other. Developmental differences in stream segregation may be due to immature auditory scene analysis abilities in children[14-15].

Amongst the many factors that influence auditory perception is musicianship. Musicians are thought to exhibit better auditory abilities than non-musicians[16-17]. As such, musical training likely enhances speech perception in noise. Several studies have demonstrated a musician advantage in speech perception in noise with multiple interfering talkers or speech-shaped noise[18-19]. Indeed, musicians are more sensitive to changes in pitch, timing, and other acoustic characteristics of speech compared to non-musicians[20-21]. Some studies suggest that music and speech processing share common neural structures, which could explain the positive impact of musical training on speech abilities[22][23]. However, the transfer of musical abilities to speech perception in noise is still debated[24].

Only a few studies have investigated the impact of musical expertise on auditory scene analysis mechanisms. They have shown that musicians perform better than non-musicians in auditory segregation tasks based on harmonicity. Musicians require a smaller frequency difference to perceive two sounds as distinct[20]. Additionally, the ability of auditory tracking was found to be correlated with the number of years of musical training[21].

One limitation of existing studies is the categorization of individuals into musicians versus non-musicians, which may not accurately reflect the full range of musical abilities within the population[25, 26]. Some individuals may have good musical perceptual skills without formal musical training. Therefore, musical abilities should be considered as a continuous auditory perceptual skill that can be developed and improved through practice, rather than a binary characteristic[27-28]. Last, the criteria for musicianship categorization varies widely across studies.

Our study aims to (i) assess young listeners' auditory scene analysis (by means of stream segregation and selective auditory tracking) and (ii) how this process relates to the development of speech perception in noise, (iii) taking musical abilities into account.

We hypothesize that there is a developmental effect on speech perception in noise and auditory scene analysis. Musical abilities may predict auditory scene analysis skills, as musicians have been shown to have superior auditory abilities and these skills may transfer to other auditory tasks, such as speech perception in noise.

2. METHOD

A total of 76 children (ages ranging from 8 to 12 years) and 73 adults (ages ranging from 18 to 28 years) took part in this experiment. Adult participants were recruited from the participant pool of Université Libre de Bruxelles (ULB). Children were recruited from several schools in Brussels. The study was approved by the ethical committee of ULB Faculty of Psychology and the ethical committee of Hopital Universitaire des Enfants Reine Fabiola. Before the experiment, informed consent was obtained from all participants.

Participants reported no hearing, neurological or developmental disorders. All participants presented normal audiometric thresholds as measured using the smartphone-based application Mimi Hearing Test. All participants had normal working memory evaluated by WAIS-IV (adults) and WISC-V (children).

All participants were native speakers of French. All completed all experimental tasks. They took part in the study within their educational institution (school or university). Stimuli was presented through headphones connected to a tablet.

2.1 Stimuli

2.1.1 Musical perceptive scale

Participants completed the MICRO-PROMS battery[29] which objectively evaluates musical perceptive abilities.

It is a shorter version (approximately 10 minutes) than the original PROMS battery[27]. The MICRO-PROMS provides an objective scale of musicality for measuring the perception of distinct musical dimensions, using a 3IAFC procedure.

2.1.2 Auditory segregation

Auditory segregation stimuli were based on a stochastic figure-ground task[30]. Trials consisted of a 2 s long sequence of chords with 0 ms of inter-chord interval. Each chord lasted for 50 ms and contained 5 to 15 pure tones components. In 2/3 of trials (40 trials) a figure was overlaid on the background. The figure consisted of the coherent repetition of 8 pure tones that were repeated over 7 chords (hence lasting 350 ms). The figure appeared between 750 ms and 1 s after the onset of the trial. The remaining 1/3 of trials consisted merely of the background presentation, without any figures.

Participants were asked to indicate whether each trial contained a figure. At the beginning of the task, participants were presented with 5 familiarization trials. Feedback was provided after every trial. This task was implemented in the online testing platform Gorilla[33].

2.1.3 Auditory tracking

The stimuli for the auditory tracking task were adapted from [31]. Participants were asked to listen to a mixture consisting of two competing synthetic voices whose fundamental frequency (F0) and first formants changed progressively over a 2 s duration.
At the beginning of each trials, a cue was displayed before the mixture. It corresponded to the initial portion (400ms) of the target stream. The cue indicated which stream the participant needed to pay attention to. The mixture was then followed by a probe, which corresponded to the final portion of either the cued or the uncued stream.
Participants were instructed to identify whether the probe corresponded to the final portion of the target stream or the interfering stream. To facilitate auditory tracking of the cued stream, it started 50 ms before the uncued stream in the mixture. The signal-to-noise ratio (SNR) was 0 dB SNR.
Participants were first presented with 5 training trials. Feedback was provided to the participants. This task was also implemented in a game-based version on the Gorilla platform.
As a familiarization, a 1-stream version of this task was presented as a baseline condition. Similarly to the "2-streams" condition, the target streams were preceded by a cue that corresponded to the initial portion of the target stream. In this condition, the probe corresponded to the final portion of the cued stream or to a random stream. Participants were asked to determine whether the probe coincided with the target voice or not.

