Too Young for the Cocktail Party?

Lori J. Leibold
Address:
Center for Hearing Research
Boys Town National Research Hospital
555 North 30th Street
Omaha, Nebraska 68131
USA

Email:
lori.leibold@boystown.org

Emily Buss
Address:
Department of Otolaryngology/Head and Neck Surgery
University of North Carolina at Chapel Hill
170 Manning Drive
Campus Box 7070
Chapel Hill, North Carolina 27599
USA

Email:
ebuss@med.unc.edu

Lauren Calandruccio
Address:
Department of Psychological Sciences
Case Western Reserve University
Cleveland Hearing and Speech Center
11635 Euclid Avenue
Cleveland, Ohio 44106
USA

Email:
lauren.calandruccio@case.edu

One reason why children and cocktail parties do not mix.

There are many reasons why children and cocktail parties do not mix. One less obvious reason is that children struggle to hear and understand speech when multiple people are talking at the same time. Cherry (1953) was not likely thinking about children when he coined the "cocktail party problem" over 60 years ago, referring to the speech perception difficulties individuals often face in social environments with multiple sources of competing sound. Subsequent research has largely focused on trying to understand how adults recognize what one person is saying when other people are talking at the same time (reviewed by Bronkhorst, 2000; McDermott, 2009). However, modern classrooms pose many of the same challenges as a cocktail party, with multiple simultaneous talkers and dynamic listening conditions (Brill et al., 2018). In contrast to the cocktail party, however, failure to recognize speech in a classroom can have important consequences for a child's educational achievement and social development. These concerns have prompted several laboratories, including ours, to study development of the ability to recognize speech in multisource backgrounds. This article summarizes findings from the smaller number of studies that have examined the cocktail party problem in children, providing evidence that children are at an even greater disadvantage than adults in complex acoustic environments that contain multiple sources of competing sounds.

For much of the school day, children are tasked with listening to their teacher in the context of sounds produced by a range of different sound sources in the classroom. Under these conditions, we would call the teacher's voice the target and the background sounds would be the maskers. All sounds in the environment, including the target and the maskers, combine in the air before reaching the child's ears. This combination of acoustic waveforms is often referred to as an *auditory scene*. An example of an auditory scene is illustrated in **Figure 1**, where sounds include the relatively steady noise produced by a projector as well as more dynamic sounds, such as speech produced by classmates who are talking at the same time as their teacher. To hear and understand the teacher, the spectral and temporal characteristics of this mixture of incoming sounds must be accurately represented by the outer ear, middle ear, cochlea, and auditory nerve. This processing is often referred to as *peripheral encoding*. Auditory perception is critically dependent on the peripheral encoding of sound and the fidelity with which this information is transmitted to the brain. Processing within the central auditory system is then needed to identify and group the acoustic waveforms that were generated by the teacher from those that were generated by the other sources (sound source segregation) and then allocate attention to the auditory "object" corresponding to the teacher's voice while discounting competing sounds (selective auditory attention). Auditory scene analysis also relies on cognitive processes, such as memory, as well as listening experience and linguistic knowledge. Collectively, these processes are often referred to as *auditory scene analysis* (e.g., Bregman, 1990; Darwin and Hukin, 1999).

