Judgments of intensity for brief sequences

Frederick J. Gallun

*VA RR&D National Center for Rehabilitative Auditory Research, Portland VA Medical Center,
3710 SW US Veterans Hospital Road, Portland, Oregon 97239
frederick.gallun@va.gov*

Abstract: The ability to make intensity judgments for sequential stimuli was examined with an intensity-discrimination task involving three 50-ms noise bursts with non-overlapping frequency ranges. Targets (single bursts) presented in three-burst sequences were required to be as much as 5 dB more intense than targets presented as single bursts in isolation, especially for the later targets. Randomizing target position in the sequence did not reliably reduce performance, nor were thresholds for younger and older listeners reliably different. These increases in increment detection threshold are indications of a specific intensity-processing deficit for stimuli occurring later in a sequence.

PACS numbers: 43.66.Fe, 43.66.Mk [QF]

Date Received: January 5, 2010 **Date Accepted:** February 18, 2010

1. Introduction

Older listeners' reduced ability understand speech in complex environments has been associated with a specific temporal processing deficit (e.g., [Gordon-Salant and Fitzgibbons, 1999](#); [Wingfield et al., 1999](#); [Salthouse, 1992](#)). Listeners of all ages asked to make judgments about stimuli embedded in rapidly changing sequences tend to do worse for longer sequences and later arriving stimuli (e.g., [Vachon and Tremblay, 2005](#); [Cousineau et al., 2009](#)). If aging is associated with specific temporal impairments, then it becomes very important to understand the baseline ability of listeners to make temporal judgments for sequential stimuli.

The ability to make intensity judgments is usually tested by asking the listener to compare an isolated standard of a known frequency and level with a target matched in duration and frequency and varying only in level (e.g., [McGill and Goldberg, 1968](#); [Jesteadt et al., 1977](#)). Such paradigms provide stable estimates of threshold and extremely precise performance (thresholds of 2 dB or less).¹ Embedding standard and target in a sequence, however, can greatly increase thresholds. [Watson \(2005\)](#) found that the intensity of targets presented in sequences of distractor tones needed to be as much as 7 dB more intense than when those same targets were presented in isolation. Uncertainty about the target or confusions about which stimulus was target and which was distractor are the usual explanations for these effects (e.g., [Watson et al., 1975](#); [Leek and Watson, 1984](#); [Watson, 2005](#)). Nonetheless, some data are consistent with a reduction in performance associated merely with temporal position. The current experiment had as its goal the establishment of the size of such effects and the dissociation of problems due to impaired temporal processing from the impacts of age and/or uncertainty. In addition, a simultaneous condition was included in order to establish baseline performance for a condition in which it was more likely that all of the elements would be perceived as a single auditory object.

2. Methods

Five "younger" listeners (4 females, mean age: 31.0 years, range: 28–36 years) and seven "older" listeners (3 females, mean age: 56.0 years, range: 51–60 years) participated. All had pure-tone hearing thresholds of 25 dB hearing level (HL) or better (re [ANSI, 2004](#)) at octave frequencies between 0.25 and 8 kHz in the left ear—the test ear in all conditions. All were paid for their participation and all procedures were reviewed and approved by the Institutional Review Boards of both the Portland VA Medical Center and the Oregon Health Sciences University.

Stimuli were 50-ms noise bursts with 5 ms $\cos^2$ onset and offset ramps. Bursts were presented in three frequency ranges: low (400–560 Hz), middle (1620–2400 Hz), or high (4080–6100 Hz). Noise was generated digitally (sampling rate of ~ 50 kHz) in the MATLAB environment (Mathworks, Inc., Natick, MA) from 20-Hz spaced equal-amplitude tones with randomized starting phase. Tucker-Davis Technologies digital-to-analog converters, anti-aliasing filters, and attenuators (TDT System 3 RP2.1, PA5, and HB7 hardware) generated an analog signal that was presented to the left ear through a Sennheiser HD280 headphone.²