2.1.4 Speech perception

The paradigm was similar to that used by Calcus et al, 2016[32]. Participants were instructed to listen to a set of 32 vowel-consonant-vowel logatomes in three acoustic environments: in quiet, in the presence of an interfering speaker (1-talker), and in the presence of speech-shaped noise (SSN). The logatomes consisted of two repetitions of 16 /aCa/ utterances. The C corresponded to consonants from the following list: /p, t , k, b, g, f, s, m, n, r, l, v, z, j, ʃ/. The logatomes were spoken by a French speaker and lasted no longer than 500ms.
The one-speaker masker consisted of recordings of French media clips produced by French male speakers. The SSN masker was a derivation of the 1-talker recordings. A new signal was created by keeping the power spectrum but randomizing the phases. To generate the envelope-modulated SSN, a fast Fourier transform was used. In both noise conditions, the masker lasted 2s.
The participants' task was to identify the consonant pronounced by the speaker. They indicated their responses on a confusion matrix. No feedback was provided during this task.

3. RESULTS

3.1 Statistical analyses

First, a univariate ANOVA was computed to evaluate the effect of age and noise condition on speech intelligibility in noise. Speech intelligibility significantly improved with age regardless of the acoustic condition [$F(1, 122) = 52.66$, $p < 0.001$].
A linear regression model was conducted to investigate the effect of age on auditory segregation. Auditory segregation significantly increased with age [$F(1, 132) = 49.99$, $p < 0.001$].
Similarly, a linear regression model was performed to investigate the effect of age on auditory tracking. Auditory tracking significantly increased with age [$F(1, 132) = 12.93$, $p < 0.001$]. For a description of the statistical data, refer to Table 1.

To determine the relationship between auditory scene analysis, speech perception in noise and musical perception abilities, we used structural equation modelling (SEM) with path analysis. Path analysis is a statistical technique that enables investigation of the strength and direction of relationships between variables. It helps to identify the ways in which one variable may influence another, and how this influence is transmitted through a system[34-35].

To simplify our model, we organized certain variables into latent variables, namely Auditory Scene Analysis mechanisms and Speech perception. Auditory segregation and auditory tracking are observable variables that reflect the intricate process of auditory scene analysis, which cannot be directly measured. Similarly, we grouped the three consonant perception conditions under the latent variable "Speech Intelligibility". By grouping observable variables under latent variables, we can succinctly and accurately represent the causal relationships between variables, facilitating a clearer and easier interpretation of our analysis[35].

A structural equation model was used to examine the relationship between musical perception abilities, auditory scene analysis, and speech perception (see Table 2 for standardized beta and p-values). The model (see Figure 1) demonstrated a good fit to the data, with values of SBS-χ2 (11) = 14.351 p = .214; CFI = .983; RMSEA = .05; SRMR = .032. The analysis revealed a significant direct path between musical perceptual abilities and auditory scene analysis (β= 0.27, p = .026). Additionally, the relationship between musical perceptual abilities and speech perception was found to be mediated by auditory scene analysis (β= 0.55, p = .028). These results suggest that there is a pathway from musical perceptual abilities to speech perception via auditory scene analysis. Last, the model revealed a significant direct path between age and auditory scene analysis (β= 0.59, p <.001) and between age and speech intelligibility (β= 0.31, p = .037).

3.2 Figures and tables

Task name	Group	Mean	Std. Deviation
Auditory tracking	Children	55.59	9.85
	Adults	62.72	15.48
Auditory Segregation	Children	62.43	11.18
	Adults	76.25	10.83
Speech-in-quiet	Children	93.35	6.38
	Adults	97.57	4.06
Speech-in-noise (1-talker)	Children	79.26	13.66
	Adults	93.44	8.21
Speech-in-noise (speech-shaped-noise)	Children	75.41	10.64
	Adults	85.05	11.23

Table 1 : Descriptive results (mean and standard deviation) of the different tasks by age groups.

Latent variables	Measured variables	Beta standardized	p-value
Auditory Scene Analysis	*Auditory Tracking*	*.53*	*< .001*
	Auditory Segregation	*.73*	*< .001*
Speech perception	*Speech in quiet*	*.59*	*< .001*
	Speech in 1-talker	*.73*	*= .002*
	Speech in speech-shaped noise	*.62*	*< .001*

Table 2: Table of standardized beta values and p-values for latent variables and their composites. The beta values represent the strength and direction of the relationship between the latent variable and the criterion variable, with positive values indicating a positive relationship and negative values indicating a negative relationship.

