The decay of pitch memory during rehearsal

Christian Kaernbach[a)]
Institut für Psychologie, Christian-Albrechts-Universität zu Kiel, Olshausenstrasse 62, 24098 Kiel, Germany

Kathrin Schlemmer
Institut für Musikwissenschaft, Martin-Luther-Universität Halle-Wittenberg, Kleine Marktstrasse 7, 06099 Halle, Germany

(Received 21 February 2007; revised 23 January 2008; accepted 24 January 2008)

The present study investigates the decay of pitch memory over time. In a delayed pitch comparison paradigm, participants had to memorize the pitch of a Shepard tone, with silent, overt, or without any rehearsal. During overt rehearsal, recordings of the rehearsing were effectuated. Performance was best for silent rehearsal and worst for overt rehearsal. The differences, although partially significant, were not marked. The voice pitch during overt rehearsal was compatible with a random walk model, providing a possible explanation of why rehearsal does not improve the retention of the pitch trace. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2875365]

PACS number(s): 43.66.Ba, 43.66.Hg, 43.66.Mk [AJO] Pages: 1846–1849

I. INTRODUCTION

Auditory sensory memory has been shown to share many characteristics with classical short-term memory, such as lifetime, capacity, and susceptibility to interference (Kaernbach, 2004). However, a major difference seems to be that rehearsal does not appear to be as effective with sensory information as with categorical information.

Keller *et al.* (1995) note that no measures are taken to prevent rehearsal in standard delayed pitch comparison tasks. However, this does not fully prevent the loss of auditory information over time. They found a slight decrease in performance in a pitch memory task, if a rehearsal-preventing distractor task is to be performed during the retention interval. They suggest that attention might help to maintain the trace in the condition without the distractor task.

In contrast to this, Demany *et al.* (2001, 2004) failed to demonstrate a beneficial influence of attention and/or covert rehearsal. They investigated the influence of a cue that directed the attention toward one out of three possible characteristics or components of a stimulus in a delayed matching experiment. If the cue fell in the retention interval, it made no difference whether it came early or late. If a rehearsal algorithm would have been effective, the early cue should have been more helpful.

The beneficial effects of rehearsal, if any, are small. Overt rehearsal has even been reported to be detrimental to the retention of a pitch trace (Massaro, 1970). This is in contrast to classical short-term memory for categorical information, where the lifetime of a trace can be lengthened, ad infinitum, by rehearsing the stored information. For a better understanding of this discrepancy it would be useful to study the temporal dynamics of the decay of the sensory trace. The voice pitch during overt rehearsal might be a tool to probe the actual state of the sensory trace during decay. Therefore, the present study compared overt, covert, and no rehearsal conditions. The recorded pitch data during overt rehearsal were compared to theoretical predictions of the time course of the sensory trace during the retention interval.

II. METHODS

Three individuals with an auditory-related profession (two female, one male, age range 29 to 30) participated in our experiment. All of them had been playing an instrument for 15 to 24 years. Two of them had had formal vocal training for 15 years. Participants had to compare two stimuli (S1 and S2) that were separated by a certain retention interval. The second stimulus would be slightly higher or slightly lower than the first stimulus, with equal probability. Participants had to indicate which of these two possibilities was the case.

In order to facilitate rehearsal, we employed Shepard tones (Shepard, 1964). These octave-complex tones elicit circular pitch percepts, with a well-defined chroma (pitch class, e.g., C vs C#), but an ambiguous pitch height (octave register, e.g., C4 vs C3). Shepard tones enable participants to rehearse the presented tone at whatever octave is most appropriate to their vocal range. The duration of the tones was 1 s, with 0.1 s ramps at the onset and the offset of the tones. We presented the tones at a level 60 dB above threshold. The chroma of the first stimulus was randomized uniformly on a logarithmic frequency scale over the range of one octave. After each single trial, we gave feedback in order to help the participants to improve their performance. The next trial started 1.2 s after the preceding trial.