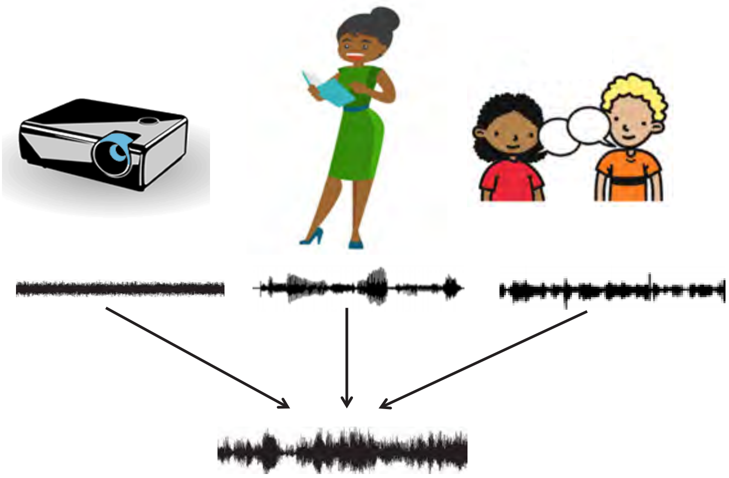

__Figure 1.__ This cartoon illustrates the cocktail party problem in the classroom. In this example, acoustic waveforms are produced by three sources: (1) noise is produced by a computer projector in the classroom; (2) speech is produced by the teacher; and (3) speech is produced by two classmates who are also talking. The fundamental problem is that the acoustic waveforms produced by all three sound sources combine in the air before arriving at the students' ears. To follow the teacher's voice, students must "hear out" and attend to their teacher while disregarding the sounds produced by all other sources.

Immaturity at any stage of processing can impact the extent to which students in the classroom hear and understand the target voice. For example, spectral resolution refers to the ability to resolve the individual frequency components of a complex sound. Degraded spectral resolution is one consequence of congenital hearing loss, specifically sensorineural hearing loss caused by damage to the outer hair cells in the cochlea. This degraded peripheral encoding may reduce audibility of the target speech, making it impossible for adults or children with sensorineural hearing loss to perform auditory scene analysis. Perhaps less obvious, immature central auditory processing could result in the same functional outcome in a child with normal hearing. For example, the perceptual consequence of a failure to selectively attend to the speech stream produced by the teacher, while ignoring classmates' speech, is reduced speech understanding, even when the peripheral encoding of the teacher's speech provides all the cues required for recognition.

Maturation of Peripheral Encoding

Accurate peripheral encoding of speech is clearly a prerequisite for speech recognition. However, sensory representation of the frequency, temporal, and intensity properties of sound does not appear to limit auditory scene analysis during the school-age years. The cochlea begins to function in utero, before the onset of visual functioning (Gottlieb, 1991). Physiological responses to sound provide evidence that the cochlea is mature by term birth, if not earlier (e.g., Abdala, 2001). Neural transmission through the auditory brainstem appears to be slowed during early infancy, but peripheral encoding of the basic properties of sound approaches the resolution observed for adults by about six months of age (reviewed by Eggermont and Moore, 2012; Vick, 2018).

A competing noise masker can interfere with the peripheral encoding of target speech if the neural excitation produced by the masker overlaps with the neural representation of the target speech. This type of masking can be more severe in children and adults with sensorineural hearing loss than in those with normal hearing. Sensorineural hearing loss is often due to the loss of outer hair cells in the cochlea (reviewed by Moore, 2007). As mentioned above, outer hair cell loss degrades the peripheral encoding of the frequency, intensity, and temporal features of speech, which, in turn, impacts masked speech recognition. Indeed, multiple researchers have demonstrated an association between estimates of peripheral encoding and performance on speech-in-noise tasks for adults with sensorineural hearing loss (e.g., Dubno et al., 1984; Frisina and Frisina, 1997).

Additional evidence that competing noise interferes with the perceptual encoding of speech comes from the results of studies evaluating consonant identification in noise by adults (e.g., Miller and Nicely, 1955; Phatak et al., 2008). Consonant identification is compromised in a systematic way across individuals with normal hearing when competing noise is present, presumably because patterns of excitation produced by the target consonants and masking noise overlap on the basilar membrane (Miller, 1947). In the classroom example shown in **Figure 1**, overlap in excitation patterns between speech produced by the teacher and noise produced by the projector can result in an impoverished neural representation of the teacher's spoken message, although this depends on the relative levels of the two sources and distance to the listener. The term *energetic masking* is often used to describe the perceptual consequences of this phenomenon (reviewed by Brungart, 2005).

