Hearing Research 255 (2009) 14–21

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research papers

Effects of recurrent tonal information on auditory working memory for pitch

Dennis T. Ries [a,*], Jeffrey J. DiGiovanni [b]

[a] Auditory Perception Laboratory, W241 Grover Center, School of Hearing, Speech, and Language Sciences, Ohio University, Athens, OH 45701, USA
[b] Auditory Psychophysics and Signal Processing Laboratory, W151A, School of Hearing, Speech, and Language Sciences, Ohio University, Athens, OH 45701, USA

ARTICLE INFO

Article history:
Received 16 January 2009
Received in revised form 1 May 2009
Accepted 1 May 2009
Available online 10 May 2009

Keywords:
Frequency
Pitch
Difference limen
Auditory working memory

ABSTRACT

This study ascertained the influence of repeating pitch information within an intervening tonal sequence upon the extent of interference for a pitch standard held within auditory working memory as measured by the difference limen for frequency (DLF). Standard and comparison tones were presented to subjects and same/different responses were obtained using a touch screen monitor and the DLF was measured using single interval adjustment matrix (SIAM) procedure [Kaernbach, C., 1990. A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing. J. Acoust. Soc. Am. 88, 2645–2655]. Estimates of the DLF were obtained in a control condition with a silent inter-comparison interval and three conditions containing intervening tones within the temporal gap between the standard and comparison stimuli. The presence of intervening stimuli produced a significant increase in the DLF when the intervening tonal sequence contained tones with pitches that differed from that of the standard (Int condition) as well as when the sequence contained a tone with a pitch identical to that of the comparison (RptCmp condition). Further, the DLFs obtained for RptCmp condition were significantly higher than those measured in the Int condition. The DLFs measured in the condition where the pitch of an intervening tone was identical to the standard were significantly lower than those for the Int and RptCmp conditions, but did not differ from the DLFs for the control condition. These results indicate that either a release from or an increase in interference in auditory working memory for pitch can occur dependent upon the frequency relationships between of the standard, comparison, and intervening tones.

1. Introduction

Tones that appear within an inter-comparison interval (intervening tones) can affect pitch memory (Bull and Cuddy, 1972; Deutsch, 1970a,b, 1972a,b, 1974, 1975; Deutsch and Feroe, 1975; Ries and DiGiovanni, 2007; Ross et al., 2003, 2004; Elliot, 1970; Massaro, 1970; Semal and Demany, 1991, 1993; Wickelgren, 1966, 1969). Musical training (Pechmann and Mohr, 1992; Berti et al., 2006) as well as lateralization/localization cues (Deutsch, 1978; Ries and DiGiovanni, 2007; Kallman et al., 1987) can lessen the degree of interference within auditory working memory (AWM). Listeners who possess absolute pitch also are less affected by the presence of intervening interference in AWM for pitch tasks (Ross et al., 2003, 2004). Introduction of the pitch of the comparison or repetition of the pitch of the standard within the intervening tonal sequence, however, can influence one's ability to identify pitch differences correctly (Deutsch, 1972b). Sound lateralization cues that differ between the standard/comparison pair and the intervening tones produce a significant decrease in the difference limen for frequency (DLF; Ries and DiGiovanni, 2007) as well as the pitch categorization (same/different) error rate (Deutsch, 1978). Similarly, the DLF for the standard/comparison pair may well decrease or increase when the pitch of the standard or comparison appears within the intervening tonal sequence.

The frequency relationship between the intervening tones and the standard can have a negative or positive influence upon the retention of the pitch of the standard within AWM (Deutsch, 1972a,b). One manner in which this occurs is dependant upon the within octave relationship between the standard, comparison, and intervening tone(s). Deutsch (1972a) found that the percent error in same/different pitch comparisons increased systematically as the pitch of the second tone in an intervening sequence of six tones approached two-thirds tone separation from the pitch of the standard (e.g., if standard = note C (259 Hz), the 2nd intervening

Abbreviations: ANOVA, analysis of variance; AWM, auditory working memory; DLF, difference limen for frequency; G–GA, Gerisser–Greenhouse Adjustment; Hz, Hertz; Int, condition containing intervening tones; MCS, method of constant stimuli; ms, milliseconds; NoInt, condition containing no intervening tones; RptCmp, condition containing an intervening tone with a frequency identical to that of the comparison tone; RptStd, condition containing an intervening tone with a frequency identical to that of the standard tone; SIAM, single interval adjustment matrix

* Corresponding author. Address: School of Hearing, Speech and Language Sciences, W221 Grover Center, Ohio University, Athens, OH 45701, USA. Tel.: +1 740 593 1420; fax: +1 740 593 0287.
E-mail addresses: ries@ohio.edu (D.T. Ries), digiovan@ohio.edu (J.J. DiGiovanni).

0378-5955/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2009.05.002

tone = note G# (411 Hz). Likewise, repetition of the standard pitch within the intervening series decreases the number of errors in same/different pitch judgments whereas introduction of the comparison pitch in the series produces an increase in the number of errors (Deutsch, 1972b).

Deutsch (1972b) proposed a three-dimensional model of pitch retention in which the strength (i.e., amplitude) of the pitch percept (*z*-axis) for the standard within AWM was influenced by the passage of time (*y*-axis) as well as interaction with subsequent pitch percepts (*x*-axis). In this model, the strength of a pitch percept slowly decreases in amplitude as time passes, leading to a broadening of the bell-shaped distribution in the temporal and pitch domains. Spread of the distribution associated with decreasing pitch strength primarily occurs in the temporal domain, leading to a pitch domain distribution that decreases in height with a negligible broadening of the distribution along the pitch continuum. As such, accurate information regarding the pitch of a tone can be retained for relatively prolonged periods (e.g., Harris, 1952; Bachem, 1954), but with the associated cost of some temporal smearing. The model calls for summation of the overlapping portions of pitch distributions over time. That is, if the later occurring comparison tone has the same pitch as the standard tone, the values of the two pitch distributions at the moment in time in which the information is accessed would combine to produce a larger amplitude at a particular pitch location on the *x*-axis than would either alone.

