Spectral and temporal integration of brief tones

Evelyn M. Hoglund and Lawrence L. Feth

Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

(Received 22 February 2008; revised 17 September 2008; accepted 19 September 2008)

Both spectral and temporal integration of tones have been explored in detail, but integration of tones varying across both dimensions has received little attention. This study explores temporal integration of tone pulses that vary over a range of frequencies. Baseline thresholds were obtained for both spectral and temporal integration with the same signals and compared with prior research. The signals were then varied on both dimensions in several ways: with equivalent spectral and temporal step sizes, different spectral and temporal step sizes, and a random pattern of frequency presentation. The data were also analyzed by spectral step size, temporal step size, frequency range, direction and slope of frequency change, and predictability. The spectral and temporal integration conditions showed that the current procedures and signals yielded the same improvement in detection thresholds as prior studies. The spectrotemporal integration conditions showed the improvement for overall detection of the signals to be limited by spectral integration, with improvement related primarily to the number of tones, regardless of timing and frequency. Surprisingly, trial-by-trial random presentation of signal frequencies did not negatively influence detection. These results support the multiple looks hypothesis [Viemeister, N. F. and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865] as applied to spectrotemporal integration. © 2009 Acoustical Society of America. [DOI: 10.1121/1.2999338]

PACS number(s): 43.66.Ba, 43.66.Cb [RLF]

Pages: 261–269

I. INTRODUCTION

The attempt to understand the nature of human auditory perception has provided a wide variety of directions for study. Important aspects of sound that have been of great interest include the temporal processing of sounds, and processing within the frequency, or spectral, domain. For both of these areas of auditory perception, there is the dilemma related to specificity versus generality. In other words, how do we achieve both fine acuity (resolution), of the "small picture," and also wide integration (summation) of the "big picture?"

The dilemma occurs because fine temporal acuity requires very brief time windows, and good frequency selectivity demands narrow auditory filters. Integration over a wide frequency range, or long signal duration, on the other hand, dictates that the processing window should be extended. Most of the research has been conducted on either temporal or spectral processing while keeping the other dimension constant. However, by holding one dimension constant, the results obtained may not reflect perception outside the laboratory since time and frequency are not independent of one another.

Early work on temporal or spectral summation of tones applied the energy detector model (Green, 1958). If the auditory system is simply summing up signal energy across N tones, the detectability of two or more tones set to be equally detectable with $d_i' = (2E/N_o)^{1/2}$ (where energy in the signal tone is E , and N_o is the spectrum level of the background noise) should improve with increasing number of tones according to $d_N' = [\sum_i^N (d_i')^2]^{1/2}$. For a fixed percentage of correct responses in a forced-choice experiment, this amounts to a detection threshold improvement of $-10 \log(N)$, or 3 dB of improvement for each doubling of N . Human listeners do not

achieve this "ideal" level performance in either temporal or spectral summation tasks, but improvements characterized by $-k \log(N)$, with $k < 10$, are common (van den Brink and Houtgast, 1990a, 1990b; Hicks and Buus, 2000).

A. Temporal processing

In their work on temporal integration, Viemeister and Wakefield (1991) introduced a "multiple looks" model for temporal integration that offers a good explanation for both temporal summation and resolution. Prior to this model, integration was studied in terms of either a long integration "window" (hundreds of milliseconds), which provided a good explanation for the summation, or a short integration window (3–5 ms), which provided a good explanation for the resolution. Unfortunately, neither type of model could offer a satisfactory explanation for the opposite extreme: long windows could only explain resolution via "leaky" integration (allowing some integration while summing over the window), and short windows offered poor prediction of summation.

The multiple looks model proposes that the auditory system actually uses short time constant windows or "looks" at the acoustic input. Thus, one look could detect a short duration signal, consistent with thresholds measured on resolution tasks. As the duration of the signal increases, the auditory system uses an increasing number of these looks consecutively, and the information from these windows is accumulated for detection, accounting for improved thresholds for longer duration signals. By measuring detection thresholds for one versus two pulse signals, Viemeister and Wakefield (1991) were able to show that listeners are able to utilize "intelligent" sampling to detect sounds. In quiet, detection thresholds for 200 μ s pulse pairs separated by 1 ms

exhibited a 4 dB improvement relative to single pulses. The threshold increased until the separation was 5 ms, with no further change for longer separations. This suggests integration in a single window at 1 ms, partial integration up to 5 ms, then independent processing with longer separations. They then introduced a noise into the middle 50 ms of a 100 ms separation. When this intervening noise was introduced, the detection thresholds for two pulses averaged 2.5 dB lower than for single pulses, regardless of the level of the noise. The consistent improvement in threshold supports the use of intelligent processing of sounds, rather than summation of the entire duration of the signal, as would be assumed for a long integration window.

