Iterated rippled noise discrimination at long durations

William A. Yost^{a)}

Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 12 March 2009; revised 30 June 2009; accepted 6 July 2009)

Iterated rippled noise (IRN) was used to study discrimination of IRN stimuli with a lower number of iterations from IRN stimuli with a higher number of iterations as a function of stimulus duration (100–2000 ms). Such IRN stimuli differ in the strength of the repetition pitch. In some cases, the gain used to generate IRN stimuli was adjusted so that both IRN stimuli in the discrimination task had the same height of the first peak in the autocorrelation function or autocorrelogram. In previous work involving short-duration IRN stimuli (<500 ms), listeners were not able to discriminate between IRN stimuli that had different numbers of iterations but the same height of the first peak in the autocorrelation function. In the current study, IRN discrimination performance improved with increases in duration, even in cases when the height of the first peak in the autocorrelation was the same for the two IRN stimuli. Thus, future studies involving discrimination of IRN stimuli may need to use longer durations (1 s or greater) than those that have been used in the past.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3192345]

PACS number(s): 43.66.Hg [RLF]

Pages: 1336–1341

I. INTRODUCTION

Iterated rippled noise (IRN) has been used to study the pitch and pitch strength of complex noise-like stimuli. IRN is generated by a cascade of add and delay networks as described by Yost *et al.* (1996). Models based on autocorrelation (see Yost *et al.*, 1996) have been successful in accounting for most of the data generated using IRN stimuli. Figure 1(a) shows the summary autocorrelogram¹ of an IRN stimulus generated with a 1.25-ms delay (d), gain (g) of 1, and four iterations (n). The reciprocal of the lag of the first (non-zero) peak in the autocorrelogram has been used to account for the pitch of IRN stimuli (Yost, 1996a), and the relative height of this peak has been used to account for the pitch strength of IRN stimuli, which is usually measured in discrimination experiments (Patterson *et al.*, 2000; Yost, 1996b; Yost *et al.*, 1996). The other peaks (at longer lags) in the autocorrelogram have usually not been used to account for either the pitch or the pitch strength of IRN stimuli (however, see Patterson *et al.*, 2000). The height of the first peak increases as gain (g) increases from 0 to 1.0 and as the number of iterations increases (i.e., the normalized height of the first peak of the autocorrelation function when $g=1.0$ is $n/(n+1)$). Of particular relevance to this paper, previous research (e.g., Patterson *et al.*, 2000; Yost *et al.*, 1996; and Yost, 1996b) has indicated that when the first peak in the autocorrelogram is the same for two different IRN stimuli, performance is at or near chance when listeners are asked to discriminate between two such IRN stimuli.

Figure 1(b) shows the auditory excitation pattern based on filtering a low pass (8000 Hz) IRN stimulus with a gammatone filter bank (see Patterson *et al.*, 1995). The stimulus for Fig. 1(b) was the same IRN stimuli used in Fig. 1(a). The excitation pattern shows spectral peaks at 800 Hz (the reciprocal of 1.25 ms) and its integer multiples out to about 6000

Hz (the upper limited of peripheral resolvability for this particular IRN stimulus). That is, the spectra of IRN stimuli have the characteristic spectral ripple with spectral peaks at one over the delay and its integer multiples and spectral valleys in between.

De Cheveigne (2007) warned that the statistics of IRN stimuli are complicated with several changing slowly over time. De Cheveigne (2007) described the slow change in spectral envelope (spectral ripples) that occurs for large number of iterations and suggested that these changes or ones related to them may be a possible cue for processing IRN stimuli. This implies that auditory processing of IRN stimuli may be different at long durations (greater than 250–500 ms, which is the typical duration of IRN stimuli used in the previous literature) as compared to shorter durations.

The main purpose of this paper was to investigate the ability of listeners to discriminate between IRN stimuli as a function of their duration. In addition, the ability of listeners to discriminate between IRN stimuli generated with different numbers of iterations was also investigated.

