

The salience of enhanced components within inharmonic complexes (L)

Andrew J. Byrne,^{a)} Mark A. Stellmack, and Neal F. Viemeister
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 30 April 2013; revised 23 August 2013; accepted 26 August 2013)

A subjective listening task was used to measure the salience of enhanced components using typical intensity-enhancement stimuli, time-reversed versions of those stimuli, and stimuli which contained a frequency shift of the target component. Twenty-five listeners judged whether or not a pitch “stood out” within an inharmonic complex. For comparison, judgments also were made for stimuli with a single segment that consisted of a simultaneously masked target. The results indicate that the perceived salience of enhanced components is greater than might be predicted by the effective magnitude of those components, and that informational masking is likely involved.

© 2013 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4820897>]

PACS number(s): 43.66.Mk [LD]

Pages: 2631–2634

I. INTRODUCTION

The detection of a “target” tone simultaneously masked by an inharmonic complex made up of flanking component frequencies improves when a precursor, consisting of the masker alone, is presented first (e.g., Viemeister, 1980). Based on our observations, when the target is near threshold, it often is perceived as standing out from the masking complex when the precursor is included; however, without the precursor, the target often is perceived as part of the larger complex and not salient as an individual component.

Three explanations have been posited for this intensity enhancement effect. First, perceptual grouping could combine the precursor and the masker as one “object,” while the target component stands out as a result of its delayed onset (e.g., Carlyon, 1989). Second, the masking components, being presented first, could undergo adaptation (produce decreasing excitation over time) and therefore have a weaker simultaneous masking effect (e.g., Viemeister, 1980). Third, inhibition produced by the masker components on the target frequency may adapt, and, if so, the target would produce greater excitation than when the masker is able to fully suppress the neural response to the target (e.g., Viemeister and Bacon, 1982). Although all three factors may contribute to the enhancement effect in various experimental paradigms, at present, adaptation of inhibition appears to be the best account for intensity enhancement across methodologies (Byrne *et al.*, 2011).

The amount of enhancement often is defined as the difference between detection thresholds measured in enhanced and unenhanced conditions. One problem with quantifying enhancement in this way is that the effect may persist across intervals or trials within a typical objective task (e.g., Cao *et al.*, 2009). In a two-interval, forced-choice task, when an interval does not contain the target component, that interval could act as a precursor for a subsequent interval containing the target, and therefore in an “unenhanced” condition the target can be effectively enhanced. To have a truly unenhanced condition, the delay between presentations of stimuli would

need to be on the order of several seconds or involve very limited exposure to the target-absent stimuli, as in the single-interval task of Viemeister (1980). The persistence of enhancement may also explain the presence of the “backward” enhancement effect seen by Kidd and Wright (1994) using time-reversed intensity enhancement stimuli.

Other work has shown that auditory enhancement may not be limited to the intensity enhancement described above. Instead of adding a target component to the masker with some onset asynchrony, the target is gated simultaneously with the masker but then is slightly shifted in frequency when included within a second presentation of the masker. As a result of this shift, the target is more salient and allows the listener to use it for pitch comparisons (frequency shift enhancement; Erviti *et al.*, 2011; Demany *et al.*, 2013). These experiments suggest the presence of frequency shift detectors (Demany and Ramos, 2005) that appear to have the same perceptual effect that adaptation of inhibition may have in intensity enhancement paradigms.

The present experiment used a subjective task to measure listeners’ judgments of the salience of a single component within inharmonic complexes. A variety of stimulus configurations were used to evaluate the differences in salience of enhanced target components to those that were unenhanced. In contrast to typical objective paradigms like those described above, both the frequencies of the tonal complexes and different stimulus conditions within a block of trials were randomized resulting in a high degree of stimulus uncertainty.

II. METHODS

A. Stimuli

The stimuli of the present experiment were based on those previously used by Byrne *et al.* (2011), where the target (enhanced) component was a 2-kHz pure tone, while the simultaneous masker and precursor stimuli were identical inharmonic complexes consisting of tones from 1149 to 3482 Hz with 0.1-octave spacing between components. These inharmonic complexes had five components removed symmetrically (in logarithmic frequency) around a center frequency of 2 kHz leaving a 0.6-octave notch, thus creating complexes with six components above and six below the

^{a)}Author to whom correspondence should be addressed. Electronic mail: byrn0050@umn.edu

target frequency. All components, including the target, were presented at 50 dB sound pressure level (SPL), and the phase of each component was randomized for every trial.

