Aging and the Perceptual Organization of Sounds: A Change of Scene?

Claude Alain, Benjamin J. Dyson, and Joel S. Snyder

The peripheral and central auditory systems undergo tremendous changes with normal aging. In this review, we focus on the effects of age on processing complex acoustic signals (such as speech and music) amid other sounds, which requires a set of computations known as auditory scene analysis. Auditory scene analysis is the process whereby the brain assigns parts of the acoustic wave derived from an amalgamation of physical sound sources into perceptual objects (such as words or notes) or streams (such as ongoing speech or music). Solving the scene analysis problem, therefore, depends on listeners' ability to perceptually organize sounds that occur simultaneously and sequentially. The perceptual organization of sounds is thought to involve low-level automatic processes that group sounds that are similar in physical attributes such as frequency, intensity, and location, as well as higher-level schema-driven processes that reflect listeners' experience and knowledge of the auditory environment. In this chapter, we review prior research that has examined the effects of age on concurrent and sequential stream segregation. We will present evidence supporting the existence of an age-related change in auditory scene analysis that appears to be limited to concurrent sound segregation. Evidence also suggests that older adults may rely more on schema-driven processes than young adults to solve the scene analysis problem. The usefulness of auditory scene analysis as a conceptual framework for interpreting and studying age-related changes in sound perception is discussed.

Introduction

Normally, our brains do a masterful job of filtering and sorting the information that flows through our ears such that we are able to function in our acoustic world, be it conversing with a particularly charming individual at a social gathering, or immersing ourselves in a joyful piece of music. Unfortunately, as we get older, our sound world becomes impoverished as a result of the normal changes

in peripheral auditory structures, which represent a significant challenge for hearing science and medicine because of their prevalence, significant impact on quality of life, and lack of mechanism-based treatments. Moreover, as average life expectancy continues to increase, auditory problems will apply to an ever-greater number of individuals. Older adults experience a general deterioration in auditory processing, the most common form of which is age-related peripheral change in hearing sensitivity or presbycusis, characterized by impaired thresholds particularly for high frequencies, which become increasingly prominent after the fourth decade of life (Morrell, Gordon-Salant, Pearson, Brant, and Fozard, 1996). Presumably, these losses in peripheral sensitivity result in impoverished and noisy signals being delivered to the central nervous system. This initial distortion of the acoustic waveform may lead to further reductions in fidelity as the incoming acoustic signals are submitted to increasingly detailed analyses, resulting in the eventual disruption of higher-order processes such as speech comprehension.

Age-related change in auditory perception varies substantially between individuals and may include: (1) impaired frequency and duration discrimination (Abel, Krever, and Alberti, 1990); (2) impaired sound localization (Abel, Giguere, Consoli, and Papsin 2000); (3) difficulties in determining the sequential order of stimuli (Trainor and Trehub, 1989); (4) difficulties in processing novel (Lynch and Steffens, 1994) or transposed melodies (Halpern, Bartlett, and Dowling, 1995); and (5) difficulties in understanding speech, especially when the competing signal is speech rather than homogeneous background noise (Duquesnoy, 1983), or when speech occurs in a reverberant environment (Gordon-Salant and Fitzgibbons, 1993). Although sensorineural hearing loss is highly correlated with speech perception deficits (Abel, Sass-Kortsak, and Naugler, 2000; Humes and Lisa, 1990) and accounts for some of the difficulties in identifying novel melodies (Lynch and Steffens, 1994), there is increasing evidence that age-related changes in the peripheral auditory system (e.g. cochlea) alone cannot adequately account for the wide range of auditory problems that accompany aging. For instance, some of the auditory deficits experienced by older adults remain even after controlling for differences in audiometric thresholds

Claude Alain, Benjamin J. Dyson, and Joel S. Snyder

(Divenyi and Haupt, 1997). In addition, some older individuals with near-normal sensitivity exhibit problems in speech discrimination (Middelweerd, Festen, and Plomp, 1990), whereas others with hearing loss receive little benefit from hearing aids despite the restoration of hearing thresholds to normal levels (Chmiel and Jerger, 1996). Moreover, hearing status does not always interact with age, suggesting that hearing impairment and age are independent sources contributing to many older listeners' difficulties with understanding speech (Gordon-Salant and Fitzgibbons, 2004).

Since reductions in hearing sensitivity fail to account completely for speech perception deficits, it is necessary to consider the additional possibility that the elderly also may suffer from impairment at one or more levels of the central auditory processing pathways. That is, the difficulties commonly observed in older individuals may in fact be related to age-related changes in the neural mechanisms critical for the perceptual separation of the incoming acoustic information to form accurate representations of our acoustic world, rather than simply the reception of a weak and noisy signal from the ear. Auditory scene analysis is one framework that guides our theorizing with respect to the putative mechanisms involved in age-related changes in auditory perception, providing novel insights about how our perception of the auditory world breaks down as a result of normal aging.

Auditory Scene Analysis

The root of the auditory scene analysis problem is derived from the inherent complexity of our everyday acoustic environment. At any given moment, we may be surrounded by multiple sound-generating elements such as a radio playing music, or a group of people speaking, with several of these elements producing acoustic energy simultaneously. In order to make sense of the environment, we must parse the incoming acoustic waveform into separate mental representations called auditory objects or, if they persist over time, auditory streams. The problem is compounded by the fact that each ear has access to only a single pressure wave that comprises acoustic energy coming from all active sound sources. The cocktail party problem, which is a classic example of speech perception in an adverse listening situation, can therefore be viewed as a real-world auditory scene analysis problem in which individuals must attempt to separate the various, simultaneously active sound sources while trying to integrate and follow a particular ongoing conversation over time. It should come as no surprise, therefore, that auditory perceptual organization commonly is described across two axes: organization along the frequency axis (i.e. simultaneous organization) involves the moment-to-moment parsing of acoustic elements according to their different frequency and harmonic relations; organization along the time axis (sequential organization) entails the grouping of successive auditory

events that occur over several seconds into one or more streams. Understanding how speech and other complex sounds are translated from the single pressure waves arriving at the ears to internal sound object representations may have important implications for the design of more effective therapeutic interventions. Moreover, it becomes essential to understand the effects of normal aging on listeners' abilities to function in complex listening situations where multiple sound sources are producing energy simultaneously (e.g. cocktail party or music concert) if we hope to provide treatments through which individuals can better cope with these changes and continue to have fulfilling auditory experiences.