As the number of iterations (n) increases, so does the normalized height of the first peak in the autocorrelation function (AC1). For instance, for $n=1$, $AC1=0.5$, and for $n=2$, $AC1=0.67$. If gain (g) is decreased to 0.525 for the $n=2$ case, then $AC1=0.5$ for the $n=1$ case. In previous studies, using IRN stimuli that had the same value of AC1 made it very difficult for listeners to distinguish one from the other (for instance, Yost, 1996b). In this paper, the authors describe discrimination results in which the two IRN stimuli to be discriminated differed in n . In some conditions, the two stimuli were both generated with $g=1$. In other conditions, the value of g for the stimulus with the larger number of iterations was generated with a value of g so that AC1 was the same as that for the stimulus generated with the smaller number of iterations. These discrimination tasks were performed as a function of IRN duration.

^{a)}Electronic mail: william.yost@asu.edu

FIG. 1. The top panel represents a summary autocorrelogram [based on the auditory image model of Patterson *et al.* (1995)] for an IRN stimulus generated with a delay (d) of 1.25 ms, a gain (g) of 1.0, and four iterations (n). The large peak at a lag of 1.25 ms is used to account for the pitch of 800 Hz for the IRN stimulus, and the relative height of the first peak (and only the first peak) has previously been used to account for the pitch strength of IRN stimuli. But note the peaks at lags of 2.5 and 3.75 ms (integer multiples of 1.25 ms). The bottom panel is an auditory spectrum or excitation pattern of the IRN stimulus used for the top panel (but low passed filtered at 8 kHz). The spectral peaks at 800 Hz and its integer multiples indicate the resolved spectral ripples for this IRN stimulus.

II. METHOD

A. Participants

Six listeners who indicated they had normal hearing participated in the experiments. They ranged in age from 19 to

32, and four were female and two were male. All procedures were approved by Loyola University's Institutional Review Board (IRB).

B. Stimuli

The IRN stimuli were generated using the add-original procedure described by Yost *et al.* (1996). The IRN stimuli were generated independently during each interval on each trial, low pass filtered at 8000 Hz, and played out of an Echo Gina sound card at 44 100 samples/s. Table I indicates the values of g and n used in the experiments. The base number of iterations was increased by the additional number of iterations, and the gain (g) of the IRN stimulus with the higher number of iterations (base plus additional iterations) was the gain shown in Table I. The actual number of additional iterations used for each base number of iterations was determined in a pilot study. A d of 1.25 ms produces an 800-Hz pitch, and d of 16 ms produces a 62.5-Hz pitch (Bilsen and Ritsma, 1967/1968; Yost, 1996a). The values of g that were less than 1 were chosen so that AC1 for the IRN stimulus with the larger value of n would be the same AC1 value as for the base IRN stimulus.² The overall level of all IRN stimuli was randomly roved between 65 and 75 dB SPL (sound pressure level) across intervals of the same-different task. Duration ranged from 100 to 2000 ms, and the stimuli were shaped with 10-ms raised cosine rise-decay times. The stimuli were presented diotically over Sennheiser HD 280 Pro headphones, while the listeners were seated in a double-walled soundproof room.

C. Procedure

A same-different task was used to estimate percent correct. Half of the trials (determined randomly) consisted of two intervals with independent IRN noise samples but with the same values of d , g , and n used to generate the IRN stimuli. The number of iterations for these "same" trials was always the base number of iterations (see Table I), and g was always 1.0. The remaining half of the trials were "different" trials in that one of the two intervals (chosen at random) contained the IRN stimulus with the larger n (base plus additional number of iterations, see Table I), and the other interval contained the stimulus with the base number of iterations (smaller number of iterations). In different conditions, the value of g for the stimuli with the larger number of

TABLE I. Base number of iterations (n), additional number of iterations, and gain (g) used to generate the IRN stimuli.

