



OPEN ACCESS

EDITED BY

István Winkler,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences
(MTA), Hungary

REVIEWED BY

Alexandra Bendixen,
Chemnitz University of Technology, Germany
Brigitta Toth,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences
(MTA), Hungary

*CORRESPONDENCE

Axelle Calcus
✉ axelle.calcus@ulb.be

RECEIVED 07 December 2023

ACCEPTED 22 February 2024

PUBLISHED 12 March 2024

CITATION

Calcus A (2024) Development of auditory scene analysis: a mini-review.
Front. Hum. Neurosci. 18:1352247.
doi: 10.3389/fnhum.2024.1352247

COPYRIGHT

© 2024 Calcus. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of auditory scene analysis: a mini-review

Axelle Calcus*

Center for Research in Cognitive Neuroscience (CRCN), ULB Neuroscience Institute (UNII), Université Libre de Bruxelles, Brussels, Belgium

Most auditory environments contain multiple sound waves that are mixed before reaching the ears. In such situations, listeners must disentangle individual sounds from the mixture, performing the auditory scene analysis. Analyzing complex auditory scenes relies on listeners' ability to segregate acoustic events into different streams, and to selectively attend to the stream of interest. Both segregation and selective attention are known to be challenging for adults with normal hearing, and seem to be even more difficult for children. Here, we review the recent literature on the development of auditory scene analysis, presenting behavioral and neurophysiological results. In short, cognitive and neural mechanisms supporting stream segregation are functional from birth but keep developing until adolescence. Similarly, from 6 months of age, infants can orient their attention toward a target in the presence of distractors. However, selective auditory attention in the presence of interfering streams only reaches maturity in late childhood at the earliest. Methodological limitations are discussed, and a new paradigm is proposed to clarify the relationship between auditory scene analysis and speech perception in noise throughout development.