Despite mature peripheral encoding, school-age children have more difficulty understanding speech in noise compared with adults. For example, 5- to 7-year-old children require a 3-6 dB more favorable signal-to-noise ratio (SNR) than adults to achieve comparable speech detection, word identification, or sentence recognition performance in a speech-shaped noise masker (e.g., Corbin et al., 2016). Speech-in-noise recognition gradually improves until 9-10 years of age, after which mature performance is generally observed (e.g., Wightman and Kistler, 2005; Nishi et al., 2010). The pronounced difficulties experienced by younger school-age children are somewhat perplexing in light of data indicating that the peripheral encoding of sound matures early in life. It has been posited that these age effects reflect an immature ability to recognize degraded speech (e.g., Eisenberg et al., 2000; Buss et al., 2017). It has also been suggested that children's immature working memory skills also play a role in their speech-in-noise difficulties (McCreery et al., 2017).

Maturation of Auditory Scene Analysis

Children are at a disadvantage relative to adults when listening to speech in competing noise, but the child/adult difference is considerably larger when the maskers are also composed of speech. Hall et al. (2002) compared word recognition for children (5-10 years) and adults tested in each of two maskers: noise filtered to have the same power spectrum as speech (speech-shaped noise; see Multimedia File 1 at acousticstoday.org/leibold-media) and competing speech composed of two people talking at the same time (see Multimedia File 2 at acousticstoday.org/leibold-media). On average, children required a 3 dB more favorable SNR relative to adults to achieve a comparable performance in the noise masker. This disadvantage increased to 8 dB in the two-talker masker. In addition to the relatively large child/adult differences observed in the two-talker masker relative to the noise masker, the ability to recognize masked speech develops at different rates for these two types of maskers (e.g., Corbin et al. 2016). Although adult-like speech recognition in competing noise emerges by 9-10 years of age (e.g., Wightman and Kistler, 2005; Nishi et al., 2010), speech recognition performance in a two-talker speech masker is not adult-like until 13-14 years of age (Corbin et al., 2016). This prolonged time course of development appears to be at least partly due to immature sound segregation and selective attention skills. Recognition of speech produced by the teacher is likely to be limited more by speech produced by other children in the classroom than by noise produced by the projector (see **Figure 1**). The term *informational masking* is often used to refer to this phenomenon (e.g., Brungart, 2005).

An important goal for researchers who study auditory development is to characterize the factors that both facilitate and limit children's ability to perform auditory scene analysis (e.g., Newman et al., 2015; Calandruccio et al., 2016). For listeners of all ages, the perceptual similarity between target and masker speech affects performance in that greater masking is associated with greater perceptual similarity. A common approach to understanding the development of auditory scene analysis is to measure the extent to which children rely on acoustic voice differences between talkers to segregate target from masker speech (e.g., Flaherty et al., 2018; Leibold et al., 2018). For example, striking effects have been found between conditions in which the target and masker speech are produced by talkers that differ in sex (e.g., a female target talker and a two-male-talker masker) and conditions in which target and masker speech are produced by talkers of the same sex (e.g., a male target talker and a two-male-talker masker). Dramatic improvements in speech intelligibility, as much as 20 percentage points, have been reported in the literature for sex-mismatched relative to sex-matched conditions (e.g., Helfer and Freyman, 2008).

School-age (Wightman and Kistler, 2005; Leibold et al., 2018) and 30-month-old (Newman and Morini, 2017) children also show a robust benefit of a target/masker sex mismatch, but infants younger than 16 months of age do not (Newman and Morini, 2017; Leibold et al., 2018). Leibold et al. (2018), for example, measured speech detection in a two-talker masker in 7- to 13-month-old infants and in adults. Adults performed better when the target word and masker speech were mismatched in sex than when they were matched. In sharp contrast, infants performed similarly in sex-matched and sex-mismatched conditions. The overall pattern of results observed across studies suggest that the ability to take advantage of acoustic voice differences between male and female talkers requires experience with different talkers before the ability emerges sometime between infancy and the preschool years.