Auditory scene analysis is particularly useful for thinking about complex acoustic environments because it acknowledges both that the physical world acts upon us as our perception is influenced by the structure of sound (i.e. bottom-up contributions), just as we are able to modulate how we process the incoming signals by focusing attention on certain aspects of an auditory scene according to our goals (i.e. top-down contributions). As for the inherent properties of the acoustic world that influence our perception, sounds emanating from the same physical object are likely to begin and end at the same time, share the same location, have similar intensity and fundamental frequency, and have smooth transitions and "predictable" auditory trajectories. Consequently, it has been proposed that acoustic, like visual, information can be perceptually grouped according to Gestalt principles of perceptual organization, such as grouping by similarity and good continuation (Bregman, 1990). Figure 64.1 shows a schematic diagram of these basic components of auditory scene analysis.

Many of the grouping processes are considered automatic or *primitive* because they can occur irrespective of listener expectancy and attention. Therefore, an initial stage of auditory scene analysis following basic feature extraction involves low-level processes in which fine spectral and temporal analyses of the acoustic waveform are employed so that distinct perceptual objects can be formed. However, the perception of our auditory world is not always imposed upon us. Our knowledge from previous experiences with various listening situations can influence how we process and interpret complex auditory scenes. These higher-level schema-driven processes involve the selection and comparison between current auditory stimulation and prototypical representations of sounds held in long-term memory. It is thought that both primitive and schema-driven processes are important for the formation of auditory objects, and these two types of mechanisms might interact with each other to constrain perceptual organization.

The auditory scene analysis framework allows for the act of listening to be dynamic, since previously heard sounds may lead us to anticipate subsequent auditory events. This is well illustrated by phenomena in which the

64. Perceptual Organization of Sounds

Figure 64.1 Schematic representation of a working model of auditory scene analysis. This model postulates low-level analyses where basic sound features are processed such as frequency, duration, intensity, and location. Scene analysis can be divided into at least two modes: primitive (bottom-up) and schema-driven (top-down). Primitive processes, the subject of most auditory scene analysis research, relies on cues immediately provided by the acoustic structure of the sensory input. These processes take advantage of regularities in how sounds are produced in virtually all natural environments (e.g. unrelated sounds rarely start at precisely the same time). Schema-driven processes, on the other hand, are those involving attention, or are based on experience with certain classes of sounds—for example, the processes employed by a listener in singling out a familiar melody interleaved with distracting tones. Our behavioral goals may further bias our attention toward certain sound attributes or the activation of particular schemas depending on the context. In everyday listening situations, it is likely that both primitive and schema-driven modes are at play in solving scene analysis problems.

brain “completes” acoustic information masked by another sound, as in the continuity illusion (see Figure 64.2A).

In this illusion, discontinuous tone glides are heard as continuous when the silences are filled by noise bursts. In another example, missing phonemes can be perceptually restored when replaced by extraneous noise, and these effects occur primarily in contexts that promote the perception of the phoneme, suggesting that phonemic restoration may be guided by schema-driven processes (Repp, 1992).

With respect to speech and music, schemata are acquired through exposure to auditory stimuli; they are consequently dependent on a listener’s specific experiences. For instance, in speech processing, the use of prior context helps listeners to identify the final word of a sentence embedded in noise (Pichora-Fuller, Schneider, and Daneman, 1995). Similarly, it is easier to identify a familiar melody interleaved with distractor sounds if the listener knows in advance the title of the melody (Dowling, Lung, and Herrbold, 1987) or if they have been presented with the same melody beforehand (Bey and McAdams, 2002).

Even though the role of both bottom-up and top-down processes are acknowledged in auditory scene analysis, the

effects of age on these collective processes are not well defined. Yet, it is clear that deficits in both simultaneous and sequential sound organization would have dramatic consequences on everyday acoustic computations such as speech perception. For example, deficits in concurrent sound organization may result in the perceiver being unable to adequately separate the spectral components of the critical speech event from the background noise. A failure to segregate overlapping stimuli may also result in false recognition based on the properties of the two different sounds. For example, in dichotic listening procedures, which involve simultaneous presentation of auditory materials to the two ears, individuals presented with “back” in one ear and “lack” in the other ear often report hearing “black,” suggesting that acoustic components from two different sources may be “miscombined” into one percept (Handel, 1989). Moreover, difficulties in integrating acoustic information over time and maintaining the focus of auditory attention may further limit the richness of acoustic phenomenology, leading to problems in social interaction and an eventual retreat from interpersonal relations.

Although auditory scene analysis has been investigated extensively for almost 30 years, and there have been several attempts to characterize central auditory deficits in older adults, major gaps between psychoacoustics and aging research remain. The goal of this chapter is to assess the scene analysis framework as a potential account for age-related declines in processing complex acoustic signals such as speech and music. This review may assist in the development of more effective ways to assess and rehabilitate older adults who suffer from common types of hearing difficulty by evaluating the role of bottom-up and top-down processes in solving scene analysis problems. We now consider auditory scene analysis with respect to the initial parsing of the acoustic signal, before going on to discuss more schema-driven and attention-based mechanisms in audition. With respect to primitive grouping, this may be performed by segregating and grouping concurrently- or sequentially-presented sounds.