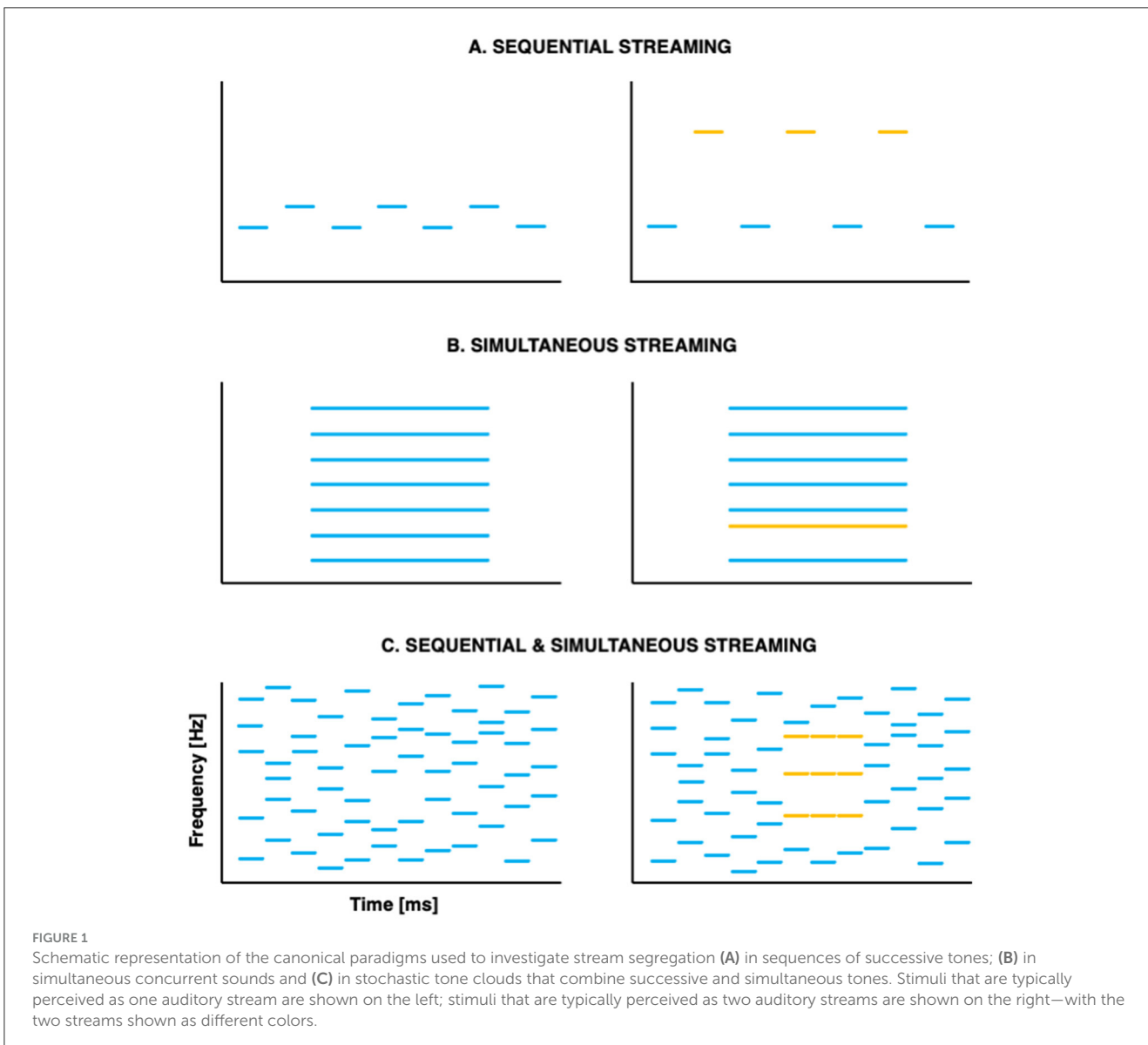
KEYWORDS

development, auditory scene analysis (ASA), streaming, selective attention, neurophysiology

1 Introduction

Contrary to appearances, lively playgrounds and business meetings have one thing in common: they are noisy. In such complex auditory environments, sound waves are mixed before reaching the ears. Listeners must disentangle individual sounds from the mixture, performing what is called the *auditory scene analysis* (ASA; [Bregman, 1990, 2015](#)). Analyzing complex auditory scenes relies on the listeners' ability to segregate acoustic events into different streams, and to selectively attend to the stream of interest.

With respect to segregation, pioneer studies used sequences of tones organized temporally in repeated ABABAB patterns, where A and B represent successive tones of different frequencies (e.g., [Miller and Heise, 1950](#); see [Figure 1A](#)). When listeners report hearing two streams, they are effectively experiencing stream segregation: they parse the *sequential* auditory events into distinct streams. At a given presentation rate, the larger the frequency distance between A and B, the more likely participants are to experience stream segregation. Later studies set out to evaluate segregation abilities in response to *simultaneous*, concurrent sounds. Listeners were presented with complex harmonic tones, of which one component had been mistuned ([Moore et al., 1986](#); see [Figure 1B](#)) or delayed ([Hedrick and Madix, 2009](#)); manipulations that contributed to segregation into distinct auditory objects. With respect to selective attention, canonical studies investigated adults' ability to focus on a specific auditory feature in the



presence of simultaneous or sequential distractors (e.g., Greenberg and Larkin, 1968).

A major limitation of these early studies is their focus on either sequential or simultaneous stimuli. However, in everyday life, broadband streams that are temporally correlated often overlap with one another. In such situations, temporal coherence between different elements of the auditory scene appears essential for auditory segregation (Elhilali et al., 2009), potentially guiding selective attention such that it binds together coherent acoustic (spectral, spatial, and/or temporal) features into streams (Shamma et al., 2011). In this view, attention contributes not only to stream selection, but also to stream formation. An interesting development of the past decade was the creation of a paradigm in which the spectral coherence varies across time, requiring listeners to perform both simultaneous and sequential streaming at once (Teki et al., 2011, 2013; see Figure 1C).

How ASA develops in the first decades of life has attracted a lot of interest over the years. So far, studies focused on paradigms that tackled either sequential or simultaneous ASA. In a comprehensive review published about a decade ago on the topic, Leibold (2011) showed that sequential stream segregation and selective attention are functional early in life, albeit not yet as efficient as they are in adulthood. At the time, the author identified several open questions regarding the development of ASA: (i) How does simultaneous ASA develop from infancy to adulthood? (ii) Which acoustic cues are used by infants/children to perform ASA? (iii) How does sensorineural hearing loss affect the development of ASA? Here, we aim to review recent developmental data that answer some of these questions or raise new interrogations. We focus on studies using non-linguistic stimuli, to illustrate the development of basic auditory perception and processing involved in ASA, without the confound of language abilities.

2 Stream segregation

2.1 First year of life

Pioneer studies of ASA development investigated sequential streaming in the 1st year of life by habituating infants to a repeating (forward) sound sequence, then measuring their dishabituation to a reversed version of the sequence. Should infants parse the auditory scene based on each individual sound of the sequence, they would show a dishabituation response to the reversed pattern. On the contrary, newborns and 3-month-olds appeared to parse the streams of complex auditory scenes using the same cues adults use (Demany, 1982; McAdams and Bertoncini, 1997; Smith and Trainor, 2011)—albeit less accurately (for a detailed review, see Leibold, 2011).

Later studies investigated the neural correlates of sequential segregation in infants, using the mismatch negativity (MMN). The brain generates a MMN when it processes a difference between an unexpected auditory stimulus (a deviant) and the neural representation of a standard, expected pattern (for a review, see Näätänen et al., 2012). In adults, this “oddball paradigm” would even entail an MMN in the presence of interleaved sounds of a different frequency, as long as the interleaved sounds are perceived as separate streams (Sussman et al., 1999). Presented with this “interleaved” oddball paradigm, newborns also show an MMN, indicating that the neural correlates of sequential stream segregation are functional from birth (Winkler et al., 2003). Seven-month-olds also show an MMN if the deviant is placed in a chord component, and successive chords are played as a sequence (Marie and Trainor, 2013). Note that in this case, infants, like adults (Fujioka et al., 2005), show larger MMN to a deviant in the high than low voice, supporting early emergence of a preference for the highest stream.