Further research by Buus (1999) considered the weighting of the pulses based on their temporal location in the signal, which revealed that detection of the pulses is approximately independent of the other pulses in the train, regardless of the noise masker used. Additionally, Buus (1999) found that the improvement in detection grows less quickly for pulse trains masked by single-band (50 Hz wide) and incoherent maskers (masking band with six flanking bands with unrelated envelopes) than for coherent maskers (masking band with six flanking bands with identical envelopes). His results support the predictions of the multiple looks theory for the single-band and incoherent masker conditions. The greater improvement in detection observed with the coherent masker condition, however, is inconsistent with the theory.

In their work with multiple looks in informational maskers, Kidd *et al.* (2003) suggested the influence of auditory streaming on thresholds for increasing number of tones. They argued that the consistency in the signal allowed streaming to facilitate detection.

B. Spectral processing

Early work involving spectral processing (Fletcher, 1940; French and Steinberg 1947; Schafer and Gales, 1949; Zwicker *et al.*, 1957; Greenwood, 1961; among others) established the critical band as the limit or look for spectral integration. In 1979, Spiegel (1979) studied the critical band and spectral integration to determine both the maximum limit of integration and the critical bandwidth in the same listeners. He found that thresholds for his noise signal increased at a rate consistent with the energy detector model, which predicts that for signals beyond the critical band, several bands will be combined to process the signal. The threshold increases, rather than decreases, as a result of increased noise through the bands along with the signal, at a rate proportional to $10 \log(N)$. Based on his use of a masker noise with a bandwidth of 100–3000 Hz, Spiegel (1979) concluded that spectral integration could potentially occur through the entire range of audibility. Additionally, for detection of single tones, Green (1961) found that when the frequency was uncertain, detectability only decreased by 3 dB in the most extreme cases (frequency varying between 500 and 4000 Hz). Thus, detection was only minimally affected by the inability to predict at what frequency the tone would occur.

van den Brink and Houtgast (1990a, 1990b) reported a series of studies in which they measured spectral integration and temporal integration for long signals (100 ms) and brief signals (4.7 cycles). They found that brief tones were integrated more efficiently, at a rate of $8 \log(N)$, than the longer tones, which were integrated at a rate of $5 \log(N)$. These improvements contrast with the $10 \log(N)$ improvement predicted by the energy detector model. Increased efficiency was also reported for temporal integration when the bandwidth of the signals was less than one critical band. Using the same signals, Hicks and Buus (2000) measured psychometric functions for long versus brief tones and tone complexes. They found steeper psychometric functions for brief tones, with integration functions corresponding to $8 \log(N)$ for brief tones and $5 \log(N)$ for long tones.

Work by Grose and Hall (1997) found that thresholds for tone complexes improved with increasing number of components, with thresholds improving at a rate of $10 \log(N)$, consistent with predictions based on the multiple looks model. They found a parallel improvement [still $10 \log(N)$], for detection of increments in narrow band noise, but with thresholds 2 dB higher than the pure tones. They then found less than $10 \log(N)$ improvement for detecting a decrement in noise. In a final experiment, they carried the principle of decremented signals to its extreme, performing a gap detection experiment with the same narrow band noise complexes, which revealed improvement in detection greater than the predicted $10 \log(N)$. A notable difference between these two decrement experiments relates to possible temporal integration differences; the decrement study used a 200 ms signal, but the gap detection study involved detection of gaps between 20 and 100 ms. Bacon *et al.* (2002) found consistent improvement in threshold for three-tone complexes compared with individual tones in an unmodulated noise masker, regardless of spectral distance between components. In modulated noise conditions, however, the amount of integration appeared to decrease with increased spacing between the components, possibly exhibiting a limit in spectral integration. Additionally, they combined pure tones at equally detectable levels for their threshold measures, an improvement over previous research that used equal physical levels.

C. Spectrotemporal processing

Since there is evidence to support the idea of multiple looks in both the time and the frequency dimensions of sound integration, the question remains about how these two processes may relate to each other. There has been little work published relating to both dimensions simultaneously. Spiegel (1979) made reference to the flexibility of the auditory system for both resolving short duration signals and summing longer duration signals. His data supported the idea that the spectral processing of sounds must be similarly flexible. In their study of spectral integration, Grose and Hall (1997) noted that the differences between their experiments on level decrement detection and gap detection may have been due in part to the differences in the temporal integration of the signals. However, no quantitative analysis of this suggestion is available.

Dai and Green (1993) compared thresholds for 3- and 21-tone complexes with signal durations ranging from 10 to 1000 ms. They reported that the threshold for the 3-component complex with closer frequency spacing was 10 dB lower than the 21-component complex at 10 ms, but the 21-component complex showed more rapid improvement in detectability with increasing duration than the 3-tone complex. They then measured the thresholds for a 3-component complex with the same frequency spacing as the 21-component complex. The results indicated that the improvement in thresholds was related to the frequency spacing of the tones rather than the total number of components. They invoked the multiple looks model for explanation of the results for the different duration conditions. However, this model would only predict the improved thresholds observed in the data for durations under 100 ms. Beyond this duration, no evidence of further integration can be seen. As a modification to the multiple looks model, Dai and Green (1993) suggested a “single look” model, which assumes that the filters are initially wide, and become increasingly narrow, with maximum tuning reached by 100 ms. This duration is the point at which their three-tone complex with the narrowest frequency spacing reached the steady portion of the threshold curve.

