Auditory stream segregation of iterated rippled noises by normal-hearing and hearing-impaired listeners^{a)}

Daniel E. Shearer,^{b)} Michelle R. Molis, Keri O. Bennett, and Marjorie R. Leek^{c)}

National Center for Rehabilitative Auditory Research, Portland VA Healthcare System, Portland, Oregon 97239, USA

(Received 16 November 2016; revised 9 August 2017; accepted 2 January 2018; published online 25 January 2018)

Individuals with hearing loss are thought to be less sensitive to the often subtle variations of acoustic information that support auditory stream segregation. Perceptual segregation can be influenced by differences in both the spectral and temporal characteristics of interleaved stimuli. The purpose of this study was to determine what stimulus characteristics support sequential stream segregation by normal-hearing and hearing-impaired listeners. Iterated rippled noises (IRNs) were used to assess the effects of tonality, spectral resolvability, and hearing loss on the perception of auditory streams in two pitch regions, corresponding to 250 and 1000 Hz. Overall, listeners with hearing loss were significantly less likely to segregate alternating IRNs into two auditory streams than were normally hearing listeners. Low pitched IRNs were generally less likely to segregate into two streams than were higher pitched IRNs. High-pass filtering was a strong contributor to reduced segregation for both groups. The tonality, or pitch strength, of the IRNs had a significant effect on streaming, but the effect was similar for both groups of subjects. These data demonstrate that stream segregation is influenced by many factors including pitch differences, pitch region, spectral resolution, and degree of stimulus tonality, in addition to the loss of auditory sensitivity.

<https://doi.org/10.1121/1.5021333>

[JFL]

Pages: 378–387

I. INTRODUCTION

For listeners with sensorineural hearing loss, decreased audibility of sound is typically associated with a degradation of spectral and temporal processing ability (Moore, 1995). Together, these factors may result in impaired perception of target signals, such as speech, in the presence of background sounds. Poor speech-in-noise understanding can be related to a listener's ability to properly group similar acoustic elements together into a coherent auditory object—a perceptual attribute known as “stream integration”—while also separating dissimilar acoustic elements into separate sources—a process known as “source separation” or “stream segregation” (see Moore and Gockel, 2002, for a review). Mackersie (2003) reported significant correlations in hearing-impaired listeners between concurrent speech sound perceptions and the tonal frequency difference at which a sequence of alternating tones was perceived as a single auditory stream. In normal-hearing (NH) or mildly hearing-impaired (HI) subjects, the ability to pull apart two simultaneous words presented with differences in pitch (“concurrent stream segregation”) was related to the tendency to segregate or integrate sequences of natural vowel sounds with different pitches (“sequential stream segregation”; Gaudrain *et al.*,

2012; Mackersie, 2003). However, neither the word nor vowel segregation results were related to measures of frequency selectivity in those studies. The focus of the current study is sequential stream segregation.

The tendency of listeners to perceptually organize signals into single or multiple streams has been tested repeatedly in experiments where sounds are presented in sequences that differ along a variety of potentially relevant acoustical dimensions. Sequential stream segregation has been shown to be influenced by differences in the frequency of pure tones (Micheyl and Oxenham, 2010; Rose and Moore, 1997; van Noorden, 1975), the fundamental frequency of harmonic complex tones (Grimault *et al.*, 2000, 2001; Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999), changes in waveform shapes (envelope) due to phase relationships within harmonic complexes (Roberts *et al.*, 2002), and spectral characteristics of narrowband noises, such as bandwidth, center frequency and the amount of frequency separation of the edges of the alternating noise bands (Bregman *et al.*, 2001). Stream segregation has also been shown to be influenced by aspects of attention, lexical, and acoustic characteristics of familiar speech sounds (Billig *et al.*, 2013; David *et al.*, 2017), including many studies using natural and synthetic speech stimuli. For example, Gaudrain *et al.* (2007, 2008) assessed the effects of frequency resolution on sequential streaming of alternating synthetic speech sounds, and David *et al.* (2017) investigated stream segregation related to changes in fundamental frequency in voiced and unvoiced natural speech consonant-vowel sounds.

The likelihood that a sequence of alternating stimuli will be perceived as two separate streams depends mainly on the acoustic relationship between the components, and

^{a)}Portions of this work were presented at the Annual Scientific and Technology Conference of the American Auditory Society, Scottsdale, AZ, USA, March, 2012.

^{b)}Current address: Alfred I. Dupont Hospital for Children, Wilmington, DE 19899, USA.

^{c)}Current address: VA Loma Linda Healthcare System, Loma Linda, CA 92357, USA. Electronic mail: Marjorie.Leek@va.gov

characteristics of the listener's auditory system, where larger perceived component differences lead to a greater probability of segregation (Bregman *et al.*, 1990; van Noorden, 1975). For example, Micheyl and Oxenham (2010) measured the effect of frequency separation between two pure tones on the likelihood that subjects would perceive one stream or two streams. When the frequency separation was less than about 3 semitones, listeners generally heard one integrated stream, but when the separation was about 6 semitones or greater, most listeners heard two streams. While slight differences between paired stimuli typically lead to stream integration and large differences lead to segregation, there is often a region of perceptual ambiguity where streaming performance is more variable within a listener. When the difference between pairs of alternating sounds are within the ambiguous region, at the beginning of a sequence there is a bias toward hearing an integrated stream, but as the sequence continues in duration the perception may begin to be more segregated (Anstis and Saida, 1985; Hupe and Pressnitzer, 2012). This pattern of buildup of streaming decisions was not supported, however, by Deike *et al.* (2012) for complex tone stimuli. The ambiguity region of frequency difference may be influenced by more cognitive than sensory factors, such as interpretation of instructions or attentional influences (Carlyon *et al.*, 2001; Micheyl and Oxenham, 2010; Deike *et al.*, 2015).