According to Deutsch's (1972b) model, an increase in amplitude for a particular pitch value compared to the volumes (amplitude $\times$ frequency spread $\times$ temporal spread) associated with other areas along the *x*-axis (pitch value) at a given moment in time would indicate that the same pitch has occurred again. In this instance, the listener would indicate that these two tones had the same pitch. If the standard and comparison differ sufficiently in pitch, however, the volume of the distribution at the point along the pitch continuum which corresponds with the comparison tone would increase as that of the standard decreases and the two tones would be judged to differ in pitch. In addition, if a tone appears within the inter-comparison interval which has a pitch identical to that of standard, the amplitude value corresponding with the pitch of the standard would increase, leading to better discrimination relative to the comparison. However, if a tone appears within the inter-comparison interval with a pitch identical to that of the comparison, provided the pitch of the tone is near the discrimination boundary, the distribution associated with the standard will be pulled towards that of the comparison (i.e., broaden the distribution) with little associated increase in peak amplitude making discrimination more difficult. The percentage of same/different judgment errors reported by Deutsch (1972b) for standard and comparison tone pairs that differed by a fixed musical distance (i.e., one semitone) match well with the model predictions.

Earlier investigations of AWM for pitch and loudness from this lab (Ries and DiGiovanni, 2007; Jump and Ries, 2008) employed the method of constant stimuli (MCS) to measure the difference limen for correct identification of a change in frequency or intensity between a standard stimulus and a later occurring comparison stimulus as this procedure is readily adaptable to a single interval task. While this method provides for accurate assessment of the underlying psychometric function, it is regarded as inefficient with respect to adaptive staircase procedures (e.g., Levitt, 1971; Schlauch and Rose, 1990). Kaernbach (1990) devised a forced-choice staircase procedure for use in a single interval task (the single interval adjustment matrix or SIAM procedure), based upon the underlying tenets of signal detection theory, which produced absolute threshold measures comparable to those obtained using a two-interval forced-choice procedure and that effectively controlled for response bias. Although this method was introduced almost two decades ago, it has received little use in psychoacoustic studies. Only three studies of human audition have employed the SIAM procedure (Bernstein and Trahiotis, 2008; Bernstein et al., 2006; Leeuw and Dreschler, 1998), however, none of these studies measured differential sensitivity or AWM, focusing instead on tone detection in noise. Although limited in use in hearing research, this method is well regarded in discussions of psychometric procedures (Brown, 1996; Kaernbach, 1991; Klein, 2001; Meese, 1995). In this light, we adopted the SIAM procedure for use in the present study in order to make use of a more efficient data collection paradigm for differential sensitivity measures involving AWM and to provide some additional commentary and assessment this psychometric method of measurement.

The focus of this study was to investigate the influence of the repetition of standard pitch as well as introduction of the comparison pitch within the intervening tonal sequence on the DLF. This information serves to quantify the extent of disruption of an actively maintained pitch standard within AWM in relation to the pitch characteristics of other similar, later occurring tones. In addition, we used a variation of the single interval adjustment matrix (SIAM) procedure proposed by Kaernbach (1990) to obtain subject data and the results were compared to those obtained earlier (Ries and DiGiovanni, 2007) using the MCS in order to quantify the differences in DLF estimates obtained using these two procedures. Based on our previous results as well as the model and response error data provided in Deutsch (1972b), it was hypothesized that (1) the DLF would increase in the presence of intervening interference; (2) the repetition of the standard pitch within the intervening tonal sequence would result in at least a partial release from interference as evidenced by a decrease in the DLF; (3) introduction of the comparison pitch within the intervening tonal sequence would produce an increase interference as evidenced by an increase in the DLF; and (4) the SIAM procedure would be more time efficient while producing a pattern of results similar to that obtained previously using the MCS.

2. Materials and methods

2.1. Subjects

Eight people (2 males, 6 females; mean age = 26.6 years; age range = 23–37 years) recruited from the Ohio University student and staff population participated in this study. Seven of the eight subjects had prior experience in similar psychoacoustic experiments and all received at least 1 h of training prior to formal data collection. All subjects had pure-tone air conduction thresholds within normal limits, defined as 15 dB HL or less at the octave frequencies from 250 to 8000 Hz (Stewart and Downs, 1984). Six of the eight subjects had little (<1 year) or no musical training; subjects S3 and S4 had 6–7 years of formal musical training prior to entering college. Subjects were not screened for absolute pitch as: (1) the prevalence rate in the general population is less than 0.0001% (Profita and Bidder, 1988), (2) none of the subjects claimed to have this attribute and only one known case of absolute pitch in persons without extensive musical experience has been reported (Ross et al., 2003), (3) all were US citizens of mostly European decent that had little or no experience with tone languages (prevalence of absolute pitch may be higher in populations speaking tone languages, Deustch et al., 2004, 2006), and (4) the existence of this attribute in one or more of the subjects would have biased the results against finding a significant difference across at least two of the three conditions (Int & RptCmp, see Section 2.3) containing tones within the inter-comparison interval (Ross et al., 2003, 2004). Subjects provided informed consent before participating in the study and they received no compensation for their involvement. The study was approved for use of human subjects by the Institutional Review Board at Ohio University.

2.2. Stimuli

Stimuli were produced using System III hardware (Tucker Davis Technologies, Alachua, FL) controlled by a Pentium 4 computer (Dell, Round Rock, TX) running Matlab (The Mathworks, Natick, MA) and RPVD (Tucker Davis Technologies) software. The signals were verified electrically using a dynamic signal analyzer (Stanford Research Systems, Sunnyvale, CA) and digital storage oscilloscope (Tektronix, Richardson, TX), as well as acoustically using a sound level meter connected to a 2cc coupler (Bruel and Kjaer, Naerum, Denmark). All stimuli were presented binaurally to each subject via ER-2 insert earphones (Etymotic Research, Elk Grove Village, IL).