Base n	Additional n	Gain (g)	Base n	Additional n	Gain (g)
1	1 (small)	1.0	4	8 (small)	1.0
1	1 (small)	0.525	4	8 (small)	0.82
1	3 (large)	1.0	4	16 (large)	1.0
1	3 (large)	0.51	4	16 (large)	0.8
2	3 (small)	1.0	8	16 (small)	1.0
2	3 (small)	0.56	8	16 (small)	0.84
2	8 (large)	1.0	8	32 (large)	1.0
2	8 (large)	0.65	8	32 (large)	0.89

FIG. 2. (Color online) Four panels, one for each duration condition, indicating average (across listeners) percent correct discrimination as a function of a small or large increment in the number of iterations (base plus additional number of iterations) over the base number of iterations. The different curves in each panel represent the different conditions of the base number of iterations. The gain (g) was always 1.0 for all conditions shown in this figure. The data are for the delay of 1.25 ms condition, and the legend indicates the base number of iterations—the base plus the small number of additional iterations and base plus the large number of additional iterations.

iterations was either 1.0 or set to the values shown in Table I. The two intervals were separated by 300 ms. Thus, the different trials contained IRN stimuli with different numbers of iterations. The listeners indicated whether a trial was the same or different. Feedback indicating the correct response was provided on each trial. Four 50-trial blocks were used to estimate percent correct for each condition.

III. RESULTS and DISCUSSION

Figures 2 and 3 show mean percent correct vs a small or a large difference in n (additional iterations, see Table I) for the ability of listeners to discriminate between IRN stimuli with different numbers of iterations (n) when the gain (g) was 1.0 (i.e., when the IRN stimulus with the higher number of iterations also had the higher value of AC1). For each base number of iterations, there was either a small increment in the number of iterations or a large increment (additional number of iterations). Each panel in each figure represents a different duration. Figure 2 shows the data for a delay of 1.25 ms (800-Hz pitch), and Fig. 3 shows the data for 16 ms (62.5-Hz pitch). Each item in the legend indicates the base number of iterations, followed by the base plus the small number of additional iterations, and then the base plus the

FIG. 3. (Color online) The same display shown in Fig. 2, but for data for a delay of 16 ms.

large number of additional iterations (e.g., a base of 1 iterations with one and three additional iterations would be 1, 2, and 4)

Figure 4 shows data when the gain for the larger number of iterations was set so that each stimulus (small and large numbers of additional iterations) had the same value of the first peak in the autocorrelation function, AC1 [i.e., adjustments were made to g to make the stimulus with the larger number of iterations have the same value of AC1 as the IRN stimulus with the lower number of iterations² (see Table I)]. The data are plotted as average percent correct (averaged across listeners and the various numbers of iteration conditions for each duration) vs duration for the 1.25-ms delay (top panel) and 16-ms delay (bottom panel).

The overall result is that performance for all of the discriminations improved with increasing duration at least over the range from 100 to 1000 ms. While there was improvement in performance in all tasks with increasing duration, the improvement was not large. That is, performance increased by 10%–20%, which was statistically significant in some cases, but even at 2 s some listeners were not always 100% correct. In Figs. 2 and 3, as the number of additional iterations increased over the base number of iterations, so did discrimination performance. The change in number of iterations required to discriminate between two IRN stimuli that only differ in the number of iterations increases substantially as the base number of iterations increased. For instance, at 100 ms an $n=2$ condition can be discriminated from an $n=1$ condition with about 75% accuracy, but for an $n=4$ condition, it takes a difference of eight iterations ($n=4$ vs n

FIG. 4. The average percent correct discrimination between the smaller (base number of iterations) and larger numbers of iterations (base plus additional number of iterations) is shown as a function of duration when the gain (g , see Table I) for the larger number of iteration conditions was set such that the value of the first peak in the autocorrelation function (AC1) was the same for both IRN stimuli (small and large numbers of additional iterations) in the same-different task. The top figure represents data for a delay of 16 ms, and the bottom represents data for a delay of 1.25 ms. Error bars are one standard deviation of the mean threshold for each listener and IRN iteration condition.