Concurrent Sound Segregation

A powerful way to organize incoming acoustic energy involves an analysis of the harmonic relations between frequency peaks. Typically, an examination of the spectra for single sound sources in the real world reveals consistencies in the spacing of frequency peaks. For example, complex sounds with a distinct pitch (e.g. a voice or a chord) corresponding to a low fundamental frequency (f_0) are accompanied by other higher frequencies that are multiples of the f_0 . This is because frequencies emanating from the same source tend to be harmonically related to one another; that is, frequency peaks are approximately integer multiples of the f_0 . If acoustic energy contains frequency elements that are not related to the dominant f_0 , then it is likely that this portion of

Claude Alain, Benjamin J. Dyson, and Joel S. Snyder

Figure 64.2 A. Continuity illusion. Schematic representation of a repeated ascending and descending glide pattern interrupted by a brief gap (left panel). When the gaps are replaced by loud broadband noise (right panel) listeners reported hearing an ascending and descending glide pattern without interruption. That is, the brain appears to offer a good guess of what might be happening “behind” the noise. **B.** An example of concurrent sound segregation. Schematic representation of harmonic series comprised of 12 pure tones with a fundamental frequency at 200 Hz. When all the harmonics are integer multiples of the fundamental, observers report hearing one buzz-like sounds (left panel). However, if one of the harmonics is mistuned by 4% or more of its original value then listeners report hearing two sounds simultaneously, a buzz plus another sound with a pure tone quality. The arrow indicates the mistuned harmonic. **C.** An example of sequential sound segregation. The top and bottom sequences show five cycles of a three-tone pattern. When the frequency separation between the adjacent tones is small, most observers report hearing one stream of sound with alternating pitch and a galloping rhythm (indicated by the dashed line in the top panel). However, when the frequency separation is large the three-tone pattern splits into two separate streams of sound with constant pitch and an even rhythm. Each bar represents a pure-tone in the three-tone pattern.

64. Perceptual Organization of Sounds

acoustic energy belongs to a separate, secondary source. Hence, one way of investigating the effects of age on concurrent sound segregation is by means of the mistuned harmonic paradigm. Here, the listener typically is presented with two successive stimuli, one comprised entirely harmonic components, and the other containing a mistuned harmonic (see Figure 64.2B). The task of the listener is to indicate which of the two stimuli contains the mistuned component. Several factors influence the perception of the mistuned harmonic, including the harmonic number in relation to the fundamental, and sound duration (Moore, Peters, and Glasberg, 1985). For example, thresholds for detecting inharmonicity are lower for long- than short-duration sounds.

In a different version of the mistuned harmonic paradigm, listeners are presented with complex sounds with either all tuned or one mistuned harmonic and indicate on each trial whether they hear a single complex sound with one pitch or a complex sound plus a pure tone, which did not “belong” to the complex. Using this more subjective assessment of listeners’ perception, it was found that when a low harmonic is mistuned to a certain degree (4% or more of its original value), listeners often report hearing two concurrent sounds (Alain, Arnott, and Picton, 2001; Moore, Glasberg, and Peters, 1986). The listeners’ likelihood of hearing two concurrent objects increases with the degree of inharmonicity and is greater for long- rather than short-duration (e.g. 100 ms or less) sounds. In terms of phenomenology, the tuned stimulus often sounds like a buzz with a pitch corresponding to the f_0 , whereas the complex sound containing a low partial mistuned by 4% (or more) contains the buzz element plus a separate sound with a pure-tone quality at the frequency of the mistuned harmonic. Therefore, within the same burst of acoustic energy, two concurrent sounds can be perceived when one component of a complex sound is mistuned such that it is not an integer multiple of the f_0 . In this regard, a sound that has a different fundamental from other concurrent sounds (i.e. the mistuned harmonic) might signal the presence of another object within the auditory scene such as a smoke alarm or a telephone ring. The auditory system is thought to achieve this type of perceptual organization by means of a pattern-matching process in which predictions regarding a potential harmonic series are made on the basis of the f_0 (Lin and Hartmann, 1998). Acoustic energy is then filtered through this harmonic sieve or template, essentially separating out the tuned components from mistuned components. When a portion of acoustic energy is mistuned to a sufficient degree, a discrepancy occurs between the perceived frequency and that expected on the basis of the template, signaling to higher auditory centers that multiple auditory objects may be present within the environment. This early registration of inharmonicity appears to be independent of attention (Alain and Izenberg, 2003) and occur in primary auditory cortex, the earliest stage of sound processing in the cerebral cortex (Dyson and Alain, 2004).

Difficulties in resolving harmonic relations may contribute to some of the speech perception problems experienced by older adults since the perception of a mistuned harmonic, like speech embedded in babble or noise, depends on the ability to parse auditory events based on their spectral pattern. Older listeners may have difficulty in detecting mistuning because auditory filters tend to broaden with age (Patterson, Nimmo-Smith, Weber, and Milroy, 1982). In terms of the harmonic template account, this would mean that inharmonic partials would have to be mistuned to a greater degree in order for segregation from the tuned components to occur, given that the harmonic filters in older adults allow for more frequencies to pass through the sieve as tuned components. Therefore, one prediction that the auditory scene analysis account makes is that older adults should exhibit higher thresholds in detecting a mistuned harmonic. Alain *et al.* (2001) tested this hypothesis using a two-interval, two-alternative forced-choice procedure in young, middle-aged, and older adults. Stimuli consisted of short (100 ms) or long (400 ms) duration sounds with either the second, fifth, or eighth harmonic tuned or mistuned. They found that older adults had higher thresholds for detecting a mistuned harmonic than young or middle-aged adults. This age-related increase in mistuning thresholds was comparable for the second, fifth, and eighth harmonic, but was greater for short- than long-duration sounds (see Figure 64.3) and remained significant even after statistically controlling for differences in audiometric threshold between age groups.