In the last decade, a number of studies have set out to investigate the early development of *simultaneous* ASA, answering one of the open questions identified by Leibold (2011). Folland et al. (2012) presented 6-month-old infants with complex tones consisting of 6 harmonic components. In half of the trials, one of the harmonic components was mistuned by 2–8% of its initial value. Infants were able to discriminate 4% mistuning or larger, whereas adults' thresholds were between 1 and 2% mistuning. Smith et al. (2017) paired in-tune and 8% mistuned complex tones with visual displays showing either one or two bouncing balls, hence being congruent or incongruent with the complex tones. Four-month-olds looked longer at incongruent audiovisual displays, indicating that they use harmonicity as a cue for stream segregation when integrating multisensory information. Whether newborns can segregate simultaneous auditory objects, or use acoustic cues to guide simultaneous streaming remains an open question.

To our knowledge, only two studies have investigated the neural correlates of simultaneous segregation in the 1st year of life, leading to contradictory results. Both studies used a similar paradigm, where half of the trials were 500 ms long complex tones of which the second harmonic was mistuned by 8% of its original value while the other half were in-tune complex tones. The object-related negativity (ORN) is an event-related potential that indexes listeners' processing of two simultaneous auditory objects (Alain et al., 2001). It is typically elicited by a mistuned component

in otherwise harmonic complex tones (see Figure 1B). Whereas, newborns (Bendixen et al., 2015) and 4- to 12-month-old infants (Folland et al., 2015) showed an ORN in response to the mistuned complex tones, 2-month-olds did not (Folland et al., 2015). Future studies are needed to determine whether this discrepancy is due to methodological differences between the studies, or whether they reflect non-linearities in the development of the neural correlates of simultaneous stream segregation.

2.2 Childhood

For the sake of this review, childhood will be defined as ranging from 3 to 12 years of age. Most behavioral studies of stream segregation in children have been reviewed in Leibold (2011). They show that the acoustic difference required to segregate sequential or simultaneous sounds into distinct streams decreases as children grow older, but remains larger in late childhood than in adulthood (Alain et al., 2003; Sussman et al., 2007; Sussman and Steinschneider, 2009). Note that 5- to 13-year-old children benefit from visual cues helping simultaneous ASA to the same extent as adults (Bonino et al., 2013). Yet 5-year-olds show less benefit from spatial cues to perform simultaneous stream segregation than adults (Wightman et al., 2003).

Electrophysiological studies are in line with the behavioral observation of immature stream segregation in children up to 12 years of age. Like infants, children show an MMN when presented with stimuli that entail sequential streaming (Sussman et al., 2001; Lepistö et al., 2009). However, the frequency separation between the successive sounds of these sequences needs to be larger in passively attending 9–12 year-olds than adults to elicit an MMN (Sussman and Steinschneider, 2009).

With respect to simultaneous ASA, Alain et al. (2003) recorded the ORN in 8- to 13-year-old children and adults. Their results indicate that children have a *larger* ORN than adults, despite having poorer behavioral performance when segregating streams in the mistuned complex tones. This was interpreted as suggesting greater neuronal activity associated with the perception of separate auditory objects in children than adults. In a recent follow-up to that study, the same team investigated the ORN of 6–12 year-olds with a moderate to severe congenital hearing loss (55–70 dB HL), who were regular hearing aid users (Mehrkian et al., 2022). Note that children with a hearing loss were tested unaided, but sounds were presented at higher sound pressure level than for age-matched children with normal hearing, thus aiming to equate sensation level across groups. Children with a hearing loss had smaller and later ORN than age-matched children with normal hearing. Congenital sensorineural hearing loss thus seem to have a pervasive effect on the central processing of simultaneous streams, that is not merely due to an audibility loss.

2.3 Adolescence

In the past decade, researchers started to investigate the maturational trajectory of ASA at adolescence. The frequency separation needed to experience streaming of successive tones did

not change between 7 and 15 years (Sussman et al., 2015). However, in the same study, there was a gradual improvement in the ability to detect an intensity deviant in one of two sequential streams. More studies are needed to investigate adolescent development of simultaneous ASA, and to explore the neural correlates of both sequential and simultaneous streaming at adolescence.

3 Selective attention in the context of ASA

3.1 First year of life

Do infants use selective attention to guide streaming in complex auditory scenes? To address this challenging question, researchers have investigated the effects of non-sensory factors on the detection of an auditory target in the presence of *simultaneous* distractors (for a detailed review, see Leibold, 2011). From 6 months of age, infants rely on temporal (Werner et al., 2009) but not spectral (Werner and Bargones, 1991; Bargones and Werner, 1994) expectations to selectively direct their attention toward a target in the presence of a simultaneous interference. Several questions remain open: are newborns able to selectively direct their attention in complex auditory scenes? Are infants able to selectively attend to a target that unfolds over time in a sequential stream? What are the neural correlates of infants' selective attention in the presence of auditory distractors?