Most studies of auditory stream segregation have focused on the effects of variable stimulus parameters and task demands on the streaming tendencies of normal-hearing listeners. Fewer studies have addressed the effects of these variables on hearing-impaired listeners. Rose and Moore (1997) reported tonal stream segregation results for listeners with either unilateral or bilateral hearing losses as well as listeners with normal hearing. Results, given in auditory filter bandwidth units [i.e., equivalent rectangular bandwidths (ERBs), see Glasberg and Moore, 1990] were mixed, with a great deal of variability observed across listeners. Listeners indicated when the perception of the alternating sequence separated into two streams rather than a single "galloping" rhythm. As expressed with the ERB metric, NH listeners demonstrated no difference in segregation judgments for stimuli in different frequency regions. Listeners with unilateral losses showed no differences on this metric between their good ear and their poorer ear. Although listeners with bilateral hearing loss sometimes showed highly abnormal ERB differences (much larger differences required than normal) they also often showed metrics within the normal range. Since the ears with hearing loss likely had poorer frequency selectivity, these results do not provide strong evidence of a relationship between auditory filter bandwidths and the perception of one or two streams in a sequence. Other authors, however, have hypothesized a relationship between the poorer-than-normal frequency selectivity commonly observed in hearing-impaired listeners and their reduced tendency to segregate streams of harmonic complex tones presented in repeating ABA_n triplets (Grimault *et al.*, 2001). However, Rose and Moore (1997) reported that listeners with unilateral hearing loss did not show consistent tendencies for segregation across their two ears, suggesting

that the good ear might be compensating in some way for the poorer ear.

Vliegen and Oxenham (1999) found that stream segregation by normal-hearing listeners did not require differences in spectral content of pairs of sounds but that it could be coded by periodicity or other temporal cues. However, Vliegen *et al.* (1999) have argued that when both types of cues are present, spectral cues influence streaming more than temporal periodicity cues. Stainsby *et al.* (2004) examined perception of harmonic complexes differing in temporal envelope (created with differing phase relationships among the harmonics) or in spectral passbands. In a subjective stream segregation task, hearing-impaired listeners segregated sequential streams that differed in temporal structure, but not those differing in spectral envelopes. Stainsby *et al.* interpreted this as evidence that frequency-place coding is not required for stream segregation. In fact, the literature on stream segregation in cochlear implant patients indicates that the loss of accurate frequency-place information does not greatly interfere with streaming, which is attributed to a more faithful representation of temporal cues in a tonal sequence (e.g., Chatterjee *et al.*, 2006; Hong and Turner, 2009; Cooper and Roberts, 2009). In general, while the stimulus differences required for hearing-impaired listeners to report subjective stream segregation are greater, there is much variability, and the exact contributions of spectral or temporal cues is not clear.

Some insight might be gained concerning the relative contributions of spectral and temporal cues in streaming by the use of selected noisy stimuli such as iterated rippled noise (IRN). The perceived pitch of these stimuli is influenced most by the quasi-periodicity in the temporal waveform, rather than spectral structure (Yost, 1996, 2009). The relative contribution of temporal pitch cues and spectral cues might be clarified by employing iterated rippled noises (IRNs) as alternating stimuli in a sequence in order to investigate factors such as frequency and periodicity regions, reduction of low frequency information through high-pass filtering, and the influence on streaming of cochlear damage with respect to stimuli whose tonal quality is not based on frequency but instead on temporally based pitch. IRN stimuli are unique in that they can be designed to systematically control their noisy versus tonal quality to have greater or lesser pitch strength. It is expected that IRN stimuli that have greater pitch strength, or greater tonality, will be more readily segregated into two streams than stimuli with less pitch strength. An IRN is created by copying a Gaussian noise and adding the copied waveform to the original noise after a given delay. The repeated addition of the copied waveform gives the aggregate waveform a temporal regularity, which is likely the primary contributor to the perceived pitch of the noise. That pitch is typically matched to a pure tone with a frequency equal to the reciprocal of the delay (e.g., a delay of 4 ms is perceived as a pitch around 250 Hz; Yost, 1996).

Figure 1(A) displays the temporal waveform for an IRN with a 4 ms delay, and eight iterations (i.e., repetitions of the delay and add process used to create the IRN). Although there is quasi-periodicity in the temporal fine structure of the

FIG. 1. Examples of temporal waveforms and frequency spectra of several IRNs: (A) Waveform of an IRN with delay of 4 ms, created with eight iterations. Gain is one on each iteration. (B) Spectrum of IRN with 4 ms delay and two iterations; (C) Spectrum of IRN shown in panel (A), with delay of 4 ms and eight iterations; (D) Waveform of an IRN with delay of 1 ms and two iterations. (E) Spectrum of IRN shown in panel (D). (F) spectrum of IRN with 1 ms delay and eight iterations.