Stimuli (i.e., standard, intervening, and comparison) were tones of equal loudness with durations of 200 ms, including 20 ms cosine-squared onset and offset ramps, generated at a sampling rate of 24,414 Hz and low pass filtered at 12,000 Hz. The frequency of the standard tone (first tone in a trial) was roved randomly over an 80-Hz range centered at 435 Hz. The frequency of the comparison tone was set relative to that of the standard and the pedestal value was adjusted adaptively using the SIAM procedure (Kaernbach, 1990; see Section 2.4). The intervening stimuli, when present, were four individual tones selected randomly from a frequency range extending from 183 to 691 Hz such that the value selected corresponded to the notes of an equal tempered scale based on international pitch wherein the note A is equal to 435 Hz (Deutsch and Feroe, 1975). The levels producing equal loudness were based on loudness matching data collected from two highly trained listeners who have participated in a wide array of psychoacoustic experiments including multiple studies on loudness. The level of the 1000-Hz standard in the matching paradigm was 70 dB SPL. None of the subjects reported any notable or useful change in the loudness of the stimuli throughout the study. In addition, we found that a binaural-intensity difference of 12 dB between the level of the intervening tones and the standard/comparison pair resulted in no significant change in the DLF in our earlier study of memory for pitch (Ries and DiGiovanni, 2007).

The timing skeleton for the experiment is illustrated in Fig. 1. The offset of the standard tone (first tone presented on a given trial) was separated from the onset of the comparison tone (last tone presented on a given trial) by 4800 ms. The duration of the inter-comparison interval (i.e., 4800 ms) was selected in order to place the task within the retention timeframe of AWM (Clément et al., 1999). Four intervening tones were presented in three of the four conditions; no intervening tones were presented in the remaining condition. The onset of the first tone in the intervening sequence occurred 1000 ms after the offset of the standard tone with 300 ms inter-stimulus intervals placed between the four tones in the sequence and a 2100 ms interval placed between the offset of the last intervening tone and the onset of the comparison tone.

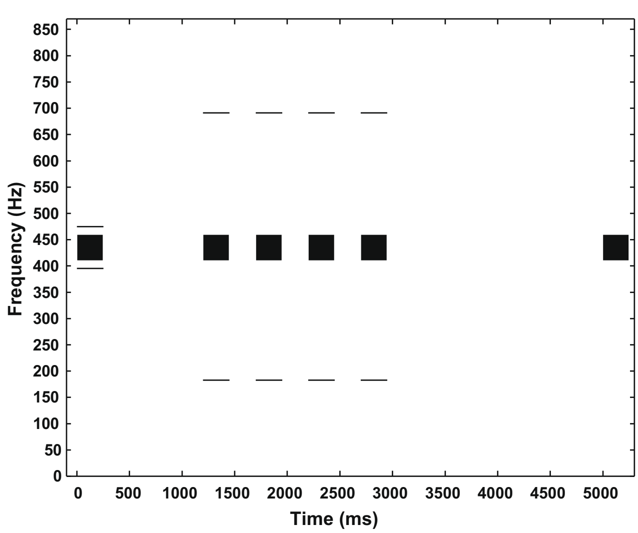

Fig. 1. The timing skeleton depicting the temporal location of the standard, comparison, and the four intervening stimuli. Intervening stimuli (middle four squares) were not present in the NoINT condition. The squares are located at the center placement along the frequency range whose extreme extent in either direction is represented by the dashed lines above and below the squares. The last square (i.e., the comparison tone) is not bounded by these lines as the frequency value of this stimulus was adjusted adaptively relative to the value for the first square (i.e., the standard tone).

2.3. Conditions

The experiment contained four conditions. Condition one (here after referred to as NoInt for no intervening tones) contained no stimuli in the inter-comparison interval and served as the baseline for comparison with the remaining conditions, all of which contained four intervening tones. Condition two (hereafter referred to as Int for intervening tones) contained intervening tones within the inter-comparison interval. In condition three (hereafter referred to as RptStd for repeated standard), the pitch of the second intervening tone (third tone presented in a trial) was set to that of the standard as this configuration produced a sizable reduction in the number of correct identification errors produced with respect to the Int condition in Deutsch's study (1972b). In condition four (hereafter referred to as RptCmp for repeated comparison), the pitch of the second intervening tone was set equal to that of the comparison as this produced a substantial increase in the number of correct identification errors with respect to the Int condition in Deutsch's study (1972b). The frequencies of the intervening tones, when present, were selected such that they did not equate with one another or with the value of the standard and comparison tones except as noted for the RptStd and RptCmp conditions.

2.4. Procedure

Each subject completed the study over several individual sessions while seated within a double walled, sound attenuating booth (Industrial Acoustics Corporation, Bronx, NY). Measures of subject performance were obtained using the SIAM procedure combined with a delayed comparison task in which the frequency of the comparison tone was varied adaptively relative to that of the standard tone. Our implementation of the procedure generally followed the method proposed by Kaernbach (1990) for the measurement of absolute threshold in a single interval task using an adaptive staircase procedure. The frequency of the standard and comparison was different 75% of the time prior to the first reversal, thereafter the chance of the standard and comparison being the same or different occurred with equal probability. The extent of the frequency difference for the comparison tone relative to that of the standard in a subsequent different trial was determined based upon the response(s) given since the last different trial as interpreted within the framework of signal detection theory (Kaernbach, 1990). More specifically, if on a different trial the subject produced a 'hit' (i.e., the standard and comparison tones differ in frequency and the subject responded 'different'), the frequency difference on the next different trial would be reduced by one step size. If on a different trial the subject produced a 'miss' (i.e., the standard and comparison tones differ in frequency and the subject responded 'same'), the frequency difference on the next different trial would be increased by one step size. If on a same trial the subject produced a 'false alarm' (i.e., the standard and comparison tones have the same frequency and the subject responded 'different'), the frequency difference on the next different trial would be increased by two step sizes. If on a same trial the subject produced a 'correct rejection' (i.e., the standard and comparison tones have the same frequency and the subject responded 'same'), the frequency difference on the next different trail would remain unchanged. It should be noted that, by chance, several "same" trials could occur in a row and the net effect of the adaptive tracking adjustments would not be heard until the occurrence of the next 'different' trial. This implementation of the SIAM procedure targets the 75% correct point on the psychometric function (Kaernbach, 1990). A step size of 4.0 Hz was used on a given run until four reversals were obtained; a step size of 1.0 Hz was employed for the remainder of that run.

