Published in final edited form as:

Hear Res. 2017 December ; 356: 118–126. doi:10.1016/j.heares.2017.10.001.

Auditory sequential accumulation of spectral information

Yi Shen

Department of Speech and Hearing Sciences, Indiana University Bloomington, Bloomington, IN 47405

Abstract

In many listening situations, information about the spectral content of a target sound may be distributed over time, and estimating the target spectrum requires efficient sequential processing. Listeners' ability to estimate the spectrum of a random-frequency, six-tone complex was investigated and the spectral content of the complex was revealed using a sequence of bursts. Whether each of the six tones was presented within each burst was determined at random according to a presentation probability. In separate conditions, the presentation probabilities (p) ranged from 0.2 to 1, the total number of bursts varied from 1 to 16, and the inter-burst interval was either 0 or 200 ms. To evaluate the information acquired by the listener, the burst sequence was followed, after a 500-ms silent interval, by the six-tone complex acting as an informational masker and the listener was required to detect a pure-tone target presented simultaneously with the masker. Greater performance in this task indicates more accurate estimation of the spectrum of the complex by the listener. Evidence for integration of information across bursts was observed, and the integration process did not significantly depend on inter-burst interval.

Keywords

temporal integration; informational masking; auditory memory; sequential processing

1. Introduction

The auditory system is highly efficient in processing sequentially presented information. For example, speech recognition in noise is known to depend on contextual cues, and these cues are often distributed across time (e.g., Moore, 2003). Moreover, the processing of this sequential acoustic information is robust against temporal interruptions and masking (e.g., Miller and Licklider, 1950). It appears that the auditory system can keep acoustic information in short-term memory and the stored information is available for retrieval and computation at a later time. In other words, the auditory system is capable of combining “multiple looks” across time (Viemeister and Wakefield, 1991). The current study

Corresponding Author: Yi Shen, Department of Speech and Hearing Sciences, Indiana University Bloomington, Bloomington, IN 47405, shen2@indiana.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

investigated the efficiency of the auditory system in combining temporally distributed information to determine the spectral content of a target sound.

How the auditory system processes acoustic information over time has been a longstanding issue in psychoacoustics. It is known that absolute threshold decreases and loudness increases with increasing stimulus duration (e.g., Garner and Miller, 1947; Plomp and Bouman, 1959; Zwislocki, 1960; Florentine et al., 1988). Moreover, listeners cannot follow envelope fluctuations at very high rates (e.g., Viemeister, 1979) or detect very brief temporal gaps (e.g., Buunen and van Valkenburg, 1979; Fitzgibbons and Wightman, 1982; Shailer and Moore, 1983). These phenomena have been explained using the concept of a temporal integration window. According to the temporal integration hypothesis, the inputs into the auditory system, after peripheral processing, are smoothed by a sliding temporal window. The duration of the integration window represents the sluggishness of the auditory system, with longer window duration corresponding to poorer temporal acuity (e.g., Viemeister, 1979; Forrest and Green, 1987; Moore et al., 1988). However, in many situations, the auditory system seems to be able to carry out more sophisticated sequential operations than just summing acoustic energy within a contiguous temporal neighborhood (Viemeister and Wakefield, 1991). This involves keeping prior stimuli in short-term memory and combining information from the memory with that from the on-going stimulus. These computations could involve both bottom-up and top-down processes (e.g., Bregman, 1990; Näätänen and Winkler, 1999; Alain et al., 2001; Sussman et al., 2002; Cusack et al., 2004; Michey and Oxenham, 2010).

Many studies have investigated high-level auditory sequential processing through the auditory stream-segregation paradigm. For example, two alternating tones of different frequencies were perceived as a single perceptual stream or two segregated streams, depending on factors such as frequency separation, alternation rate, and sequence duration (van Noorden, 1975; Bregman and Capbell, 1971; Bregman, 1978; Moore and Gockel, 2002; Carlyon and Gockel, 2008; Michey and Oxenham, 2010; Richards et al., 2012). Sequential grouping or segregation has also been studied using stimuli other than pure tones, and it has been demonstrated that perceptual grouping across time can be based on fundamental frequency, spatial location, spectral cues, component phase, and level (e.g., Darwin, 1997; Cusack and Roberts, 1999; Vliegen and Oxenham, 1999; Grimault et al., 2002; Roberts et al., 2002; Stainsby et al., 2004; Gaudrain et al., 2007). These stream segregation studies have demonstrated that the auditory system is capable of assessing the consistency of acoustic information over time and grouping the consistent components into the same perceptual stream.

In a realistic auditory scene, the acoustic information associated with individual sound sources may vary over time and exhibit uncertainty. In such situations, the auditory system may have to progressively estimate the properties of each sound source based on acoustic information that is distributed across time, to limit the adverse influences of uncertainty. This process has been revealed using informational-masking paradigms. Informational masking refers to the masking effects that are not related to peripheral interactions between the target and masker, but rather to the similarity between the target and masker or the

uncertainty associated with them (e.g., Watson and Yost, 1987; Leek et al., 1991; Durlach et al., 2003).

Kidd et al. (1994) investigated whether listeners can take advantage of temporal consistency in the target sound to reduce the informational masking effect from a random-frequency masker. In the Multiple-Burst-Same (MBS) condition, the masker consisted of a sequence of 50-ms bursts and each burst contained multiple simultaneous pure tones. The frequencies of these tones were randomly drawn for each trial, and were fixed across bursts on the same trial. The target to be detected was a sequence of 1-kHz tones, gated on and off with the masker bursts. Due to the uncertainty in the frequencies of masker components, this task was very challenging even when the masker frequencies were chosen to minimize interactions between the target and masker at the auditory periphery. On the other hand, in the Multiple-Burst-Different (MBD) condition, the frequency components in the masker were redrawn for each burst. This led to a significant improvement in the detection threshold for the target compared to the MBS condition.