The purpose of the current study was to measure quiet thresholds for tone complexes that differed in both frequency and duration, in order to quantify the integration across both dimensions, and to examine the potential expansion of the multiple looks hypothesis into spectrotemporal integration. If the integration of increasing numbers of tones differing in both temporal and spectral location is similar to that of each domain individually, support may be evident for the multiple looks model with spectrotemporally varying signals. This study was designed to systematically measure quiet thresholds for tone complexes that vary either in spectral or temporal composition, or along both dimensions, in order to consider the potential application of multiple looks to the integration for both dimensions in the same subjects. The results should provide information about the relative salience of perception in the two dimensions and how discrete changes in both dimensions of the signals are integrated.

II. METHODS AND RESULTS

A. Subjects

Six normal hearing young adult listeners were included. Normal hearing was defined as air conduction thresholds ≤ 20 dB hearing level (HL) at the standard audiometric frequencies (250–8000 Hz) with normal otoscopic findings. Five female subjects and one male subject were included, ranging in age from 20 to 37 years old. No prior experience with psychoacoustics research was required for participation; however, introductory training was provided to assure familiarity with the procedure and signals and to assure stable quiet threshold measurements. The same subjects participated in all experiments and completed all experiments in four to five 2 h sessions per week for four to five weeks.

B. Procedures

All experiments were conducted in a sound attenuated room. Signals were generated digitally with MATLAB (version 2006b), processed through Digital Audio Labs CardDeluxe sound cards, attenuated with a TDT PA4 programmable attenuator, and presented monaurally over Sennheiser HD580 headphones. All subjects used the right ear, with the exception of subject No. 6, who used the left ear due to a slight increase in pure tone threshold in the right ear at one frequency. The experiments were conducted using an adaptive tracking [one up three down two interval forced choice (2IFC)] procedure to target a threshold at the 79% level (Levitt, 1971). The step size for signal level was initially set at 5 dB and was reduced to 2 dB after the first reversal, then reduced again to 1 dB after the next reversal. The starting level for each run was at 10 dB of attenuation, with signal generation levels set to target a pure tone threshold of 35 dB of attenuation at each frequency. Each run consisted of 50 trials and included between 5 and 10 reversals, with the first 4 reversals discarded in the calculation of the threshold estimate. Thresholds were calculated using three separate runs for each condition, with each threshold estimate based on a total of 150 trials. Occasionally, subjects ran more than three runs for a condition, if one or more run yielded an inconsistent result for the threshold estimate. These inconsistencies were generally a result of errors early in the run, causing the step size to decrease too quickly and not allowing the run to stabilize, or variation in levels for the retained reversals, causing a wide variance around the eventual calculated threshold. Results from these runs were discarded.

C. Stimuli

Stimuli consisted of one to eight 10 ms sinusoidal tone bursts. Tone frequencies were 356, 494, 663, 870, 1125, 1442, 1838, and 2338 Hz. These frequencies were selected based on the work by Grose and Hall (1997), with tones centered in alternating equivalent rectangular bands (ERBs) (Moore and Glasberg, 1983). Each 10 ms burst was gated on and off with quarter sine wave 5 ms rise and decay times, with no steady state component (Viemeister and Wakefield, 1991). Tones were separated by 10 ms silent intervals in order to minimize interaction between tones. The temporal separation was chosen to separate them into alternating equivalent rectangular durations (ERDs). ERDs were reported by Moore *et al.* (1988) to range between 8 and 10.8 ms for brief tones, depending on the conditions. While the two dimensions cannot be equated in a quantitative measure, the spacing in both dimensions was made as equivalent as possible. Overall duration of signals and silent intervals were specified to ensure that all temporal integration be contained within a 150 ms maximum window. The nominal stimulus duration represented in the stimulus matrix (Fig. 1) was 160 ms for all the experimental conditions, but the final 10 ms interval was always silent. A 300 ms interstimulus interval was used. All signals were presented in quiet. Figure 1 represents the matrix of time and frequency cells used in the experimental signals. Rows in the figure represent the temporal domain, while columns represent the spectral domain.

FIG. 1. TFR of the stimuli used in the study. The time axis is divided into 10 ms windows. The frequency axis is divided into one ERB_n filter band. Signals were separated in time and frequency so that empty squares indicate combinations for which signals were not presented. Brackets along the top and right side reflect the spacing for two-tone signals. Four-tone signals were created using consecutive or alternating tones and time intervals for close or mid conditions. Spectrotemporal signals are represented along diagonal lines.

Thus the time frequency representation (TFR) for each spectrotemporal condition, as represented by a line linking the relevant cells across the matrix, would have a slope of ± 1 , $< \pm 1$, or $> \pm 1$. The diagonal to be used was determined by the conditions within the context of each particular experiment.