The subject, on any given run, was to determine whether the last tone they heard in a trial (i.e., the comparison) had the same pitch or a different pitch as the first tone they heard in that same trial (i.e., the standard). The subjects were informed that they could ignore any tones that may appear between the first and the last tone. They registered their response by pressing either the box labeled "S" for same or the box labeled "D" for different on a touch screen monitor. Correct answer feedback was provided by changing the color of the box corresponding to the correct answer from blue to yellow for 1000 ms after the subject's response for a given trial was recorded by the computer. The next trial was presented 1000 ms after the end of the feedback sequence. A run was terminated after 14 reversals were obtained and 5 runs were obtained in each condition. The initial DLF values were taken as the average of the frequency difference values over the last 10 reversals (see results sections for more details regarding alternate DLF estimates). The runs across all conditions were obtained in pseudo-random order. There were a base total of 20 runs completed across all the conditions (5 runs per condition × 4 conditions) for each subject. An extra run was completed if the DLF for the individual run within a given condition differed from mean of the remaining four runs by more than two standard deviations. When this occurred the DL for the extra run replaced that of the outlying original only if it were closer to mean of the four remaining original values. An extra run was required roughly 13% of the time per condition with no individual condition requiring substantially more additional runs than any of the other conditions.

3. Results

It took 55 trials on average (standard deviation = 15) to obtain 14 reversals for a given run. This was approximately half the number of trials as used in the MCS procedure reported previously (Ries and DiGiovanni, 2007). Use of the SIAM procedure clearly resulted in an increase in efficiency with respect to data collection time. This outcome was confirmed by Grossmann and Ries (2008) who made a direct comparison of results obtained using the MCS and SIAM procedure obtained using an identical response format.

The reversal points across each individual run in all four conditions were examined for every subject to ascertain whether the threshold calculation method proposed by Kaernbach (1990) (i.e., drop the first 4 reversals and average the values of the remaining 10 reversals) was the most rational rule (i.e., reliably and efficiently provided coherent and relevant results) to use for DLF measures obtained using the SIAM procedure. Fig. 2 shows the frequency difference value (Hz) between the standard and comparison that was tracked for presentation in each trial of a single run in the RptCmp condition completed by subject 5. The frequency difference indicated was only presented when a difference trial occurred (black squares), but could be adjusted adaptively even on the same trials (open squares; see Section 4, for details). It is evident that dropping more than 4 reversals may provide a more stable and less variable estimate of threshold. For example, taking the mean of the values

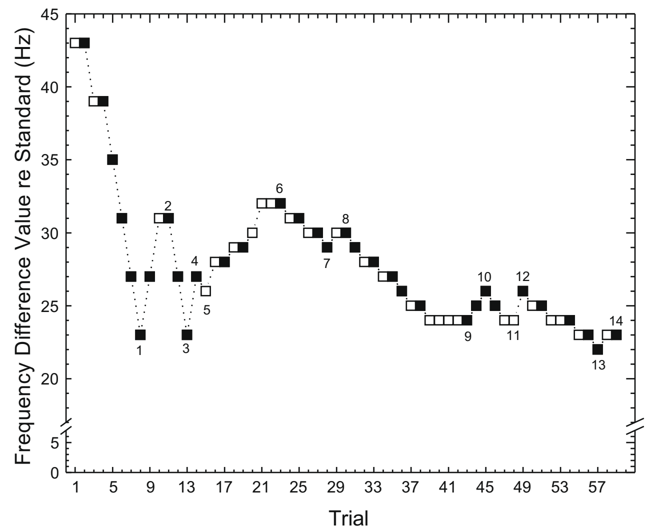

Fig. 2. Representative run in the SIAM procedure for subject 5. The filled boxes represent different trials and the open boxes represent same trials. The frequency difference value in Hz between the standard and comparison stimuli (*y*-axis) are given for each trial in the run (*x*-axis).

of reversals 11 through 14 would provide a lower DLF (23.8 Hz) with a lower standard deviation (1.7 Hz) that would averaging reversals 5 through 14 (DLF = 26.2 Hz, standard deviation = 3.2 Hz). This general trend was apparent across subjects and conditions, although in some instances the run would stabilize about the final DLF value after the first 4–6 reversals.

Counts of the lowest DLF and standard deviation values across runs were obtained for various threshold calculation rules. Fig. 3 shows the percentage of the time that the lowest DLF and standard deviation values occurred for the threshold calculation rules under consideration (*Note.* Total # reversals − # dropped reversals = # of reversals used to calculate the mean and standard deviation). As displayed in Fig. 3, the drop 10 reversal rule (i.e., using only the values at which the final 4 reversals occurred) resulted in the lowest DLF and standard deviations values substantially more often than any of the other rules (i.e., dropping first 4 reversals). The mean of the difference in DLF and standard deviation for each condition,

