

# Recognition memory for pitch of fixed and roving stimulus tones\*

ALAN R. BULL and LOLA L. CUDDY†  
Queen's University, Kingston, Ontario, Canada

The method of delayed comparison was used to measure pitch memory under two conditions of presentation, either the same standard tone on every trial (fixed S) or one of four equally likely standard tones (roving S), six durations of interference interval ranging from 0.0 to 4.0 sec, and two types of interference interval, either blank or filled with a pure tone. Listeners were instructed not to rehearse the S tone during the interference interval. For both fixed-S and roving-S conditions, accuracy of performance, as measured by area under the ROC curve, declined as interval duration increased and declined more rapidly following an interference-filled than following a blank interval. In general, forgetting proceeded at a slower rate under fixed-S conditions than under roving-S conditions. A second experiment showed that very little forgetting is obtained over 4 sec for roving-S conditions if instructions permit rehearsal of the S tone and suggests that the conclusions from the first experiment be restricted to nonrehearsal instructional conditions.

The present study used the delayed comparison paradigm to study factors affecting memory for pitch. For this paradigm, a standard tone (S) was presented, followed by a variable delay or interference (I) interval, followed by a comparison tone (C), which was either the same frequency as, higher than, or lower than the standard tone. The listener's task was to judge whether the S and C frequencies were the same or different.

The decline of accuracy of judgment as the I interval increases in duration may be regarded as a map of the memory process (Massaro, 1970a, b, c; Moss, Myers, & Filmore, 1970; Wickelgren, 1966, 1969). However, not all Es (e.g., Aiken & Lau, 1966; Thomas et al, 1970, Experiment I) have found evidence of such a decline. As was typical of earlier experiments demonstrating little or no forgetting, Aiken and Lau used a single S tone and a blank I interval. Harris (1952) concluded that the single or fixed-S method facilitated the development of a long-term memory for S that was little affected by the passage of time. Following Harris's argument, investigators wanting to study forgetting over short

time delays have employed the roving-S method—at least 2 (Moss et al, 1970), 3 (Massaro, 1970a; Wickelgren, 1966), or up to 20 (Wickelgren, 1969) S tones.

The purpose of the present experiment was to reexamine pitch memory under fixed- and roving-S conditions. A major methodological change was the use of bias-free measures of memory strength derived from the theory of signal detection rather than from the classical threshold measures employed by Harris. Pitch memory was studied over short (0 to 4 sec) I intervals under two interference conditions—a blank interval, as in Harris's study, and an I interval containing an interference tone (I-tone interval). With practiced listeners, it was expected to find reliable evidence of forgetting under roving-S conditions, with a faster rate of forgetting for the I-tone interval than for the blank interval (Massaro, 1970a). If the assumption is correct that a stable long-term memory for a single S tone is formed, very little forgetting should be evidenced under fixed-S conditions, either with a blank or an I-tone interval. Intratrial events may produce a degraded (i.e., noisy) trace for the S tone in short-term memory, but an ideal observer would compare the C tone with the relatively noise-free trace in the long-term store (cf. Sorkin, 1962).

There are alternative possibilities, however. For instance, auditory interference may adversely affect long-term retention of the S tone, leading to a decline in performance as a function of time delay under fixed-S combined with I-tone-interval conditions. The decline would not be as rapid as that found under roving-S conditions if partial information concerning the absolute value of the

fixed-S tone is retained in long-term storage.

In addition, investigators have noted that the pitch memory function is sensitive to the listener's strategy or set to rehearse the S tone (Massaro, 1970a; Wickelgren, 1969). If different groups of listeners are tested under fixed-S and roving-S conditions (as in Harris's study), differences in strategies between groups will contribute to experimental differences. A listener tested only under fixed-S conditions may be more likely to rehearse the S tone during the I interval than a listener tested only under roving-S conditions. In Experiment 1 all listeners were tested under all experimental conditions under identical instructions to refrain from rehearsing the S tone during the I interval. If, under these conditions, a long-term trace for the fixed-S tone is not formed, performance under fixed-S and roving-S conditions will estimate the same memory function.

## EXPERIMENT 1

### Method

Listeners were four paid female volunteers aged 20-26. They were given extensive training on experimental procedures before data for the main experiment were collected.