=12) for this same level of performance. There was only a very small effect of delay (related to the pitch of the IRN stimuli) on the results for any of these conditions. Slightly better performance was obtained for the 1.25-ms condition over the 16-ms condition. There was no apparent change in performance based on the various base number of iteration conditions.

Yost (1996b) argued that the z -score transform of $(10^{AC1h} - 10^{AC1l})/2$ describes listeners' performance in discriminating between IRN stimuli based on different heights of AC1.³ In the equation, AC1h is the height of the first peak of the autocorrelation function for a condition with a high AC1 peak (e.g., the base plus additional number of iteration conditions), and AC1l is the height of the first peak of the autocorrelation function for a condition with a low peak height (e.g., base number of iteration conditions). In these experiments, the duration of the IRN stimuli was 500 ms. Table II shows a comparison of the results of the present study for an average of the 500-ms duration data when $g = 1$ to the outcome of the function described by Yost (1996b). Except for perhaps the condition involving discrimination between $n=8$ and $n=40$, the data and the output of the equation agree reasonably well. There are at least two differences

TABLE II. Comparison of the percent correct discriminations for all of the 500-ms data of the present paper when $g=1$ as compared to the predictions of Yost (1996b). Predictions are the cumulative standard normal probabilities assuming standard scores (z -scores) determined by $(10^{AC1h} - 10^{AC1l})/2$, where AC1h is the height of AC1 for the base plus additional number of iteration conditions and AC1l is the height for the base number of iteration conditions.

	Base vs base plus additional iterations			
	1 vs 2	2 vs 5	4 vs 12	8 vs 24
Data (%)	73%	83%	73%	76%
Predictions (%)	77%	83%	79%	75%

	1 vs 4	2 vs 10	4 vs 20	8 vs 40
Data (%)	89%	96%	91%	80%
Predictions (%)	94%	96%	91%	92%

in the experiments of Yost (1996b) and the present experiments that may impact comparisons between the two studies: (1) no feedback was used in Yost (1996b), and feedback was employed in the experiments of this paper; (2) the height of the first peak of AC1 was never as great in the Yost (1996b) experiments as it was for the $n=24$ and 40 conditions of the present paper. Given the difference in performance as a function of duration indicated in the current paper, the function used by Yost (1996b) would have to be modified to account for the results as a function of duration.

Two separate analysis of variance (ANOVA) tests were conducted; one when $g=1$ (Figs. 2 and 3) and one when g was not 1.0. (Fig. 4). When $g=1$, the ANOVA had two levels of delay (16 and 1.25 ms), four levels of base number of iteration (1, 2, 4, and 8), two levels of additional iterations (small and large), and four levels of duration (100, 500, 1000, and 2000 ms). The main effect of additional number of iterations and the main effect of duration were both significant ($p < 0.05$), while the main effect of base number of iterations and the main effect of delay were not significant ($p > 0.05$), and no interactions were significant ($p > 0.05$). When g was not equal to 1, the ANOVA had four levels of duration (100, 500, 1000, and 2000 ms) and two levels of delay (16 and 1.25 ms). The main effect of duration was significant ($p < 0.05$), while the main effect of delay and the interaction were not significant ($p > 0.05$).

These data show that increasing duration improved performance in these IRN discrimination experiments. These results also indicate that at longer durations, IRN stimuli with the same height of the first peak in the autocorrelation (Fig. 4) can be discriminated from one another at performance levels above chance. This result is in contrast with previous studies (see Yost, 1996b) obtained at shorter durations (500 ms or less) that suggested that when the first peak of the autocorrelation was the same, discrimination was at or near chance. While the average discrimination performance was near chance in the present study, the average data at 100 and 500 ms were statistically different from chance using a T-test ($p < 0.05$). No statistical tests to determine if the near chance discrimination performance was statistically different from chance were calculated by Yost (1996b). While the data

reported by Yost (1996b) appear to be closer to chance than those in the current paper, feedback was not employed in the discrimination tasks used by Yost (1996b) as it was in the present paper.