3.2 Childhood

Behavioral studies of simultaneous ASA in children have been reviewed by Leibold (2011), and suggest a progressive improvement in selective auditory attention throughout the primary school years (Greenberg et al., 1970; Allen and Wightman, 1995; Stellmack et al., 1997; Leibold and Neff, 2007). A recent psychoacoustic study aimed at understanding the mechanism underlying this progressive improvement (Jones et al., 2015). Reverse correlations were used to estimate which spectral region children and adults paid attention to when asked to detect a 1 kHz target embedded in an unpredictable noise. Results confirmed that 4- to 7-year-olds had poorer thresholds than 8- to 11-year-olds and adults. In fact, younger children were less efficient at analyzing the spectral content of the stimuli than older children. Their poorer thresholds in noise thus likely reflect an inability to selectively attend to the target while ignoring the distractor. How selective attention to sequential sound streams develops during childhood remains so far unexplored.

Neural correlates of selective attention to sequential streams can be investigated using a variation of the "oddball paradigm" described above (Sussman et al., 1999; Winkler et al., 2003). Participants are presented with two streams of interleaved sounds, differing in frequency (see Figure 1A, right panel). They are asked to focus on one of the streams, and to indicate when they detect a deviant within this target stream, while ignoring deviants that appear in the distracting stream. This allows to compare the neural response of the to-be-attended deviant to that of the to-be-ignored deviant, which typically leads to an early frontal positivity followed by a difference negativity (Nd, for a review see Näätänen et al.,

2001). Nds were recorded in a group of 9-year-olds, a group of 12-year-olds, and a group of adults (Gomes et al., 2007). Both groups of children exhibited a later Nd than adults, indicating persistent processing immaturities in sequential streaming in late childhood. Whether persistent processing immaturities would also be observed in the neural correlates of simultaneous streaming, despite the seemingly mature behavioral performance (Jones et al., 2015) remains an open question.

3.3 Adolescence

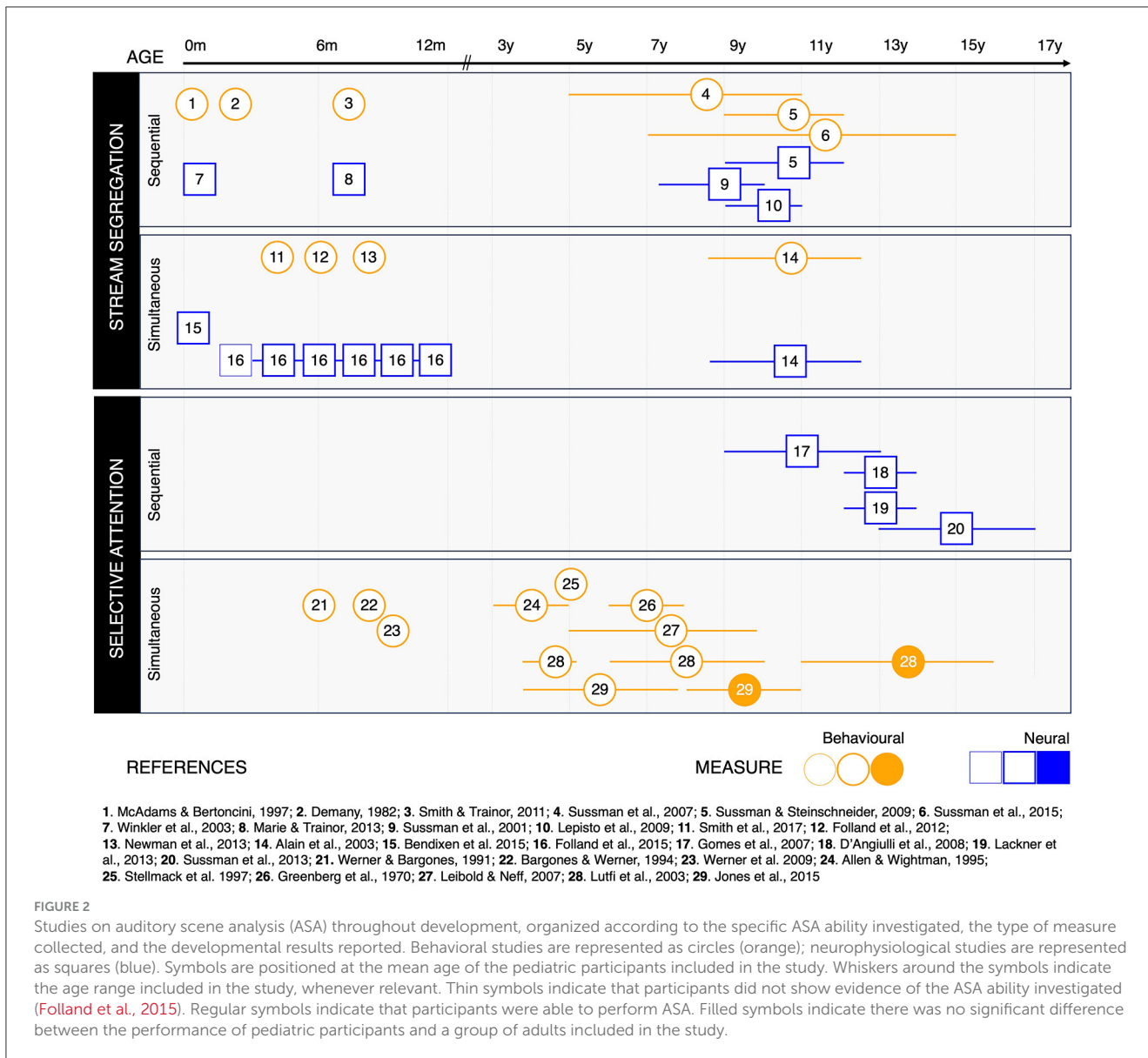
Selective auditory attention to a target in the presence of a simultaneous multitone masker seems to be mature by late childhood (Jones et al., 2015). This observation is consistent with earlier results collected in a small cohort of children as well as adolescents and adults (Lutfi et al., 2003). Whereas, 4- to 10-year-old children showed more masking than adults, there was no difference between adolescents (11–16 years) and adults. A principal component analysis was performed on the variance in masking performance, to investigate whether different age groups and/or individuals use different detection strategies. If so, several components would be identified as significantly contributing to the variance observed in masking performance. On the contrary, a single principal component was found to account for more than 80% of variance in masking performance, both across and within age groups. This suggests that children use similar target detection strategies to adults, but that they vary in their selective attention abilities.