Sinusoidal tones were produced by a system of oscillators and transient-free sound switches adapted for automatic control by a BRS electronics logic system. The system was especially designed for high-frequency stability (better than  $\pm 1$ -Hz drift short-term) and low distortion characteristics (less than .2%). After selection and amplification, tones were presented binaurally to listeners through headphones (Sharpe Model HA-10). The intensity of each tone was set so that tones appeared equally loud to each listener, at a comfortable listening level.

Each trial consisted of a 2-sec S tone, followed by a variable I interval, followed by a 2-sec C tone. The C tone differed from the S tone by 0, +8, or -8 Hz, with the proportion of trials assigned to each difference being .50, .25, and .25, respectively. Within this constraint, the S-C difference on a particular trial was determined randomly. During the 5-sec intertrial interval, Ss were instructed to rate their judgment of the S-C difference on a 6-point scale ranging from "very sure same" to "very sure different."

Experimental variables were: (1) presentation condition—for the fixed-S condition, the S tone was always 800 Hz, while for the roving-S condition, the S tone was one of four equally likely stimulus tones—780,

\*The present study was supported by grants to the second author from the Defence Research Board, Grant No. 9425-17, and the National Research Council, Grant No. APA-333. The data of Experiment 1 were presented in a masters thesis submitted to Queen's University, Kingston, Ontario, Canada, by the first author. The authors thank Norman Chapman and Brian Holt, of Electrocraft of Canada Ltd., and E. T. Griffiths, of the Department of Psychology, Queen's University, for the design, construction and maintenance of acoustical equipment. Catherine Grimes assisted in the data collection of Experiments 2A and 2B.

†Requests for reprints should be sent to Lola L. Cuddy, Queen's University, Kingston, Ontario, Canada.

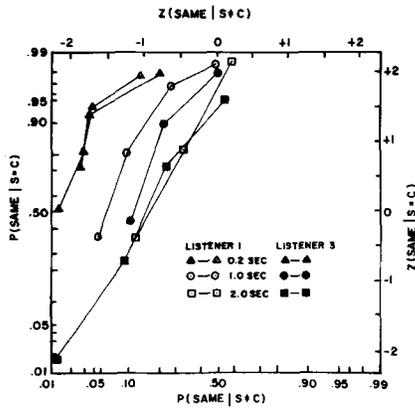


Fig. 1. Operating characteristics for two listeners for roving-S and I-tone interval conditions (Experiment 1).

795, 810, or 825 Hz; (2) I-interval duration—duration of the I interval under both fixed- and roving-S conditions was randomly selected from one of six equally likely alternatives—0.0, 0.2, 0.5, 1.0, 2.0, or 4.0 sec; (3) I-interval type—the I interval was randomly determined either to be blank or to contain a 930-Hz I tone.

Each listener was trained and tested under all possible combinations of the experimental variables. At each session, listeners were presented with one block of 96 fixed-S trials and one block of 384 roving-S trials, with order of blocks counterbalanced across sessions and listeners. Each block was preceded by a short series of at least 10 practice trials. Listeners were given a brief rest break between blocks and halfway through the roving-S block. Each of the 96 fixed-S trials contained a random combination of one of the two I-interval types and one of the six I-interval durations, with the restrictions that each combination occur eight times over the 96 trials and no particular value of I-interval duration or I-interval type occur more than three times in succession. Each of the 384 roving-S trials contained a random combination of one of four S tones, one of two I-interval types, and one of six I-interval durations. There were, in all, eight replications of each combination and no more than three successive identical values of I-interval type or I-interval duration.

Listeners underwent 12 practice sessions followed by 18 experimental sessions; the experimental sessions represented a total of 144 observations per experimental condition. At each session, listeners were reminded, "Do not rehearse, vocally or otherwise, the S tone during the I interval."