IRN, the temporal envelope of the waveform is basically flat. Its spectrum, shown schematically in Fig. 1(C), has relatively noisy, but defined, peaks at frequency intervals equal to the reciprocal of the delay. These spectral peaks are similar to those present in harmonic complexes—though they are usually broader and their width depends on the number of iterative add-delay cycles used in their creation. Decreasing the number of iterations reduces the perceived pitch strength, or tonal quality, of the stimulus, and produces a noisier and less tonal percept. Additionally, the spectral peaks of an IRN are broadened and their peak-to-valley amplitude is decreased with smaller numbers of iterations of the copy, delay, and add processing [Fig. 1(B)]. Though pitch identity is more likely achieved through auditory processing of the IRN temporal fine-structure regularity, the strength of the pitch is also reflected in the characteristics of its spectral peaks (i.e., peak locations, peak-to-valley amplitudes, and peak bandwidths). Figures 1(D)–1(F) display the waveform and schematic spectra for IRNs with a 1 ms delay and either two or eight iterations).

Similar to harmonic complexes, IRNs have been shown to have a pitch dominance region between the third and sixth spectral peaks (Leek and Summers, 2001; Yost, 1982). High-pass filtering may be used to reduce the contribution of the lower-frequency spectral peaks of IRNs while leaving the fine structure regularities intact. If IRNs are filtered to remove the spectral peaks within the dominance region, or those peaks that are spectrally resolvable by listeners, it is expected that the perceived pitch strength (i.e., tonal quality) will be weakened. The pitch identity, however, remains.

In the present study, IRN pairs created with different delay times (and therefore, different pitches) were presented in repeating ABA_n sequences to investigate the subjectively reported stream segregation perceptions of normal-hearing and hearing-impaired listeners. Two pitch regions were evaluated, and the pitch strengths of the stimuli were modified by changing the number of iterations and the spectral content of the stimuli.

Based on the literature, we hypothesized that hearing-impaired listeners would be less likely than normal-hearing listeners overall to segregate sequentially alternating IRNs into two streams, due mainly to relatively inadequate

processing of the temporal and/or spectral cues to pitch present within the stimuli. Because the degree of tonality may be reduced in the presence of hearing loss, as has been shown in frequency discrimination tasks (e.g., Moore and Peters, 1992), the noisy quality of the stimuli is more likely to interfere with the perceived tonality, particularly when fewer cues are available (i.e., when IRNs have fewer iterations or are high-pass filtered above critical spectral components). For both groups of listeners, increasing the number of iterations used to create an IRN should improve the perceived tonality of the stimulus and presumably increase the likelihood of stream segregation. We also expect high-pass filtering at 2000 Hz to reduce critical spectral cues to pitch, particularly for lower-frequency IRNs (250 Hz) in which the spectral peaks within the dominance region (~1000–1750 Hz) fall below the filter cutoff frequency. The result should be a decreased likelihood of stream segregation for both groups of listeners. Finally, we expect that all listeners will more readily segregate streams of low pitch IRNs (250 Hz), based on the observation that the strength of IRN pitches around 250 Hz is stronger than those around 1000 Hz (Leek and Summers, 2001; Yost and Hill, 1978).

II. MATERIALS AND METHODS

A. Subjects

Subjects in this study were recruited from among patients and visitors at the VA Medical Center who had either participated in earlier studies and indicated they wished to be considered for other studies, or who had observed a flyer posted around the medical center advertising the need for subjects, or by word of mouth among the staff. Subjects were divided into two groups: normal-hearing (NH: $n = 5$, aged 25–61, mean = 35.6) and hearing-impaired (HI: $n = 5$, aged 49–78, mean = 63.2). Subjects in the NH group had pure tone audiometric thresholds better than 25 dB hearing level (HL) at octave frequencies between 250 and 8000 Hz, plus interoctave frequencies of 3000 and 6000 Hz. Thresholds of the HI subjects fell within the range of mild-to-moderate hearing loss, with no more than two thresholds worse than 60 dB HL at 250–8000 Hz. Immittance measures in addition to bone conduction testing were used to rule out a conductive component to any hearing

FIG. 2. Mean audiograms for two groups of subjects ($N = 5/\text{group}$). Error bars are standard deviations.

loss. Figure 2 shows the mean audiograms of the NH and HI subjects. Error bars on the figure show standard deviations within each group. Table I presents the individual audiograms of the subjects.

All methods and procedures used in this study were approved by the VA Portland Institutional Review Board, and all subjects provided written informed consent prior to their participation in the experiment. Subjects typically completed the study within two or three 2-h sessions, and they were compensated for their time for testing.

B. Stimuli

The IRNs were constructed by copying, delaying, and iteratively adding a random Gaussian noise to itself. Two or eight iterations were used to produce IRNs that were perceived as more noisy or more tonal, respectively. Two different pitch delays were used, producing pitches that would typically be matched to around 1000 Hz (1 ms delay) or 250 Hz (4 ms delay). (Remembering that these stimuli are broadband noises and are not defined by their frequency region, for the sake of simplicity, the pitch regions will be referred to by their typically matched frequencies.)

TABLE II. Perceptual characteristics evaluated and associated stimulus factors used in creating sequences of IRNs.