A few studies have investigated the neural correlates of selective attention during sequential streaming at adolescence. Nds did not change between 11- and 14 year-olds as they were asked to detect a deviant in a target stream while ignoring those in the distracting stream (D'Angiulli et al., 2008). Interestingly, the early frontal positivity evoked by the to-be-ignored targets was larger in adolescents with poorer executive functioning skills than in those with higher executive function skills (Lackner et al., 2013). Last, an oddball paradigm was presented with different instructions, directing adolescents' attention toward different auditory cues, or away from the auditory modality and toward visual information (Sussman, 2013). The morphology of adolescents' event-related potentials and MMN varied with the instructions, like adults' (Sussman et al., 2002).

Overall, studies did not find developmental effects on the neural correlates of selective attention to sequential streams, which may indicate mature attentional responses at adolescence. Note however that none of the studies reviewed in the above paragraph included a group of adults, which limits interpretation in terms of the maturational trajectory at adolescence. Additionally, how the neural correlates of selective attention in simultaneous segregation tasks develop throughout adolescence remains unexplored.

4 Discussion

Figure 2 shows the studies reviewed in this paper, with respect to the age range of their pediatric population, the type of measure collected, and the specific ASA ability investigated. To sum up,



cognitive and neural mechanisms supporting both simultaneous and sequential stream *segregation* are functional from birth. Yet, their efficiency keeps improving throughout childhood and adolescence (Alain et al., 2003; Sussman et al., 2007, 2015; Sussman and Steinschneider, 2009).

Developmental studies of selective auditory *attention* in the context of ASA paint a seemingly contradictory picture. From 6 months of age, infants benefit from some (but not all) auditory cues to orient their attention toward a target in the presence of simultaneous interferers (Werner and Bargones, 1991; Bargones and Werner, 1994; Werner et al., 2009), in line with neurophysiological data showing developmental changes in arousal over the first 2 years of life (Richards et al., 2010). Yet, the existent developmental data on selective attention in ASA suggest that behavioral performance reaches maturity by late childhood (Lutfi et al., 2003; Jones et al., 2015), whereas its neural correlates keep maturing until adolescence (Gomes et al., 2007). Two

explanations might account for this apparent discrepancy. First, children may perform similarly to adults by recruiting different cognitive resources (Trau-Margalit et al., 2023). Future studies are thus warranted to investigate the development of the neural markers of listening effort in noise. Second, the literature on selective attention in the context of ASA seems to present a blind spot. Indeed, to the best of our knowledge, behavioral studies all used simultaneous streaming tasks, whereas neurophysiological studies used sequential streaming tasks. Discrepancies between behavioral and neural results may thus stem from different maturational trajectories between simultaneous and sequential streaming tasks. Note however that speech-in-speech perception inherently requires *both* simultaneous and sequential ASA abilities, whereas the bulk of the literature reviewed here has focused on one or the other. Noteworthy, studies investigating selective attention to speech in the presence of distractors indicate a protracted development of neurophysiological attentional responses from

childhood until adulthood (Berman and Friedman, 1995; Karns et al., 2015).