Several processes could contribute to this MBD advantage. First, similar to auditory stream segregation, the temporal consistency of the target may contribute to the formation of an auditory stream that is distinct from the masker bursts (Kidd et al., 1994; Huang and Richards, 2006; Micheyl et al., 2007). Second, the listeners may be able to accumulate statistical evidence with regard to the long-term spectrum of the masker (Kidd et al., 2003). These different processes are relatively difficult to dissociate in informational masking (Kidd et al., 1994; Kidd et al., 2003; Micheyl et al., 2007) or stream segregation (Elhilali et al., 2009; Akram et al., 2014) experiments using the MBS/MBD stimuli, because the two mechanisms (i.e. stream segregation and information accumulation) would lead to similar effects for simultaneously presented target and masker stimuli.

The current study focused on the ability to accumulate statistical evidence over time. Information with regard to the spectrum of a random-frequency, multi-tone complex was conveyed using multiple brief bursts. Each burst in the sequence contained a subset of the frequency components from the complex, determined at random and separately for each burst. To obtain an accurate estimate of the spectrum of the complex, accumulation of information across multiple bursts is required. Efficiency in estimating the spectral content of the complex was estimated indirectly with an experimental paradigm inspired by previous informational masking studies.

It is known that random-frequency, multi-tone complexes give rise to informational masking (e.g., Neff and Green, 1987; Neff and Callaghan, 1988; Oh and Lutfi, 1998). When a preview of the masker is presented as a precursor to the target and masker, the detection threshold often improves significantly. This phenomenon is often referred to as the auditory enhancement effect (Viemeister, 1980). Sensory mechanisms have been invoked as explanations of the enhancement effect. For example, according to the adaptation of suppression/inhibition hypothesis (Viemeister and Bacon, 1982; Nelson and Young, 2010; Byrne et al., 2011; Shen and Richards, 2012), the presence of the precursor gives rise to suppression or inhibition in the spectral region of the target, and the suppression or inhibition effect adapts over time. When the masker and target are later presented, the

response to the target is enhanced since it is subjected to less suppression/inhibition. Recent findings have suggested that sensory mechanisms may be insufficient to fully account for the enhancement effect (e.g., Byrne et al., 2013). Rather, the benefit of the precursor may partially occur because it provides knowledge about the spectral content of the masker (Neff and Green, 1987; Richards and Neff, 2004; Richards et al., 2004; Kidd et al., 2011). Frequency cuing effects have been observed not only in detection tasks, but also in tasks that involve judgements based on pitch (Demany and Ramos, 2005; Erviti et al., 2011; Carcagno et al., 2012) or amplitude modulation (Shen, 2016).

A core assumption of the current study is that if a listener can estimate the spectrum of a multi-tone complex well, then the complex will provide little masking. Thus, higher target detectability indicates higher fidelity of the estimated masker spectrum. Using this approach, the ability to accumulate statistical evidence across precursor bursts was measured as a function of the number of bursts and inter-burst interval (IBI). These stimulus manipulations were intended to alter the way the statistical information was sequentially distributed. When the number of bursts increases, the precursor sequence contains more information regarding the spectrum of the upcoming masker. Therefore, the effectiveness of the masker is expected to decrease with increasing number of bursts.

Besides sequential information processing, manipulations of the total number of bursts and the IBI may change the amount of frequency cuing via sensory mechanisms. For example, increasing the number of bursts increases the total energy of the precursor. The increased total energy of the precursor may then affect target detection through mechanisms related to neural adaptation (Viemeister, 1980). To investigate the effects of stimulus manipulations unrelated to information processing, Exp. I used precursor bursts that were copies of the complex masker in the detection task. Because the precursors were always informative about the masker as the number of bursts and IBI were varied, the effects of these stimulus manipulations reflected mainly sensory processes. In Exp. II, the precursors with reduced informativeness were used, the effects of number of bursts and IBI observed in Exp. II reflected both sensory processes and non-sensory, information-processing mechanisms.

2. Experiment I: cueing effect with deterministic precursor sequences

2.1. Method

2.1.1. Listeners—Nine listeners (six females) were recruited from the student population at Indiana University. The listeners were between 19 and 29 years of age and had audiometric thresholds equal to or better than 15 dB HL between 250 and 8000 Hz in both ears. For each listener, the ear with the lower pure tone average (PTA) threshold (mean of the hearing level at 0.5, 1, and 2 kHz) was tested. If the PTA thresholds were the same for the two ears, the left ear was tested.

2.1.2. Stimuli—Listeners detected the presence of a 1-kHz tonal target in a simultaneous masker in a single-interval, Yes/No task. A schematic of the stimuli is shown in the top panel of Fig. 1. The masker consisted of six equal-amplitude frequency components. The frequencies of the components were randomly drawn on every trial from a uniform distribution between 200 and 5000 Hz on a logarithmic frequency scale, but excluding a

“protection region” between 841 and 1189 Hz (indicated by the shaded area in the top panel of Fig. 1). The protection region was included to limit energetic masking. The masker duration was 150 ms, including 10-ms raised-cosine onset and offset ramps. The masker components had the same level and the component level was randomly drawn for each trial from a uniform distribution between 45 and 55 dB SPL. The roving was intended to limit the use of intensity cues for the detection of the target. The target tone was present with a probability of 50%. When present, the target tone was gated on and off simultaneously with the masker, also using 10-ms raised-cosine ramps.

In all experimental conditions, except for the No-Precursor condition, a precursor sound was presented prior to the masker and target. The precursor consisted of a sequence of bursts, each burst being a tone complex that consisted of the same frequencies as the masker but with a shorter duration of 50 ms. The component level for the precursor bursts was 50 dB SPL. In separate conditions, the IBI was either 0 or 200 ms and the total number of bursts was 1, 2, 4, 8, or 16. The total number of bursts will be referred to as “sequence length” for simplicity. The offset of the precursor occurred 500 ms prior to the onset of the masker. The relatively long precursor-masker interval was chosen to ensure that the facilitation effect of the precursor on target detection was dominated by non-sensory mechanisms (Cao and Richards, 2012; Shen, 2016).