Prior to the main experiment, we determined the just noticeable difference (JND) for these stimuli (interstimulus interval 0.5 s) for each participant. Using an unforced-choice adaptive procedure (Kaernbach, 2001), we estimated the point of the psychometric function where 75% of the judgments were correct. The JNDs for the three participants were close to 3, 4, and 5 cents. We rounded them off to these integer values. A JND of 4 cents corresponds to a frequency difference of 0.23%. This is compatible with what is known

[a)]Electronic mail: www.kaernbach.de.

 0001-4966/2008/123(4)/1846/4/$23.00

about the JND for the pitch of complex tones (see, e.g., Hoekstra, 1979; Shackleton and Carlyon, 1994).

In the main experiment, we tested four different conditions. These conditions were tested blockwise in blocks of 20 trials of the same type. Participants performed several training blocks until they felt at ease with the different tasks. Then, they performed 20 blocks of each condition in rotating order.

In one condition, the duration of the retention interval was 0.5 s. In this condition, no specific rehearsal instruction was given. The second stimulus S2 differed from the first stimulus S1 by −1.67, −1, −0.33, +0.33, +1, or +1.67 JNDs. The actual difference was taken randomly from these six possibilities. This condition was tested 400 times per participant, i.e., approximately 400/6=67 times per difference.

In three other conditions, the duration of the retention interval was 6 s. The difference between S2 and S1 was taken randomly from −3, −1, +1, or +3 JNDs. Each difference was tested approximately 400/4=100 times per participant in each condition. The three different conditions differed only by the rehearsal instruction.

The first rehearsal instruction was "no rehearsal." It was symbolized by a mandala appearing on the computer screen, and the participants were told not to pay attention to the pitch of the first stimulus until the second stimulus appeared. The other two conditions were rehearsal conditions, with silent versus overt rehearsal. They were also symbolized by icons on the computer screen, but this time the participants had to direct their attention toward the pitch of S1. All participants stated that they were able to act differently in these conditions as compared to the no rehearsal condition, and that they believed that this would improve their performance. In the overt rehearsal condition, we recorded the voice during the retention interval.

III. PITCH DISCRIMINATION PERFORMANCE

The data analysis was performed by calculating maximum likelihood fits and likelihood ratios, in order to test the various hypotheses against each other.

The data of the three participants were similar if compared on a JND scale. A likelihood ratio test revealed that they could be pooled into a single data set. Figure 1 shows the pooled psychometric functions for all four conditions. The psychometric function for the 0.5 s retention condition is much steeper than the other three psychometric functions, illustrating that the precision of the trace is much higher after 0.5 s than after 6 s. The differences among the other three psychometric functions are not marked.

Cumulative normal distributions were fitted to the four psychometric functions. Table I shows the maximum likelihood estimates for the means and the standard deviations of the trace for these four conditions. The means are compatible with zero for the 0.5 s retention condition and for the 6 s silent rehearsal condition. For the other two conditions, the means are significantly different from zero ($p<0.05$). Note that the mean is lower than zero in the overt rehearsal condition, which would result from a decreasing voice pitch, as is often observed in singing.

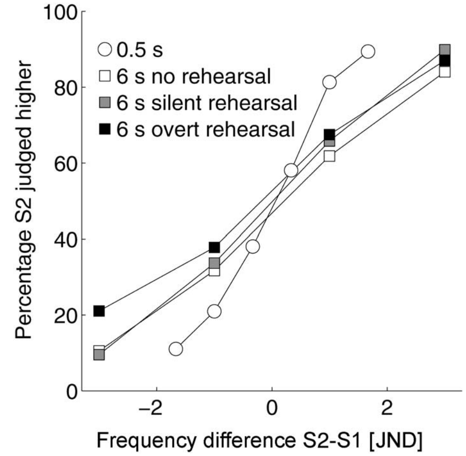

FIG. 1. Psychometric functions for the four experimental conditions.

The standard deviations for the three 6 s retention conditions differ slightly from each other, with the silent rehearsal condition resulting in the lowest spread of the trace, and the overt rehearsal condition resulting in the highest spread of the trace. The difference in the standard deviations between silent and overt rehearsal is significant ($p<0.05$), whereas the other two differences are not significant.