The major deviation from the stimuli of [Byrne et al. \(2011\)](#) is that, in the present experiment, the tonal complex was roved in frequency. In each single-interval trial, all components in the stimulus were uniformly roved (rectangularly on logarithmic frequency) over a 2-octave range, i.e., the target could take on any value from 1 to 4 kHz. Due to this frequency rove, it would not be possible to listen for the presence of one specific frequency within the stimulus and persistence of enhancement across trials presumably would be minimized.

The intensity enhancement conditions (labeled “FWD” in the figure, for “forward” intensity enhancement) consisted of a 250-ms precursor (the masker alone) followed by a 250 ms masker-plus-target segment. Each segment was gated with 20-ms raised-cosine, on-and-off ramps. For the FWD condition, the two segments were temporally contiguous such that there was no silence between precursor and masker other than that due to the gating, while for the FWD:GAP condition, a 250-ms silent delay was added between the segments. This latter condition was included because increasing the precursor-signal delay has been shown to reduce the effect of the precursor (e.g., [Viemeister, 1980](#)).

In the preceding two conditions, the stimuli were presented monaurally to the left ear. Given that presenting the precursor contralaterally results in a greater reduction for intensity enhancement than for frequency enhancement ([Ervti et al., 2011](#)), in a third condition (with a 0ms precursor-masker delay) the first segment was moved to the right ear, creating a contralateral precursor (FWD:CONTRA). [Refer to [Carcagno et al. \(2012\)](#) for more on contralateral enhancement.] Because of the possibility of backward enhancement ([Kidd and Wright, 1994](#)), time-reversed versions of the FWD conditions were used, resulting in a masker-plus-target segment followed by a postcursor segment that did not contain the target (BWD). In addition, in separate control conditions, stimuli that did not contain the target were presented to obtain baseline measures of listeners’ tendencies to report a salient component within the masking complexes alone.

The effect of including a frequency shift across segments ([Ervti et al., 2011](#); [Demany et al., 2013](#)) was measured using conditions with a semitone change to the target frequency. In the precursor segment, a 2119-Hz component was added, while in the second segment, the 2119-Hz component was replaced by a 2-kHz tone. The stimuli were then roved over the same range as the intensity enhancement stimuli. The appropriate controls for these frequency shift enhancement (FSE) conditions were not, strictly speaking, “target absent,” but instead had no frequency change present. Both segments in these conditions were identical and were comprised of the masker plus the target component.

All of the preceding conditions were used to estimate the salience of a single tonal component that was either added, removed, or shifted in frequency between two otherwise identical inharmonic complexes. For comparison, additional conditions were run with a single temporal segment in which the target component was gated on and off with the masker, but the level of the target component was varied

(across conditions). The idea was to determine how much an unenhanced target component must be increased in level to be as salient as an enhanced component. In these masked conditions (MSK), unlike the previous stimuli with all components presented at 50 dB SPL, the target component was presented either at 50 dB SPL (MSK:T) or at 55, 60, 65, or 70 dB SPL (MSK:T+5 dB through MSK:T+20 dB). An additional condition with the target absent served as a control for all five target-present conditions.

B. Procedure and apparatus

The salience of the target component within the various stimuli was measured using a subjective task. Prior to participating in the experiment, the general (and purposefully minimal) description of the task given to the listeners was that they would hear various complex sounds (a sample of the MSK stimuli was presented to them), and that they should judge whether or not a particular pitch stood out, or seemed to pop-out, within those sounds. The stimuli were generated digitally with MATLAB on a personal computer equipped with a 24-bit sound card. Stimuli were presented over stereo headphones (Sony MDR-V6) while listeners were seated in a sound-attenuating chamber.

Each trial began with the presentation of the stimulus for a given condition. The listeners could then replay that exact stimulus by clicking a button on the graphical user interface (GUI); however, they were instructed to replay the stimulus only when necessary (e.g., after a lapse of attention), not repeatedly in an attempt to hear out a pitch that was not initially salient. Displayed on the GUI after each trial was the question: “Did one pitch seem to stand out within the sound?” along with two buttons labeled “Yes” and “No.” After the listener responded to the question, that trial was complete and the next began 1 to 2 s later. No feedback was given, and, with the exception of the authors, the listeners were not aware of the various types of conditions presented. Each of the 21 conditions described above was presented within a set of trials, but the order of the conditions was randomized within each new set. Testing was complete after a total of 20 trials per condition were obtained (typically a single 1-h session).