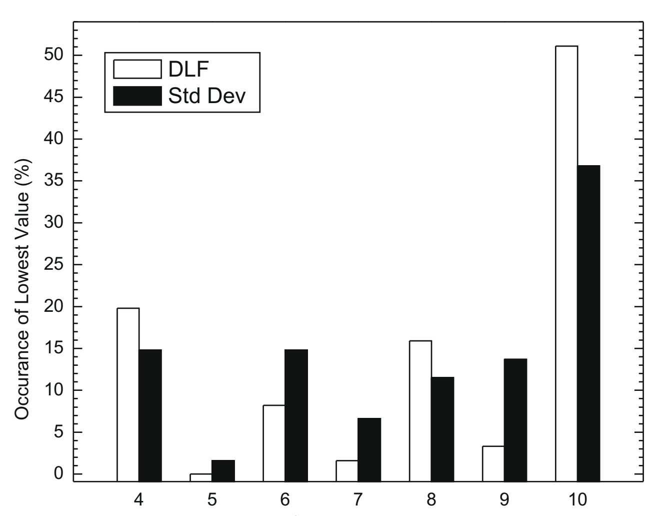

Fig. 3. Percentage of times the lowest DLF (open bars) and standard deviation (dark bars) values occurred (*y*-axis) for each reversal drop rule evaluated. Higher values (i.e., bars) are indicative of better performance.

Table 1
Comparison of the differences in the DLF and standard deviation values across subjects determined by dropping the first 4 vs. the first 10 of the 14 DLF values at which a reversal occurred. Positive values in the second, third, and fourth columns indicate that the measure specified was greater by the extent given when calculated dropping the first 4 reversals.

Condition	Mean of DLF differences (Hz)	Percent runs w/DLF differences ⩾ +2.0 Hz	Percent runs w/DLF differences ⩾ +4.0 Hz
NoInt	0.8	15.0	0.0
Int	0.6	22.5	7.5
RptStd	1.5	30.0	20.0
RptCmp	1.3	40.0	12.5
	Mean of SD differences (Hz)	Percent runs w/SD differences ⩾ 2.0 Hz	Percent runs w/SD differences ⩾ 4.0 Hz
NoInt	0.8	10.0	0.1
Int	0.7	12.5	0.0
RptStd	0.9	17.5	12.5
RptCmp	1.9	27.5	17.5

along with the percentage of the time this values exceeded 2.0 (two times the small step size) and 4.0 Hz (four times the small step sized), are given in Table 1. The range of mean differences across condition for the DLF and standard deviation between the drop 4 and drop 10 reversals rules were 0.6–1.5 Hz and 0.7–1.9 Hz, respectively. The drop 10 reversal rule resulted in lower DLFs within a run approximately 67% (107/160 runs) of the time and lower standard deviations within a run approximately 74% (119/160 runs) of the time across all conditions with respect to those calculated using the drop 4 reversal rule. Use of the drop 10 reversal rule also resulted in less within subject variability across DLF estimates for a given condition as well as lower within run standard deviations (see columns two and three in Table 1). Given the above information, we used the DLF values calculated using the drop 10 reversal rule for statistical analyses and all values given throughout the remainder of this article were determined as such.

The DLF values for each subject and condition are depicted in Fig. 4 The boxes labeled “A”, “B”, “C”, and “D” represent the DLF values obtained in the NoInt, Int, RptStd, and RptCmp conditions, respectively. The basic pattern of results was similar across most subjects, although the absolute DLF values differed to some extent. The lowest values occurred in the NoInt and RptStd conditions. The results for the Int and RptCmp conditions were always higher than those in the NoInt and RptStd conditions with the DLF for the RptCmp being greatest for 6 of the 8 subjects. For the two instances in which the value for the RptCmp was lower than that in the Int condition, the difference was smaller (−5.3 Hz, S1; −1.1 Hz, S3) than the 9.7 Hz mean level of increase in the value for the RptCmp data with respect to the Int data across the remaining 6 subjects (this mean level of increase was 5.6 Hz excluding S6 who exhibited a very large increase in the DLF in the RptCmp condition).

Fig. 5 shows the relative mean (filled square) and median (horizontal line within box) DLF across all subjects for each condition relative to the DLF from each subject’s NoInt condition. The relative values were plotted to allow for ease of comparison of the differences in the DLF values across conditions. The vertical extent of the boxes is 1.96 standard errors in either direction from the mean, which approximates the 95% confidence interval. The general trend found within subjects was preserved in the mean data. The NoInt and RptStd conditions exhibited the lowest DLFs on average, the Int condition exhibited a larger DLF, and the RptCmp condition produced an even higher mean value.

Mean d' prime values for each condition were calculated based on the number of hits, misses, false alarms, and correct rejections a subject produced over the trials encompassing the last four reversals of each run in each condition. These counts were summed across all four runs for a given condition for each subject, the d' value for each subject was calculated, and the mean of the d' prime

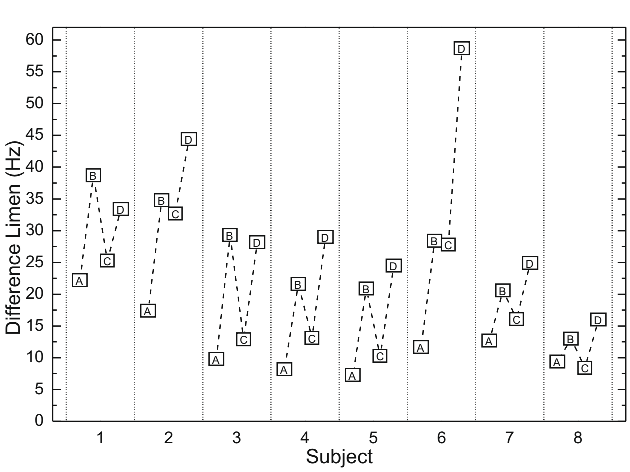

Fig. 4. Mean DLF values (y-axis) for each subject (x-axis) for each condition. The mean values are depicted by squares and the condition is denoted by the letter within each symbol: “A” for the NoInt condition, “B” for the Int condition, “C” for the RptStd condition, and “D” for the RptCmp condition. The dashed lines connecting the results for a given participant are provided to aid in visualization of the pattern of performance across conditions both within a particular listener as well as across subjects.