Although children as young as 30 months of age benefit from a target/masker sex mismatch, the ability to use more subtle and/or less redundant acoustic voice differences may take longer to develop. Flaherty et al. (2018) tested this hypothesis by examining whether children (5-15 years) and adults benefited from a difference in voice pitch (i.e., fundamental frequency; F0) between target words and a two-talker speech masker, holding other voice characteristics constant. As previously observed for adults (e.g., Darwin et al, 2003), adults and children older than 13 years of age performed substantially better when the target and masker speech differed in F0 than when the F0 of the target and masker speech was matched. This improvement was observed even for the smallest target/masker F0 difference of three semitones. In sharp contrast, younger children (<7 years) did not benefit from even the most extreme F0 difference of nine semitones. Moreover, although 8-12 year olds benefitted from the largest F0 difference, they generally failed to take advantage of more subtle F0 differences between target and masker speech. These data highlight the importance of auditory experience and maturational effects in learning how to segregate target from masker speech.

In addition to relying on acoustic voice differences between talkers when listening in complex auditory environments, adults with normal hearing take advantage of the differences in signals arriving at the two ears. These differences provide critical information regarding the location of sound sources in space, which, in turn, facilitates segregation of target and masker speech (e.g., Bregman, 1990; Freyman et al., 2001). The binaural benefit associated with separating the target and masker on the horizontal plane is often called spatial release from masking (SRM). In the laboratory, SRM is typically estimated by computing the difference in speech recognition performance between two conditions: the co-located condition, in which the target and masker stimuli are presented from the same location in space, and the spatial separation condition, in which the target and masker stimuli are perceived as originating from different locations on the horizontal plane. For adults with normal hearing, SRM is substantially larger for speech recognition in a masker composed of one or two streams of speech than in a noise masker (reviewed by Bronkhorst, 2000).

Several studies have evaluated SRM in young children and demonstrate a robust benefit of spatially separating the target and masker speech (e.g., Litovsky, 2005; Yuen and Yuan, 2014). Results are mixed, however, regarding the time course of development for SRM. Although Litovsky (2005) observed adult-like SRM in 3-year-old children, other studies have reported a smaller SRM for children compared with adults, a child/adult difference that remains until adolescence (e.g., Yuen and Yuan, 2014; Corbin et al., 2017). In a recent study, Corbin et al. (2017) assessed sentence recognition for children (8-10 years) and adults (18-30 years) tested in a noise masker and in a two-talker masker. Target sentences were always presented from a speaker directly in front of the listener, and the masker was either presented from the front (co-located) or from 90° to the side (separated). Although a comparable SRM was observed between children and adults in the noise masker, the SRM was smaller for children than adults in the two-talker masker. In other words, children benefitted from binaural difference cues less than adults in the speech masker. This is important from a functional perspective because it means that not only are children more detrimentally affected by background speech, but they are also less able to use spatial cues to overcome the masking associated with speech.

In addition to sound source segregation, auditory scene analysis depends on the ability to allocate and focus attention on the target. Findings from studies using behavioral shadowing procedures provide indirect evidence that selective auditory attention remains immature well into the school-age years (e.g., Doyle, 1973; Wightman and Kistler, 2005). In a typical shadowing task, listeners are asked to repeat speech presented to one ear while ignoring speech or other sounds presented to the opposite ear. Children perform more poorly than adults on these tasks, with age-related improvements observed into the adolescent years (e.g., Doyle, 1973; Wightman and Kistler, 2005). Moreover, children's incorrect responses tend to be intrusions from speech presented to the ear they are supposed to disregard. For example, Wightman and Kistler (2005) asked children (4-16 years) and adults (20-30 years) to attend to target speech presented to the right ear while disregarding masker speech presented to both the right and left ears. Most of the incorrect responses made by adults and children older than 13 years of age were due confusions with the masker speech that was presented to the same ear as the target speech. In contrast, incorrect responses made by the youngest children (4-5 years) tested were often the result of confusions with the masker speech presented to the opposite ear as the target speech. This result is interpreted as showing that young children do not reliably focus their attention on the target even in the absence of energetic masking.

