CHAPTER 8

REMEDIATION OF SPATIAL PROCESSING ISSUES IN CENTRAL AUDITORY PROCESSING DISORDER

SHARON CAMERON and HARVEY DILLON

Spatial Processing

When we are trying to listen to speech in noisy environments, auditory processes in the brain help us to focus on the person we want to hear while simultaneously suppressing competing sounds coming from different locations. The target speech appears to *pop out* from the competition, so to speak. The technical term for this process is spatial release from masking—or spatial processing—and it allows us to take in the vital information we need to be able to comprehend speech and participate in conversations. But what if we didn't have this ability? What if we when we were listening to speech in noise nothing seemed to *pop out*, but instead all we could hear was a jumble of sounds? We would most likely fail to hear key information, limiting our ability to communicate effectively. This is exactly what happens to children and adults with spatial processing disorder (SPD).

In this chapter we discuss how spatial processing assists in communication and the underlying mechanisms involved. We also discuss how deficits in spatial processing ability impact listeners, particularly children who, despite normal hearing thresholds and cognitive ability, have difficulty understanding speech in the classroom when background noise is present. Difficulty understanding speech when there is competing speech or other types of background noise is a commonly reported symptom of central auditory processing disorder (CAPD) (Bamiou, Musiek, & Luxon, 2001; Jerger & Musiek, 2000; Vanniasegaram, Cohen, & Rosen, 2004). We are certainly not suggesting that spatial processing disorder is the only cause of difficulty understanding speech in background noise for children with normal hearing thresholds, but it is an important cause. For many children, it is the only cause. The main focus of the chapter involves the remediation of spatial processing disorder using the LiSN & Learn, a deficit-specific computer-based auditory training program.

Spatial Processing and Communication

Normal hearing listeners effortlessly communicate in very complex acoustic environments that may contain multiple sound sources, as well as room reverberation. In such adverse conditions, the auditory system takes advantage of the temporal-spectral dynamics of the acoustic input at the two ears to analyze the spatial acoustic scene and thus, to understand speech. For example, listeners can use differences in sound source directions to perceptually separate target speech from one or more interfering sources (Cherry, 1953; Hirsch, 1950). This can result in a significant improvement in speech intelligibility.

As previously mentioned, the benefit gained from spatially separating distracting noise from a target signal is known as spatial release from masking (SRM), or alternatively *spatial advantage* (Bronkhorst, 2000; Cameron, Dillon, & Newall, 2006a; Darwin, 2008; Yost, 1997; Zurek, 1993). Spatial advantage is particularly large (as much as 14 dB depending on age) when maskers are also speech signals (Behrens, Neher, & Johannesson, 2008; Cameron & Dillon, 2007a; Jones & Litovski, 2011; Marrone, Mason, & Kidd 2008a). As shown in Figure 8–1, spatial advantage improves with increasing age until late adolescence and remains stable until at least age 60 (Brown, Cameron, Martin, Watson, & Dillon, 2010; Cameron & Dillon, 2007a; Cameron, Dillon, & Newall, 2006b; Cameron et al., 2009; Cameron, Glyde, & Dillon, 2011; Glyde, Cameron, Dillon, Hickson, & Seeto, 2013).

Crandell and Smaldino (1995) reported that the accurate perception of speech—which is essential for academic achievement—is particularly degraded by noises with spectra similar to the speech spectrum, as these are most effective at masking speech cues (although this effectiveness is influenced by fluctuations in the intensity of the noise over time). Noise generated within a classroom, including children talking, is said to be the most detrimental to a child's ability to perceive speech, because the frequency content of the noise is spectrally similar to the teacher's voice. Thus, the ability of children to utilize spatial processing mechanisms to separate their teacher's voice from background noise is critical to their ability to understand speech in the classroom.

Mechanisms Underlying Spatial Processing

Sensing sounds in two ears is referred to as binaural hearing. Binaural hearing makes it possible for a person to locate the source of sounds in the horizontal plane (Dillon, 2012). However, the main benefit of binaural hearing to humans is to aid the detection of sounds in noisy environments (Moore, 1991). Accurate horizontal localization of sounds coming from a particular location is made