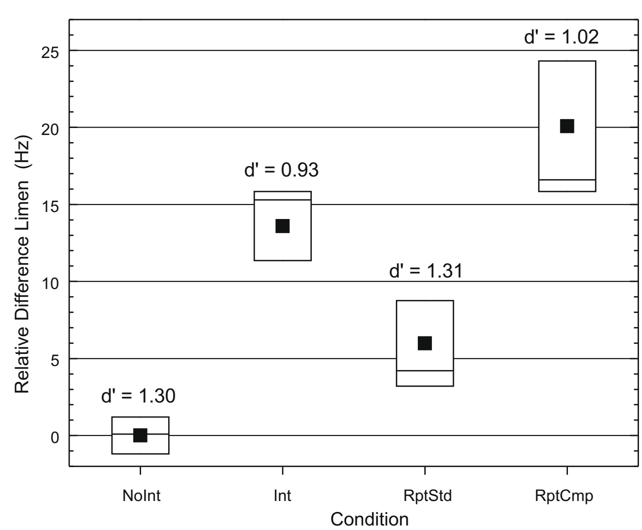

Fig. 5. Box plot of the relative DLF with respect to that obtained for each subject in the NoInt condition averaged across subjects (y-axis) for each condition (x-axis). The filled square and dark horizontal line with each box represent the mean and median values for each condition, respectively. Each box extends 1.96 standard errors above and below the average values for each condition thereby depicting the 95% confidence interval about the mean in each case. The sensitivity index d' is provided above the box plot for each condition.

values across subjects was then calculated for each condition. The resultant values are displayed above the corresponding box plot in Fig. 5. The resultant d' values fit the trend evident in the mean DLF values across conditions. Specifically, d' was larger for the NoInt and RptStd conditions, consistent with lower DLFs, and d' was lower for the Int and RptCmp, consistent with higher DLFs. A larger d' prime value for a targeted level of performance (e.g., 75% correct) is indicative of a greater perceptual distance between the standard and comparison. The increase in perceptual distance can be accomplished either by increasing the frequency difference between the means of the standard and comparison distributions or by a reduction in the spread of the distributions. As the DLF values are lowest in the conditions with the highest d' values (i.e., the NoInt and RptStd conditions), it can be concluded that the improvement in pitch discrimination for these two conditions was a result of a reduced amount of spread of the standard distribution relative to that produced in the Int and RptCmp conditions.

The mean estimates of the subjects' DLF values across runs were evaluated within each condition via calculation of the Mahalanobis distance (T^2). The T^2 value was then evaluated against the F-distribution to determine the probability of occurrence to determine whether any subject's result exercised undue leverage upon the overall results for each condition. As none of the mean DLF estimates for a given individual differed significantly ($p > 0.05$) from the overall mean for any of the conditions, subsequent statistical analyses included the data from all subjects.

The experimental outcomes were analyzed using a repeated measures analysis of variance (ANOVA) test where condition and DLF served as the independent and dependent variables, respectively. The outcomes of the analysis, using the Geisser–Greenhouse Adjustment (G–GA) in order to correct for violation of sphericity, found a significant difference for condition ($F(3, 21) = 17.25$, $p = 0.001$) revealing that the addition of intervening tones significantly increased the frequency difference between the standard and comparison tones required for correct identification. As the overall F-statistic with the G–GA for condition was significant, post hoc pairwise comparisons of the four conditions were made using Fisher's LSD multiple comparison test (Ryan, 1959) at an α level of 0.05. Results from the post hoc comparisons revealed that the DLFs obtained in the Int condition were significantly larger than those for the NoInt and RptStd conditions as predicted in hypotheses 1 and 2. The DLFs for the RptCmp condition were significantly higher than those measured in the Int condition as predicted in hypothesis 3. Additionally, the DLF values for the NoInt and RptStd conditions did not diff significantly while, as would be expected, the DLF values measured in the RptCmp were significantly higher than those obtained in the RptStd condition.

4. Discussion

The present results provide quantitative measures of the effect of various changes in an intervening tonal sequence upon the ability of a listener to use cues to aid in maintenance of the pitch standard held within AWM. In this study, we measured the DLF in order to determine the extent of change in the frequency of the stimulus required to approximate constant performance (i.e., 75% correct performance) in various conditions. Similar to our previous report (Ries and DiGiovanni, 2007), the present results showed that listeners required a substantially larger difference in frequency to maintain performance when intervening tones were present. This result, however, was mitigated significantly when the second tone of the intervening sequence had a pitch identical to that of the standard (RptStd condition) as evidenced by a reduction in the DLF relative to that obtained in the Int condition. In contrast, the DLF increased over that measured in the Int condition when the pitch of the second tone of the intervening sequence was identical to that of the comparison. As in our previous work (Ries and DiGiovanni, 2007) and that of König (1957), the absolute DLF values obtained differed across subjects. More importantly, however, the pattern of results across subjects was similar.

The outcomes of this study support the qualitative model proposed by Deutsch (1972b) in which presentation of the standard or comparison pitch within the intervening sequence significantly impacts a listener's ability to maintain information about the pitch of the standard within AWM. In support of her model, Deutsch (1972b) reported percent correct identification of a fixed frequency separation (i.e., a semitone). The results of the present study quantify the extent to which the repetition of the pitch of the standard within the intervening sequence promotes improved performance in pitch judgments in a delayed comparison task. According to the model proposed by Deutsch (1972b), the reoccurrence of the standard within the intervening sequence adds to the amplitude of the waning standard maintained within AWM thereby strengthening the pitch standard within working memory which can help the listener overcome the interference produced by the remainder of the intervening sequence. If we extend her model to include predicted changes in the DLF, the DLF in the RptStd condition with respect to the Int condition is predicted to be significantly lower. Analysis of experimental data from this investigation revealed this exact outcome. That is, the frequency difference between the standard and comparison tone needed for a listener to reliably detect a change in pitch in the presence of intervening tones was reduced by approximately 44% (6.0 Hz lower) when the pitch of the standard was repeated within the intervening sequence.