This supports the need to better understand the development of ASA in more ecological situations that require both simultaneous and sequential streaming abilities. The stochastic figure-ground paradigm (Teki et al., 2011, 2013; see Figure 1C) offers a unique opportunity in this respect. The paradigm consists in a series of identical chords (the figure) presented against a background of random chords. Adults are remarkably sensitive to the appearance of such figures in stochastic noise backgrounds—discrimination performance even improves as figure coherence increases. Additionally, the ORN and a later positive wave (P400) have been elicited in adults listening to such stochastic sequences, providing “neural signatures” of figure-ground discrimination (Tóth et al., 2016). Adapting this task to children and adolescents would further our understanding of the development of combined simultaneous and sequential streaming, as is often required in real-life.

Other limitations are that most studies focused on narrow age ranges, and a number of them did not include a group of adult participants. In addition, most of the results reported here stem from single studies that addressed a specific question. In the few cases where more than one study was conducted to address a research question in a specific age range, results were partly contradictory. This points toward the need for comprehensive developmental investigations, including replication studies. This would allow to examine the transition toward adult-like performance, and the factors that contribute to this transition, including those that relate to individual differences in maturation. Cognitive (executive functions and working memory), neurochemical (modulation of serotonin, dopamine and gamma-aminobutyric acid) and environmental factors (exposure to music and language) should be included as potential predictors of maturation, as they are thought to contribute to stream segregation and/or speech perception in noise in adults (Moore et al., 2008; Kondo et al., 2012; Lackner et al., 2013; van Loon et al., 2013; Chabal et al., 2015; Tierney et al., 2020; Porto et al., 2023). Last, future studies are warranted to disentangle the relationship between selective attention and auditory streaming throughout development.

Together, this would pave the way toward a model of ASA development from infancy to adulthood. This would contribute

to understand typical development, and to better grasp the difficulties faced by clinical populations in noisy environments. Many children seem to be disproportionately affected by the presence of background noise (Calculus et al., 2018; Sharma et al., 2019). Adding insult to injury, classrooms are notoriously noisy (Brill et al., 2018). A better understanding of ASA development may therefore have a significant societal impact on the academic performance of children/adolescents in noisy environments.

Author contributions

AC: Writing—original draft, Writing—review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. We gratefully acknowledge funding from the H2020 European Research Council under grant 101076968.

Acknowledgments

We are very thankful to two reviewers for their comments on a previous version of this manuscript.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Alain, C., Arnott, S., and Picton, T. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol.* 27, 1072–1089. doi: 10.1037//0096-1523.27.5.1072
- Alain, C., Theunissen, E., Chevalier, H., and Batty, M. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cogn. Brain Res.* 16, 210–218. doi: 10.1016/S0926-6410(02)00275-6
- Allen, P., and Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *J. Speech Hear. Res.* 38, 503–511. doi: 10.1044/jshr.3802.503
- Bargones, J. Y., and Werner, L. (1994). Adults listen selectively; infants do not. *Psychol. Sci.* 5, 170–174. doi: 10.1111/j.1467-9280.1994.tb00655.x
- Bendixen, A., Haden, G., Nemeth, R., Farkas, D., Torok, M., and Winkler, I. (2015). Newborn infants detect cues of concurrent sound segregation. *Dev. Neurosci.* 37, 172–181. doi: 10.1159/000370237
- Berman, S., and Friedman, D. (1995). The development of selective attention as reflected by event-related brain potentials. *J. Exp. Child Psychol.* 59, 1–31. doi: 10.1006/jecp.1995.1001
- Bonino, A. Y., Leibold, L. J., and Buss, E. (2013). Effect of signal-temporal uncertainty in children and adults: tone detection in noise or a random-frequency masker. *J. Acoust. Soc. Am.* 134, 4446–4457. doi: 10.1121/1.4828828
- Bregman, A. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sounds.* Cambridge, MA: MIT Press.