There was a fair amount of individual differences in overall performance, but all subjects showed the same pattern of results, as indicated in Figs. 2–4. Two subjects could perform at nearly 100% correct at 1000 and 2000 ms for the conditions in which $g=1.0$ and when there was a large increment in additional number of iterations. These listeners were close to 100% correct at 1000 and 2000 ms for the cases in which the value of the AC1 were the same for both stimuli (Fig. 4). The other four listeners always performed at less than 100% percent correct in all conditions.

It is the case that with increasing duration there is less variability from trial to trial in the spectral peaks and valleys and in the autocorrelation function peaks of the physical stimuli. Thus, the improvement in performance with increasing duration may simply reflect this change in IRN stimulus variability. Two analyses were performed to estimate this variability. Normalized autocorrelation functions were generated for all of the various numbers of iteration conditions used in this paper for the IRN stimulus generated with 1.25-ms delay, and the height of the peak at a lag of 1.25 ms was tabulated for 1000 50-trial blocks (50 000 trials). The comparison to be made was between the mean height over the 1000 blocks and the standard deviations over the 1000 blocks for the 100-ms conditions vs the 1000-ms condition. Across all of the various iterations, there was never a difference of more than 1% between the mean heights of the first peak in the autocorrelation function of the 100-ms and 1000-ms IRN stimulus conditions. While the variability (standard deviations) in the mean heights of the first peak in the autocorrelation function was always less for the 1000-ms conditions than for the 100-ms conditions, the differences were always less than 3%. A 3% change in the height of the first peak in the autocorrelation function is not detectable (Yost, 1996b). Thus, it is unlikely that a difference in the variability in the autocorrelation functions can account for the increase in performance with increasing duration.

To analyze the results spectrally, a one-equivalent-rectangular-filter wide gammatone filter was centered at 800 Hz (the first peak in the spectrum of the 1.25-ms IRN stimulus) and a similar gammatone filter were centered at 1200 Hz (the valley above the first spectral peak). The peak-rms to valley-rms ratio expressed in decibels was calculated for 1000 50-trial blocks for the 100-ms duration condition and then for the 1000-ms duration condition. There was less than a decibel difference on average between the peak-to-valley ratios for the 100-ms condition and for the 1000-ms condition. The variability in the peak-to-valley ratios was always less for the 1000-ms condition than for the 100-ms condition, but the difference in the standard deviations between the 100- and 1000-ms duration conditions was always less than 1.39 dB. It is unlikely that this difference in spectral variability could account for the improved performance measured as a function of duration.

One basis for distinguishing between stimuli with different numbers of iterations, especially when the stimuli have

the same height of the first peak in the autocorrelation function, is the height of the higher-order peaks [peaks at lags at higher integer multiples of the reciprocal of the delay (see Fig. 1) and the peaks at lags of 2.5 and 3.75 ms]. For instance, if g is adjusted so that the height of the first peak for the $n=10$ IRN stimulus is the same as that for the $n=2$ stimulus, then the height of the second peak for the $n=10$ case is close to 0.68, while the height of this second peak for the $n=2$ case is 0.33. Thus, the height of the second peak (and peaks at longer lags in the autocorrelation function) clearly differs when n differs, even when the heights of the first peak are adjusted to be equal. Thus, these peaks at longer lags are a possible cue for discrimination. The intervals in the IRN stimulus that lead to the correlations of 0.68 and 0.33 are at lags of twice d , i.e., 32 or 2.5 ms (for $d=16$ and 1.25 ms, respectively). Perhaps it takes a longer-duration stimulus to be able to sample enough of these longer intervals to aid in discrimination. No matter what cue is used for discriminating one IRN stimulus from another, performance does improve somewhat as the duration of IRN stimulation is increased out to a second or more.