Although behavioral data suggest that selective auditory attention remains immature throughout most of childhood, a key limitation of existing behavioral paradigms is that we cannot be certain to what a child is or is not attending. Poor performance on a shadowing task might reflect a failure of selective attention to the target but is also consistent with an inability to segregate the two streams of speech (reviewed by Sussman, 2017). This issue is further complicated by the bidirectional relationship between segregation and attention; attention influences the formation of auditory streams (e.g., Shamma et al., 2011). Researchers have begun to disentangle the independent effects of selective auditory attention by measuring auditory event-related brain potentials (ERPs) to both attended and unattended sounds (e.g., Sussman and Steinschneider, 2009; Karns et al., 2015). The pattern of results observed across studies indicates that adult-like ERPs associated with selective auditory attention do not emerge until sometime after 10 years of age, consistent with the time course of maturation observed in behavioral speech-recognition data and improvements in executive control (reviewed by Crone, 2009).

Role of Linguistic Experience and Knowledge

It has been suggested that the ability to use the information provided by the peripheral auditory system optimally requires years of experience with sound, particularly exposure to spoken language (e.g., Tomblin and Moeller, 2015). In a recent study, Lang et al. (2017) tested a group of 5- to 6-year-old children and found a strong relationship between receptive vocabulary and speech recognition when the masker was two-talker speech masker. As shown in **Figure 2**, children with larger vocabularies were more adept at recognizing sentences presented in a background of two competing talkers than children with more limited vocabularies. Results from previous studies investigating the association between vocabulary and speech recognition in a steady noise masker have been somewhat mixed (e.g., Nittrouer et al., 2013; McCreery et al., 2017). The strong correlation observed by Lang et al. (2016) between vocabulary and speech recognition in a two-talker masker may reflect the greater perceptual and linguistic demands required to segregate and attend to target speech in a speech masker or to the spectrotemporally sparse cues available in dynamic speech maskers.

A second line of evidence that immature language abilities contribute to children's increased difficulty recognizing speech when a few people are talking at the same time comes from studies that have compared children's and adults' ability to recognize speech based on impoverished spectral and/or temporal information (e.g., Eisenberg et al., 2000; Buss et al., 2017). For example, adults are able to recognize band-pass-filtered speech based on a narrower bandwidth than children (e.g., Eisenberg et al., 2000; Mlot et al. 2010). One interpretation for this age effect is that children require more information than adults in order to recognize speech because they have less linguistic experience.

This hypothesis was recently tested by assessing speech recognition in a two-talker masker across a wide age range of children (5-16 years) and adults using speech that was digitally processed using a technique designed to isolate the auditory stream associated with the target speech (Buss et al., 2017). Children and adults showed better performance after the signal processing was applied, indicating that sound source segregation negatively impacts children's speech recognition in a speech masker. The child/adult difference in performance persisted, however, providing evidence of developmental effects in the ability to reconstruct speech based on sparse speech cues.

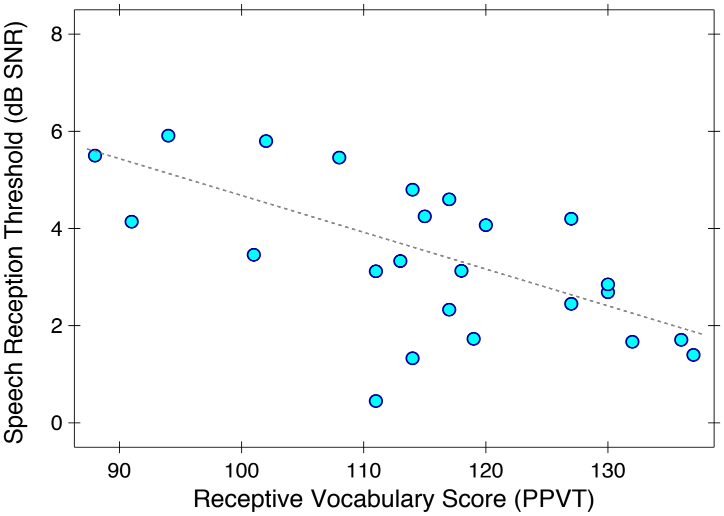

Figure 2. Receptive vocabulary scores and thresholds for sentence recognition in a two-talker masker are shown for 30 young children (5-6 years) tested by Lang et al. (2017). There was a strong association between performance on these two measures (r = −0.75; P < 0.001), indicating that children with larger vocabularies showed better speech recognition performance in the presence of two competing talkers than children with smaller vocabularies. SNR, signal-to-noise ratio; PPVT, Peabody Picture Vocabulary Test.