- Bregman, A. (2015). Progress in understanding auditory scene analysis. *Music Percept.* 33, 12–19. doi: 10.1525/mp.2015.33.1.12
- Brill, L., Smith, K., and Wang, L. (2018). Building a sound future for students: considering the acoustics in occupied active classrooms. *Acoust. Tod.* 14, 14–21.
- Calculus, A., Deltenre, P., Colin, C., and Kolinsky, R. (2018). Peripheral and central contribution to the difficulty of speech in noise perception in dyslexic children. *Dev. Sci.* 21:12558. doi: 10.1111/desc.12558
- Chabal, S., Schroeder, S., and Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Atten. Percept. Psychophys.* 77, 2684–2693. doi: 10.3758/s13414-015-0973-7
- D'Angiulli, A., Herdman, A., Stapells, D., and Hertzman, C. (2008). Children's event-related potentials of auditory selective attention vary with their socioeconomic status. *Neuropsychology* 22, 293–300. doi: 10.1037/0894-4105.22.3.293
- Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behav. Dev.* 5, 261–276. doi: 10.1016/S0163-6383(82)80036-2
- Elhilali, M., Ma, L., Micheyl, C., Oxenham, A., and Shamma, S. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317–329. doi: 10.1016/j.neuron.2008.12.005
- Folland, N. A., Butler, B. E., Payne, J. E., and Trainor, L. J. (2015). Cortical representations sensitive to the number of perceived auditory objects emerge between 2 and 4 months of age: electrophysiological evidence. *J. Cogn. Neurosci.* 27, 1060–1067. doi: 10.1162/jocn_a_00764
- Folland, N. A., Butler, B. E., Smith, N. A., and Trainor, L. J. (2012). Processing simultaneous auditory objects: infants' ability to detect mistuning in harmonic complexes. *J. Acoust. Soc. Am.* 131, 993–997. doi: 10.1121/1.3651254
- Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* 17, 1578–1592. doi: 10.1162/089892905774597263
- Gomes, H., Duff, M., Barnhardt, J., Barrett, S., and Ritter, W. (2007). Development of auditory selective attention: event-related potential measures of channel selection and target detection. *Psychophysiology* 44, 711–727. doi: 10.1111/j.1469-8986.2007.00555.x
- Greenberg, G. Z., Bray, N. W., and Beasley, D. S. (1970). Children's frequency-selective detection of signals in noise. *Percept. Psychophys.* 8:BF03210199. doi: 10.3758/BF03210199
- Greenberg, G. Z., and Larkin, W. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *J. Acoust. Soc. Am.* 44, 1513–1523. doi: 10.1121/1.1911290
- Hedrick, M. S., and Madix, S. G. (2009). Effect of vowel identity and onset asynchrony on concurrent vowel identification. *J. Speech Lang. Hear. Res.* 52, 696–705. doi: 10.1044/1092-4388(2008/07-0094)
- Jones, P. R., Moore, D. R., and Amitay, S. (2015). Development of auditory selective attention: why children struggle to hear in noisy environments. *Dev. Psychol.* 51, 353–369. doi: 10.1037/a0038570
- Karns, C. M., Isbell, E., Giuliano, R. J., and Neville, H. J. (2015). Auditory attention in childhood and adolescence: an event-related potential study of spatial selective attention to one of two simultaneous stories. *Dev. Cogn. Neurosci.* 13, 53–67. doi: 10.1016/j.dcn.2015.03.001
- Kondo, H., Kitagawa, N., Kitamura, M., Koizumi, A., Nomura, M., and Kashino, M. (2012). Separability and commonality of auditory and visual bistable perception. *Cerebr. Cortex* 22, 1915–1922. doi: 10.1093/cercor/bhr266
- Lackner, C. L., Santesso, D. L., Dywan, J., Wade, T. J., and Segalowitz, S. J. (2013). Electrocortical indices of selective attention predict adolescent executive functioning. *Biol. Psychol.* 93, 325–333. doi: 10.1016/j.biopsycho.2013.03.001
- Leibold, L. J. (2011). Development of auditory scene analysis and auditory attention. *Hum. Audit. Dev.* 42, 137–161. doi: 10.1007/978-1-4614-1421-6_5
- Leibold, L. J., and Neff, D. L. (2007). Effects of masker-spectral variability and masker fringes in children and adults. *J. Acoust. Soc. Am.* 121, 3666–3676. doi: 10.1121/1.2723664
- Lepistö, T., Kuitunen, A., Sussman, E., Saalasti, S., Jansson-Verkasalo, E., Wendt, T. N., et al. (2009). Auditory stream segregation in children with Asperger syndrome. *Biol. Psychol.* 82, 301–307. doi: 10.1016/j.biopsycho.2009.09.004
- Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Percept. Psychophys.* 65, 396–406. doi: 10.3758/BF03194571
- Marie, C., and Trainor, L. J. (2013). Development of simultaneous pitch encoding: infants show a high voice superiority effect. *Cerebr. Cortex* 23, 660–669. doi: 10.1093/cercor/bhs050
- McAdams, S., and Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *J. Acoust. Soc. Am.* 102, 2945–2953. doi: 10.1121/1.420349
- Mehrkian, S., Moossavi, A., Gohari, N., Nazari, M. A., Bakhshi, E., and Alain, C. (2022). Long latency auditory evoked potentials and object-related negativity based on harmonicity in hearing-impaired children. *Neurosci. Res.* 178, 52–59. doi: 10.1016/j.neures.2022.01.001
- Miller, G., and Heise, G. (1950). The trill threshold. *J. Acoust. Soc. Am.* 22, 637–638. doi: 10.1121/1.1906663
- Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* 80, 479–483. doi: 10.1121/1.394043
- Moore, D., Ferguson, M., Halliday, L., and Riley, A. (2008). Frequency discrimination in children: perception, learning and attention. *Hear. Res.* 238, 147–154. doi: 10.1016/j.heares.2007.11.013
- Näätänen, R., Alho, K., and Schröger, E. (2001). "Electrophysiology of attention," in *Steven's Handbook of Experimental Psychology, 3rd Edn, vol. 4*, ed. H. Pashler (New York, NY: John Wiley & Sons), 601–653.
- Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., et al. (2012). The mismatch negativity (MMN)—a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin. Neurophysiol.* 123, 424–458. doi: 10.1016/j.clinph.2011.09.020
- Porto, L., Wouters, J., and van Wieringen, A. (2023). Speech perception in noise, working memory, and attention in children: a scoping review. *Hear. Res.* 439:108883. doi: 10.1016/j.heares.2023.108883
- Richards, J., Reynolds, G., and Courage, M. (2010). The neural bases of infant attention. *Curr. Direct. Psychol. Sci.* 19, 41–46. doi: 10.1177/0963721409360003
- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123. doi: 10.1016/j.tins.2010.11.002
- Sharma, M., Purdy, S., and Humburg, P. (2019). Cluster analyses reveals subgroups of children with suspected auditory processing disorders. *Front. Psychol.* 10:2481. doi: 10.3389/fpsyg.2019.02481
- Smith, N. A., Folland, N. A., Martinez, D. M., and Trainor, L. J. (2017). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition* 164, 1–7. doi: 10.1016/j.cognition.2017.01.016
- Smith, N. A., and Trainor, L. J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distracters. *Infancy* 16, 655–668. doi: 10.1111/j.1532-7078.2011.00067.x
- Stellmack, M., Willihnganz, M., Wightman, F., and Lutfi, R. (1997). Spectral weights in level discrimination by preschool children: analytic listening conditions. *J. Acoust. Soc. Am.* 101, 2811–2821. doi: 10.1121/1.419479
- Sussman, E. (2013). Attention matters: pitch vs. pattern processing in adolescence. *Front. Psychol.* 4:333. doi: 10.3389/fpsyg.2013.00333
- Sussman, E., Ceponiene, R., Shestakova, A., Näätänen, R., and Winkler, I. (2001). Auditory stream segregation processes operate similarly in school-aged children and adults. *Hear. Res.* 153, 108–114. doi: 10.1016/S0378-5955(00)00261-6
- Sussman, E., Ritter, W., and Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36, 22–34. doi: 10.1017/S0048577299971056
- Sussman, E., and Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia* 47, 771–785. doi: 10.1016/j.neuropsychologia.2008.12.007
- Sussman, E., Steinschneider, M., Lee, W., and Lawson, K. (2015). Auditory scene analysis in school-aged children with developmental language disorders. *Int. J. Psychophysiol.* 95, 113–124. doi: 10.1016/j.ijpsycho.2014.02.002
- Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., and Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cogn. Brain Res.* 13, 393–405. doi: 10.1016/S0926-6410(01)00131-8
- Sussman, E., Wong, R., Horváth, J., Winkler, I., and Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5-11-year-old children. *Hear. Res.* 225, 117–127. doi: 10.1016/j.heares.2006.12.013
- Teki, S., Chait, M., Kumar, S., Shamma, S., and Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife* 2:9. doi: 10.7554/eLife.00699.009
- Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci.* 31, 164–171. doi: 10.1523/JNEUROSCI.3788-10.2011
- Tierney, A., Rosen, S., and Dick, F. (2020). Speech-in-speech perception, nonverbal selective attention, and musical training. *J. Exp. Psychol.* 46, 968–979. doi: 10.1037/xlm0000767
- Tóth, B., Kocsis, Z., Háden, G. P., Szerafin, Á., Shinn-Cunningham, B. G., and Winkler, I. (2016). EEG signatures accompanying auditory figure-ground segregation. *NeuroImage* 141, 108–119. doi: 10.1016/j.neuroimage.2016.07.028