The stimuli were generated digitally and presented at a sampling rate of 22.05 kHz using custom software running in Matlab (The MathWorks, Inc.), which was also used for experimental control and the presentation of visual feedback following each trial. The stimuli were presented monaurally to the test ear through a 24-bit sound card (MOTU Microbook II, Mark of the Unicorn, Inc.) and headphones (HD-280 Pro, Sennheiser Electronic). During the experiment, listeners were seated in a sound-attenuating booth.

2.1.3. Procedure—Before data collection, all listeners practiced on the experimental task for at least four hours. The practice sessions included all experimental conditions. During data collection, the two IBIs were tested in random order, and for each IBI the different sequence lengths were tested in separate experimental blocks and in random order. Note that for the sequence lengths of 0 (no-precursor) and 1, the IBI did not affect the stimuli. These two conditions were labeled under the 0-ms IBI for convenience and they were not tested under the 200-ms IBI. This leads to a total of ten conditions. Once all ten conditions were tested, the above-described process was repeated five times with independent random orders.

For each experimental block, the target level, relative to the masker component level (the relative target level), was varied on a trial by trial basis. The relative target level at the beginning of a block was 10 dB. The relative target level was decreased following three consecutive correct responses (including both hits and correct rejections) and increased following a single incorrect response (including both false alarms and misses). The initial step size for the manipulation of the relative target level was 10 dB, which was reduced to 5 dB after two reversals of the adaptive track. The relative target level was limited to be between -40 and 20 dB. The experimental block terminated after a total of 12 reversals or when the relative target level remained at either the lower or upper limit for three consecutive trials.

A Yes/No task design was adopted in the current experiment instead of a two-alternative, forced-choice design commonly used with adaptive up-down procedures. This was to prevent potential interactions between the stimuli in multiple intervals of the same trial (Richards et al., 2004; Feng and Oxenham, 2015). From a practical perspective, the use of a Yes/No task shortened the total data-collection time compared to a two-alternative, forced-choice design because the stimuli for the current experiment could be fairly long (e.g., 4.65 s for a IBI of 200 ms and a sequence length of 16).

Once data collection was complete, the experimental trials collected with the same IBI and sequence length were pooled together. For these trials, the hit rate was calculated for each target level and the false alarm rate was calculated based on all target-absent trials regardless of what the target level would be on those trials. The hit and false alarm rates were used to estimate the target detectability d' as a function of the relative target level (Cao and Richards, 2012). These data were fitted using a weighted linear regression using the number of target-present trials as the weights. The fitted line was the psychometric function relating the relative target level to d' . The detection threshold was identified as the relative target level that corresponded to a d' of 2 on the psychometric function.

2.2. Results

The top panels of Fig. 2 plot the target detection thresholds as functions of sequence length for the 0-ms (left) and 200-ms (right) IBIs for eight of the nine listeners. As expected from informational masking experiments, large individual differences were observed. From the best to the worst performer, the range of thresholds was as large as 40 dB in certain conditions. The performance for listener S3 was much poorer than for all other listeners and the adaptive tracks stayed near the upper limit for the relative target level. Consequently, the detection thresholds could not be reliably estimated for this listener. Similar difficulties in reliably estimating the threshold occurred for listener S9 for the no-precursor condition. Therefore no results are shown for S3 or for S9 for the no-precursor condition in the top panels of Fig. 2.

For the seven listeners with complete data (see the bottom panels of Fig. 2 for average thresholds across all listeners except for S3 and S9), a paired t -test based on the thresholds for the 0-burst (no-precursor) and 1-burst conditions confirmed the presence of the frequency cueing effect [$t(6) = 2.47, p = .024$, one-tailed]. With the no-precursor condition omitted, a two-way repeated measures ANOVA treating IBI and sequence length as the two independent variables and the threshold as the dependent variable was conducted. The data from the 1-burst condition were reused for the 0-ms and 200-ms IBIs. The effect of sequence length approached significance [$F(4, 24) = 2.67, p = .057$], suggesting a trend for threshold improvement with increasing sequence length. There was no significant effect of IBI [$F(1, 6) = 0.34, p = .583$], and the was not significant [$F(4, 24) = 0.70, p = .598$].

The results from Exp. I replicated results from previous informational masking studies in showing a significant enhancement effect. The target threshold did not vary significantly with IBI, but tended to decrease gradually with increasing sequence length. An additional analysis was carried out to estimate how much, on average, target detectability d' changed as the sequence length increased. For each of the two IBIs, a fixed target level was chosen as

the threshold for a sequence length of 16. At this fixed target level, the expected d' was estimated from the fitted psychometric function for each sequence length. Figure 3 plots d' as a function of sequence length for the 0-ms (left panel) and 200-ms (right panel) IBIs. As expected from the choice of the fixed target levels, d' estimates were 2 when the sequence length was 16 for both IBIs and for all listeners (see the individual results in the top panels). As the sequence length decreased from 16, d' decreased gradually for some listeners. When averaged across all listeners but S3 (see the average results in the bottom panels), d' decreased by 0.060 and 0.058 when the sequence length was reduced by half for the 0- and 200-ms IBIs, respectively. For convenience, these two values will be referred to as r_0 and r_{200} (i.e. $r_0 = 0.060$, $r_{200} = 0.058$) in the following discussions.

Since the informativeness of the precursor was independent of sequence length, the effect of sequence length observed in the current experiment was mainly due to sensory mechanisms. For example, increasing the total energy of the precursor may have caused the precursor's suppression/inhibition effect near the target frequency to adapt more, leading to a more pronounced auditory enhancement effect (Viemeister and Bacon, 1982). The potential effect of the total energy of the precursor will be considered when discussing the results of Exp. II.