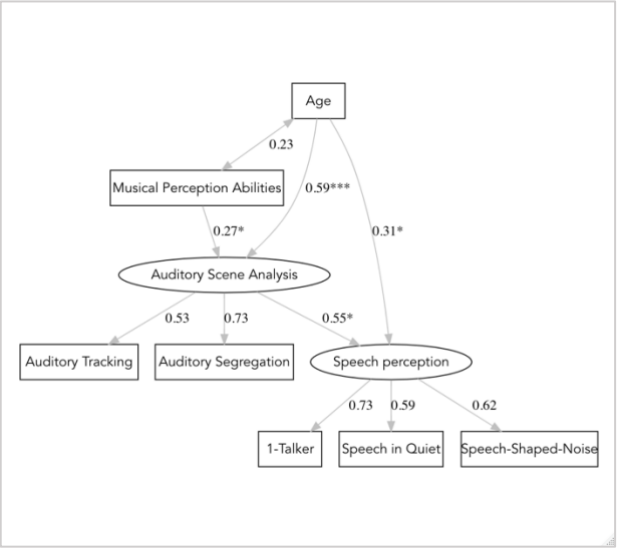

Figure 1: Structural equation modeling (SEM) path analysis diagram representing the relationships between musical perception, auditory scene analysis skills and speech intelligibility. The solid lines represent the direct effects of each variable. Coefficients along the lines indicate the strength and significance of the relationships, with values closer to 1 indicating a strong positive relationship and values closer to 0 indicating a weaker relationship. The circles represent latent variables, and the squares represent observed variables.

4. DISCUSSION

Our results are consistent with our initial hypotheses. Age significantly predicts performance in both auditory tracking and segregation tasks, as well as speech perception in noise: children are overall poorer than adults in these tasks. Thus, the mechanisms involved in auditory scene analysis and speech perception in noise likely continue to develop throughout adolescence.

Second, musical abilities do not directly impact speech intelligibility in noise, but the relationship between the two is mediated by auditory scene analysis abilities. This finding could explain the contradictory results in previous studies regarding the musician advantage for speech perception in noisy environments[24]. Our study, like others[36-37], does not find a direct advantage for people with good musical abilities in speech perception. In fact, high musical abilities predict good auditory scene analysis abilities, which in turn contribute to speech intelligibility in noise.

The direct relationship between musical abilities and auditory scene analysis abilities in our study is in line with the limited literature on this topic [20, 38]. While some previous research has found a correlation between musical skills and auditory segregation, only one study has examined the relationship between musical abilities and auditory tracking.
The direct relationship between the ability to analyze auditory scenes and the ability to understand speech in noise is coherent with existing theories suggesting that the ability to differentiate and analyze the sounds in the presence of interferers enhances speech intelligibility[39].

Our study provides evidence for the role of development, auditory scene analysis abilities, and musical abilities on speech perception in noise. These findings have implications for understanding the factors that influence speech perception in challenging listening conditions.

5. ACKNOWLEDGMENTS

This work was supported by a grant from Belgian Kids' Fund for Pediatric Research awarded to Elena Benocci. We are grateful to Luna Leonardy and Esperance Moutikila Kanama for their help with data collection.

6. REFERENCES

[1] Bregman, A. (2015) Progress in understanding auditory scene analysis. *Music Perception* 17, 106–109.

[2] Cramer, K. S., Coffin, A. B., Fay, R. R.; Popper, A. N. (2017) *Auditory Development and Plasticity : In Honor of Edwin W Rubel Springer Handbook of Auditory Research;* Springer international publishing: Cham, Switzerland.

[3] Leibold, L. J. & Buss, E. (2019) Masked Speech Recognition in School-Age Children. *Frontiers in Psychology* 1–8.

[4] Cabrera, L., Varnet, L., Buss, E., Rosen, S. & Lorenzi, C. (2019). Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity. *J Acoust Soc Am,* 146, 2415–2429.

[5] Bonino, A. Y., Leibold, L. J. & Buss, E. (2013). Release From Perceptual Masking for Children and Adults. *Ear & Hearing* 34, 3–14.

[6] Buss, E., Leibold, L. J., Porter, H. L. & Grose, J. H. (2017). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *J Acoust Soc Am* 141, 2650–2660.

[7] Baker, M., Buss, E., Jacks, A., Taylor, C. & Leibold, L. J. (2014). Children's Perception of Speech Produced in a Two-Talker Background. *J Speech Lang Hear Res* 57, 327–337.