D. Baseline

1. Procedure

A baseline experiment was conducted to determine the appropriate presentation levels for the individual sine wave tones for each subject. Initial thresholds were obtained, then the amplitude of each tone was adjusted to equalize the detectability of the tones across the frequencies. This level adjustment ensured that the detection of the overall signal complex in the remaining experiments was not based solely on the signal(s) with the lowest individual threshold(s). Thresholds obtained for each frequency at equal presentation levels ranged from 27.1 to 50.0 dB of attenuation, with the lowest thresholds measured for the higher frequencies, as would be expected based on the shape of the normal audibility curve. After these thresholds were obtained, the presentation levels were adjusted until all the measured thresholds were within a range of less than 2.5 dB, and the overall mean was between 34.75 and 35.25 dB of attenuation. With the range of initial thresholds, it was expected that without adjustment, the subjects would have likely been able to detect signals on the basis of the higher frequency components.

2. Results

As seen in Fig. 2, the mean threshold for all frequencies ranged from 34.56 to 35.64 dB, with an overall mean of 35.14 dB. A repeated measures analysis of variance (ANOVA) indicated no significant difference between frequency thresholds, $F(1,7)=0.891$, $p>0.05$. The thresholds for individual frequencies were averaged across subjects, and

FIG. 2. Thresholds for baseline frequencies averaged across subjects. The abscissa is the frequency for which the threshold was obtained, and the ordinate is the decibel of attenuation for that threshold. The dotted line is set at 35 dB of attenuation (the target threshold), and the dashed line indicates the obtained average for all eight frequencies, 35.14. Error bars represent one standard deviation around the mean.

then averaged across frequency. Variance for these averaged thresholds was 1.08 dB. Just noticeable differences (JNDs) in intensity for signals with a duration of 15 ms are between 3 and 4 dB (Oxenham and Buus, 2000) when presented at a level of 65 dB sound pressure level (SPL). It was assumed in the current study that the JNDs at absolute threshold would not be substantially different than those at higher presentation levels. For this reason, the variance found within individual subjects' thresholds was believed to be acceptable for the assumption of equal detectability. To check the accuracy of this assumption, the calculated improvement in thresholds based on the obtained data was compared with the improvement predicted by the energy detection model. The model predicts an improvement of 9.03 dB [i.e., $10 \log(8)$] over all eight of the frequencies in an ideal listener. When the average thresholds were submitted to the same calculation, the improvement in threshold was 8.9 dB, with the most disparate overall anticipated threshold being 8.76 dB of improvement (for subject No. 4).

E. Experiment 1 (single dimension)

1. Procedure

The first experiment was conducted to establish a reference for spectral and temporal integration within the context of the frequencies and time window used in this study. Using the adjusted presentation levels for the individual frequencies obtained in the baseline experiment for each listener, thresholds were determined for spectral integration for combinations of two, four, and eight tones. The configuration of these signals can be seen in the TFR shown in Fig. 1. The terms "high," "mid," and "low" were used to describe the location within the frequency range used for the two-tone signal conditions. For example, the high condition included the frequencies at the upper end of the spectrum used in the study (e.g., 1838 and 2338 Hz). Additionally, the terms "close," mid, and "distant" were used to describe the step size between components. Two-tone combinations were presented at 1838 and 2338 Hz for the close-high condition, at 870 and 1125 Hz for the close-mid condition, and at 356 and 494 Hz for the close-low condition. These conditions use tones next to each other within the experimental matrix and were separated by one intervening ERB_n . The tones used for the mid step size condition were 663 and 1442 Hz and were separated by five ERB_n s. The distant condition consisted of 356 and 2338 Hz, the most disparate frequencies in the matrix, and separated by 13 ERB_n s. The four-tone combinations consisted of 663, 870, 1125, and 1442 Hz for the close condition, again including adjacent tones and having only one intervening ERB_n between components, and 356, 663, 1125, and 1838 Hz for the mid condition, with three ERB_n s between components. The eight-tone combination included all the identified frequencies. All temporal and spectral integration conditions were presented randomly in order to eliminate order effects.

Temporal integration was measured for two, four, and eight signals, again using the adjusted presentation levels obtained in the baseline experiment. The signals were presented at 356, 1125, or 2338 Hz. These frequencies were

FIG. 3. Integration of multicomponent tones for experiments 1 and 2. Open circles are the thresholds for spectrotemporal signals with the same step sizes. The open triangles are the thresholds for the spectral dimension alone, and the open squares are the thresholds for temporal integration alone. Data points represent two, four, and eight tone signals collapsed across all other conditions. The dotted line is the average threshold for single tones. The long dashed line illustrates $10 \log(N)$, the medium dashed line represents $8 \log(N)$, and the short dashed line represents $5 \log(N)$.

selected to represent the high, middle, and low frequency ranges in the study. Again, the configuration of the signals can be seen in Fig. 1. Two-tone signals were presented with an intervening silent interval of 10 ms for the close condition, 50 ms for the mid condition, and 130 ms for the distant condition. Four-tone signals were presented with a 10 ms intertone interval in the close condition and a 30 ms intertone interval for the mid condition. The temporal integration for eight tones was measured with 10 ms intervals between signals.