3. Experiment II: cueing effect with probabilistic precursor sequences

3.1. Method

Eight of the nine listeners (all but S7) from Exp. I participated in Exp. II. The stimuli used in Exp. II were identical to those for Exp. I, unless otherwise described. For each listener and for each combination of IBI and sequence length, the relative target level was fixed at the threshold identified in Exp. I for the corresponding condition. Since the masker component level was roved between 45 and 55 dB SPL from trial to trial, the target level was varied accordingly. For listener S3, from whom the detection thresholds were not estimated in Exp. I, the relative target level was set to 20 dB. In all conditions of Exp. II, the precursor sequence was present. The sequence lengths were 1, 2, 4, 8, and 16 and the IBIs were 0 and 200 ms in separate conditions. In contrast to Exp. I, each burst in the precursor sequence consisted of a subset of the masker components (see the middle and bottom panels of Fig. 1 for schematics of the stimuli in Exp. II). Specifically, for each burst in the sequence, whether or not a masker frequency component was presented in the burst was determined at random according to a presentation probability p . The bursts in a given precursor sequence were constructed from independent draws of masker components. When the value of p was 1, the precursor bursts were previews of the masker as in Exp. I, and d' was expected to be close to 2 for all sequence lengths and IBIs. When the value of p was less than 1, each precursor burst carried only partial information regarding the masker spectrum, hence the informativeness of the precursor was reduced. This was expected to lead to poorer performance and d' estimates less than 2.

During the experiment, for each listener the 0- and 200-ms IBIs were tested in random order. For each IBI, the various sequence lengths were tested in random order. Once one experimental block was completed for each combination of the sequence length and IBI, the above process was repeated five more times with independent random sequences. The 1-burst condition was not tested when the IBI was 200 ms, since this condition was the same

as the 1-burst condition when the IBI was 0 ms. For each experimental block, target detection was measured for p values of 0.2, 0.4, 0.6, 0.8, and 1, using the method of constant stimuli with each p value repeated 10 times.

3.2. Results

The top panels of Fig. 4 plots the average d' values for the detection of the 1-kHz target tone across listeners as a function of sequence length for the 0-ms (left panel) and 200-ms (right panel) IBIs. For both IBIs, d' increased as the p value increased. The value of d' ranged from about 1, when the p value was 0.2, to above 2 when the p value was 1. The value of d' also increased as the sequence length increased from 1 to 16. A repeated measures ANOVA was conducted, treating IBI, sequence length and p as the three independent variables and d' as the dependent variable. The data from the 1-burst condition were reused for the 0-ms and 200-ms IBIs. There were significant main effects of p [$F(1.74, 12.17) = 24.31, p < .001$, Greenhouse-Geisser corrected] and sequence length [$F(4, 28) = 5.96, p = .001$], while the effect of IBI was not significant [$F(1, 7) = 1.59, p = .248$]. None of the interactions was significant ($p > .05$). *Post hoc* pairwise comparisons (Bonferroni corrected) suggested that d' was significantly higher for 4-burst sequences than 1-burst sequences. Other pairs of sequences lengths did not lead to significant differences in d' .

As demonstrated in Exp. I, target detectability d' depended on sequence length even when the precursor bursts were highly informative about the masker. This means that the total duration or total energy of the precursor sequence may affect the amount of frequency cueing effect due to sensory mechanisms. In the current experiment, the potential effect of sequence length due to sensory processing was controlled for, because the target level used in Exp. II was chosen to reach an expected d' of 2 when $p = 1$ for all listeners and across all IBIs and sequence lengths. That is, the potential sensory effect associated with sequence length was counterbalanced by the choices of the target level.

On the other hand, the potential sensory effect associated with p may confound the interpretation of the current results, because lower values of p were not only associated with degraded informativeness but also decreased total precursor energy. Since the expected total number of tones was proportional to both sequence length and p , the average effect of sequence length observed from Exp. I (i.e. d' increase by r_0 and r_{200} for each doubling of sequence length for the 0- and 200-ms IBIs, respectively) may be used to correct for the potential sensory effect associated with p . Specifically, compared to $p = 1$, d' was expected to be decreased by $r_0 \log_2(1/p)$ for the 0-ms IBI and $r_{200} \log_2(1/p)$ for the 200-ms IBI. This was used to correct the estimated d' (see the corrected d' data in the lower panels of Fig. 4), and the ANOVA was repeated on the corrected data. There were significant main effects of p [$F(1.74, 12.17) = 18.83, p < .001$, Greenhouse-Geisser corrected] and sequence length [$F(4, 28) = 6.02, p = .001$], while the effect of IBI was not significant [$F(1, 7) = 0.46, p = .236$]. None of the interactions was significant ($p > .05$). *Post hoc* pairwise comparisons (Bonferroni corrected) suggested that d' was significantly higher for 4-burst sequences than 1-burst sequences. Other pairs of sequences lengths did not lead to significant differences in d' .

The significant effect of sequence length indicates that information degradation introduced by lowering p could be counterbalanced by increasing the sequence length. This reflected an information accumulation process across at least four precursor bursts. The lack of an interaction between sequence length and IBI suggests that the span for the information accumulation process was not strongly dependent on absolute time but the number of precursor bursts.

4. Discussion

4.1. Rate of Information Accumulation

The current study demonstrated that listeners are capable of estimating the spectral content of an acoustic stimulus even when the spectral information is presented in an unreliable fashion. With longer sequence lengths, a pronounced enhancement effect and high target detectability can be maintained even for low p values.

To investigate the rate of information accumulation, the p values at performance threshold (p_{thre}) were estimated for each individual listener by fitting a linear regression model to the d' data from Exp. II and identifying the p values associated with a criterion d' of 1.8. This criterion value (i.e. d' of 1.8) was chosen to minimize the cases in which the estimated p_{thre} values were lower than 0 or greater than 1. The linear model can be written as:

$$d' = a \times p + b_{\text{IBI, sequence length}} + \varepsilon, \quad (1)$$

where the intercept term $b_{\text{IBI, sequence length}}$ was assumed to be different for different combinations of IBI and sequence length, while the coefficient a was assumed to be consistent across conditions. These assumptions were based on the significant main effect of p and the lack of significant interactions involving p from the ANOVAs described previously. The estimation of p_{thre} was repeated for uncorrected, raw d' data (plotted in the top panels of Fig. 4) and the d' data after correcting for the potential sensory effect (plotted in the lower panels of Fig. 4).