C. Listeners

Twenty-five normal hearing listeners participated in the experiment (9 males and 16 females, ages 18–48). Two were the first and second authors, while the others were students and staff from the University of Minnesota who were paid to participate. Five of the listeners had extensive experience with psychoacoustical tasks, including various enhancement experiments, while the other 20 listeners had no previous experience or training.

III. RESULTS

The average salience judgments of the 25 listeners are shown in Fig. 1. In the upper panel, the percent of trials that the listeners designated as having a salient pitch present are shown for the target-present conditions (light gray bars), as well as for the target-absent control conditions (dark gray bars). Each listener’s percentage of “salient” responses was

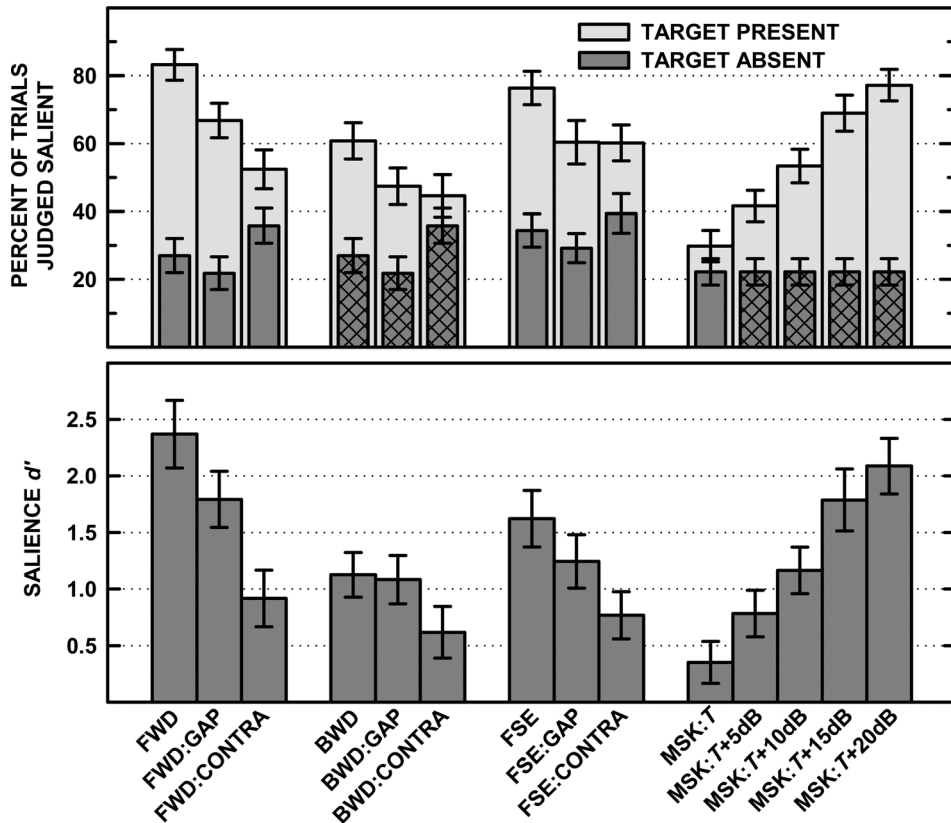


FIG. 1. The average results across 25 listeners for each condition (see text for condition descriptions), with the error bars representing the standard errors of the means. Upper panel: The percentage of trials which the listeners designated as having a salient pitch. The light gray bar condition in each pair included the target tone, while for the darker bar the target was absent (crosshatched bars indicate duplicated target-absent conditions, see text). Lower panel: A target salience metric (d') using the target-present values for each listener as the hit rates and the target-absent values as the false alarm rates.

calculated from the 20 trials of each condition, and those percentages were then averaged across listeners to obtain the means and standard errors shown.

In the bottom panel of Fig. 1, a d' -like salience metric was calculated using each individual listener's target-present percentage values as the hit rates, while the target-absent values were considered as the false alarm rates for this d' calculation. (In cases where the rates were either 1 or 0, they were adjusted to 0.99 or 0.01, respectively.) Finally, the individual d' results were averaged across listeners. The d' measure seems intuitively more appealing than the salience responses, because the d' measure takes into account the differences in the responses across stimuli configurations when the target component was not even present. (Note the differences between the dark gray bars of the upper panel.)