2. Results

Spectral integration. The spectral integration measures show an improvement in threshold between approximately 1 and 2.5 dB for each doubling of the number of tones. The overall improvement in threshold from one to eight tones was 5.17 dB. In Fig. 3, this can be seen with the open triangles representing the average of the thresholds of all conditions plotted against the number of tones. The data were collapsed across all conditions since all comparisons based on frequency range and spectral spacing were found to be statistically nonsignificant. Linear regression revealed that the best fit to this function was the $5 \log(N)$ predicted curve ($R^2=0.688$, slope=1.1). A repeated measures ANOVA indicated a significant difference for increasing numbers of tones, $F(1, 3)=187.40$, $p<0.05$. Differences seen in thresholds on the basis of frequency range and spectral spacing were not significant.

Temporal integration. The temporal integration measures show an improvement in threshold between approximately 1.98 and 2.47 dB for each doubling of the number of tones. The overall improvement in threshold from one to eight tones was 6.84 dB. In Fig. 3, this can be seen with the open squares representing the average of the thresholds of all conditions plotted against the number of tones. Linear re-

gression revealed that the best fit to this function was the $8 \log(N)$ curve for these data ($R^2=0.805$, slope=0.95). Repeated measures ANOVA revealed that increasing number of tones was significant, $F(1,3)=287.42$, $p<0.05$.

The temporal integration conditions were run at three frequencies to sample potential differences on the basis of the spectral range. The change in detection threshold on the basis of the frequency was not significant, $F(2,4)=0.119$, $p>0.05$. This result confirms that the remaining conditions can readily be centered around any of the included frequencies without concern about the influence of any spectral differences confounding the perception of these signals.

F. Experiment 2 (equivalent spectral-temporal spacing)

1. Procedure

The second experiment measured thresholds for signals differing in both the temporal and spectral domains, with the signals configured in equally spaced “steps” in both the spectral dimension and the temporal dimension. The resulting signals followed the major and minor diagonals in the TFR. Signals included complexes of two, four, and eight sine wave tones varying in either an “up” or “down” frequency direction. For example, the two-tone stimuli were presented in conditions defined as close-close (adjacent frequencies and adjacent time intervals), mid-mid (steps that spanned approximately one-half the total possible range both in frequencies and temporal intervals), and distant-distant (end-point values for frequency and time). The close-close conditions were presented in both the high frequency range, that is, with 1838 and 2338 Hz tones, and the low frequency range, with 356 and 494 Hz tones. The mid-mid conditions also used the high frequency range (1125 and 2338 Hz) and the low frequency range (356 and 870 Hz). The close conditions presented tones in the 1st and 3rd time windows, with starting times of 0 and 20 ms, the mid conditions presented the tones in the 1st and 4th time windows, with starting times of 0 and 60 ms, and the distant conditions presented the tones in the 1st and 15th time windows, with starting times of 0 and 140 ms. In similar fashion, four-tone stimuli were presented with close conditions in consecutive frequency and time windows and presented in the high and low frequency ranges. The high frequency conditions included 1125, 1442, 1838, and 2338 Hz tones presented in time windows with starting times of 0, 20, 40, and 60 ms. The four-tone mid signals included alternate frequencies starting with an endpoint. For example, the up condition included the 356, 663, 1125, and 1838 Hz tones presented in the time windows with starting times of 0, 40, 80, and 120 ms. The eight-tone stimuli were all presented in close intervals since this was the only possibility within the spectral and temporal constraints of the experiment. A total of 10 two-tone conditions, 6 four-tone conditions, and 2 eight-tone conditions was used.

2. Results

For the spectrotemporal signals, the overall improvement in threshold from one to eight tones was 3.86 dB. Linear regression revealed that the best fit to this function was

the $5 \log(N)$ curve ($R^2=0.765$, slope=0.872). Repeated measures ANOVA revealed that the thresholds were significantly different than for temporal integration alone, $F(1,5)=43.184$, $p<0.05$. The differences between spectrotemporal and spectral integration were not significant, $F(1,5)=1.667$, $p>0.05$. The open circles in Fig. 3 show the overall thresholds for these equal step size signals plotted by the number of tones.

The data were also considered on the basis of spectral and temporal spacing. Because of the design of this experiment, the step size in both dimensions was equivalent in each condition. The improvement from one to two tones was nearly identical in all spacing conditions. Across all comparisons, the differences in thresholds were not statistically significant on the basis of spectral and temporal spacing or direction of frequency change.