Figure 5 plots the average p_{thre} estimates across listeners for each combination of sequence length and IBI and separately for the uncorrected (top panels) and corrected (bottom panels) d' estimates. If the listeners based their estimates of the masker spectrum on only the last burst, then p_{thre} would be independent of sequence length and it would equal that for the 1-burst condition. Instead, the average p_{thre} tended to decrease gradually as the sequence length increased for both IBIs and both uncorrected and corrected data.

To assist the interpretation of the benefit from multiple bursts, two theoretical boundaries were established according to the statistical characteristics of the stimuli. First, a longer burst sequence is associated with a higher probability that at least one burst in the sequence is a preview of the masker. That is,

$$P(\text{preview available}) = 1 - (1 - p^6)^{\text{sequence length}}. \quad (2)$$

If a listener's performance is associated with this preview probability, then this listener does not need to integrate or accumulate information across bursts, and task performance could be a reflection of efficiency in retrieving the preview from memory. The relationship between p and sequence length specified in Eq. 2 provides a boundary, namely the Preview boundary. If a listener performs better than the Preview boundary (e.g. a lower p_{thre} for the same sequence length), then it is theoretically impossible to achieve that performance without sequentially integrating information across bursts. Moreover, a real listener who does not integrate information across bursts would probably perform much worse than predicted by the Preview boundary because the listener does not know which of the bursts was a preview and the burst stored in memory are subject to decay over time. Therefore, worse performance than the Preview boundary does not exclude the possibility of sequential integration.

Second, a longer burst sequence is associated with a higher probability that all six components of the masker are contained in the precursor sequence with each frequency component appearing at least one time anywhere in the sequence. That is,

$$P(\text{no missing frequency}) = [1 - (1 - p)^{\text{sequence length}}]^6. \quad (3)$$

If a listener's performance is associated with this probability of not missing any frequencies, then this listener accumulates information in an optimal fashion. The relationship between p and sequence length specified in Eq. 3 sets another theoretical boundary, namely the Integration boundary. Better performance than the Integration boundary is theoretically impossible. Figure 5 shows Preview and Integration boundaries, together with the thresholds derived from the experiment, as dashed curves with two different dash styles. These two curves were obtained by setting $P(\text{preview available})$ in Eq. 2 and $P(\text{no missing frequency})$ in Eq. 3 so that the curves passed through the average p_{thre} for the 1-burst condition.

For the 0-ms IBI, the average p_{thre} decreased as the sequence length increased at a faster rate than the Preview boundary. After correcting for the potential sensory effect of p , the average p_{thre} followed closely to the Integration boundary. Therefore, it is unlikely that the current group of listeners in Exp. II relied only on the previews of the masker in the precursor sequence and they were likely to have combined information across multiple bursts. For the 200-ms IBI, the average p_{thre} estimated using the uncorrected data followed the Preview boundary closely, while the average p_{thre} estimated using the corrected data fell between the Preview and Integration boundaries. As explained previously, a listener who does not accumulate information across bursts would likely perform worse than the Preview boundary due to limitations on short-term memory and *a priori* knowledge. The fact that the average experimental data for the 200-ms IBI were, for the most part, close to the Preview boundary

provides support for the information accumulation process in this condition among the listeners of Exp. II. Although large inter-subject variability was observed for p_{thre} (see the standard errors of the means, indicated in Fig. 5 by the error bars), the likelihood of the average p_{thre} being below the Preview boundary was greater than being above the Preview boundary for most sequence lengths above 4 and for both IBIs.

4.2. Comparison to Previous Studies

The results are similar to those of studies that used MBD/MBS stimuli. Kidd *et al.* (2003) systematically manipulated the sequence length and IBI in an informational masking study. Their listeners were instructed to detect a target tone sequence at 1 kHz while the masker was a burst sequence with independent random frequency components (the MBD masker). As the sequence length increased, the detection threshold improved and the threshold improvements were more evident for shorter IBIs. Kidd *et al.* (2003) pointed out that the effect of sequence length is not expected to depend on IBI if performance is based on lossless information integration across bursts. The observed interaction in the study of Kidd *et al.* (2003) between sequence length and IBI may result from at least two factors. First, there may be information loss associated with longer IBIs. That is, information accumulation is subject to the decay of a short-term/working memory store. Second, the processes underlying target detection may be related to the formation of segregated target and masker streams (Micheyl *et al.*, 2007). For example, shorter pauses between tones in a tone sequence with a galloping frequency pattern (e.g., HLH-HLH-HLH-... where H and L indicate high and low frequencies, respectively) promote hearing the sequence as two segregated auditory streams (e.g., Bregman *et al.*, 2000).

An explanation based on perceptual segregation is not applicable to the current results because the current experimental paradigm did not involve perceptual segregation during the presentation of the precursor sequence. Therefore, the lack of an interaction between sequence length and IBI in the current study suggests that the information loss associated with the IBI may be quite limited for IBIs up to 200 ms.

Carcagno *et al.* (2013) conducted a study using similar stimuli to those of the current study. In their study, the listeners detected a target tone in a six-tone complex masker. Three of the masker components had frequencies above that of the target tone, forming an upper masker band, while the other three components had frequencies below that of the target tone, forming a lower masker band. The frequency spacing between the three tones within each band was 100 cents and the spacing between the target and the nearest component of each band was 350 cents. A precursor burst sequence preceded the target and masker on every trial by at least 200 ms. There were five bursts in the sequence, and each burst in the precursor sequence consisted of six tones at the masker frequencies. Similar to the current study, Carcagno *et al.* (2013) included stimulus manipulations intended to degrade the information about the masker spectrum carried in the precursor bursts. In the SYNCH condition, the six tones in each burst were gated on and off simultaneously, so the bursts were previews of the masker. In the ASYNCH condition, the six tones in each burst were gated on asynchronously, making the bursts less similar to the masker. Asynchronies between components also discouraged their perceptual fusion during periods of temporal

overlap. If the target detection threshold reflects efficiency in estimating a spectral template of the masker, then detection performance should be different for the SYNCH and ASYNCH conditions. However, similar performance thresholds were obtained in the SYNCH and ASYNCH conditions. The authors argued that the listener may not have used the precursor bursts to estimate the spectrum of the masker, since the similarity between the precursor bursts and the masker did not significantly affect performance.