The mean salience d' value for the FWD condition (left-most bar) is 2.37 [standard error (S.E.)=0.30] and is the highest value of all the conditions, while the BWD and FSE conditions had mean values of 1.13 (S.E. = 0.20) and 1.62 (S.E. = 0.25), respectively. When the conditions included a 250-ms gap between the segments, the salience d' decreased, while the CONTRA conditions had the lowest value for each set of three conditions. To support these observations, a 3×3 repeated-measures analysis of variance was performed for the dependent variable of salience d' . There were significant differences between enhancement types (FWD, BWD, and FSE), $F(2,48) = 11.0, p < 0.001$ and sub-condition types [(ipsilateral precursor and no gap), GAP, and CONTRA], $F(2,48) = 13.4, p < 0.001$, as well as a significant interaction, $F(4,96) = 3.2, p = 0.02$.

Post hoc comparisons with the Bonferroni correction for multiple comparisons were performed and confirmed that

the FWD condition showed greater enhancement than the FSE [$t(24) = 2.87, p = 0.009$]. With contralateral precursors, there was not a significantly greater reduction of intensity enhancement compared to frequency enhancement [(FWD FWD:CONTRA) vs (FSE - FSE:CONTRA): $t(24) = 2.05, p = 0.023$], an effect which was significant in the results of [Erviti et al. \(2011\)](#), but could suggest adaptation occurring more centrally in the auditory system ([Carcagno et al., 2013](#)). Finally, the FWD:GAP condition had a significantly higher d' than the BWD:GAP [$t(24) = 5.91, p < 0.001$], consistent with what could be predicted from [Cao et al. \(2009\)](#).

As would be expected, the MSK conditions yielded increasing d' values as the intensity of the target increased. One can then estimate the effect of enhancement by comparing the d' values of the various enhanced conditions to those of the unenhanced (MSK) conditions. For instance, in the MSK:T+20 dB condition, in which the target was 20 dB greater than the components in the masker, the target is still not quite as salient as the target of the FWD condition where all components are at equal amplitude. In other words, the salience of the target in the FWD condition was greater than that of the MSK target, despite the fact that the MSK target was actually 20 dB more intense. After performing a linear regression on the salience values of the MSK conditions ($r^2 = 0.991$), it was estimated that the level of the unenhanced target would need to be raised by 23, 9, and 14 dB to equal the salience of the FWD, BWD, and FSE conditions, respectively.

IV. DISCUSSION

Although the present experiment was purely subjective, the salience values obtained may be a better measure of the

perceptual “pop-out” effect from auditory enhancement than typical objective detection tasks. These effects were observed with a few highly trained listeners as well as with many other participants with no prior experience with psychoacoustical tasks or enhancement stimuli. Given that it was a subjective task, there was considerable variability between the results of individual listeners; however, the trends shown in Fig. 1 were generally observed in most of the listeners. Some seemed to have a strict criterion for saying that a pitch was salient, and thus reported hearing a salient pitch on a low percentage of trials, while others apparently had more lax criteria and high salience responses overall.

Since the enhancement effect has been shown to produce an increase in the effective level of the target component (e.g., Viemeister and Bacon, 1982), it was reasonable to question whether the higher salience responses of the FWD conditions could have resulted from the target being at an effectively higher level. Supplemental experimentation on five of the 25 listeners examined the magnitude of this increased gain using a typical objective task. Two-interval, forced-choice (target) detection thresholds were measured for the FWD and MSK stimuli without roving frequency (i.e., the masker frequencies were fixed and the target was always a 2-kHz tone). [The threshold procedure was an exact replication of the “signal enhancement” task used in Byrne *et al.* (2011).] The difference in detection thresholds between these enhanced (FWD) and unenhanced (MSK) conditions was only 2.3 dB (S.E. = 0.4).

When using comparable stimuli, the magnitude of the intensity enhancement effect, and presumably the increase in the effective target level, is similar across different methodologies and actual (dB SPL) target levels (Byrne *et al.*, 2011); therefore, the magnitude measured with the detection task is inconsistent with the much larger effect seen in the salience results. Although this direct comparison between two very different tasks is probably inappropriate, the magnitude discrepancy could be partially due to the rove used for the salience stimuli. When the same rove is applied to the detection task (the complexes are roved in frequency independently on each interval), the enhancement effect is increased to 14.3 (S.E. = 1.6) for those five listeners, closer to the magnitude seen in the salience measures.