G. Experiment 3 (different spectral-temporal spacing)

1. Procedure

The third experiment measured the thresholds for signals with different step sizes in the spectral and temporal dimensions. For example, the steps in the spectral domain may be distant, while the temporal steps may be mid or close. The alternative difference was also presented, with steps in the spectral domain being smaller than those in the temporal domain. In this way, the perceptual slope for the signals may be either steeper than 1 (as in the first example) or shallower than 1 (as in the second example), depending on the specific condition. These conditions were included in order to allow for the possibility of determining the relative importance of the two dimensions (spectral and temporal) for threshold. If one of these dimensions has a greater influence on the overall integration than the other, then the thresholds in this configuration should show an influence of slope. Because the signals required a difference in step size between the two dimensions, no eight-tone signals could be used, as the constraints of the experimental design required that these signals were always at a close step size for both dimensions. As for equivalent spacing, the signals were presented in both upward and downward slopes. A total of eight two-tone conditions and four four-tone conditions was used. These conditions were centered around the middle frequencies in the matrix.

2. Results

The magnitude of integration for these signals considered all together is less than the comparable conditions with equivalent spacing. The threshold for signals with different step sizes showed an overall improvement from one to four tones of 1.74 dB. This is different from the improvement in threshold for the same step size signals, which showed an improvement of 2.80 dB for the same change from one to four tones. Linear regression for these data showed a poor fit to any of the predicted threshold improvements, with the best fit to the $5 \log(N)$ curve ($R^2=0.575$, slope=0.577). Regression would suggest a possible fit to a slope somewhat less than $5 \log(N)$. These thresholds are represented in Fig. 4 with open squares and compared with the same step size thresh-

FIG. 4. Spectrotemporal integration across experiments 2–4. Open triangles are the signals with the same step sizes, open squares are signals with different step sizes, and open circles are random signals. Data points represent two, four, and eight tone signals collapsed across all other conditions. The dotted line is the average threshold for integration of single tones and the long dashed line is $10 \log(N)$, the medium dashed line represents $8 \log(N)$, and the short dashed line represents $5 \log(N)$.

olds (open triangles). Recall that there were no conditions that included eight tones since the constraints of the TFR would not allow a difference in the step size for the temporal and spectral domains, while still remaining within the matrix.

As in experiment 2, the data were analyzed for differences in spectral and temporal spacing and direction, and these were not significant. Additionally, the data were compared with those of equivalent spacing on the basis of the slope of the TFR of the signals. The data for signals with different spacings were all represented by slopes of greater than or less than 1, while all the signals with equivalent spacing from experiment 2 had a slope equal to 1. The thresholds for signals with a slope of <1 are plotted with open circles and those with a slope of >1 are filled circles in Fig. 5. While the thresholds for signals with different slopes appeared to be different, the difference was not statistically significant.

H. Experiment 4 (random presentation order)

1. Procedure

The fourth experiment was conducted to measure the thresholds for the complex signals when the component tones were randomly selected. The signals were presented in random order without replacement for each trial. Every tone complex presented was different. The temporal and spectral step size constraints were limited to equal step sizes between the dimensions and defined as in experiment 2. Thus, the two-tone close condition consisted of two neighboring frequencies presented in consecutive temporal windows, but the two tones were randomly selected from among the eight possible. Similarly, the two-tone mid condition consisted of two frequencies spaced approximately half the distance of the frequency range, and in the first and middle temporal windows, with the specific frequencies randomly selected. The conditions used in this experiment included three two-tone

FIG. 5. Integration of signals with different slopes. Filled circles are the thresholds for signals with greater spectral distance (slope >1), open circles are thresholds for signals with greater temporal distance (slope <1), and open triangles are threshold for signals with equivalent step sizes (slope $=1$). The dotted line is the average threshold for integration of single tones and the long dashed line illustrates $10 \log(N)$, the medium dashed line represents $8 \log(N)$, and the short dashed line represents $5 \log(N)$.

conditions, two four-tone conditions, and one eight-tone condition. All different spectral and temporal distances (close, mid, and distant) were represented in the conditions.

2. Results

The overall improvement in threshold for signals with random frequencies showed a very close match to the thresholds for the signals with equivalent spacing. This can be seen in Fig. 4 by comparing the open circles with the open triangles. Thus, there appears to be little difference in the spectrotemporal integration for signals when there is no ability to predict the frequency range in which the tones will occur. Linear regression showed the best fit to the $5 \log(N)$ predicted curve ($R^2=0.568$, slope= 0.765). Repeated measures ANOVA revealed that the thresholds for signals with random frequency presentation did not differ significantly from those of the signals with equivalent spacing, $F(1,5)=1.667$, $p > 0.05$, nor from those with different spacing, $F(1,5)=2.486$, $p > 0.05$.

When the random frequency signal conditions were analyzed on the basis of the spectral and temporal spacings, the only significant difference appeared between the two-tone conditions with close and mid spacings. All other comparisons were nonsignificant.

I. General results

A significant difference can be seen between integration for the spectral domain and the temporal domain individually. The slope of spectral integration closely approximated $5 \log(N)$, while the slope of temporal integration was closer to $8 \log(N)$. Spectrotemporal integration was not significantly different than the spectral dimension alone. The signals throughout the study were analyzed on the basis of spectral distance between tones, temporal distance between tones, direction of slope representing the signal change, and steep-

ness of that slope. The signals were also randomized for the equal spectral and temporal step size conditions. None of these analyses were statistically significant.