[8] Corbin, N. E., Bonino, A. Y., Buss, E. & Leibold, L. J. (2016). Development of Open-Set Word Recognition in Children. *Ear Hearing* 37, 55–63.

[9] Elliott, L. L., Connors, S., Kille, E. & Levin, S. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. *J Acoust Soc Am* 66, 12.

[10] Wightman, F. & Kistler, D. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *J. Acoust. Soc Am* 118, 3164–3176.

[11] Evans, S. & Rosen, S. (2021). Who is Right? A Word-Identification-in-Noise Test for Young Children Using Minimal Pair Distracters. *J Speech Lang Hear Res* 1–10 .

[12] Caballero, A., Granberg, R. & Tseng, K. Y. (2016). Mechanisms contributing to prefrontal cortex maturation during adolescence. *Neurosci Biobehav Rev* 70, 4–12.

[13] Leibold, L. J. (2011). Development of Auditory Scene Analysis and Auditory Attention. in vol. 42 137–161.

[14] Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., & Taylor, M. J. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cognitive Brain Research*, *16*(2), 210-218.

[15] Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, *225*(1-2), 117-127.

[16] Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiology of Aging*, *33*(7), 1483.e1-1483.e4.

[17] Alain, C., Zendel, B. R., Hutka, S., & Bidelman, G. M. (2014). Turning down the noise : The benefit of musical training on the aging auditory brain. *Hearing Research*, *308*, 162-173.

[18] Slater, J. & Kraus, N. (2016). The role of rhythm in perceiving speech in noise : A comparison of percussionists, vocalists and non-musicians. *Cogn Process*, 17, 79-87.

[19] Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound. *The Journal of Neuroscience*, 29(45), 14100-14107

[20] Zendel, B. R., & Alain, C. (2009). Concurrent Sound Segregation Is Enhanced in Musicians. *Journal of Cognitive Neuroscience*, *21*(8), 1488-1498.

[21] Tierney, A., Rosen, S., & Dick, F. (2019). Speech-in-speech perception, non-verbal selective attention, and musical training. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 46 (5), 968- 979

[22] Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus Song : Multiple Pitch-Sensitive Areas Revealed by a Naturally Occurring Musical Illusion. *Cerebral Cortex*, *23*(2), 249-254.

[23] Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98-108.

[24] Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians : A review. *Hearing Research*, *352*, 49-69.

[25] Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape

the neural encoding of speech. *Proceedings of the National Academy of Sciences*, *115*(51), 13129-13134.

[26] Swaminathan, S., Kragness, H. E., & Schellenberg, E. G. (2021). The Musical Ear Test : Norms and correlates from a large sample of Canadian undergraduates. *Behavior Research Methods*, *53*(5), 2007-2024.

[27] Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively : Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE*, *7*(12), e52508.

[28] Rajan, A., Shah, A., Ingalhalikar, M., & Singh, N. C. (2021). Structural connectivity predicts sequential processing differences in music perception ability. *European Journal of Neuroscience*, ejn.15407.

[29] Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively : Development and validation of the Short-PROMS and the Mini-PROMS: Assessing musical ability. *Annals of the New York Academy of Sciences*, *1400*(1), 33-45.

[30] Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife*, *2*, e00699.

[31] Woods, K. J. P., & McDermott, J. H. (2015). Attentive Tracking of Sound Sources. *Current Biology*, *25*(17), 2238-2246.

[32] Calcus, A., Lorenzi, C., Collet, G., Colin, C., & Kolinsky, R. (2016). Is There a Relationship Between Speech Identification in Noise and Categorical Perception in Children With Dyslexia? *Journal of Speech, Language, and Hearing Research*, *59*(4), 835-852.

[33] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. & Evershed, J. K. Gorilla in our midst: An online behavioral experiment builder. *Behav Res Methods* 52, 388–407 (2020).

[34] Hair, J. F., Ringle, C. M., & Sarstedt, M. (2012). Partial Least Squares : The Better Approach to Structural Equation Modeling? *Long Range Planning*, *45*(5-6), 312-319.

[35] Kline, R. B. (2013). Assessing statistical aspects of test fairness with structural equation modelling. *Educational Research and Evaluation*, *19*(2-3), 204-222.

[36] Boebinger, Dana, Evans, Samuel, Rosen, Stuart, Lima, C esar F., Manly, Tom, Scott, Sophie K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J. Acoust. Soc Am* 137 (1), 378e387.

[37] Ruggles, Dorea R., Freyman, Richard L., Oxenham, Andrew J., 2014. Influence of musical training on understanding voiced and whispered speech in noise. *PloS One* 9 (1), e86980.

[38] Zendel, B. R., & Alain, C. (2014). Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiology of Aging*, *35*(1), 55-63.

[39] Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114-123.