Detection thresholds (DTs) were obtained for each noise burst (low, middle, and high frequencies) in a same/different task. Two intervals were presented, each marked with a visual display and separated by 500 ms of silence. Each interval had a 50% chance of containing a signal. Listeners were to respond “same” when a signal or silence had been presented on both intervals (50% chance) and “different” when one interval contained silence and one interval contained a signal (50% chance). Feedback was given after each trial, and the rms level of the noise burst was increased or decreased following a three-down/one-up adaptive tracking procedure (Levitt, 1971). The initial level of 40 dB sound pressure level (SPL) was raised or lowered in 4-dB steps until the first reversal, 2-dB steps until a second reversal, and then 1-dB steps for eight more reversals, which were averaged to give the DT estimate. Three estimates of DT were collected for each frequency.

Increment-detection thresholds (ITs) were obtained for each listener for bursts presented alone and in sequences. Two intervals were presented with the same visual marking and 500 ms separation, and listeners reported “same” if the stimulus in the second interval had the same intensity (or set of intensities) as the stimulus in the first interval (as was done in Cousineau *et al.*, 2009). Differences were always an increment in the intensity of only one of the bursts in the second interval. ITs were estimated using a same/different tracking procedure similar to that employed in the detection task.

Baseline burst level (to which the increment could be added) was roved on each trial from a nominal level of 35 dB sensation level (SL) by ± 5 dB in order to encourage comparison of the two intervals rather than judgments of the second (see Green, 1988). When multiple bursts were presented in a single interval, each was roved independently to discourage listeners from comparing across bursts in a sequence. The increment was initially set to 10 dB and changed by 4 dB until the first reversal, 2 dB until the second reversal, and 1 dB for eight more reversals, which were averaged to estimate threshold. Increments never exceeded 25 dB and in general the maximum level of the incremented stimulus rarely exceeded 85 dB SPL.

Four conditions were tested in an interleaved fashion to preclude learning effects. In the “single-burst” condition, one fixed frequency burst was presented in each interval for the entire adaptive track. In the “simultaneous” condition, all three bursts were presented with simultaneous onsets. In the “adjacent” condition, the onsets of the bursts were delayed by 50 ms (the duration of the preceding burst), while in the “delayed” condition, the onsets of the bursts were delayed by 200 ms (the duration of the preceding burst plus 150 ms). For each of the three conditions with multiple bursts, either the burst to which the increment was added was fixed for an entire adaptive track (low, middle, or high) or was randomly changed on each trial (“random”). In the sequential conditions with multiple bursts (adjacent and delayed), the order of the bursts was fixed (low, then middle, then high) in order to reduce uncertainty in the fixed target conditions. This leads to a potential confound of frequency and order effects, and thus the second condition (simultaneous presentation) was included to examine the effects of frequency alone.

It was hypothesized that the uncertainty introduced by randomizing the target burst would have the greatest effect for conditions 3 and 4, where both temporal and frequency uncertainty were introduced. Listeners were expected to require larger increments to reliably detect that a change had occurred when the target was random, and the impacts were expected to be greatest for the older listeners.

Fig. 1. Panels show average ITs for five listeners younger than 37 years (filled circles) and seven listeners older than 50 years (unfilled circles) in condition 1 [single burst, panel (A)], condition 2 [simultaneous bursts, panel (B)], condition 3 [50-ms delay between onsets, adjacent, panel (C)], and condition 4 [200-ms delay between onsets, delayed, panel (D)]. Unfilled squares indicate the mean across all listeners for that condition.

3. Results

Mean DTs in dB SPL were 20.3, 15.8, and 24.5 dB SPLs for low, middle, and high frequency bursts, respectively. All listeners had individual thresholds between 3 and 36 dB SPL. These values are consistent with the audiometric thresholds for these listeners after conversion to dB SPL. A mixed-measures analysis of variance (ANOVA)³ with age group as a between-subjects variable and frequency as a within-subjects variable showed that the differences among thresholds for the three frequencies were statistically significant ($F_{2,20}=6.07$, $p<0.01$), but there was no significant main effect of age group ($F_{1,10}=2.26$, $p=0.16$) or interaction ($F_{2,20}=0.68$, $p=0.52$). *Post-hoc* t-tests showed that the effects of frequency were due to the difference between the middle frequency burst and the high frequency burst.