This indicates that age-related problems in detecting multiple objects as defined by inharmonicity do not solely depend on peripheral factors and must involve an age-related decline in central auditory functioning. In each group, a subset of participants also completed the

Figure 64.3 Thresholds for detecting inharmonicity for 100 ms and 400 ms sounds in young (YNG), middle-aged (MA), and older (OLD) adults. The threshold for inharmonicity detection is expressed as percent mistuning. The error bars indicate standard error of the mean. Adapted and reprinted with permission from Alain, C. *et al.* (2001). Copyright 2001, Acoustical Society of America.

Claude Alain, Benjamin J. Dyson, and Joel S. Snyder

Speech In Noise (SPIN) test, which requires identifying the last word of a sentence embedded in multitalker babble. The SPIN threshold, estimated by calculating the signal-to-noise ratio at which participants correctly identify 50% of words, was higher in older adults than in young adults. The increased SPIN threshold in individuals who also showed deficits in detecting a mistuned harmonic is consistent with the hypothesis that age-related changes in processes critical for primitive sound segregation may contribute to the speech perception problems found in older adults.

We propose one possible explanation for the effect of concurrent sound segregation on eventual speech identification. Listeners who find it difficult to detect mistuning may assimilate one or more frequency components coming from secondary sound sources into the target signal. The inclusion of extraneous frequency components into the target signal would lead to subsequent errors in perception, given the blend of acoustic energy incorrectly conjoined at an earlier stage of processing. However, since previous investigations used a task that did not require participants to identify the frequency of the mistuned harmonic, it remains an empirical question whether the age-related increase in thresholds for detecting a mistuned harmonic contributes to speech identification problems. Furthermore, it is worth exhibiting caution when generalizing from relatively artificial sounds such as pure-tone complexes to over-learned stimuli such as speech sounds. For instance, perception of ecologically valid or over-learned stimuli may be less sensitive to aging because they are more likely to activate schema-based representations or allow the use of multiple redundant cues.

Another example of concurrent sound segregation that to some extent overcomes the artificiality of the mistuned harmonic paradigm is the double-vowel task. The additional benefits of this task are that it provides a more direct assessment of the effects of age on speech separation, and also evokes the processes involved in acoustic identification as opposed to detection. Here, listeners are presented with a mixture of two phonetically different synthetic vowels either having the same or different f_0 , and are required to indicate which two vowels were presented. Psychophysical studies have shown that the identification rate improves with increasing separation between the f_0 of the two vowels (Assmann and Summerfield, 1994). In addition to this basic finding, Summers and Leek (1998) also reported an age effect, in that older individuals failed to take advantage of the difference in f_0 between the two vowels and that age accounted for most of the variance in vowel identification performance. This effect of age was observed in both normal and hearing-impaired listeners, once again emphasizing the need to consider both deficits in peripheral and central auditory systems.

More recently, Snyder and Alain (in press) found that normal-hearing older adults had greater difficulty than

young adults in identifying two concurrently presented vowels and that older adults took more time to identify the nondominant vowel, especially when the difference between vowel f_0 was small. However, it is unlikely that the age effects on accuracy during the task were due to general perceptual and cognitive decline because the young and older adults did not differ in identifying just the dominant vowel within the acoustic blend. Rather, age effects on performance manifested themselves only when attempting to extract the nondominant vowel. Hence, it appears that age impairs listeners' ability to parse and identify concurrent speech signals. This effect of age on concurrent sound segregation suggests an impairment in performing a detailed analysis of the acoustic waveforms and is consistent with previous research showing age-related declines in spectral (Peters and Moore, 1992) and in temporal (Schneider and Hamstra, 1999) acuity.

In everyday situations, it is fairly unusual to have concurrent sound objects emanating from the same physical location, as was the case in the vowel segregation study. Sounds that radiate from different physical sources are also likely to be coming from different directions and the auditory system must decide whether several sound sources are active or whether there is only one sound source with the second source being generated by the first's reflection from nearby surfaces. Hence, in addition to spectral cues, directional cues alone or in conjunction with spectral and temporal cues can also be used to break apart the incoming acoustic wave into separate concurrent auditory objects. For example, early and now classical work has shown that increasing the spatial separation between two concurrent messages improves performance in identifying the task-relevant message (Cherry, 1953; Treisman, 1964). Another task that has been helpful in investigating a binaural advantage in sound segregation is the masking level difference paradigm, where a masking noise and a target signal are presented either at the same or at different locations. Typically, thresholds for detecting the signal decrease when it is presented at a different location than the masking noise and the magnitude of this threshold reduction is termed the masking level difference (MLD). Pichora-Fuller and Schneider (1991) measured the MLD in young and older adults and showed that thresholds for detecting the signal (e.g. 500 Hz tone) was comparable in both groups when the target and the masking noise were presented at the same location. When binaural cues were introduced, both groups' thresholds decreased although the magnitude of this threshold reduction was smaller in older than in young adults. Similar age-related differences have also been observed using speech sounds (Grose, Poth, and Peters, 1994). Hence, it appears that normal aging impairs the binaural processing necessary to effectively separate a target signal from masking noise (Grose, 1996; Warren, Wagener, and Herman, 1978). However, other studies using whole sentences embedded in noise rather than single brief targets failed to replicate

64. Perceptual Organization of Sounds

the age-related decline in MLD (Helfer, 1992), suggesting that environmental support provided by whole sentences may alleviate age differences in binaural processing. In other words, older adults may compensate for deficits in binaural processing by calling upon schema-driven processes in order to solve the scene analysis problem.