IV. VOICE PITCH ANALYSIS

In each single trial of the overt rehearsal condition, voice pitch during the retention interval was determined with the YIN algorithm (de Cheveigné and Kawahara, 2002) for eleven 1 s segments. These started with the interval $[0,1]$ s, going forward in steps of 0.5 s, and ending with the interval $[5,6]$ s. In total, there were 1200 trials of this condition, giving 13 200 1 s intervals. YIN could determine the voice pitch in 10 027 of these intervals. In the other 3173 intervals, the voice was either too soft or absent.

The voice pitch was then set into relation to the nearest octave component of the S1 stimulus. On average, it was 1.2 cents lower than this component. The standard deviation was 43 cents, i.e., nearly half of a semitone. In order to reduce the effect of outliers, voice pitches, which lay farther apart from the nearest component of S1 than 2 s.d., were excluded from further processing.

V. APPLYING A RANDOM-WALK MODEL OF SENSORY RETENTION

Kinchla and Smyzer (1967) have proposed and successfully tested a random walk model of sensory retention. Although this model was recently challenged by Gold *et al.*

TABLE I. Maximum likelihood estimates for the mean and the standard deviation of the trace expressed in just noticeable differences (3–5 cents). To obtain estimates in cents multiply by a factor of about 4.

	0.5 s	6 s no rehearsal	6 s silent rehearsal	6 s overt rehearsal
Mean	0.01	0.27	0.03	−0.34
Standard deviation	1.29	2.65	2.34	3.05

(2005) for visual information, and by Demany *et al.* (2005), concerning its dependence on the sensory noise, it should apply to the present data.

In this model, the stimulus S1 is encoded to trace S1′. The encoding process is affected by the encoding noise N_e with standard deviation S_e. During retention, the trace undergoes degradation and the standard deviation increases. The variance of the trace increases linearly with time with diffusion rate φ. Stimulus S2 is encoded to trace S2′ with the encoding noise having the same variance as for the encoding process for S1. The total variance of the comparison variable S2′−S1′ is hence:

$$\mathrm{Var}(S2' - S1') = S_e^2 + \varphi t + S_e^2. \qquad (1)$$

In an unbiased model and for an infinite number of random walks, the mean across all random walks stays equal to the starting point. However, if the set of random walks to be averaged is selected depending on its end point (comparison with S2′), the subsets [S1′(end) > S2′ vs S1′(end) < S2′] will have drifting means. In this case, the mean of the subsets will drift linearly away from the starting point.

Overt rehearsal suffers from voicing noise, and the voice pitch feeds back on the trace S1′ and might possibly degrade it further. The standard deviation of the voice error (43 cents) is very large compared to the standard deviation derived from the judgments (5 cents after 0.5 s, about 12 cents after 6 s) and to the effect of the voice on the trace (difference between overt and silent rehearsal about 3 cents). This tells us two things: The audible feedback supplied by the voice can definitively not be helpful for the maintenance of the trace; on the other hand, it does not change much for the retention process.

While the voice is not helpful for the maintenance of S1′, it is a valuable probe into the state of S1′. It represents a noisy image of S1′, but this noise can be reduced by averaging across large numbers of recorded voice segments. It can thus help to understand what is going on during retention. By averaging across successful versus unsuccessful trials, we can test whether the voice data are compatible with the linearity predicted by the random walk model; and we can derive estimates for S_e and φ. Figure 2 shows the mean voice error during overt rehearsal as a function of time for successful and unsuccessful trials. The data of S2 < S1 trials have been mirrored so that Fig. 2 can be interpreted in terms of S2 > S1 trials. The direction of the voice error is correlated with the judgment of the participant: A large deviation of the voice in the direction of S2 (upwards in Fig. 2) is typical for an error of judgment.

The data show a clear tendency of increasing voice error over time. There is a contrary effect of large initial errors only for the first two time segments (center at 0.5 and 1 s); the smallest errors are found at 1.5 s. This initial error has certainly to do with the difficulties that the participants had in matching the internal trace S1′ with their voice. Once they achieved the match, the voice error increases due to the internal random walk process. Also plotted in Fig. 2 are the model predictions, with the two parameters of the model ($S_e = 1.3$ JNDs, $\sqrt{(\varphi \times 6\ \mathrm{s})} = 2.0$ JNDs) optimized, so as to minimize the χ^2 distance over all points of all curves, except the first two points. For these parameters, χ^2 is 3.18, indicating a very good fitting of the model. Note that the slopes and intercepts of all four straight lines in Fig. 2 (eight parameters) are not fitted individually but follow from only two parameters, S_e and φ.