Perceptual effects	Stimulus parameters to vary
Tonality/Noisiness	Iterations (2 vs 8)
Pitch identity region	Delay (1 vs 4 ms)
Spectral shape	High-pass cutoffs of filters (50 vs 2000 Hz)
Stream segregation in alternating sequences	Pitch difference of IRN components (0.5, 1, 2, 3, 7, 10 semitones)

Additional IRNs were also generated at six pitch decrements of 0.5, 1, 2, 3, 7, and 10 semitones relative to each of the nominal pitches noted above. To clarify the perceptual aspects of the stimuli, if these were harmonic complexes, the lower pitch range would correspond to fundamental frequencies over a range of 148–250 Hz, and the higher pitch range would be from 558 to 1000 Hz. Finally, two different high-pass filter cutoff frequencies (50 or 2000 Hz) were tested to either include or eliminate resolvable “harmonics” in the frequency spectra of the stimuli. For the purposes of this study, the 50-Hz high-pass stimuli will be considered full bandwidth. In total, the variation of the three stimulus parameters [number of iterations, delay (pitch region), and high-pass filter cutoff] produced eight unique IRN conditions, to test the effects of pitch clarity or noisiness, perceived pitch region, and spectral resolvability on the perception of auditory streams. Table II shows the stimulus parameters that were varied and their respective perceptual effects. The outcomes to be compared among these factors for the two groups of listeners are functions relating the probability of subjects responding that they perceived two streams (stream segregation) rather than only one cohesive stream (stream integration).

IRN stimuli were presented in sequences of eight ABA₃ triplets, where IRN-A had a fixed pitch (corresponding to either 250 or 1000 Hz) and IRN-B varied across trials by a selected semitone difference below IRN-A. Figure 3 illustrates the temporal presentation pattern used (only half a trial is shown on the figure to conserve space). Each IRN within the sequence was 60 ms in duration, including 20-ms $\cos^2$ ramps at the beginning (onset) and end (offset) of the stimulus. There were eight triplets within a presentation sequence;

TABLE I. Audiometric thresholds (dB HL) for individual participant’s right and left ears at each tested frequency (Hz). NH = Normal-hearing listener, HI = hearing-impaired listener.

Participant	Right ear threshold (dB HL) by frequency (Hz)								Left ear threshold (dB HL) by frequency (Hz)							
	250	500	1000	2000	3000	4000	6000	8000	250	500	1000	2000	3000	4000	6000	8000
HI1	30	35	35	45	45	45	55	65	30	35	35	45	45	45	50	60
HI2	25	30	35	30	35	45	45	65	25	30	30	30	20	50	30	60
HI3	35	30	30	35	35	30	20	10	35	30	30	30	35	30	20	15
HI4	15	15	20	40	40	45	60	55	15	30	30	45	45	45	55	50
HI5	15	30	20	25	15	15	25	40	15	25	20	15	20	20	35	45
NH1	5	5	5	5	0	0	5	0	0	−5	5	5	0	5	0	5
NH2	5	10	10	5	10	5	5	15	5	5	10	10	10	5	0	5
NH3	5	10	10	10	0	0	5	0	5	10	5	5	−5	−5	−5	−5
NH4	5	10	10	10	5	5	10	5	5	5	10	10	10	10	25	20
NH5	5	5	0	10	10	10	10	10	5	5	5	15	5	10	10	10

FIG. 3. (Color online) Schematic of presentation sequence (only half the sequence is shown in this display). IRN A represents tones with a fixed IRN delay; IRN B shows a sequence with variable IRN delays from 0.5 to 10 semitones different from sequence A. Within each ABA triplet, durations of the components were 60 ms, separated by a silent interval of 40 ms.

the triplets were separated by 80 ms, and the within-triplet intervals were 40 ms, producing trials with a total duration of 2640 ms.

C. Procedure

Eight different IRN conditions were presented, which together evaluated the effects on stream segregation of number of iterations, IRN delay, and high-pass filtering. Each condition varied from the others by its unique combination of IRN stimulus parameters, which included two levels of iterations (two or eight iterations), two levels of IRN delay (1 or 4 ms, producing pitches of 1000 and 250 Hz, respectively), and two levels of high-pass filter cutoff frequency (50 or 2000 Hz). Each block of 60 trials consisted of one of the eight IRN conditions, chosen randomly with the constraint that all eight conditions would be completed before beginning a second replication. IRN-A of the ABA_ triplets was fixed in regards to the stimulus parameters described above. Within a 60-trial block, six levels of semitone difference between IRN-A and IRN-B were presented 10 times each, all in random order. Each IRN condition was presented to subjects three times, for a total of 24 blocks (1440 IRN sequences) per subject. In total, each semitone difference within a condition was presented 30 times.

IRNs were presented monaurally at 85 dB sound pressure level (SPL) through Sennheiser HD 580 circumaural headphones. An intensity rove of ± 3 dB was applied on each presentation in order to reduce the possibility of intensity cues being used by the subjects. A low-intensity, broadband masker was presented continuously in the test ear to mask possible combination tones (Yost *et al.*, 1998).

Subjects were instructed to respond by touching a touchscreen monitor to report their perception of the IRN sequence after each presentation. A single response was made at the end of the sequence. Subjects were asked to respond as quickly as comfortable, but were allowed as much time as they needed to make their response. The touchscreen displayed two boxes, one labeled “Gallops” and one labeled “Beats.” The subject was trained to touch the “Gallops” box if a coherent stream was perceived, and to touch “Beats” if IRN-A and IRN-B were heard as two

separate streams (see Sec. IID). Subjects were told that if they could not decide on a response, they should select whichever response reflected the stronger percept. Feedback was not provided. There was a 1-s silent delay following each subject response and before the onset of the next trial presentation.