The average IT in the single-burst condition was similar across noise bursts: 7.8 (low), 8.8 (middle), and 9.3 dB (high). Average performance across listeners is plotted as the large open squares in Fig. 1, panel (A), with the smaller filled circles indicating mean performance for the five younger listeners and the smaller open circles the seven older listeners. These values are higher than those reported for tonal stimuli (3.5 dB; Cousineau *et al.*, 2009) or even for roving-level noise bursts (4 dB; Heinz and Formby, 1999). A mixed-measures analysis of variance showed that the differences among bursts were statistically significant ($F_{2,20}=4.66$, $p<0.05$) but there was no significant main effect of age group ($F_{1,10}=2.26$, $p=0.164$) or interaction ($F_{2,20}=0.32$, $p=0.73$). *Post-hoc* t-tests showed that the effects of frequency were due to the significant differences ($p<0.05$) between the low burst and the middle and high bursts, while the middle and high did not differ significantly ($p>0.05$). Average IT for each burst was not significantly correlated with detection threshold (for all correlation analyses throughout, each subject contributed one pair of values per frequency; thus 12 points per correlation).

When the bursts were all presented simultaneously, with no differences in onset or offset, average IT for the low was 7.7 dB, while the middle was 11.5 dB and the high was 10.2 dB. Randomizing the target burst within an adaptive track resulted in an increment threshold of 11.6 dB. Performance is plotted in Fig. 1, panel (B). A mixed-measures analysis of variance revealed that the differences among the burst thresholds were statistically significant ($F_{1,8,18,1}=14.19$, $p<0.001$) but there was no significant main effect of age group ($F_{1,10}=0.07$, $p=0.80$) or interaction ($F_{1,8,18,1}=0.34$, $p=0.70$). *Post-hoc* t-tests showed that the thresholds of all of the frequency bursts differed significantly from each other ($p<0.05$), with the exception of the middle burst, which was not significantly different from the high burst or the random frequency condition ($p>0.05$). Average IT for each burst was not significantly correlated with detection threshold. This suggests that frequency range alone had an impact on performance when all three bursts were presented simultaneously in each interval. In the absence of a clear model distinguishing sequential and simultaneous masking (which are known to differ in important ways), it is still useful

to consider the two as simply additive in order to reveal the minimum changes in performance that are likely to be due to the onset differences alone. This analysis is applied below.

When all three bursts were presented on each interval, delayed in onset by the duration of the preceding burst so that onsets and offsets were adjacent, average IT for the low was 10.0 dB, while the middle was 14.3 dB and the high was 14.8 dB. Randomizing the target burst resulted in an increment threshold of 13.0 dB. Performance for the younger and older listeners is plotted in Fig. 1, panel (C). A mixed-measures analysis of variance revealed that the differences among the burst thresholds were statistically significant ($F_{3,30}=8.82$, $p<0.001$) but there was no significant main effect of age group ($F_{1,10}=2.19$, $p=0.18$) or interaction ($F_{3,30}=0.71$, $p=0.55$). *Post-hoc* t-tests showed that the thresholds of all of the frequency bursts differed significantly from each other ($p<0.05$), with the exception of the middle burst, which was not significantly different from the high burst or the random frequency condition ($p>0.05$), and the random frequency threshold, which did not differ from the thresholds for the middle or high frequency bursts ($p>0.05$). Average IT for each burst was not significantly correlated with detection threshold. Even assuming that the changes in threshold found in the simultaneous condition would have occurred in the burst conditions, these ITs represent increases of 2–4.5 dB for each burst that were a direct result of the sequential presentation. Thus, delaying the onsets of the bursts *hurt* performance, even when the target burst was fixed for the entire adaptive run.