III. DISCUSSION

A. Spectral integration

The spectral integration conditions of the first experiment were conducted to replicate the work by [Grose and Hall \(1997\)](#), with minor adjustments to the method to allow for comparison with the remainder of the experiments within this study. The current study showed improvements of 1.02–2.44 dB per doubling of tones, comparable to their improvement of 1.5–2 dB per doubling of tones. These improvements were a good match for their results, despite different durations, as well as the adjustment to the signal levels to make them equally detectable. The wider variation in the threshold improvements may be related to the brief duration of the current tones, as little data are available related to the stability of threshold measures for these brief tones. These threshold improvements are inconsistent with the results of [van den Brink and Houtgast \(1990a, 1990b\)](#) for brief tones, which showed thresholds for brief duration tones to improve more rapidly than longer duration tones. While [Grose and Hall \(1997\)](#) used only signals centered around the 1125 Hz signal, this study also included two-tone signals in the high and low frequencies in order to consider the potential for differential integration on the basis of frequency range. The differences found in these data were not significant.

To consider integration across the spectrum, the concept of multiple looks could be considered in the frequency domain. The auditory system may be combining the signals across independent filters in a spectral multiple look in much the same way that [Viemeister and Wakefield \(1991\)](#) proposed for the temporal domain and selectively responding to the bands that include energy. In the current study, the distance between component tones did not affect spectral integration, which was consistent with the unmodulated noise conditions reported by [Bacon *et al.* \(2002\)](#). The multiple looks hypothesis would suggest that the auditory system can “intelligently” select the time intervals containing information for integration. It appears with the current results that this may be the case in the spectral domain.

B. Temporal integration

The temporal integration conditions of the first experiment were also conducted to replicate earlier research. In their work, [Viemeister and Wakefield \(1991\)](#) demonstrated an improvement in threshold of 2.5 dB from one tone to two tones, while the current study showed an improvement of 2.47 dB for the same condition. The extension beyond their work yielded improvements of 2.39 dB from two to four tones and 1.98 dB from four to eight tones. No significant difference can be seen for different temporal spacings, from 10 ms of silence between tones to 130 ms of silence. All of these distances are greater than the 5 ms gap required for independence of the observations. Thus the integration ap-

pears to be the same throughout the overall time window used in all the signals, supporting the multiple looks hypothesis for temporal integration.

The similarity in the detection of the signals presented at the three frequencies (356, 1125, and 2338 Hz) additionally confirms that the temporal integration task does not vary with the spectral range being used. As a result, further study can reliably be pursued in any frequencies in the range included here, with good assurance that the results will be the same.

C. Spectrotemporal integration

The combination of the two domains yields additional information about the function of the auditory system that has previously not been explored. The results of this study indicate that the limits of spectrotemporal integration are due to the limits in the ability to integrate spectral information, with very similar detection thresholds between the two sets of integration conditions. The spectrotemporal signals did not show significant differences on the basis of spectral and temporal step sizes or direction of frequency change. This, again, supports the multiple looks hypothesis. The auditory system appears to be capable of monitoring the filters throughout the entire range of frequencies included here and detecting signals that occur in any of the time windows. The information contained in those tones is accumulated over the course of the longer, overall, time window for use in detection.

While the same step size signals were processed in a way that was very similar to those varying in only the spectral dimension, the processing of the different step size signals appeared to be dissimilar. The thresholds for these signals were consistently poorer than for the same step size signals. This came as a surprise after the data collection was initiated. Prior expectation was that there should be no difference between the two types of spectrotemporal signals. One would expect that this prediction would be further supported based on the nonsignificant differences related to step sizes in experiment 2. Separate analysis of slope showed that the thresholds for signals that changed over greater spectral than temporal range showed essentially no improvement with the increase from two tones to four. While these differences were not statistically significant, they suggest the possibility that the auditory system may be near the limit for tracking changes in signals at this rate, as these signals consisted of a total duration of 30 ms, with spectral changes from 7 to 15 ERB_s. Further study with different signal durations and different spectral ranges may provide additional insight into this uncertainty.

D. Random presentation order

If a phenomenon such as auditory streaming is involved in the detection of these signals, then trial-by-trial random order of presentation should have a negative effect on the integration of the tones. Listeners should be unable to use streaming as a cue to detect the signal in the random presentation order conditions, which should result in higher thresholds for these signals. Alternatively, if selective monitoring

of critical bands is important for this detection, then again, the random order of presentation of frequencies should have a negative effect on integration. The results showed that the detection of the randomly generated signals was closer to that of the same step size signals than was detection of the different step size signals. This also supports the multiple looks hypothesis, in that clearly the listeners are not simply establishing which critical bands contain the component tones and then monitoring those in order to detect the total signals. Rather, they are monitoring all the critical bands that may be included and detecting the tones wherever they may be present. In fact, auditory streaming is not supported as a factor in detection for these signals, as the poorest overall detection in the random frequency signals was for the closely spaced tones. Even in the predictable conditions, the thresholds showed no difference on the basis of spectral spacing. However, since the differences in the thresholds are quite small overall, and the signals are brief, the analysis of the potential influence of streaming in spectrotemporal integration cannot be fully addressed here.