D. Training and anchoring

Prior to data collection, subjects were introduced to the concept of auditory streaming by listening to Track 36, “Pitch Streaming” from *Auditory Demonstrations* (Houtsma *et al.*, 1989), which presents tones in an ABA_ pattern, where tone A is at a fixed frequency and tone B varies incrementally in frequency above or below tone A. The track additionally introduces the descriptor of the “galloping” percept heard when the two tones are fairly close together in frequency, and the “beating” when two separate streams are heard. The description of “gallops” has been used in many auditory streaming studies (e.g., van Noorden, 1975; Mackersie, 2003; Stainsby *et al.*, 2004; Grimault *et al.*, 2001). The term “beats” to signify the alternative response, that the sequence had perceptually broken into two streams, each beating at its own rate, was used to provide a description of what the listeners heard, just as “gallop” was used to signify one stream. Other studies have used the term “fission” rather than “beats” to refer to stream segregation (e.g., Bregman, 1990; Rose and Moore, 1997; Moore and Gockel, 2002).

Once subjects expressed understanding of the concept after listening to the compact disk, a training program began, employing pure tones to teach them what was meant by a response of “gallop” or of “beats.” It should be noted that auditory stream segregation of alternating pure tones is typically a compelling percept given enough separation of tone frequencies and speed of presentation (van Noorden, 1975), so the purpose of the training was to verify for the listener what responses were to be used for a given percept. The training stimuli consisted of two pure tones in sequences of eight ABA_ triplets. The training program included three stimulus conditions based on the constant frequency of tone A (500, 1000, or 1500 Hz). Pilot data suggested that a frequency difference of 9 semitones was large enough for ABA_ sequences to be perceived consistently as two streams. Within each frequency condition, 10 presentations consisted of tone B differing from tone A by 1 semitone and 10 presentations had a tonal difference of 9 semitones within a sequence for a total of 60 sequences presented in random order. Following each sequence of tonal triplets, subjects expressed their perception of the sequence by pressing either “Gallops” (one stream) or “Beats” (two separate streams) on a touchscreen monitor. For the training only (i.e., not in the main experiment itself), visual feedback on the touchscreen monitor was provided. This was done to establish “anchor” points as to what was meant by “gallops” and what the perception was that was clearly “beats.” At the end of the training period, the concepts of “gallops” and “beats” were clearly understood by all the subjects, as demonstrated by the outcomes discussed below.

At the end of a training block, the number of times the subject gave the “correct” answer (either “Gallop” or “Beats”) was displayed for the six conditions of Tone A frequency and Tone B semitone differences. A totally “correct” score would be 0% “Beats” responses when the sequence had a 1-semitone difference and 100% “Beats” response when the difference was 9 semitones. If this pattern of responses had three or more “errors” on three or more of the six training conditions, the subject was re-instructed and could run up to four additional training blocks. However, all subjects met the training criteria in no more than two training blocks. This training presented the subjects with unambiguous differences between integrated and segregated tonal sequences and was intended to evaluate their understanding of auditory streaming and the experimental task. In order to be assured that subjects still had their understanding of the task and established anchor points, before each experimental session, an additional training block was presented. All subjects continued to meet the training criteria above throughout their participation, typically only having one training block at the start of the session on which they scored better than 90% on these pure tone sequences. Note that feedback was not given during the experimental trials using IRN stimuli.

III. RESULTS

The perception of streams of IRN stimuli by NH and HI listeners was evaluated as the percent of responses indicating that sequences were perceived as two streams [“Beats” (i.e., segregated)] averaged across three blocks of 60 trials for each subject and condition. These individual subject scores were then further averaged within the two groups to determine a group mean at each semitone difference. Figure 4 shows group means for the conditions with IRN delay of

1 ms (pitch of 1000 Hz), while Fig. 5 shows mean data for an IRN delay of 4 ms (pitch of 250 Hz). For all conditions and both groups shown in these two figures, with the exception of the HI group in the condition of 4 ms delay and high pass filtering at 2000 Hz, segregation responses rise steeply as the differences between IRN_A and IRN_B increased from 0.5 semitones to 3 semitones, after which the percentage of segregation responses plateaus. Also for all conditions, when there is a group difference in perception of streaming, both groups perceived the sequences similarly for the smaller semitone differences, although the NH listeners segregated the sequence more often and generally at lower semitone differences than the HI group. In other words, for larger “tone” separations, the HI listeners tended to continue to perceive the sequences as one integrated stream. The average perceptions for NH listeners typically rose to at least 80% segregation responses for pitch differences corresponding to at least 3 semitones with the exception of high-pass filtered conditions for the delay of 4 ms [Fig. 5, panels (B) and (D)]. When the full bandwidth was presented for the 1-ms delay [Fig. 4, panels (A) and (C)], the two groups performed similarly regardless of the number of iterations in the IRNs. This was not the case when the stimuli were high-pass filtered at 2000 Hz [Fig. 4, panels (B) and (D)].

A repeated measures analysis of variance (ANOVA) of the percent of segregated responses was performed, with number of iterations, IRN delay, high-pass filter cutoff frequency, and semitone difference as within-subject factors. Subject hearing status was a between-subjects factor.

All main effects were significant. There was a significant effect of group membership [$F(1,8) = 11.20$, $p = 0.01$, $\eta^2 = 0.58$]. Listeners were less likely to segregate stimuli with a delay of 4 ms (Fig. 5), compared to 1 ms (Fig. 4) [$F(1,8) = 28.10$, $p = 0.001$, $\eta_p^2 = 0.78$]. Additionally, a

FIG. 4. For IRN delay of 1 ms, mean percent segregation responses (“beats”) and standard deviations as a function of difference between A and B IRN’s. Panel variable is group; each panel represents different IRN characteristics.