In contrast to the findings of Carcagno et al. (2013), when the reliability of the precursor bursts as previews of the masker was systematically degraded in the current study (by decreasing the value of p), the target detection threshold increased. The discrepancy between the two studies could be due to several differences in the stimuli and procedures. First, although the temporal asynchrony among the frequency components did introduce differences between the precursor bursts and the masker, this effect may have been modest compared to eliminating certain masker components from the precursor bursts. Second, the target and masker frequencies were transposed to different spectral regions from trial to trial in the study of Carcagno et al. (2013), while the logarithmic-frequency distances from the six masker components to the target were fixed. In the current study, on the other hand, the six masker frequencies were randomly drawn for every trial. For example, it was possible for the number of masker components above and below the target frequency to be unequal. This means that the information required to estimate the masker spectrum from the precursor sequence may have differed between the two studies. Lastly, the total energy of the precursor was not affected by the synchronicity manipulations in the Carcagno et al. (2013) study, while in this experiment, decreasing the value of p did decrease the total energy. This could increase the influences of sensory processes such as adaptation of suppression/inhibition, partially accounting for the observed effect of sequence length.

4.3. Potential Involvement of Top-Down Processes

Like other high-level auditory processes, such as auditory streaming, it is possible that both bottom-up and top-down processes are involved in the sequential accumulation of spectral information. Evidence for the potential involvement of top-down processes can be obtained by comparing the results of Exps. I and II.

In Exp. II, trials during which the precursor bursts were previews of the masker ($p = 1$), were identical to those for Exp. I. In addition, except for one listener (S3), the relative target level in Exp. II was set at the performance threshold ($d' = 2$) obtained from Exp. I. If performance were solely governed by stimulus-driven factors, d' estimates should have been close to 2 when p was 1 for all sequence lengths and IBIs in Exp. II. However, d' estimates from these conditions (upward-pointing triangles in Fig. 3) often deviated from 2 and depended on sequence length. A repeated measures ANOVA was conducted treating IBI and sequence length as the independent variables and d' as the dependent variable. Data from listener S3 were not included in this analysis and the data for the 1-burst condition were reused for IBIs of 0 and 200 ms. There was a significant effect of sequence length [$F(4, 24) = 4.07, p = .012$]. That effect of IBI was not significant [$F(1, 6) = 0.24, p = .639$], nor was the interaction between IBI and sequence length [$F(4, 24) = 1.03, p = .413$]. Post hoc analysis (Bonferroni corrected) showed that the expected d' value of 2 was lower than the

95-% confidence intervals of average d' estimates for sequence lengths of 8 (2.03 ~ 2.75) and 16 (2.22 ~ 3.19). For all other combinations of sequence length and IBI, the 95-% confidence interval enclosed the expected d' value.

Since all listeners included in Exp. II completed Exp. I first, one may speculate that the observed differences between the two experiments mainly reflected either a practice effect or fatigue of the listeners. However, such time-dependent effects were not expected to affect various sequence lengths differently. It is possible that the listeners adopted different strategies for these two experiments. For instance, these listeners, through practice, may have learned to more effectively integrate information over time, leading to improvement in performance for sequence lengths of 8 and 16. Alternatively, the strategic shift may have been due to how the experimental conditions were blocked. In Exp. II, the trials with $p = 1$ were mixed with trials with $p < 1$ within each block. The listeners may have expected reduced informativeness in the precursor even when $p = 1$. As a result, they may have adopted a listening strategy that promoted information integration in Exp. II. Regardless of the cause for the strategic change, the information integration process is unlikely to be invariant and solely stimulus-driven, and future investigations on the potential involvement of top-down processes are warranted.

5. Summary

This study investigated whether the auditory system is capable of accumulating spectral information over time. The spectrum of a complex masker was cued using a sequence of bursts where each burst contained only partial information about the masker spectrum. The fidelity of the estimated masker spectrum was revealed by the detectability of a tonal target in the masker. Detection performance improved as the number of bursts increased, indicating information integration across bursts. On the other hand, no significant effect of inter-burst interval was observed, suggesting that the sequential integration of information may not be based on absolute time.

Acknowledgments

This work was supported by the NIH Grants No. R21 DC013406 (MPIs: Virginia M. Richards and Yi Shen). Elizabeth A. Pugh, Carolyn J. Herbert, and Allison B. Kern provided assistance with data collection. Audrey B. Hiner, Allison B. Kern, and Dylan V. Pearson provided editorial support. The authors would also like to thank the associate editor, Brian C. J. Moore, and two anonymous reviewers for helpful suggestions that substantially improved the manuscript.

References

- Akram S, Englitz B, Elhilali M, Simon JZ, Shamma SA. Investigating the neural correlates of a streaming percept in an informational-masking paradigm. *PloS one*. 2014; 9(12):e114427. [PubMed: 25490720]
- Alain C, Arnott SR, Picton TW. Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*. 2001; 27(5):1072–1089. [PubMed: 11642696]
- Bregman AS, Campbell J. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*. 1971; 89(2):244–249. [PubMed: 5567132]
- Bregman AS. Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*. 1978; 4(3):380–387. [PubMed: 681887]