The findings reviewed in this section indicate that the aging brain exhibits difficulty in parsing and identifying concurrent auditory events. This age-related decline in concurrent sound segregation remains even after controlling for age difference in hearing sensitivity. Older adults appear to have difficulty in using both spectral and directional cues in solving the scene analysis problem, indicating general rather than specific deficits in concurrent sound segregation. Further research is needed to clarify the link between spectral and temporal acuity and the age-related decline in concurrent sound segregation. We now turn to the effects of age on sequential sound organization.

Sequential Sound Segregation

A second form of primitive grouping takes place along the time axis, and auditory streaming acts as a striking psychophysical demonstration of how sequential sound segregation works (see Figure 64.2C). In a typical experiment, participants are presented with ABA—ABA—sequences in which A and B are sinusoidal tones of different frequencies and — is a silent interval. The frequency separation between the A and B tones is manipulated to promote either the perception of a single gallop-like rhythm (ABA— ABA—) or the perception of two distinct perceptual streams (A—A—A—A— and B—B—). Listeners are required to indicate when they can no longer hear the two tones as separate streams, but instead hear a single galloping rhythm. The fusion or coherence boundary refers to the frequency separation where individuals perceive a single stream of sounds with a galloping rhythm, whereas the fission or segregation boundary refers to the point where listeners can no longer hear the galloping rhythm and report hearing the sounds as coming from two separate sources. The area between the fusion and fission boundaries is typically ambiguous and leads to a bistable percept in which listeners' perception alternates between hearing the sounds with a galloping rhythm or hearing two concurrent streams. Using such a paradigm, Mackersie, Prida, and Stiles (2001) observed only a tendency for older listeners to have increased threshold at which the single stream became two streams. This weak relationship between fission threshold and age may be related to a lack of power given that only five participants with normal hearing formed the sample. However, Mackersie *et al.* also reported a strong relationship between fusion threshold and simultaneous sentence perception, suggesting that sequential sound segregation does play an important role in higher auditory and cognitive functions. Rose and Moore (1997) showed that

bilaterally hearing impaired listeners required greater frequency separation than normal hearing listeners for stream segregation to occur. Moreover, the impact of hearing loss on sequential stream segregation has been observed in a number of other studies using a similar paradigm to that of Rose and Moore (Grimault, Micheyl, Carlyon, Arthaud, and Collet, 2001; Mackersie, Prida, and Stiles, 2001). These changes in sequential stream segregation for hearing-impaired listeners and older adults with normal hearing may once again be related to a broadening of auditory filters, in that wider frequency regions are processed in the same frequency channels (i.e. grouped together) in the older individuals. However, Mackersie *et al.* (2001) and Rose and Moore (1997) did not find a significant relationship between stream segregation and frequency selectivity. Furthermore, Stainsby, Moore, and Glasberg (2004) showed that older hearing-impaired listeners could segregate sounds using temporal cues and suggested that stream segregation does not depend solely on the frequency resolution of the peripheral auditory system. Thus, it appears that both place and temporal coding of acoustic information contribute to sound segregation and that although older adults with either normal or impaired hearing can segregate sequential acoustic information, they nevertheless show a tendency to report hearing only one perceptual stream.

One limitation of these studies, however, was that the effects of age on sequential stream segregation were assessed using subjective reports only. Other studies using more objective measures of auditory streaming suggest that aging per se may have little impact on sequential stream segregation. For example, Alain, Ogawa, and Woods (1996) used a selective attention task to examine the effects of age on auditory streaming. Young and older adults with normal hearing were presented with four-tone patterns presented repetitively at a high rate for several minutes. Participants were asked to focus their attention to either the lowest or highest frequency in order to detect infrequent changes in pitch (i.e. targets). In one condition, the frequencies composing the pattern were equally spaced at three semitone intervals along the musical scale (e.g. 1000, 1189, 1414, and 1682 Hz). In another condition, the two middle frequency sounds were clustered with the lowest and highest frequency, respectively (1000, 1059, 1587, and 1682 Hz). Frequency clustering was expected to favor auditory streaming because sounds that are near one another in frequency tend to be perceptually grouped with one another. Alain *et al.* found that accuracy as well as response time to targets improved in situations that promoted auditory stream segregation. Although older adults were overall slower than young adults, the effects of streaming on accuracy and response time were comparable in both age groups, suggesting that aging has little impact on sequential sound segregation when assessed by objective measures such as reaction time and accuracy.

Claude Alain, Benjamin J. Dyson, and Joel S. Snyder

Additional aspects of sequential sound segregation in older adults were explored by Trainor and Trehub (1989). Here, temporal order judgments were measured across age groups for sequences of tones presented at various speeds and in various contexts designed to encourage or discourage stream segregation. Participants were exposed to alternating frequency patterns and asked either to identify the order of the stimuli or to make same/different judgments. They found that older adults were less able than young adults in distinguishing between tone sequences with contrasting order, regardless of the speed of presentation, nature of the task (identification vs. same/different), amount of practice, or the frequency separation of the tones. Therefore, these findings are consistent with a temporal sequencing impairment in older listeners, but do not suggest age differences in streaming processes.

Although the putative central acoustic deficits associated with aging do not appear to have an effect on sequential sound segregation per se, sensory hearing loss does appear to impair listeners' ability to separate out sequentially occurring stimuli. These findings should be interpreted with caution, however, as the sample size in each age group tends to be relatively small in the case of the gallop paradigm, whereas the other tasks used (i.e. the selective attention task, and judgment order task) may not have been optimal to assess auditory streaming. Moreover, a direct link between these age-related changes and the speech perception problems of older adults has yet to be established since the relation between task performance and speech perception has not been fully explored. The Trainor and Trehub, (1989) study reminds us that although older adults appear to show no deficit in segregating multiple acoustic streams, the integration of any one particular stream over time may be a problem for the elderly.