Additionally, subjects' performance can be affected negatively dependant upon the changes made in an intervening sequence. Specifically, presentation of the comparison pitch within an intervening sequence results in an increase in the DLF over that in the general interference condition (i.e., Int condition). Again, if Deutsch's model (1972b) were extended to include predictions of the DLF, the frequency difference required to reliably detect a change in pitch is predicted to be significantly greater. This occurs because the pitch of the comparison was heard previously and thereby induced a summation of the two previously heard pitches of interest (i.e., the pitch of the standard and the intervening tone with a pitch identical to that of the comparison). This summation effectively results in an increase in the spread of the standard distribution maintained within AWM due to the separation in frequency between the two now amalgamated pitch percepts (i.e., the pitch of the standard tone and that of the intervening tone with a pitch equal to that of the comparison). In addition, the amplitude of the standard is lower due to decay over time. As a result, a greater pitch distance (i.e., a larger DLF) is required to maintain a constant level of performance. The results of present study follow this pattern. That is, the frequency difference between the standard and comparison tone needed for a listener to reliably detect a change in pitch in the presence of intervening tones was increased by approximately 48% (6.5 Hz higher) when the pitch of the comparison was repeated within the intervening sequence.

The DLF measures obtained in the present study via the SIAM procedure show the same pattern in the NoInt and Int conditions as we have reported previously (Ries and DiGiovanni, 2007), but the absolute values differ. In addition the d' and DLF values obtained here are consistent with the findings of Ross and colleagues (2003, 2004) who found that, for retention intervals lasting more than 2 s, the mean of pitch matches decreases as the distribution of the responses becomes more uniform (i.e., a wider distribution of responses with roughly equal occurrence) when more than one intervening tone was present. Table 2 lists results from four separate studies (Harris, 1952; Elliot, 1970; König, 1957; Ries and DiGiovanni, 2007) with conditions similar to the NoInt (column 2)

Table 2
DLF values from three previous studies (rows 3–5) with those measured in the present study (bottom row) for the NoInt (column 2) and Int conditions (column 3).

Study	Average DLF (Hz)	
	NoInt condition	Int condition
Harris (1952)	≈5.3	N/A
König (1957)*	≈6.5	N/A
Elliot (1970)	≈3.0	≈6.0
Ries and DiGiovanni (2007)	≈6.0	≈18.0
Current study	≈12.3	≈25.9

* This study presented DLF results obtained from several different procedures. The results listed in the table were estimated from the middle panel of Fig. 6 (p. 610) of König's study. The roving comparison standard procedure used in König's study was the one that most closely matches the standard and comparison setup in the present study.

and/or Int conditions (column 3). The DLF values measured in the present study are higher for a given condition than are the others in the table. Probable causes for this result include differences in procedure and response format. The results from the four other studies in Table 2 used a "higher/lower" response format whereas the present study employed a "same/different" response following that used by Deutsch (e.g., 1972a,b). This could lead to higher DLF values as subjects tend to adopt a more conservative criterion in "same/different" tasks, responding "different" only when they are certain that the stimuli differ along a particular perceptual continuum (Gerrits and Schouten, 2004); as opposed to a "higher/lower" response format wherein subjects assume a difference exists, can adopt a very liberal response criterion as even the slightest perceived change can be presumed to be a true change, and they simply judge in which direction the difference occurred. In addition, the studies of Harris (1952), Elliot (1970), and Ries and DiGiovanni (2007) used the method of constant stimuli, whereas the present study employed an adaptive tracking procedure (i.e., SIAM). While a change in procedure (i.e., SIAM versus MCS) could have lead to the higher DLF values obtained in the present study (König, 1957), the extent of the increase resultant from this change in method would be surprising given that the SIAM procedure produces absolute threshold results similar to those obtained via 2- and 3-interval forced choice procedures (Kaernbach, 1990). Last, recent work presented by Grossmann and Ries (2008) show that DLF outcomes obtained via the MCS do not differ significantly from those obtained using the MCS procedure when both methods employ a "same/different" response format.

Overall, the later occurrence of a tone with a pitch identical to that of a standard held within AWM improves a listener's ability to maintain an accurate representation of the said tone in the presence of other tones interspersed between the standard and a later occurring comparison stimulus. The extent of this improvement was quantified as the reduction in the DLF when the results measured in the condition containing the pitch of the standard within the intervening tone sequence were compared to those measured in the general intervening tone condition in which the pitch of these stimuli never matched that of the standard. In contrast, the DLF increased significantly when a pitch identical to that of the comparison was introduced within the intervening sequence. In addition, the use of the SIAM procedure in the current study produced a pattern of results similar to that reported by Ries and DiGiovanni (2007), but which differ in absolute value. The difference in the absolute values likely results from the different response formats used in the two studies and not from the procedure itself. Future investigations of alternative response formats across different methods of measurement that facilitate the use of a more liberal response criterion by listeners will clarify the extent to which the differences in the absolute values obtained are due to method and response type. Implementation of the SIAM procedure allowed for more efficient data collection than was possible using the method of constant stimuli in our earlier study while maintaining the trends in performance obtained previously for similar conditions.

References

Bachem, A., 1954. Time factors in relative and absolute pitch determination. J. Acoust. Soc. Am. 26, 751–753.

Berti, S., Münzer, S., Schröger, E., Pechmann, T., 2006. Different interference effects in musicians and a control group. Exp. Psychol. 53, 111–116.

Bernstein, L.R., Trahiotis, C., 2008. Binaural signal detection, overall masking level, and masker interaural correlation: revisiting the internal noise hypothesis. J. Acoust. Soc. Am. 124, 3850–3860.