Delay=4; 2 iterations; HP @50 Hz

Delay=4; 2 iterations; HP @2000 Hz

Delay=4; 8 iterations; HP @50 Hz

Delay=4; 8 iterations; HP @2000 Hz

FIG. 5. For IRN delay of 4 ms, mean percent segregation responses (“beats”) and standard deviations as a function of difference between A and B IRN’s. Panel variable is group; each panel represents different IRN characteristics.

high-pass filter with a cutoff frequency at 2000 Hz reduced mean segregation responses across listeners [$F(1,8) = 48.09$, $p < 0.001$, $\eta_p^2 = 0.86$], and more segregated responses occurred as the semitone difference between IRN-A and IRN-B increased [$F(5,40) = 43.81$, $p < 0.001$, $\eta_p^2 = 0.85$]. The “noisiness” of the IRN stimuli, as influenced by their number of iterations used in their creation, had a significant effect across listeners as well [$F(1,8) = 10.00$, $p = 0.01$, $\eta_p^2 = 0.586$], reducing segregation responses for two iterations (noisier) compared with eight iterations (more tonal).

The variables selected for investigation are all fairly strong determinants of stream perception, but these main effects are modulated by the interactions among the variables. Two-way interactions were observed between the listening group and some stimulus factors. Interactions were significant between listening group and IRN delay [$F(1,8) = 7.49$, $p < 0.03$, $\eta_p^2 = 0.48$], high-pass filter cutoff [$F(1,8) = 6.046$, $p < 0.04$, $\eta_p^2 = 0.43$], and semitone difference [$F(5,40) = 4.89$, $p = 0.001$, $\eta_p^2 = 0.38$]. However, there was not a significant interaction between group membership and number of iterations in the stimuli [$F(1,8) = 0.02$, $p = 0.89$, $\eta_p^2 = 0.0$].

The mean data shown in Figs. 4 and 5 are plotted to highlight the effects of IRN iterations, delays, and high-pass filtering across the range of semitone differences for both NH and HI listeners. From both of these figures, the importance of semitone difference between IRNs (shown on the abscissa) is critical. This difference is not unexpected given that frequency difference is well-known to be a very strong factor in the tendency of two streams to separate from a unified sequence, and that finding probably is represented here by pitches of the IRNs. All two-way interactions that included semitone differences were significant (all $p < 0.05$).

Generally, the effect of semitone difference is stronger for NH listeners, as seen by the greater tendency for segregation to occur at smaller semitone differences, and with greater percentages of stream segregation overall for NH listeners. The only combination of variables which did not show this effect was for the full-bandwidth stimuli at the shorter IRN delay (1 ms) [Fig. 4, panels (A) and (C)]. In these conditions, the two listener groups performed identically, on average. All other interaction conditions showed a stronger segregation tendency for NH than for HI listeners.

IV. DISCUSSION

A. Hearing loss and stream segregation

The results of the present study indicate that the tendency to segregate two streams of IRNs is dependent on the listener’s hearing sensitivity, as well as stimulus-dependent factors such as noisiness, pitch region, spectral filtering, and relative pitches of the rapidly alternating components. With the exception of IRN tonality or noisiness, the effects of all tested stimulus variables interacted strongly with the hearing ability of the listener. Generally, the smaller semitone differences were largely heard as integrated, as would be expected when the two pitches in a sequence were similar to one another.

The central finding of this study was an overall main effect of listener group (hearing-impaired vs normal-hearing), showing that, with few exceptions, hearing-impaired listeners were less likely to perceive IRN sequences as segregated, compared to normal-hearing listeners. This may be related to less pitch clarity or pitch strength in HI listeners (Leek and Summers, 2001) given that the differing pitches of the component IRNs are strong determinants of the

tendency to hear a sequence as separated into two streams versus integrated as one stream of similar-pitched components.

There was a strong tendency for subjects with hearing loss to hear two alternating stimuli as an integrated stream (i.e., as a “gallop”), rather than as two separate streams (“beats”). This may be a secondary effect of reduced spectral resolution in these listeners due to broadened auditory filters that typically accompany hearing impairment (Glasberg and Moore, 1986; Leek and Summers, 1993). It has been suggested in several earlier investigations that if the two alternating tones are processed within a single auditory channel (i.e., not resolved), they will be heard as one stream, whereas the pattern will segregate as the tones are separated in frequency and become resolved in the listener’s auditory system (e.g., Rose and Moore, 1997; Grimault *et al.*, 2001). The finding of integrated responses for greater semitone differences within stimulus pairs is consistent with less spectral resolution in listeners with hearing loss. Both groups of subjects showed an increase in segregation responses as the semitone difference grew to three or greater, which is about a 19% difference in frequency. This indicates that as two alternating pitches separated by more than 19% the perception changed from one of an integrated stream to a segregation of two streams. Depending on other stimulus conditions and subject hearing status, however, the tendency to segregate either stayed constant for larger semitone differences or increased slightly.

High-pass filtering had perhaps the largest effect on stream segregation. Lower, resolved frequencies provide the strongest perception of pitch in complex sounds. When the lower frequencies are filtered out, listeners must rely on higher frequency cues which are typically less salient for pitch perception (Yost, 2009). The segregation of a sequence into two streams relies on a separation in pitch of the alternating components, and for sounds with only higher frequencies, the two separate pitches would be less distinct, perhaps resulting in more integration responses than when lower frequencies are present in full-bandwidth stimuli.