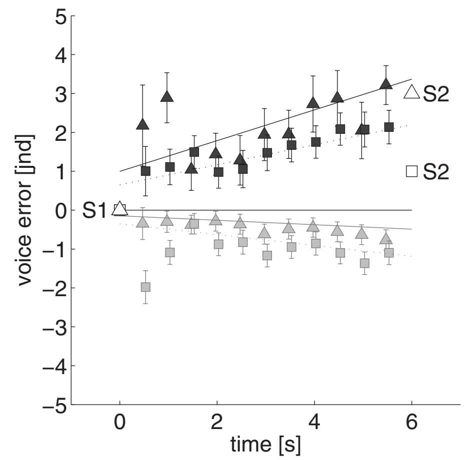

FIG. 2. Mean voice error during overt rehearsal for successful (light gray symbols) and unsuccessful (dark symbols) trials. Triangles are for trials with 3 JNDs difference, and squares are for trials with 1 JND difference between S2 and S1. Voice data of trials with S2 < S1 have been mirrored at the nearest matching component of S1. Solid lines give the model prediction for a 3 JND difference, and dotted lines for a 1 JND difference.

The voice data allow the parameters of the random walk to be estimated. The estimates of these parameters give an estimate for the final judgment performance. Following the model, the decision of the participant should be determined by the total variance [see Eq. (1)], which, with the present data, yields a standard deviation of 2.72 JNDs, slightly better than the actual performance in the overt rehearsal condition (3.05 JNDs), and quite comparable to the performance in the no-rehearsal condition (2.65 JNDs).

VI. DISCUSSION

In our study, we have compared performance in a pitch memory task for three rehearsal conditions: no rehearsal, silent, and overt rehearsal. We found that the different rehearsal conditions had little effect on the performance in the memory task. This is quite in line with previous research, where rehearsal and/or attention effects on the retention of a sensory trace were either small or absent.

In our study, the different rehearsal conditions were created by instruction. There was no control of whether the participants did what they were instructed to do, except for the condition of overt rehearsal where we recorded the voice. It may be argued that this could be the cause of the similarity of the data. This “instruction only” approach may, on the other hand, be seen as an attempt to minimize other sources of variance, such as attention effects, or interfering noises produced by movements of the participants when performing the distractor tasks. To the best of our knowledge, our study is the first attempt to compare no rehearsal and silent rehearsal based on an instruction only approach. Given that the participants reported that they were able to follow the instructions, and that they were convinced that they would perform markedly better in the rehearsal conditions, the result may be seen as adding further evidence to the uselessness of rehearsal for the retention of a sensory trace.

The main focus of our study was, however, to compare overt and covert rehearsal and to test the explanatory power of voice recordings during overt rehearsal. The discrepancy between the large effect of rehearsal on categorical information and the small or absent effect of rehearsal on the retention of a sensory trace remains intriguing. The present study provides evidence in support of the random walk model of Kinchla and Smyzer (1967) for sensory retention. This model could offer an explanation: In this model of sensory retention, there is no place for rehearsal as a retention-supportive mechanism. The (overt or covert) rehearsal can, at best, mirror the actual state of the sensory trace as it undergoes the random walk. It is not at all helpful in maintaining this trace. For the rehearsal of categorical information (Baddeley and Hitch, 1974), however, the random walk model does not apply. It is assumed that the trace is perfectly reestablished at regular intervals.

In our study, the analysis of the voice recordings suggests that this view is plausible. The voice error was related to the judgment of the participants, which would not be true if the recorded voice had no relation to the sensory trace. The voice could hence be considered a mirror of the internal trace. At first view, the quality of this mirror is not excellent: The standard deviation of the voice was 43 cents, much larger than the quality of the sensory trace (about 12 cents). After averaging across many trials, however, the voice error reflected the linear drifts predicted for unsuccessful versus successful trials. Moreover, the judgment performance predicted from the voice errors (2.72 JNDs) was in the range of the actual judgment performance of the participants (3.05 JNDs). In other words, 80% $[(2.72/3.05)^2]$ of the internal variance leading to judgment errors is explained by analyzing averaged voice data. After all, voice data mirror the internal trace quite well.