Bernstein, L.R., Trahiotis, C., Freyman, R.L., 2006. Binaural detection of 500-Hz tones in broadband and in narrowband masking noise: Effects of signal/masker duration and forward masking fringes. J. Acoust. Soc. Am. 119, 2981–2993.

Brown, L., 1996. Additional rules for the transformed up-down method in psychophysics. Percept. Psychophys. 58, 959–962.

Bull, A.R., Cuddy, L.L., 1972. Recognition memory for pitch of fixed and roving stimulus tones. Percept. Psychophys. 11, 105–111.

Clément, S., Demany, L., Semal, C., 1999. Memory for pitch versus memory for loudness. J. Acoust. Soc. Am. 106, 2805–2811.

Deutsch, D., 1970a. Dislocation of tones in a musical sequence: a memory illusion. Nature 226, 286–288.

Deutsch, D., 1970b. Tones and numbers: specificity of interference in immediate memory. Science 168, 1604–1605.

Deutsch, D., 1972a. Mapping of interactions in the pitch memory store. Science 175, 1020–1022.

Deutsch, D., 1972b. Effect of repetition of standard and comparison tones on recognition memory for pitch. J. Exp. Psychol. 93, 156–162.

Deutsch, D., 1974. Generality of interference by tonal stimuli in recognition memory for pitch. Q. J. Exp. Psychol. 26, 229–234.

Deutsch, D., 1975. Facilitation by repetition in recognition memory for tonal pitch. Mem. Cogn. 3, 263–266.

Deutsch, D., 1978. Interference in pitch memory as a function of ear input. Q. J. Exp. Psychol. 30, 283–287.

Deutsch, D., Feroe, J., 1975. Disinhibition in pitch memory. Percept. Psychophys. 17, 320–324.

Deustch, D., Henthorn, T., Dolson, M., 2004. Absolute pitch, speech, and tone language: some experiments and a proposed framework. Music Percept. 21, 339–356.

Deutsch, D., Henthorn, T., Marvin, E., Xu, H., 2006. Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period (L). J. Acoust. Soc. Am. 119, 719–722.

Elliot, L.L., 1970. Pitch memory for short tones. Percept. Psychophys. 8, 379–384.

Gerrits, E., Schouten, ME., 2004. Categorical perception depends on the discrimination task. Percept. Psychophys. 66, 363–376.

Grossmann, A., Ries, D.T., 2008. Pitch interference in working memory: effects of timbre and method. Paper presented to the American Auditory Society, Scottsdale, Arizona.

Harris, J.D., 1952. The decline of pitch discrimination with time. J. Exp. Psychol. 43, 96–99.

Jump, R.L., Ries, D.T., 2008. Effect of interaural level and phase cues on intervening interference in auditory working memory for loudness. Hear. Res. 240, 112–115.

Kaernbach, C., 1990. A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing. J. Acoust. Soc. Am. 88, 2645–2655.

Kaernbach, C., 1991. Simple adaptive testing with the weighted up-down method. Percept. Psychophys. 49, 227–229.

Kallman, H.J., Cameron, P.A., Beckstead, J.W., Joyce, E., 1987. Ear of input as a determinant of pitch-memory interference. Mem. Cogn. 15, 454–460.

Klein, S., 2001. Measuring, estimating, and understanding the psychometric function: a commentary. Percept. Psychophys. 63, 1421–1455.

König, E., 1957. Effect of time on pitch discrimination thresholds under several psychophysical procedures; comparison with intensity discrimination thresholds. J. Acoust. Soc. Am. 29, 606–612.

Levitt, H., 1971. Transformed up-down methods in psychophysics. J. Acoust. Soc. Am. 49, 467–477.

Leeuw, A.R., Dreschler, W.A., 1998. The relation between otoacoustic emissions and the broadening of auditory filter for higher levels. Hear. Res. 126, 1–10.

Massaro, D.W., 1970. Retroactive interference in short-term recognition memory for pitch. J. Exp. Psychol. 83, 32–39.

Meese, T., 1995. Using the standard staircase to measure the point of subjective equality – a guide based on computer simulations. Percept. Psychophys. 57, 267–281.

Pechmann, T., Mohr, G., 1992. Interference in memory for tonal pitch: implications for a working memory model. Mem. Cogn. 20, 314–320.

Profita, J., Bidder, T.G., 1988. Perfect pitch. Am. J. Med. Genet. 29, 763–771.

Ross, D.A., Olson, I.R., Marks, L.E., Gore, J.C., 2004. A nonmusical paradigm for identifying absolute pitch possessors. J. Acoust. Soc. Am. 116, 1793–1799.

Ross, D.A., Olson, I.R., Gore, J.C., 2003. Absolute pitch does not depend on early musical training. Ann. N.Y. Acad. Sci. 999, 522–526.

Ries, D.T., DiGiovanni, J.J., 2007. Release from interference in auditory working memory for pitch. Hear. Res. 230, 64–72.
Ryan, T.A., 1959. Comments on orthogonal components. Psychol. Bull. 56, 394–396.
Schlauch, R.S., Rose, R.M., 1990. Two-, three-, and four-interval forced-choice staircase procedures: estimator bias and efficiency. J. Acoust. Soc. Am. 88, 732–740.
Semal, C., Demany, L., 1991. Dissociation of pitch from timbre in auditory short-term memory. J. Acoust. Soc. Am. 89, 2404–2410.
Semal, C., Demany, L., 1993. Further evidence for an autonomous processing of pitch in auditory short-term memory. J. Acoust. Soc. Am. 94, 1315–1322.
Stewart, J.M., Downs, M.P., 1984. Medical management of the hearing-handicapped child. In: Northern, J.L. (Ed.), Hearing disorders, second ed. Little, Brown, Boston, pp. 267–278.
Wickelgren, W.A., 1966. Consolidation and retroactive interference in short-term recognition memory for pitch. J. Exp. Psychol. 72, 250–259.
Wickelgren, W.A., 1969. Associative strength theory of recognition memory for pitch. J. Math. Psychol. 6, 13–61.