Schema-Driven and Attention-Dependent Processes

Current models of auditory scene analysis postulate both low-level automatic processes and higher-level controlled or schema-based processes (Alain and Arnott, 2000; Bregman, 1990) in forming an accurate representation of the incoming acoustic wave. Whereas automatic processes use basic stimulus properties such as frequency, location, and time to segregate the incoming sounds, controlled processes use previously learned criteria to group the acoustic input into meaningful sources and hence require interaction with long-term memory. Therefore, in addition to bottom-up mechanisms, it is also important to assess how aging affects top-down mechanisms of auditory scene analysis.

The use of prior knowledge is particularly evident in adverse listening situations such as a cocktail party scenario. For example, a person could still laugh in all the

right places at the boss's "humorous" golfing anecdote as a result of having heard the tale numerous times before, even though only intermittent segregation of the speech is possible (indeed, the adverse listening conditions in this example may be a blessing in disguise). In an analogous laboratory situation, a sentence's final word embedded in noise is more easily detected when it is contextually predictable (Pichora-Fuller, Schneider, and Daneman, 1995), and older adults appear to benefit more than young adults from contextual cues in identifying the sentence's final word (Pichora-Fuller, Schneider, and Daneman, 1995). Since words cannot be reliably identified on the basis of the signal cue alone (i.e. without context), stored knowledge must be applied to succeed. That is to say that the context provides environmental support, which narrows the number of possible alternatives to choose from, thereby increasing the likelihood of having a positive match between the incoming sound and stored representations in working and/or longer-term memory. There is also evidence that older adults benefit more than young adults from having words spoken by a familiar than unfamiliar voice (Yonan and Sommers, 2000), suggesting that older individuals are able to use learned voice information to overcome age-related declines in spoken word identification. Although familiarity with the speaker's voice can occur incidentally in young adults, older adults need to focus attention on the stimuli in order to benefit from voice familiarity in subsequent word identification tasks (Church and Schacter, 1994; Pilotti, Beyer, and Yasunami, 2001). Thus, schema-driven processes provide a way to resolve perceptual ambiguity in complex listening situations and, consequently, older adults appear to rely more heavily on controlled processing in order to solve the scene analysis problem.

Musical processing provides another real-world example that invokes both working memory representations of current acoustic patterns and long-term memory representations of previous auditory structures. Evidence suggests that young and older adults perform equally well in processing melodic patterns that are presented in a culturally familiar musical scale, whereas older adults perform worse than young adults when the patterns are presented in culturally unfamiliar scales (Lynch and Steffens, 1994). The age difference in processing melodic patterns from unfamiliar cultural contexts again suggests that older adults may rely more heavily on long-term knowledge of musical grammar than young adults. This age effect may be related to impairment either in the processing of ongoing melodic patterns and/or working memory given that a long-term representation of the unfamiliar melodic structure is unavailable. Other studies have shown that aging impairs listeners' ability to recognize melodies (Andrews, Dowling, Bartlett, and Halpern, 1998), and this age-related decline is similar for musicians and nonmusicians (Andrews, Dowling, Bartlett, and Halpern, 1998), suggesting that musical training does not necessarily alleviate age-related decline

64. Perceptual Organization of Sounds

in melodic recognition. The use of musical stimuli in aging research offers a promising avenue for exploring the role of long-term representation and its relation with schema-driven processes involved in solving the scene analysis problem. Moreover, tasks involving musical stimuli are likely to be more engaging for participants than tasks using typical laboratory stimuli (e.g. white noise, pure tones, harmonic series) that may be less pleasant to listen to for extended periods of time. Furthermore, the results possess a higher degree of ecological validity in terms of everyday, meaningful acoustic processing.

Methodological Issues

In reviewing the preceding literature, a number of methodological procedures suggest themselves for future research on aging and auditory scene analysis. First, and somewhat obvious, it is essential that participants be screened for hearing impairment in both ears. In particular, the integrity of the cochlea should be assessed in a more comprehensive way than just measuring pure tone thresholds in an effort to dissociate nonthreshold-changing peripheral deficits from true central deficits in auditory processing. One test that could provide some information is a distortion-product otoacoustic emission (OAE). Although outer hair cells must be functioning to some degree in order to measure normal thresholds, OAE's are a more sensitive measure of outer hair cell health than pure-tone audiometry and may prove useful in assessing cochlear damage in general. Another potentially useful test would be fine-scale audiometry, which consists of obtaining thresholds with a greater degree of frequency resolution. This method can provide evidence for hearing changes that are not reflected in octave-wide measurements. There are obviously a number of deficits that cannot be easily accounted for by any peripheral problems, but the age-related problems in a wide range of auditory tasks might be accounted (at least partly) by peripheral degradation. Distortion-product OAE assessment might be useful in addressing this issue since it is readily available (and fast), and because it is generally accepted as a sensitive measure of cochlear integrity.

Whenever possible, speech discrimination scores should also be obtained for each ear. Given that the nature of auditory scene analysis research can often rely upon somewhat artificial stimuli, the acquisition of SPIN test scores (or another comparable test) becomes critical. This allows the researcher to establish a correlation between performance in experimental tasks thought to tap into the operation of low-level acoustic mechanisms and speech comprehension in noise, which is an important end product of auditory scene analysis. Also, since musical expertise is likely to influence auditory scene analysis, participants should be asked about the duration of any musical training and groups should be matched for musical training and listening habits. All of this should be in addition to using standard exclusion screening

criteria such neurological or psychiatric illness, and drug or alcohol abuse.