One use of IRN stimuli where long-duration stimulation might need to be considered is the use of IRN stimuli in neural imaging studies (e.g., functional magnetic resonance imaging), where a considerable stimulus duration may be required to average responses. If long-duration IRN stimuli are used rather than the repeated shorter-duration pulsed stimuli, the neural image data might reflect that same ability for processing shown in this paper. That is, better processing might be suggested than would be obtained for shorter-duration stimuli.

Thus, De Cheveigne's (2007) caution seems warranted. Slow statistical changes in IRN stimuli (perhaps those associated in some way with peaks at longer lags in the autocorrelation function) may be usable in discriminating one IRN stimulus from another. As such, long-duration (of a second or more) IRN stimuli may be required to accurately determine the ability to process IRN stimuli in some conditions.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD). The research was conducted while the author was at the Parmly Hearing Institute of Loyola University Chicago, and he is grateful for the interactions with Toby Dye, Dick Fay, and Stan Sheft. The author is also grateful to Dr. Chris Brown and Farris Walling at ASU for their comments on the paper.

¹An autocorrelogram is computed by generating a neural activation pattern (NAP) using the auditory image model of Patterson *et al.* (1995). Then, each frequency channel of the NAP is subjected to an autocorrelation computation. These channel by channel autocorrelations are summed across channels and presented as the summary autocorrelogram. Summary autocorrelograms are similar, but not identical, to the autocorrelations of the stimulus.

²The equation for the normalized autocorrelation function of an ensemble average of an infinite wideband noise, described in footnote 3 of Yost *et al.* (1996), was used numerically to compute the value of g for a higher

number of iterations to be the same as that for a lower number of iterations. This equation is $AC1 = [g^*(1 - g^{2n})] / [1 - g^{2(n+1)}]$. For $g = 1.0$ conditions, g equaled 0.999 when this equation was used.

³Yost (1996b) described the function 10^{AC1} as a “pitch strength” function. Shofner and Selas (2002) used a power function to describe a pitch strength function and showed that it fit their data somewhat better than the function used by Yost (1996b). Since Shofner and Selas (2002) did not use discrimination procedures such as those used in Yost (1996b) and in this paper, the function (10^{AC1}) is used in this paper. The values of n used for all conditions of the present paper were chosen from a pilot study using 1000-ms long stimuli rather than the 500-ms long stimuli used by Yost (1996b).

Bilsen, F. A., and Ritsma, R. J. (1967/68). “Repetition pitch mediated by temporal fine structure at dominant spectral regions,” *Acustica* **19**, 114–115.

De Cheveigne, A. (2007). “Comment by de Cheveigne,” in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V.

Hohmann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, New York), pp. 90 and 91.

Patterson, R. D., Allerhand, M., and Giguere, C. (1995). “Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,” *J. Acoust. Soc. Am.* **98**, 1890–1895.

Patterson, R. P., Yost, W. A., Handel, S., and Datta, J. A. (2000). “The perceptual tone/noise ratio of merged iterated rippled noises,” *J. Acoust. Soc. Am.* **107**, 1578–1588.

Shofner, W. P., and Selas, G. (2002). “Pitch strength and Stevens’ power law,” *Percept. Psychophys.* **64**, 437–450.

Yost, W. A. (1996a). “Pitch of iterated rippled noise,” *J. Acoust. Soc. Am.* **100**, 511–518.

Yost, W. A. (1996b). “Pitch strength of iterated rippled noise,” *J. Acoust. Soc. Am.* **100**, 3329–3335.

Yost, W. A., Patterson, R. D., and Sheft, S. (1996). “A time domain description for the pitch strength of iterated rippled noise,” *J. Acoust. Soc. Am.* **99**, 1066–1078.