- Trau-Margalit, A., Fostick, L., Harel-Arbeli, T., Nissanholtz-Gannot, R., and Taitelbaum-Swead, R. (2023). Speech recognition in noise task among children and young-adults: a pupillometry study. *Front. Psychol.* 2023:1188485. doi: 10.3389/fpsyg.2023.1188485
- van Loon, A., Knapen, T., Scholte, S., St. John-Saaltink, E., Donner, T., and Lamme, V. (2013). GABA shapes the dynamics of bistable perception. *Curr. Biol.* 23, 823–827. doi: 10.1016/j.cub.2013.03.067
- Werner, L., and Bargones, J. (1991). Sources of auditory masking in infants: distraction effects. *Percept. Psychophys.* 50, 405–412. doi: 10.3758/BF03205057
- Werner, L. A., Parrish, H. K., and Holmer, N. M. (2009). Effects of temporal uncertainty and temporal expectancy on infants' auditory sensitivity. *J. Acoust. Soc. Am.* 125, 1040–1049. doi: 10.1121/1.3050254
- Wightman, F., Callahan, M., Lutfi, R., Kistler, D., and Oh, E. (2003). Children's detection of pure-tone signals: informational masking with contralateral maskers. *J. Acoust. Soc. Am.* 113, 1–9. doi: 10.1121/1.1570443
- Winkler, I., Kushnerenko, E., Horváth, J., Ceponiene, R., Fellman, V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11812–11815. doi: 10.1073/pnas.2031891100