The similarity of the performance in the overt and the covert rehearsal condition suggests that the results obtained for the dynamics of the internal trace, in the case of overt rehearsal, is also valid for the covert rehearsal condition. This would be different if the performance had been better for overt rehearsal than for covert rehearsal. In this case, one would have to assume that overt rehearsal involves retention algorithms different from those involved in covert rehearsal. As it is, however, the most parsimonious assumption is that both conditions involve very similar retention algorithms. If one accepts that participants did follow the instructions, this conclusion is also valid for the no rehearsal condition. The slight disadvantage for overt rehearsal is in line with previous findings (Massaro, 1970).

It is puzzling that the participants performed so well in the overt rehearsal condition (12 cents), given that the voice error (43 cents) was much larger than the judgment error in the silent rehearsal condition (9 cents). It is well known, from many experiments on pitch memory, that interfering tones during the retention interval degrade the performance in a pitch memory task (e.g., Deutsch, 1970). Why would it be that the sound of the participant's own voice, being sometimes nearly a semitone off, interferes less than sounds introduced by the experimenter? The sound of the participant's voice is surely different from the sound of the Shepard tones, however, experiments by Semal and Demany 1991, suggest that timbre differences do not explain this effect. We can only speculate that the fact that the participants produce these sounds themselves is crucial to this effect.

Voice data support the view that the sensory trace undergoes a random walk process during the retention interval. This would explain why rehearsal is of little help: It would at best reflect the increasingly distorted internal representation of the sensory trace. The present study has demonstrated that voice recordings can improve our understanding of the dynamics of sensory retention.

ACKNOWLEDGMENTS

We would like to thank Laurent Demany and an anonymous reviewer for helpful comments on the manuscript. Part of this work was presented at the ISH (International Symposium on Hearing) 2006 and will be published in a forthcoming book.

Baddeley, A. D., and Hitch, G. J. (**1974**). "Working memory," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, edited by G. A. Bower (Academic, New York), Vol. **8**, pp. 47–89.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.

Demany, L., Clément, S., and Semal, C. (**2001**). "Does auditory memory depend on attention?," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 461–467.

Demany, L., Montandon, G., and Semal, C. (**2004**). "Pitch perception and retention: Two cumulative benefits of selective attention," Percept. Psychophys. **66**, 609–617.

Demany, L., Montandon, G., and Semal, C. (**2005**). "Internal noise and memory for pitch," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 136–142.

Deutsch, D. (**1970**). "Tones and Numbers: Specificity of interference in short-term memory," Science **168**, 1604–1605.

Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J., and Sekuler, R. (**2005**). "Visual memory decay is deterministic." Psychol. Sci. **16**, 769–774.

Hoekstra, A. (**1979**). "Frequency discrimination and frequency analysis in hearing," Institute of Audiology, University Hospital, Groningen.

Kaernbach, C. (**2001**). "Adaptive threshold estimation with unforced-choice tasks," Percept. Psychophys. **63**, 1377–1388.

Kaernbach, C. (**2004**). "The memory of noise," J. Exp. Psychol. **51**, 240–248.

Keller, T. A., Cowan, N., and Saults, J. S. (**1995**). "Can auditory memory for tone pitch be rehearsed?," J. Exp. Psychol. Learn. Mem. Cogn., **21**, 635–645.

Kinchla, R. A., and Smyzer, F. (**1967**). "A diffusion model of perceptual memory," Percept. Psychophys. **2**, 219–229.

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

Semal, C., and Demany, L. (**1991**). "Dissociation of pitch from timbre in auditory short-term memory," J. Acoust. Soc. Am. **89**, 2404–2410.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

Shepard, R. N. (**1964**). "Circularity in judgments of relative pitch," J. Acoust. Soc. Am. **36**, 2346–2353.