It is well recognized that hearing sensitivity diminishes with age, especially in the high frequency range. Despite screening for hearing impairment, it is not uncommon to find residual differences in hearing thresholds between young and old adults. Hence, several steps should be taken to dissociate the contribution of age-related changes in peripheral processing from those in central auditory processing. First, stimuli should be presented at a suprathreshold level (e.g. 80 dB sound pressure level) rather than at a given intensity increment above hearing threshold. This approach controls for loudness recruitment associated with sensorineural hearing loss (Moore, 1995). Second, stimuli generated using the lower frequency range should be used wherever possible as presbycusis affects primarily higher frequency signals. Third, analyses of covariance should be performed using participants' audiometric levels (pure tone and speech reception thresholds) as covariates to adjust for the contribution of age-related changes in hearing thresholds on behavioral measurements. Whenever possible, three or more age groups should be tested, thereby enabling the examination of when specific maturation processes begin, as well as observing differences in the rate of changes across the various age groups. Such experimental design considerations provide information that goes beyond merely identifying areas where older people perform worse than young people. Taken as a whole, the earlier research strategy provides a means to assess central auditory processing while taking into account differences in hearing thresholds.

Conclusion

Our auditory environment is inherently complex. Hence, it should not come as a surprise that successful identification of simultaneous and sequentially occurring acoustic events depends on listeners' ability to adequately parse sound elements emanating from different physical sounds, and at the same time, combine those arising from a particular source of interest across time. Solving this problem involves low-level mechanisms that take advantage of physical properties in the acoustic environment as well as high-level mechanisms that take advantage of our knowledge about the auditory world. The evidence reviewed in this chapter indicates that aging impairs listeners' ability to parse concurrent sounds based on both spectral and directional cues. Further research is nevertheless needed to clarify the link between age-related decline in low-level sound segregation and higher-level processes such as the perception of speech in noise. It is also unclear whether extended training could be used to alleviate some of the age effects on simultaneous sound and speech separation.

The robust and reliable age-related decline in concurrent sound segregation contrasts with the more subtle

Claude Alain, Benjamin J. Dyson, and Joel S. Snyder

age effects observed in sequential sound segregation. An important difference between concurrent and sequential sound segregation is that the former relies on a more precise temporal and frequency representation. The studies reviewed in this chapter suggest that the likelihood of deficits in auditory stream segregation increases with hearing loss, in that the greater the hearing loss the less likely the individual perceptually segregates the incoming stimuli into two distinct streams of sounds. This suggests that sequential sound segregation involves a relatively coarse temporal and frequency representation, which may be more resilient to the aging process than concurrent sound segregation.

We propose that speech and music perception can be viewed as complex auditory scene analysis problems that engage both low-level and schema-driven processes. It is likely to be the interaction between low-level and high-level mechanisms that lead to successful speech identification in adverse listening situations since primitive grouping mechanisms alone are not sufficient to fully account for perceptual organization of speech (Remez, Rubin, Berns, Pardo, and Lang, 1994). In solving the scene analysis problem, older adults appear to rely more on schema-driven than on low-level processes. Although knowing what to listen for helps both young and older adults, older adults appear to benefit the most from prior learning in understanding their acoustic environment. Once again, these results leave many questions unanswered and therefore further research is needed to understand the extent to which schema-driven processes are affected by age. Nevertheless, by using the scene analysis framework scientists can bridge various areas of aging and hearing research that encompass signal detection and the formation of more abstract representations of our auditory environment. Only by considering both bottom-up and top-down influences, as the current framework allows us to do, may we arrive at a complete picture as to how the auditory scene changes as a function of aging.

Recommended Resources

Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sounds*. London, England: The MIT Press.

Carlyon, R.P. (2004). How the brain separates sounds. *Trends Cogn Sci*, 8(10), 465–471.

ACKNOWLEDGMENTS

The preparation of this chapter was supported by grants from the Canadian Institutes for Health Research, the Natural Sciences and Engineering Research Council of Canada, and the Hearing Foundation of Canada. We are grateful to the volunteers who participated in the experiments reviewed here from our

laboratory. Special thanks to Steve Aiken, Lori Bernstein, and Kelly McDonald for helpful comments on earlier versions of this chapter.

REFERENCES

- Abel, S.M., Giguere, C., Consoli, A., and Papsin, B.C. (2000). The effect of aging on horizontal plane sound localization. *J Acoust Soc Am*, 108(2), 743–752.
- Abel, S.M., Krever, E.M., and Alberti, P.W. (1990). Auditory detection, discrimination and speech processing in ageing, noise-sensitive and hearing-impaired listeners. *Scandinavian Audiology*, 19, 43–54.
- Abel, S.M., Sass-Kortsak, A., and Naugler, J.J. (2000). The role of high-frequency hearing in age-related speech understanding deficits. *Scand Audiol*, 29(3), 131–138.
- Alain, C. and Arnott, S.R. (2000). Selectively attending to auditory objects. *Front Biosci*, 5, D202–D212.
- Alain, C., Arnott, S.R., and Picton, T.W. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J Exp Psychol Hum Percept Perform*, 27(5), 1072–1089.
- Alain, C. and Izenberg, A. (2003). Effects of Attentional Load on Auditory Scene Analysis. *J Cogn Neurosci*, 15(7), 1063–1073.
- Alain, C., McDonald, K.L., Ostroff, J.M., and Schneider, B. (2001). Age-related changes in detecting a mistuned harmonic. *J Acoust Soc Am*, 109(5 Pt 1), 2211–2216.
- Alain, C., Ogawa, K.H., and Woods, D.L. (1996). Aging and the segregation of auditory stimulus sequences. *J Gerontol B Psychol Sci Soc Sci*, 51(2), 91–93.
- Andrews, M.W., Dowling, W.J., Bartlett, J.C., and Halpern, A.R. (1998). Identification of speeded and slowed familiar melodies by younger, middle-aged, and older musicians and nonmusicians. *Psychol Aging*, 13(3), 462–471.
- Assmann, P. and Summerfield, Q. (1994). The contribution of waveform interactions to the perception of concurrent vowels. *Journal of Acoustical Society of America*, 95(1), 471–484.
- Bey, C. and McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Percept Psychophys*, 64(5), 844–854.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sounds*. London, England: The MIT Press.
- Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5), 975–979.
- Chmiel, R. and Jerger, J. (1996). Hearing aid use, central auditory disorder, and hearing handicap in elderly persons. *J Am Acad Audiol*, 7(3), 190–202.
- Church, B.A. and Schacter, D.L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *J Exp Psychol Learn Mem Cogn*, 20(3), 521–533.
- Divenyi, P.L. and Haupt, K.M. (1997). Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation. *Ear Hear*, 18(3), 189–201.
- Dowling, W.J., Lung, K.M., and Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Percept Psychophys*, 41(6), 642–656.