Implications

The negative effects of environmental noise on children's speech understanding in the classroom are well documented, leading to the development of a classroom acoustics standard by the Acoustical Society of America (ASA) that was first approved by the American National Standards Institute (ANSI) in 2002 (ANSI S12.60). Although this and subsequent standards recognize the negative effects of environmental noise in the classroom on children's speech understanding, they focus exclusively on noise sources measured in unoccupied classrooms (e.g., heating and ventilation systems, street traffic). The additional sounds typically present in an occupied classroom, such as speech, are not accounted for. As argued by Brill et al. (2018) in an article in *Acoustics Today*, meeting the acoustics standards specified for unoccupied classrooms might not be adequate for ensuring children's speech understanding in occupied classrooms, in which multiple people are often talking at the same time. This is problematic because, as anyone who has spent time in a classroom can attest, children spend most of their days listening and learning with competing speech in the background (e.g., Ambrose et al., 2014; Brill et al., 2018).

Conclusion

Emerging results from investigation into how children listen and learn in multisource environments provide strong evidence that children do not belong at cocktail parties. Despite the more obvious reasons, children lack the extensive linguistic knowledge and the perceptual and cognitive abilities that help adults reconstruct the auditory scene.

Acknowledgments

This work was supported by the Grants R01-DC-011038 and R01-DC-014460 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References

Abdala, C. (2001). Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels. *The Journal of the Acoustical Society of America* 110, 1465-1476.

Ambrose, S. E., VanDam, M., and Moeller, M. P. (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear and Hearing* 35, 139-147.

American National Standards Institute (ANSI). (2002). *S12.60-2002 Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools.* Acoustical Society of America, Melville, NY.

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

Brill, L. C., Smith, K., and Wang, L. M. (2018). Building a sound future for students – Considering the acoustics in occupied active classrooms. *Acoustics Today* 14(3), 14-22.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica* 86, 117-128.

Brungart, D. S. (2005). Informational and energetic masking effects in multitalker speech perception. In P. Divenyi (Ed.), *Speech Separation by Humans and Machines.* Kluwer Academic Publishers, Boston, MA, pp. 261-267.

Buss, E., Leibold, L. J., Porter, H. L., and Grose, J. H. (2017). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America* 141, 2650-2660.

Calandruccio, L., Leibold, L. J., and Buss, E. (2016). Linguistic masking release in school-age children and adults. *American Journal of Audiology* 25, 34-40.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America* 25, 975-979.

Corbin, N. E., Bonino, A. Y., Buss, E., and Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing* 37, 55-63.

Corbin, N. E., Buss, E., and Leibold, L. J. (2017). Spatial release from masking in children: Effects of simulated unilateral hearing loss. *Ear and Hearing 38,* 223-235.

Crone, E. A. (2009). Executive functions in adolescence: Inferences from brain and behavior. *Developmental Science* 12, 825-830.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 114, 2913-2922.

Darwin, C. J., and Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance* 25, 617-629.

Doyle, A. B. (1973). Listening to distraction: A developmental study of selective attention. *Journal of Experimental Child Psychology* 15, 100-115.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America* 76, 87-96.

Eggermont, J. J., and Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In L. A. Werner, R. R. Fay, and A. N. Popper (Eds.), *Human Auditory Development*. Springer-Verlag, New York, pp. 61-106.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J. and Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America* 107, 2704-2710.

Flaherty, M. M., Buss, E., and Leibold, L. J. (2018). Developmental effects in children's ability to benefit from F0 differences between target and masker speech. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000673.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America* 109, 2112-2122.

Frisina, D. R., and Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research* 106, 95-104.

Gottlieb, G. (1991). Epigenetic systems view of human development. *Developmental Psychology* 27, 33-34.

Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing* 23, 159-165.

Helfer, K. S., and Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing* 29, 87-98.

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. *Developmental Cognitive Neuroscience* 13, 3-67.