When all three bursts were presented on each interval delayed by 200 ms (i.e., creating a 150 ms gap between bursts), average IT for the low was 8.4 dB, while the middle was 14.8 dB and the high was 13.0 dB. Randomizing the target burst resulted in an increment threshold of 11.8 dB. Performance is plotted in Fig. 1, panel (D). A mixed-measures analysis of variance revealed that the differences among the burst thresholds were statistically significant ($F_{1,7,16.9}=10.67$, $p<0.001$) but there was no significant main effect of age group ($F_{1,10}=0.25$, $p=0.63$) or interaction ($F_{1,7,16.9}=0.79$, $p=0.45$). *Post-hoc* t-tests showed that the thresholds for the low, high, and random frequency bursts did not differ significantly from each other ($p>0.05$), but that the middle burst was significantly higher than the low and random burst ($p<0.05$). The middle and high burst thresholds were not significantly different ($p>0.05$). Average IT for each burst was not significantly correlated with detection threshold. As with the previous data, even assuming that these ITs represent the additive combination of frequency and order effects, they represent increases of 2–3 dB for each burst that were a direct result of the sequential presentation, suggesting that delaying the onsets of the bursts by 200 ms did not remove the impacts of sequential presentation shown in condition three.

As a final statistical test of the differences among conditions, the ITs for conditions 2–4 were entered into a repeated-measures ANOVA with condition and frequency as within-subject variables. The main effect of condition was significant ($F_{2,22}=14.50$, $p<0.001$) as was the main effect of frequency ($F_{1,99,21.94}=15.33$, $p<0.001$) and the interaction ($F_{6,66}=3.965$, $p<0.01$). *Post-hoc* t-tests showed that the simultaneous and delayed conditions differed significantly from each other ($p<0.01$), but that the 50- and 200-ms delays were not significantly different from each other ($p=0.23$). The random frequency ITs were significantly different from the low burst ITs ($p<0.01$), but not from the middle or high ITs. The low ITs were significantly lower than the middle and high burst ITs ($p<0.05$), but the middle and high were not different from each other ($p=1.00$).

4. Discussion

The hypothesis that listeners have a reduced ability to make intensity judgments about targets embedded in a temporal sequence is supported by these results. The detrimental effect of presenting three bursts and asking listeners to judge only one of them was shown in the absence of temporal variation (in condition 2), but this could not account for all of the results. The impacts of age and uncertainty reported previously (e.g., Salthouse, 1992; Watson, 2005; Kidd *et al.*, 2008) were not found for this particular combination of listeners and conditions. Previous data

(Hafter *et al.*, 1998) have shown that a sensory-trace representation appears to be resistant to the costs of dividing attention, and listeners in this experiment may have been relying on a sensory-trace representation, which reduced the effects of age and uncertainty.

Thresholds for the single-burst stimuli were much greater than usually reported, even for noise stimuli roved in level (8 dB rather than the 4 dB shown by Heinz and Formby, 1999). One possibility is that listeners made level judgments of the second stimulus as a group rather than comparing the levels of the individual bursts to memories of the first presentation. A simple model simulating the impact of roving on ITs based on an overall loudness strategy predicts single-burst thresholds of roughly 8 dB. This model also predicts a reduction in performance for the multiple burst conditions and no impact of randomizing the target. However, it predicts no difference based on whether the multiple bursts are presented sequentially or simultaneously, and so cannot account for these results. The finding that performance was not worse in the simultaneous case, where subjects would have been more likely to hear all three bursts as a single object with a single intensity, suggests that auditory grouping might actually be able to reduce interference for stimuli like these.

Sequential processing deficits have primarily been modeled in the visual domain, where the suggestion is that sequential presentation impairs processing due to an inherent time-dependence of the underlying short-term memory consolidation process (Vogel and Luck, 2002; Tremblay, *et al.*, 2005; Vachon and Tremblay, 2005). Essentially, early elements of a to-be-remembered sequence are encoded accurately, but later elements must wait in a sensory “buffer” where they degrade in a time-dependent manner until they can be encoded. The general pattern of results described here, in which the first element was more accurately discriminated than the later elements, is consistent with the general predictions of such a model. One potential explanation for the lack of an effect of onset timing is that the integration time into short-term memory is on the order of 200–600 ms (as suggested by Vogel and Luck, 2002), and that the delays used were simply too short to show a release from temporal interference.