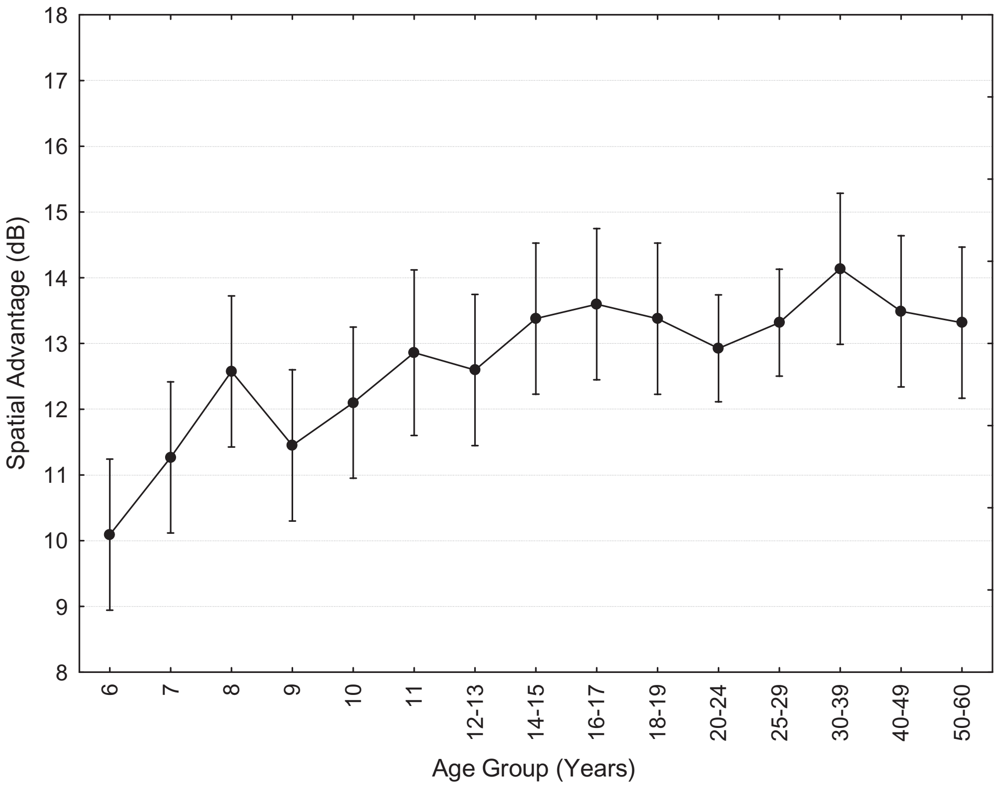

Figure 8–1. Normative data for the spatial advantage measure of the LiSN-S (*n* = 202). Error bars represent the 95% confidence intervals from the mean. (Adapted from Cameron et al., 2011, with permission.)

possible by analysis of differences in the arrival time and the intensity of such sound between the two ears. Sounds arrive at the ear closer to the source before they arrive at the ear farther away. The resulting difference in arrival time at the two ears is called the interaural time difference (ITD). ITD is zero for sounds located directly in front of the listener (i.e., 0° aximuth) and increases to a maximum of about 0.7 ms for sounds coming from 90°, relative to the front. Because any time delay leads to a phase delay, an ITD results in an interaural phase delay. Furthermore, head diffraction produces an attenuation of sound on the side of the head farther from the sound source and a boost on the side of the head nearer to the sound source, referred to as interaural level differences (ILDs).

The initial detection of interaural time and intensity differences occurs at the superior olivary complex (SOC; Reuss, 2000), which is located bilaterally at the base of the brainstem in the caudal portion of the pons, ventral and medial to the cochlear nuclei (CN). Neurons within the medial superior olivary nuclei (MSO) receive phase-locked excitatory input to low-frequency stimuli (and the envelopes of high-frequency stimuli) bilaterally from the CN. Responses to ITD similar to those recorded at the MSO are also recorded in the lateral superior olivary nuclei (LSO), except that the input from the contralateral CN is changed from excitatory to inhibitory at the trapezoid body (Fitzpatrick, Kuwada, & Batra, 2002). The LSO also is implicated in the detection of ILDs (Grothe, 2000). Inhibitory and excitatory responses from the CN that are used to code ITD in the MSO and LSO of the SOC are preserved in the inferior colliculus (IC). Cohen and Knudsen (1999) stated that a *space map* is formed in the nontonotopic subdivisions of the IC, where information about spatial cues is combined across frequency channels, yielding neurons that are broadly tuned for frequency and finely tuned for sound source location. Afferents from the IC are relayed to the primary (A1) and secondary (A2) auditory cortex via the medial geniculate body (MGB) (Pickles, 1988). Animal research has shown that the locations of sound sources are represented in a distributed fashion within individual auditory cortical areas and among multiple cortical areas with similar degrees of location sensitivity, including A1 and A2 (Middlebrooks, Xu, Furukawa, & Macpherson, 2002). However, these authors suggest that the special role of the auditory cortex is only in distributing preprocessed information about sound-source location to appropriate perceptual and motor stations, not actual computation of source locations. Other cortical areas might utilize auditory spatial information from A1 and A2 to perform functions that are not overtly spatial, but the spatial information might assist those functions by helping to segregate multiple sound sources.