Lang, H., McCreery, R. W., Leibold, L. J., Buss, E., and Miller, M. K. (2017). Effects of language and cognition on children's masked speech perception. *Presented at the 44th Annual Scientific and Technology Meeting of the American Auditory Society*, Scottsdale, AZ, March 2-4, 2017, poster #171 - SP12. Available at https://aas.memberclicks.net/assets/2017_posters.pdf. Accessed October 25, 2018.

Leibold, L. J., Buss, E., and Calandruccio, L. (2018). developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear and Hearing* 39, 935-945.

Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America* 117, 3091-3099.

McCreery, R. W., Spratford, M., Kirby, B., and Brennan, M. (2017). Individual differences in language and working memory affect children's speech recognition in noise. *International Journal of Audiology* 56, 306-315.

McDermott, J. H. (2009). The cocktail party problem. *Current Biology* 19, R1024-R1027.

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin* 44, 105-129.

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America* 27, 338-352.

Mlot, S., Buss, E., and Hall, J. W., III. (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear and Hearing* 31, 56-62.

Moore, B. C. J. (2007). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. John Wiley & Sons, Hoboken, NJ.

Newman, R. S., and Morini, G. (2017). Effect of the relationship between target and masker sex on infants' recognition of speech. *The Journal of the Acoustical Society of America* 141, EL164-EL169.

Newman, R. S., Morini, G., Ahsan, F., and Kidd, G., Jr. (2015). Linguistically-based informational masking in preschool children. *The Journal of the Acoustical Society of America* 138, EL93-EL98.

Nishi, K., Lewis, D. W., Hoover, B. M., Choi, S., and Stelmachowicz, P. G. (2010). Children's recognition of American English consonants in noise. *The Journal of the Acoustical Society of America* 127, 3177-3188.

Nittrouer, S., Caldwell-Tarr, A., Tarr, E., Lowenstein, J. H., Rice, C., and Moberly, A.C. (2013). Improving speech-in-noise recognition for children with hearing loss: Potential effects of language abilities, binaural summation, and head shadow. *International Journal of Audiology* 52, 513-525.

Phatak, S. A., Lovitt, A., and Allen, J. B. (2008). Consonant confusions in white noise. *The Journal of the Acoustical Society of America* 124, 1220-1233.

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

Sussman, E. S. (2017). Auditory scene analysis: An attention perspective. *Journal of Speech, Language, and Hearing Research* 60, 2989-3000.

Sussman, E., and Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia* 47, 771-785.

Tomblin, J. B., and Moeller, M. P. (2015). An introduction to the outcomes of children with hearing loss study. *Ear and Hearing* 36, 4S-13S.

Vick, J. C. (2018). Advancing toward what will be: Speech development in infancy and early childhood. *Acoustics Today* 14(3), 39-47.

Wightman, F. L., and Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America* 118, 3164-3176.

Yuen, K. C., and Yuan, M. (2014). Development of spatial release from masking in Mandarin-speaking children with normal hearing. *Journal of Speech, Language, and Hearing Research* 57, 2005-2023.

BioSketches

Lori Leibold directs the Center for Hearing Research at the Boys Town National Research Hospital, Omaha, NE. She received her BSc in biology from McMaster University, Hamilton, ON, Canada, her MSc in audiology from the University of Western Ontario, London, ON, Canada, and her PhD in hearing science from the University of Washington, Seattle. Her research focuses on auditory development, with a particular interest in understanding how infants and children learn to hear and process target sounds such as speech in the presence of competing background sounds.

Emily Buss is a professor at the University of North Carolina, Chapel Hill, where she serves as chief of auditory research. She received a BA from Swarthmore College, Swarthmore, PA, and a PhD in psychoacoustics from the University of Pennsylvania, Philadelphia. Her research interests include auditory development, the effects of advanced age and hearing loss, auditory prostheses, speech perception, and normative psychoacoustics.

Lauren Calandruccio is an associate professor in the Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH. She received a BA in speech and hearing science and an MA in audiology from Indiana University, Bloomington. After working as a pediatric audiologist, she attended Syracuse University, Syracuse, NY, where she earned her PhD in hearing science. Her research focuses on speech perception, with a focus on noisy or degraded auditory environments, bilingual speech perception, and improving clinical auditory assessment techniques.