Acknowledgments

This work was supported by the Department of Veterans Affairs and by the NIH. Stephen Fausti, Anna Diedesch, Michelle Molis, Marjorie Leek, Chris Mason, Virginia Best, Gerald Kidd, and Barbara Shinn-Cunningham provided helpful comments and guidance. Matthew Marble, Kelly Reavis, and Marc Caldwell helped with data collection. The greatest debt is to the participants.

References and links

¹Level difference is reported as the decibel change in the overall level of the stimulus ($[10 \log(\Delta I + I)] - 10 \log I$) that would result from a given increment in intensity (ΔI) being added to a standard level (I). This measure has been shown to be very compressive for small values of ΔI (Green, 1988), despite the good detectability of such increments in some paradigms, leading to the use of the ratio of increment to standard ($10 \log \Delta I / I$) when small increments are being detected. As these two units are approximately linearly related for changes greater than 3 dB (Green, 1988), the more familiar measure has been used here.

²Frequency response of the headphones was not flat, but the overall level of each stimulus was digitally calibrated prior to amplification and attenuation such that the rms level of each stimulus was equated at the output of the headphone speaker. Any variations in the relative levels of frequency components within each stimulus were allowed to remain, since randomization of component phase introduces such variations in effective level.

³All mixed-measures tests were corrected for significant violations of the assumption of sphericity (when they occurred) by adjusting the degrees of freedom using a Greenhouse–Geisser correction. All *post-hoc* tests were Bonferroni corrected.

ANSI (2004). *Methods for manual pure-tone threshold audiometry* (ANSI S3.21-2004), American National Standards Institute, New York, NY.

Cousineau, M., Demany, L., and Pressnitzer, D. (2009). “What makes a melody: The perceptual singularity of pitch sequences,” *J. Acoust. Soc. Am.* **126**, 3179–3187.

Gordon-Salant, S., and Fitzgibbons, P. J. (1999). “Profile of temporal processing in older listeners,” *J. Speech Lang. Hear. Res.* **42**, 300–311.

Green, D. M. (1988). *Profile Analysis* (Oxford University Press, New York).

- Haft, E. R., Bonnel, A.-M., Gallun, E., and Cohen, E. (1998). "A role for memory in divided attention between two independent stimuli," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis, (Whurr, London), pp. 228–238.
- Heinz, M. G., and Formby, C. (1999). "Detection of time- and bandlimited increments and decrements in random-level noise," *J. Acoust. Soc. Am.* **106**, 313–326.
- Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 169–177.
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 143–190.
- Leek, M. R., and Watson, C. S. (1984). "Learning to detect auditory pattern components," *J. Acoust. Soc. Am.* **76**, 1037–1044.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- McGill, W. J., and Goldberg, J. P. (1968). "Pure-tone intensity discrimination and energy detection," *J. Acoust. Soc. Am.* **44**, 576–581.
- Salthouse, T. A. (1992). "Influence of processing speed on adult age differences in working memory," *Acta Psychol.* **79**, 155–170.
- Tremblay, S., Vachon, F., and Jones, D. M. (2005). "Attentional and perceptual sources of the auditory attentional blink," *Percept. Psychophys.* **67**, 195–208.
- Vachon, F., and Tremblay, S. (2005). "Auditory attentional blink: Masking the second target in necessary, delayed masking is sufficient," *Can. J. Exp. Psychol.* **59**, 279–286.
- Vogel, E. K., and Luck, S. J. (2002). "Delayed working memory consolidation during the attentional blink," *Psychon. Bull. Rev.* **9**, 739–743.
- Watson, C. S. (2005). "Some comments on informational masking," *Acta Acust. Acust.* **91**, 502–512.
- Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (1975). "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals," *J. Acoust. Soc. Am.* **57**, 1175–1185.
- Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (1999). "Regaining lost time: Adult aging and the effect of time restoration on the recall of time-compressed speech," *Psychol. Aging* **14**, 380–389.