Thus, although both localization and spatial release from masking rely on intensity and time differences between the two ears, there is no reason to believe that the two phenomena rely on the same brain processes using these cues. Based on observations of patients with damage to specific brain regions, it seems unlikely that the same brain processes are responsible for both abilities (Litovsky, Fligor, & Tramo, 2002; Thiran & Clarke, 2003). When the task of a listener is to understand a speech signal presented in noise, the improvement in speech reception threshold (SRT) relative to diotic stimulation is referred to as the binaural intelligibility level difference (BILD). Whereas both ITDs and ILDs contribute to BILD, recent studies have shown that in people with normal hearing, ILDs are the dominant mechanism enabling spatial release from masking when speech maskers are symmetrically positioned around the listener and a target talker is in front of the listener (Glyde, Buchholz, Dillon, Cameron, & Hickson, 2013). Moment-by-moment fluctuations in the amplitude and spectrum of each masker cause one masker to dominate over the other at each specific frequency and point in time. At that frequency and point in time, the ear on the side of the head opposite to the dominant masker has a better signal-to-noise ratio (SNR) than the ear closer to the dominant masker. Referred to as *cross-ear dip listening* (Brungart & Iyer, 2012; Glyde et al., in press), this dynamic process, originally hypothesized by Zurek (1993), involves integrating information across the two ears, by selecting, separately for each frequency band, the signal from the ear with the better SNR at each point in time. Cross-ear dip listening effectively creates an optimal signal that has a better SNR than that available at either ear.

Diagnosing Deficits in Spatial Processing

As for other types of CAPD, it is essential that any test of SPD not spuriously indicate the presence of SPD as a consequence of the child having a memory, attention, or language disorder. The LiSN-S is an adaptive speech-in-noise test conducted under headphones that has been designed to avoid such confusions. The target and distracter (i.e., masker) sounds are speech materials that have been synthesized with head-related transfer functions in order to create a three-dimensional effect (Brown et al., 2009; Cameron & Dillon, 2007; Cameron et al., 2011a). A simple repetition-response protocol is used to assess a listener's speech reception threshold (SRT), which is defined as the SNR that yields 50% intelligibility. The target stimuli (sentences) are spoken by a female speaker and always appear to emanate from 0° azimuth (directly in front of the listener). The distracters (looped children's stories) are manipulated so that they appear to come from either 0° azimuth (collocated) or ±90° azimuth simultaneously (spatially separated). The distracter stories are spoken by either the same female speaker as the target sentences or two different female speakers. This test configuration results in four listening conditions: same voice at 0° (or low cue SRT); same voice at ±90°; different voices at 0°; and different voices at ±90° (or high cue SRT), as shown in Figure 8–2.

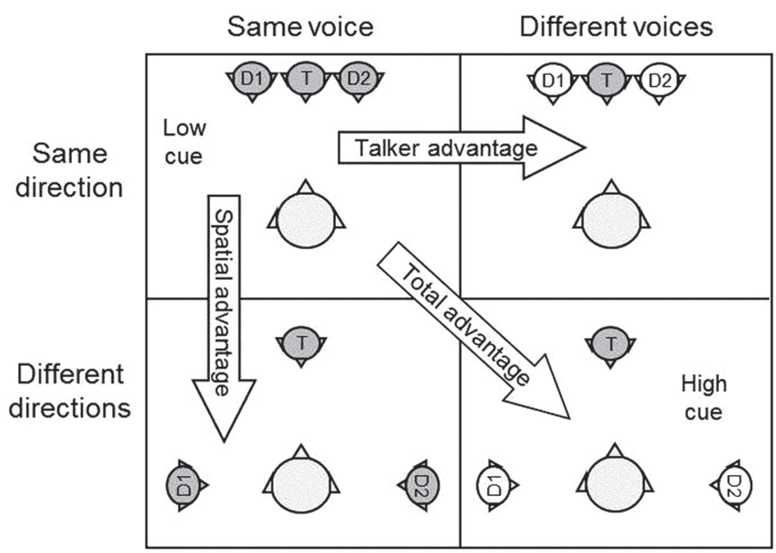

Figure 8–2. The four subtests of the LiSN-S test, and the three difference scores (advantage measures) that can be derived from them. The target speech, T, always comes from the front, whereas the two distracter stories, D1 and D2, come from the front or the sides, in different conditions. D1 and D2 can be the same voice as T or different voices. (Adapted from Cameron et al., 2011, with permission.)