- Bregman, AS. Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT press; 1990.
- Bregman AS, Ahad PA, Crum PA, O'Reilly J. Effects of time intervals and tone durations on auditory stream segregation. *Perception & psychophysics*. 2000; 62(3):626–636. [PubMed: 10909253]
- Buunen TJF, Van Valkenburg DA. Auditory detection of a single gap in noise. *The Journal of the Acoustical Society of America*. 1979; 65(2):534–537. [PubMed: 489820]
- Byrne AJ, Stellmack MA, Viemeister NF. The enhancement effect: evidence for adaptation of inhibition using a binaural centering task. *The Journal of the Acoustical Society of America*. 2011; 129(4):2088–2094. [PubMed: 21476664]
- Byrne AJ, Stellmack MA, Viemeister NF. The salience of enhanced components within inharmonic complexes. *The Journal of the Acoustical Society of America*. 2013; 134(4):2631–2634. [PubMed: 24116400]
- Cao X, Richards VM. Enhancement in informational masking. *Journal of Speech, Language, and Hearing Research*. 2012; 55(4):1135–1147.
- Carcagno S, Semal C, Demany L. Auditory enhancement of increments in spectral amplitude stems from more than one source. *Journal of the Association for Research in Otolaryngology*. 2012; 13(5):693–702. [PubMed: 22766695]
- Carcagno S, Semal C, Demany L. No need for templates in the auditory enhancement effect. *PloS one*. 2013; 8(6):e67874. [PubMed: 23826348]
- Carlyon, RR., Gockel, HE. Effects of Harmonicity and Regularity on the Perception of Sound Sources. In: Yost, WA, Popper, AN., Fay, RR., editors. *Auditory perception of sound sources*. Boston, MA: Springer; 2008. p. 191–213.
- Cusack R, Decks J, Aikman G, Carlyon RP. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of experimental psychology: human perception and performance*. 2004; 30(4):643–656. [PubMed: 15301615]
- Cusack R, Roberts B. Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception*. 1999; 28(10):1281–1289. [PubMed: 10694975]
- Darwin CJ. Auditory grouping. *Trends in cognitive sciences*. 1997; 1(9):327–333. [PubMed: 21223942]
- Demany L, Ramos C. On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *The Journal of the Acoustical Society of America*. 2005; 117(2):833–841. [PubMed: 15759703]
- Durlach NI, Mason CR, Kidd G Jr, Arbogast TL, Colburn HS, Shinn-Cunningham BG. Note on informational masking (L). *The Journal of the Acoustical Society of America*. 2003; 113(6):2984–2987. [PubMed: 12822768]
- Elhilali M, Xiang J, Shamma SA, Simon JZ. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol*. 2009; 7(6):e1000129. [PubMed: 19529760]
- Erviti M, Semal C, Demany L. Enhancing a tone by shifting its frequency or intensity. *The Journal of the Acoustical Society of America*. 2011; 129(6):3837–3845. [PubMed: 21682407]
- Feng L, Oxenham AJ. New perspectives on the measurement and time course of auditory enhancement. *Journal of Experimental Psychology: Human Perception and Performance*. 2015; 41(6):1696–1708. [PubMed: 26280269]
- Fitzgibbons PJ, Wightman FL. Gap detection in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*. 1982; 72(3):761–765. [PubMed: 7130534]
- Forrest TG, Green DM. Detection of partially filled gaps in noise and the temporal modulation transfer function. *The Journal of the Acoustical Society of America*. 1987; 82(6):1933–1943. [PubMed: 3429731]
- Florentine M, Fastl H, Buus SR. Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *The Journal of the Acoustical Society of America*. 1988; 84(1):195–203. [PubMed: 3411048]
- Garner WR, Miller GA. The masked threshold of pure tones as a function of duration. *Journal of Experimental Psychology*. 1947; 37(4):293. [PubMed: 20258425]
- Gaudrain E, Grimault N, Healy EW, Béra JC. Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing research*. 2007; 231(1):32–41. [PubMed: 17597319]

- Grimault N, Bacon SP, Micheyl C. Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*. 2002; 111(3):1340–1348. [PubMed: 11931311]
- Huang R, Richards VM. Coherence detection: Effects of frequency, frequency uncertainty, and onset/offset delays. *The Journal of the Acoustical Society of America*. 2006; 119(4):2298–2304. [PubMed: 16642843]
- Kidd G Jr, Mason CR, Deliwala PS, Woods WS, Colburn HS. Reducing informational masking by sound segregation. *The Journal of the Acoustical Society of America*. 1994; 95(6):3475–3480. [PubMed: 8046139]
- Kidd G Jr, Mason CR, Richards VM. Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *The Journal of the Acoustical Society of America*. 2003; 114(5):2835–2845. [PubMed: 14650018]
- Kidd G Jr, Richards VM, Streeter T, Mason CR, Huang R. Contextual effects in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*. 2011; 130(6):3926–3938. [PubMed: 22225048]
- Leek MR, Brown ME, Dorman MF. Informational masking and auditory attention. *Attention, Perception, & Psychophysics*. 1991; 50(3):205–214. [PubMed: 1754361]
- Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B, Wilson EC. The role of auditory cortex in the formation of auditory streams. *Hearing research*. 2007; 229(1):116–131. [PubMed: 17307315]
- Micheyl C, Oxenham AJ. Objective and subjective psychophysical measures of auditory stream integration and segregation. *Journal of the Association for Research in Otolaryngology*. 2010; 11(4):709–724. [PubMed: 20658165]
- Miller GA, Licklider JC. The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*. 1950; 22(2):167–173.
- Moore BCJ, Glasberg BR, Plack CJ, Biswas AK. The shape of the ear's temporal window. *The Journal of the Acoustical Society of America*. 1988; 83(3):1102–1116. [PubMed: 3356815]
- Moore BCJ. Temporal integration and context effects in hearing. *Journal of Phonetics*. 2003; 31(3):563–574.
- Moore BCJ, Gockel H. Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*. 2002; 88(3):320–333.
- Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychological bulletin*. 1999; 125(6):826. [PubMed: 10589304]
- Neff DL, Callaghan BP. Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *The Journal of the Acoustical Society of America*. 1988; 83(5):1833–1838. [PubMed: 3403798]
- Neff DL, Green DM. Masking produced by spectral uncertainty with multicomponent maskers. *Attention, Perception, & Psychophysics*. 1987; 41(5):409–415.
- Nelson, PC., Young, ED. Enhancement in the marmoset inferior colliculus: neural correlates of perceptual “Pop-Out”. In: Lopez-Poveda, E. Palmer, A., Meddis, R., editors. *The Neurophysiological Bases of Auditory Perception*. Springer; New York, NY: 2010. p. 155-165.
- Oh EL, Lutfi RA. Nonmonotonicity of informational masking. *The Journal of the Acoustical Society of America*. 1998; 104(6):3489–3499. [PubMed: 9857508]
- Plomp R, Bouman MA. Relation between hearing threshold and duration for tone pulses. *The Journal of the Acoustical Society of America*. 1959; 31(6):749–758.
- Richards VM, Huang R, Kidd G Jr. Masker-first advantage for cues in informational masking. *The Journal of the Acoustical Society of America*. 2004; 116(4):2278–2288. [PubMed: 15532659]
- Richards VM, Carreira EM, Shen Y. Toward an objective measure for a “stream segregation” task. *The Journal of the Acoustical Society of America*. 2012; 131(1):EL8–EL13. [PubMed: 22280734]
- Richards VM, Neff DL. Cuing effects for informational masking. *The Journal of the Acoustical Society of America*. 2004; 115(1):289–300. [PubMed: 14759022]
- Roberts B, Glasberg BR, Moore BCJ. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *The Journal of the acoustical society of America*. 2002; 112(5):2074–2085. [PubMed: 12430819]