64. Perceptual Organization of Sounds

- Duquesnoy, A.J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J Acoust Soc Am*, 74(3), 739–743.
- Dyson, B. and Alain, C. (2004). Representation of concurrent acoustic objects in primary auditory cortex. *Journal of Acoustical Society of America*, 115, 280–288.
- Gordon-Salant, S. and Fitzgibbons, P.J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research*, 36, 1276–1285.
- Gordon-Salant, S. and Fitzgibbons, P.J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *J Acoust Soc Am*, 115(4), 1808–1817.
- Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., and Collet, L. (2001). Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *Br J Audiol*, 35(3), 173–182.
- Grose, J.H. (1996). Binaural performance and aging. *J Am Acad Audiol*, 7(3), 168–174.
- Grose, J.H., Poth, E.A., and Peters, R.W. (1994). Masking level differences for tones and speech in elderly listeners with relatively normal audiograms. *J Speech Hear Res*, 37(2), 422–428.
- Halpern, A.R., Bartlett, J.C., and Dowling, W.J. (1995). Aging and experience in the recognition of musical transpositions. *Psychol Aging*, 10(3), 325–342.
- Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. London: The MIT Press.
- Helfer, K.S. (1992). Aging and the binaural advantage in reverberation and noise. *J Speech Hear Res*, 35(6), 1394–1401.
- Humes, L.E. and Lisa, R. (1990). Speech-recognition difficulties of the hearing-impaired elderly: the contribution of audibility. *Journal of Speech and Hearing Research*, 33, 726–735.
- Lin, J.Y. and Hartmann, W.M. (1998). The pitch of a mistuned harmonic: Evidence for a template model. *J Acoust Soc Am*, 103(5 Pt 1), 2608–2617.
- Lynch, M.P. and Steffens, M.L. (1994). Effects of aging on processing of novel musical structure. *J Gerontol*, 49(4), P165–P172.
- Mackersie, C.L., Prida, T.L., and Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *J Speech Lang Hear Res*, 44(1), 19–28.
- Middelweerd, M.J., Festen, J.M., and Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology*, 29(1), 1–7.
- Moore, B.C. (1995). *Perceptual Consequences of Cochlear Damage*. New York: Oxford University Press.
- Moore, B.C., Glasberg, B.R., and Peters, R.W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J Acoust Soc Am*, 80(2), 479–483.
- Moore, B.C., Peters, R.W., and Glasberg, B.R. (1985). Thresholds for the detection of inharmonicity in complex tones. *J Acoust Soc Am*, 77(5), 1861–1867.
- Morrell, C.H., Gordon-Salant, S., Pearson, J.D., Brant, L.J., and Fozard, J.L. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *J Acoust Soc Am*, 100(4 Pt 1), 1949–1967.
- Patterson, R.D., Nimmo-Smith, I., Weber, D.L., and Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J Acoust Soc Am*, 72(6), 1788–1803.
- Peters, R.W. and Moore, B.C.J. (1992). Auditory filter shapes at low center frequencies in young and elderly hearing impaired subjects. *Journal of Acoustical Society of America*, 91(1), 256–266.
- Pichora-Fuller, M.K. and Schneider, B.A. (1991). Masking-level differences in the elderly: A comparison of antiphasic and time-delay dichotic conditions. *J Speech Hear Res*, 34(6), 1410–1422.
- Pichora-Fuller, M.K., Schneider, B.A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of Acoustical Society of America*, 97(1), 593–608.
- Pilotti, M., Beyer, T., and Yasunami, M. (2001). Encoding tasks and the processing of perceptual information in young and older adults. *J Gerontol B Psychol Sci Soc Sci*, 56(2), P119–P128.
- Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., and Lang, J.M. (1994). On the perceptual organization of speech. *Psychol Rev*, 101(1), 129–156.
- Repp, B.H. (1992). Perceptual restoration of a “missing” speech sound: Auditory induction or illusion? *Percept Psychophys*, 51(1), 14–32.
- Rose, M.M. and Moore, B.C. (1997). Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. *J Acoust Soc Am*, 102(3), 1768–1778.
- Schneider, B.A. and Hamstra, S.J. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. *J Acoust Soc Am*, 106(1), 371–380.
- Snyder, J.S. and Alain, C. (in press). Age-related changes in neural activity associated with concurrent vowel segregation. *Cognitive Brain Research*.
- Stainsby, T.H., Moore, B.C., and Glasberg, B.R. (2004). Auditory streaming based on temporal structure in hearing-impaired listeners. *Hear Res*, 192(1–2), 119–130.
- Summers, V. and Leek, M.R. (1998). FO processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J Speech Lang Hear Res*, 41(6), 1294–1306.
- Trainor, L.J. and Trehub, S.E. (1989). Aging and auditory temporal sequencing: Ordering the elements of repeating tone patterns. *Percept Psychophys*, 45(5), 417–426.
- Treisman, A. (1964). The effect of irrelevant material on the efficiency of selective listening. *The American Journal of Psychology*, 77, 533–546.
- Warren, L.R., Wagener, J.W., and Herman, G.E. (1978). Binaural analysis in the aging auditory system. *J Gerontol*, 33(5), 731–736.
- Yonan, C.A. and Sommers, M.S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychol Aging*, 15(1), 88–99.

AU: Kindly update.