IV. SUMMARY AND CONCLUSIONS

In all experiments in this study, the one variable that resulted in improved thresholds was the number of tones. Regardless of spectral or temporal spacing, the thresholds showed improvement with every increase, in spectral integration alone, temporal integration alone, and spectrotemporal integration. The same improvement occurred when signals were presented with random frequency selection. Thus, detection appears to be based on the use of multiple looks for signals even when changing in both dimensions trial to trial. The auditory system appears to be able to monitor all the critical bands and select the windows that include the information in order to detect the signals.

All of the results presented here are consistent with the multiple looks hypothesis, with some limits in the spectral domain. It is unclear why spectral integration is poorer than temporal integration, but the analysis of the predicted threshold improvement calculated with baseline measures indicates that it is not due to differences in the detectability of each frequency. The results are also consistent with prior separate studies of these dimensions. Kidd *et al.* (2003) suggested an auditory stream coherence explanation for the results in their temporal integration study. This cannot be accepted directly on the basis of the current results since the random presentation order conditions were not significantly different than the predictable conditions. In fact, the thresholds for random frequency signals were more similar to those for equal step size signals than were the unequal step size thresholds. While it seems reasonable that auditory stream coherence may play a role in the detection of these signals, further study is re-

quired to determine what that role may be, particularly in light of the potential lower limit on temporal detection seen here, and similarity in thresholds for all temporal and spectral spacing.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institutes of Health, NIH/NIDCD R01-DC006879. This article is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at The Ohio State University, Columbus, OH. Thanks also to Walt Jesteadt and an anonymous reviewer for helpful comments on an earlier draft. Additional thanks to Eric Healy and Robert Fox for their input.

- Bacon, S. P., Grimault, N., and Lee, J. (2002). "Spectral integration in bands of modulated or unmodulated noise," *J. Acoust. Soc. Am.* **12**, 219–226.
- Buus, S. (1999). "Temporal integration and multiple looks, revisited: Weights as a function of time," *J. Acoust. Soc. Am.* **105**, 2466–2475.
- Dai, H., and Green, D. M. (1993). "Discrimination of spectral shape as a function of stimulus duration," *J. Acoust. Soc. Am.* **93**, 957–965.
- Fletcher, H. (1940). "Auditory patterns," *Rev. Mod. Phys.* **12**, 47–65.
- French, N. R., and Steinberg, J. C. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **9**, 90–119.
- Green, D. M. (1958). "Detection of multiple component signals in noise," *J. Acoust. Soc. Am.* **30**, 904–911.
- Green, D. M. (1961). "Detection of auditory sinusoids of uncertain frequency," *J. Acoust. Soc. Am.* **33**, 897–903.
- Greenwood, D. D. (1961). "Auditory masking and the critical band," *J. Acoust. Soc. Am.* **33**, 484–502.
- Grose, J. H., and Hall, J. W. (1997). "Multiband detection of energy fluctuations," *J. Acoust. Soc. Am.* **102**, 1088–1096.
- Hicks, M. L., and Buus, S. (2000). "Efficient across-frequency integration: Evidence from psychometric functions," *J. Acoust. Soc. Am.* **107**, 3333–3342.
- Kidd, G., Mason, C. R., and Richards, V. M. (2003). "Multiple bursts, multiple looks, and stream coherence in the release from informational masking," *J. Acoust. Soc. Am.* **114**, 2835–2845.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (1988). "The shape of the ear's temporal window," *J. Acoust. Soc. Am.* **83**, 1102–1116.
- Oxenham, A. J., and Buus, S. (2000). "Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions," *J. Acoust. Soc. Am.* **107**, 1605–1614.
- Schafer, T. H., and Gales, R. S. (1949). "Auditory masking of multiple tones by random noise," *J. Acoust. Soc. Am.* **21**, 392–397.
- Spiegel, M. F. (1979). "The range of spectral integration," *J. Acoust. Soc. Am.* **66**, 1356–1363.
- van den Brink, W. A., and Houtgast, T. (1990). "Efficient across-frequency integration in short-signal detection," *J. Acoust. Soc. Am.* **87**, 284–291.
- van den Brink, W. A., and Houtgast, T. (1990). "Spectro-temporal integration in signal detection," *J. Acoust. Soc. Am.* **88**, 1703–1711.
- Viemeister, N. F., and Wakefield, G. H. (1991). "Temporal integration and multiple looks," *J. Acoust. Soc. Am.* **90**, 858–865.
- Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). "Critical bandwidth in loudness summation," *J. Acoust. Soc. Am.* **29**, 548–557.