- Shailer MJ, Moore BCJ. Effects of modulation rate and rate of envelope change on modulation discrimination interference. *The Journal of the Acoustical Society of America*. 1993; 94(6):3138–3143.
- Shen Y, Richards VM. Investigating the auditory enhancement phenomenon using behavioral temporal masking patterns. *The Journal of the Acoustical Society of America*. 2012; 132(5):3363–3374. [PubMed: 23145617]
- Shen Y. The effect of frequency cueing on the perceptual segregation of simultaneous tones: Bottom-up and top-down contributions. *The Journal of the Acoustical Society of America*. 2016; 140(5):3496–3503. [PubMed: 27908095]
- Stainsby TH, Moore BCJ, Medland PJ, Glasberg BR. Sequential streaming and effective level differences due to phase-spectrum manipulations. *The Journal of the Acoustical Society of America*. 2004; 115(4):1665–1673. [PubMed: 15101645]
- Sussman E, Winkler I, Huotilainen M, Ritter W, Näätänen R. Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*. 2002; 13(3):393–405. [PubMed: 11919003]
- Van Noorden, LPAS. PhD Dissertation. Institute for Perception Research, Eindhoven; The Netherlands: 1975. Temporal coherence in the perception of tone sequences.
- Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*. 1979; 66(5):1364–1380. [PubMed: 500975]
- Viemeister, NF. Adaptation of masking. In: van den Brink, G., Bilsen, FA., editors. *Psychophysical, Physiological and Behavioural Studies in Hearing*. Delft, The Netherlands: Delft University Press; 1980. p. 190-199.
- Viemeister NF, Bacon SP. Forward masking by enhanced components in harmonic complexes. *The Journal of the Acoustical Society of America*. 1982; 71(6):1502–1507. [PubMed: 7108024]
- Viemeister NF, Wakefield GH. Temporal integration and multiple looks. *The Journal of the Acoustical Society of America*. 1991; 90(2):858–865. [PubMed: 1939890]
- Vliegen J, Oxenham AJ. Sequential stream segregation in the absence of spectral cues. *The Journal of the Acoustical Society of America*. 1999; 105(1):339–346. [PubMed: 9921660]
- Watson CS, Yost WA. Uncertainty, informational masking, and the capacity of immediate auditory memory. *Auditory processing of complex sounds*. 1987:267–277.
- Zwislocki J. Theory of temporal auditory summation. *The Journal of the Acoustical Society of America*. 1960; 32(8):1046–1060.

Highlights

Listeners' efficiency in integrating spectral information over time was investigated using sequences of complex bursts.

Evidence for sequential accumulation of spectral information was observed.

In an auditory scene with a great amount of spectral uncertainty, listeners may adopt top-down strategies to enhance information accumulation.

Figure 1.
Schematics of the stimuli used in Exp. I (top panel) and Exp. II (middle and bottom panel).

Figure 2.

Individual (top) and mean (bottom) target detection thresholds for Exp. I as functions of sequence length for IBIs of 0 (left) and 200 (right) ms. The left-most data points within each panel indicate thresholds obtained without the precursor burst sequence (NP). The thresholds are expressed relative to the masker component level. Error bars in the bottom panels indicate $\pm$ one standard error of the mean. Data from one listener (S3) are not plotted because thresholds could not be reliably estimated for this listener. Data from another listener (S9) were not included when computing the average threshold due to the unreliable threshold estimate in the no-precursor condition for this listener. Thresholds from individual listeners shown in the top panels are offset slightly along the horizontal axis to improve visibility.

Figure 3. Individual (top) and mean (bottom) target detectability d' for Exp. I as functions of sequence length for IBIs of 0 (left) and 200 (right) ms. For each listener and each IBI, d' estimates are plotted for a fixed target level, which is the threshold for the 16-burst sequence for the corresponding listener. Error bars in the bottom panels indicate $\pm$ one standard error of the mean. Data from one listener (S3) are not plotted. Thresholds from individual listeners shown in the top panels are offset slightly along the horizontal axis to improve visibility.

Figure 4.

Average detectability d' as a function of sequence length for presentation probabilities p of 0.2, 0.4, 0.6, 0.8, and 1 (different symbols) and for IBIs of 0 (left) and 200 (right) ms from Exp. II. The raw, uncorrected d' estimates are shown in the top panels, while the estimates corrected for the potential sensory effect associated with p are plotted in the bottom panels. Error bars indicate $\pm$ one standard error of the mean.

Figure 5.

The p values at detection threshold (p_{thre}) as a function of sequence length for IBIs of 0 (left) and 200 (right) ms. The results based on the uncorrected d' estimates are shown in the top panels, while the estimates based on the corrected d' data are plotted in the bottom panels. The short and long dashed curves in each panel indicate the Preview and Integration boundaries, respectively (see text for details).