For the full-bandwidth stimuli at a delay of 1 ms [Fig. 4, panels (A) and (C)], the two groups performed almost identically. This suggests that stream segregation was highly similar across groups when the peaks in the spectrum were spaced far apart and the frequencies in the region of the lowest peak were present. That common perception did not remain when the frequencies below 2000 Hz were removed, which also eliminated the spectral peak corresponding to the pitch. In that case, the NH group performed essentially the same for full-bandwidth and high-pass filtered stimuli, but HI listeners responded with fewer segregation responses for the high-pass stimuli indicating that HI listeners were more reliant on the temporal or spectral information between 1000 and 2000 Hz and were not as effective in their use of higher frequencies for segregation. NH listeners were not affected by the loss of information below 2000 Hz, but apparently used higher frequency information present in the stimuli to hear two streams.

The story is quite different for the lower pitched stimuli with IRN delays of 4 ms (Fig. 5). Here, although the full-bandwidth stimuli supported a change in perception from

one stream to two streams for both groups, the HI listeners gave more integrated responses than NH listeners. Both groups of listeners were affected by the high-pass filtering of these stimuli at 2000 Hz, again with a greater tendency to hear integrated streams by the HI listeners. For this condition, the NH listeners still increased their segregation responses as semitone differences increased, but they never demonstrated greater than 80% such responses, while the HI listeners maintained integration responses greater than 60% (i.e., <40% segregation responses) for these sequences. For delays of 4 ms, the perceived pitch was low and the spectral components were closely spaced. Loss of frequency resolvability could account for the increased tendency for integration by the HI listeners for the full-bandwidth stimuli. For the stimuli filtered at 2000 Hz, segregation was less likely for the NH listeners and was further reduced for the HI listeners. In this case, both groups are forced to make subjective segregation judgments based on less resolved frequency information.

It was expected that more tonal stimuli would be segregated more readily than noisier stimuli because the noisier stimuli would sound more similar, and therefore would tend to hold together perceptually. There was a significant main effect of tonality but the lack of interaction between hearing status and tonality (number of iterations) was somewhat surprising. Given a particular choice of stimulus characteristics, sometimes more segregation responses were observed more frequently for the more tonal eight-iteration stimuli than the noisier two-iteration stimuli. This was most evident as the increased segregation exhibited by NH listeners for high-pass filtered, 4 ms delay stimuli. The effect of tonality may be observed in the other conditions as well, however, and had a relatively strong effect on segregation of these sequences, but that effect was similar between the two groups of listeners.

Finally, the pitch of IRNs is mostly mediated by temporal processing generating from the regularity in the temporal fine structure of the waveform (Yost *et al.*, 1998). Hearing-impaired listeners have generally weaker pitch percepts from IRN stimuli than normal-hearing listeners (Leek and Summers, 2001), so the reduction in segregation may arise from the weaker pitch percepts generated by the IRNs. They therefore might sound less distinct for the HI listeners than for the NH listeners, leading overall to greater integration into one stream by the impaired group.

B. Age and stream segregation

Aging and hearing loss are often confounded with one another in auditory studies. It must be noted that, in this study, although there was overlap in ages of the two groups, the mean ages were quite different (NH: 35.6 yr; HI: 63.2 yr). Therefore, age as a relevant variable cannot be ruled out. Stainsby *et al.* (2004) used hearing-impaired subjects with ages from 51 to 88 in their study addressing streaming due to the component phase relationships of alternating harmonic complexes. They reported that the temporal processing deficits usually accompanying aging did not materially affect the ability to use temporal aspects of the

stimuli to stream sequences. [Grimault et al. \(2001\)](#) also did not find a direct effect of aging on stream segregation but rather attributed the differences they saw between NH and HI subjects to reduced frequency selectivity. However, [Summers and Leek \(1998\)](#) did find an age effect for separation of concurrent vowels based on fundamental frequency differences, a task related to concurrent stream segregation. They attributed the reduction in performance with increasing subject age to a decline in cognitive processing needed for the separation of speech sounds. Similar cognitive effects would not be in play with these IRN stimuli and a sequential streaming paradigm. Generally, the effects of age on many psychoacoustics tasks are less apparent than the effects of hearing loss, particularly related to impaired frequency resolution and audibility. Therefore, we believe the contribution of age to the differences seen here between NH and HI listeners is less than the characteristics of the hearing loss.

V. CONCLUSIONS

- Hearing-impaired listeners generally showed less tendency to segregate sounds in rapidly alternating patterns. This suggests that the components of the sequence sounded more similar to listeners with hearing loss, compared to perceptions of normal-hearing listeners. This may be due to the alternating pitches being processed through single channels because of a loss of spectral resolution accompanying hearing loss.
- Low-pitched IRNs (250 Hz) were generally less likely to segregate than high-pitched IRNs, especially for hearing-impaired listeners.
- High-pass filtering at 2000 Hz had the greatest effect on segregation ability in both groups, particularly for low-pitched IRNs. This effect may result from the removal of dominant frequencies in the stimuli that are most important for pitch salience ([Yost, 1982](#); [Leek and Summers, 2001](#)).
- Tonality of the stimuli had a significant effect overall, with noisier stimuli leading to less segregation than more tonal stimuli, but the effect did not differ across the two subject groups.

ACKNOWLEDGMENTS

This research was supported by the NIDCD T-35 AUD Research Training Grant DC008764, NIDCD RO1 DC 00626, the VA Rehabilitation Research & Development Grant C6116W and a VA RR&D Senior Research Career Scientist award to M.R.L. (C4042L).