The effect of the rove [FWD thresholds raised by 10.6 dB (S.E. = 0.7) compared to 22.6dB (S.E. = 1.7) for the MSK thresholds] could be partially attributed to different frequency regions affecting the magnitude of enhancement (Carcagno *et al.*, 2013), the absence of enhancement persisting across intervals and trials, or possibly to stimulus uncertainty and informational masking. The relative contribution of these factors is unfortunately difficult to determine from the present methodology and data.

As suggested by Kidd *et al.* (2011), when stimulus uncertainty is involved in an enhancement paradigm, the precursor allows a spectral comparison between the two segments of the stimulus (see also Richards and Neff, 2004). Such a cue is not available in the unenhanced conditions; therefore, stimulus uncertainty is greater. Across-segment comparisons may partially account for the salience responses

of the typical enhancement stimuli (FWD) as well as the presence of enhancement in conditions where adaptation-of-inhibition should not be as effective, e.g., the BWD conditions. [For the BWD conditions, a salient pitch at the edge of the spectral notch (using harmonic complexes; Hartmann and Goupell, 2006) may have also somewhat affected the salience responses.]

In summary, a single target tone simultaneously masked by an inharmonic complex can be enhanced and made more salient through the use of different spectral and temporally based enhancing features. Without such features, the target tone would need to be much more intense to be rated salient as often. The results also indicate that enhancement can be affected by stimulus uncertainty in addition to changes in effective target level that may result from adaptation of inhibition. Finally, the use of naive listeners, roved-frequency complexes, and a randomized assortment of different conditions likely provides the best estimate of the large *practical* magnitude of enhancement on the salience of sounds in everyday listening situations.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute on Deafness and Communication Disorders of the National Institutes of Health under award number R01DC00683. The authors would like to thank Dr. Laurent Demany who provided helpful comments and suggestions for improving this manuscript.

- Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (2011). “The enhancement effect: Evidence for adaptation of inhibition using a binaural centering task,” *J. Acoust. Soc. Am.* **129**, 2088–2094.
- Cao, X., Huang, R., and Richards, V. M. (2009). “Sequential effects on the detectability of a tone added to a multitone masker,” *J. Acoust. Soc. Am.* **125**, EL20–EL26.
- Carcagno, S., Semal, C., and Demany, L. (2012). “Auditory enhancement of increments in spectral amplitude stems from more than one source,” *J. Assoc. Res. Otolaryngol.* **13**, 693–702.
- Carcagno, S., Semal, C., and Demany, L. (2013). “No need for templates in the auditory enhancement effect,” *PLoS ONE* **8**, e67874.
- Carlyon, R. P. (1989). “Changes in the masked thresholds of brief tones produced by prior bursts of noise,” *Hear. Res.* **41**, 223–236.
- Demany, L., Carcagno, S., and Semal, C. (2013). “The perceptual enhancement of tones by frequency shifts,” *Hear. Res.* **298**, 10–16.
- Demany, L., and Ramos, C. (2005). “On the binding of successive sounds: Perceiving shifts in nonperceived pitches,” *J. Acoust. Soc. Am.* **117**, 833–841.
- Erviti, M., Semal, C., and Demany, L. (2011). “Enhancing a tone by shifting its frequency or intensity,” *J. Acoust. Soc. Am.* **129**, 3837–3845.
- Hartmann, W. M., and Goupell, M. J. (2006). “Enhancing and unmasking the harmonics of a complex tone,” *J. Acoust. Soc. Am.* **120**, 2142–2157.
- Kidd, G., Jr., Richards, V. M., Streeter, T., Mason, C. R., and Huang, R. (2011). “Contextual effects in the identification of nonspeech auditory patterns,” *J. Acoust. Soc. Am.* **130**, 3926–3938.
- Kidd, G., Jr., and Wright, B. A. (1994). “Improving the detectability of a brief tone in noise using forward and backward masker fringes: Monotic and dichotic presentations,” *J. Acoust. Soc. Am.* **95**, 962–967.
- Richards, V. M., and Neff, D. L. (2004). “Cueing effects for informational masking,” *J. Acoust. Soc. Am.* **115**, 289–300.
- Viemeister, N. F. (1980). “Adaptation of masking,” in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Noordwijkerhout, The Netherlands), pp. 190–199.
- Viemeister, N. F., and Bacon, S. P. (1982). “Forward masking by enhanced components in harmonic complexes,” *J. Acoust. Soc. Am.* **71**, 1502–1507.