Anstis, S., and Saida, S. (1985). "Adaptation to auditory streaming of frequency-modulated tones," *J. Exp. Psychol.: Hum. Percept. Perform.* **11**, 257–271.

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., and Carlyon, R. P. (2013). "Lexical influences on auditory streaming," *Curr. Biol.* **23**, 1585–1589.

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

Bregman, A. S., Ahad, P. A., and Van Loon, C. (2001). "Stream segregation of narrow-band noise bursts," *Percept. Psychophys.* **63**, 790–797.

Bregman, A. S., Levitan, R., and Liao, C. (1990). "Fusion of auditory components: Effects of the frequency of amplitude modulation," *Percept. Psychophys.* **47**, 68–73.

Carlyon, R. P., Cusack, J. M., Foxton, J. M., and Robertson, I. H. (2001). "Effects of attention and unilateral neglect on auditory stream segregation," *J. Exp. Psychol.: Hum. Percept. Perform.* **27**, 115–127.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (2006). "Auditory stream segregation with cochlear implants: A preliminary report," *Hear. Res.* **222**, 100–107.

Cooper, H. R., and Roberts, B. (2009). "Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition," *J. Acoust. Soc. Am.* **126**, 1975–1987.

David, M., Lavandier, M., Grimault, N., and Oxenham, A. J. (2017). "Sequential stream segregation of voiced and unvoiced speech sounds based on fundamental frequency," *Hear. Res.* **344**, 235–243.

Deike, S., Heil, P., Böeckmann-Barthel, M., and Brechmann, A. (2012). "The build-up of auditory stream segregation: A different perspective," *Front. Psychol.* **3**, 461.

Deike, S., Heil, P., Böeckmann-Barthel, M., and Brechmann, A. (2015). "Decision making and ambiguity in auditory stream segregation," *Front. Neurosci.* **9**, 1–6.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (2007). "Effect of spectral smearing on the perceptual segregation of vowel sequences," *Hear. Res.* **231**, 32–41.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (2008). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," *J. Acoust. Soc. Am.* **124**, 3076–3087.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (2012). "The relationship between concurrent speech segregation, pitch-based streaming of vowel sequences, and frequency selectivity," *Acta Acust. united Acust.* **98**, 317–327.

Glasberg, B. R., and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.* **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (2000). "Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency," *J. Acoust. Soc. Am.* **108**, 263–271.

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**, 173–182.

Hong, R. S., and Turner, C. W. (2009). "Sequential stream segregation using temporal periodicity cues in cochlear implant recipients," *J. Acoust. Soc. Am.* **126**, 291–299.

Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (1989). "Auditory demonstrations" (Compact Disk) (Acoustical Society of America Publications, Melville, NY).

Hupe, J. M., and Pressnitzer, D. (2012). "The initial phase of auditory and visual scene analysis," *Philos. Trans. R. Soc. London B Biol. Sci.* **367**(1591), 942–953.

Leek, M. R., and Summers, V. (1993). "Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise," *J. Acoust. Soc. Am.* **94**, 3127–3137.

Leek, M. R., and Summers, V. (2001). "Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners," *J. Acoust. Soc. Am.* **109**, 2944–2954.

Mackersie, C. L. (2003). "Talker separation and sequential stream segregation in listeners with hearing loss: Patterns associated with talker gender," *J. Speech Lang. Hear. Res.* **46**, 912–918.

Micheyl, C., and Oxenham, A. J. (2010). "Objective and subjective psycho-physical measures of auditory stream integration and segregation," *J. Assoc. Res. Otol.* **11**, 709–724.

Moore, B. C. J. (1995). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, Oxford, UK).

Moore, B. C. J., and Gockel, H. (2002). "Factors influencing sequential stream segregation," *Acta Acust. Acust.* **88**, 320–332.

Moore, B. C. J., and Peters, R. W. (1992). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," *J. Acoust. Soc. Am.* **91**, 2881–2893.

- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," *J. Acoust. Soc. Am.* **112**, 2074–2085.
- Rose, M. M., and Moore, B. C. J. (1997). "Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **102**, 1768–1778.
- Stainsby, T. H., Moore, B. C. J., and Glasberg, B. R. (2004). "Auditory streaming based on temporal structure in hearing-impaired listeners," *Hear. Res.* **192**, 119–130.
- Summers, V., and Leek, M. R. (1998). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," *J. Speech Hear. Res.* **41**, 1294–1306.
- van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation, University of Technology, Eindhoven, The Netherlands.
- Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (1999). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," *J. Acoust. Soc. Am.* **106**, 938–945.
- Vliegen, J., and Oxenham, A. J. (1999). "Sequential stream segregation in the absence of spectral cues," *J. Acoust. Soc. Am.* **105**, 339–346.
- Yost, W. A. (1982). "The dominance region of ripple noise pitch: A test of the peripheral weighting model," *J. Acoust. Soc. Am.* **72**, 416–425.
- Yost, W. A. (1996). "Pitch of iterated rippled noise," *J. Acoust. Soc. Am.* **100**, 511–518.
- Yost, W. A. (2009). "Pitch perception," *Atten. Percept. Psychophys.* **71**, 1701–1715.
- Yost, W. A., and Hill, R. (1978). "Strength of pitches associated with ripple noise," *J. Acoust. Soc. Am.* **64**, 485–492.
- Yost, W. A., Patterson, R. D., and Sheft, S. (1998). "The role of the envelope in processing iterated rippled noise," *J. Acoust. Soc. Am.* **104**, 2349–2361.