#### Results

The technique for constructing

receiving operating characteristic (ROC) curves from recognition data has been thoroughly described by Wickelgren (1966, 1969). Recognition probabilities were cumulated across rating categories. The normalized cumulative probabilities for the condition  $S=C$  were treated as inferred hit proportions,  $P(\text{same} | S=C)$ , and were plotted against normalized cumulative probabilities for the condition  $S \neq C$  treated as inferred false-alarm proportions,  $P(\text{same} | S \neq C)$ . Twenty-four ROC curves were plotted for each listener, one for each combination of presentation condition, I-interval type, and I-interval duration. Inspection of the ROC plots yielded the following information: Under all conditions, discrimination was very high when the I interval was 0.0 and 0.2 sec. Ratings assigned for these conditions were generally restricted to "very sure same" or "very sure different," with minimal overlap between distributions. Figure 1 shows sample ROC curves obtained from Listeners 1 and 3 under roving-S and I-tone interval conditions at 0.2, 1.0, and 2.0 sec duration. Five of the six ROC functions plotted on normal-normal coordinates show a curvilinear trend of peaking near the midpoint of the function. Peaking was noted in almost all ROC curves based on at least three points (27 out of 33 cases). The data for Listener 3 at 2.0 sec appear to fall along a straight line of slope greater than one. Over all listeners, the only relation between shape and slope of the ROC curve and experimental condition was a tendency for the function to flatten as performance approached the lower

limit, i.e., the chance diagonal.

The curvilinear trend of the ROC functions suggested that the assumption of normal functions for the underlying probability distributions and the calculation of the normal deviate measure  $d'$  was not entirely suitable for the data. A nonparametric measure of sensitivity, area under the ROC curve, was geometrically computed from the ROC curve plotted on linear coordinates. The area measure can be shown to be mathematically equivalent to percent correct in a two-alternative forced-choice paradigm (Green & Swets, 1966). For chance performance, area = .50; for errorless performance, area = 1.00.

Area under the ROC curve, averaged across four listeners, is plotted against I-interval duration for the four combinations of presentation condition and I-interval type in Fig. 2. Analysis of variance of the data indicated that the main effect of I-interval duration and the interaction between I-interval duration and I-interval type was significant. For I-interval duration,  $F(5,15) = 26.71$ ,  $p < .001$ ; for the interaction between I-interval duration and I-interval type,  $F(5,15) = 7.70$ ,  $p < .001$ . In Fig. 2, it can be seen that all functions commence at area = 1.00 and decrease regularly as I-interval duration increases, with a more rapid decrease for the I-tone-interval conditions than for blank-I conditions.

Neither the main effect of presentation condition nor any of the interactions with presentation condition achieved statistical significance at the .05 level. Linear and quadratic components of interactions

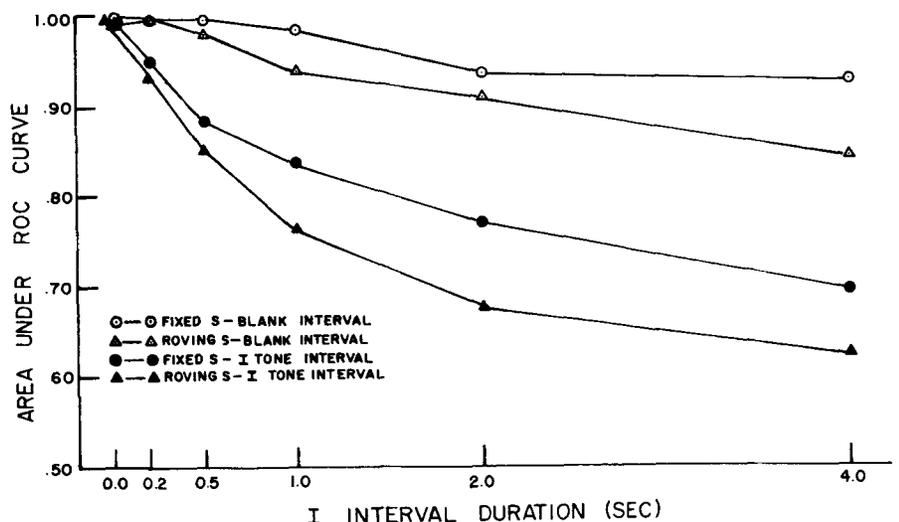


Fig. 2. Area under the ROC curve, averaged across four listeners, as a function of I-interval duration for four combinations of presentation condition and I-interval type (Experiment 1).

were also not significant. However, inspection of Fig. 2 indicates that the data are in the direction of a slower forgetting rate for fixed-S conditions, and suggests further examination of individual listener functions.

Figure 3 presents individual forgetting functions for each of the four listeners. All functions show that memory for both fixed and roving stimulus tones is adversely affected by the passage of time and by auditory interference. It is also evident from Fig. 3 that there are substantial individual differences in rate of forgetting as a function of presentation condition; the presence of Subject by Condition interactions reduces the power of the F test of significance (Myers, 1966). The data show divergence of fixed- and roving-S functions under all conditions for Listeners 2 and 3 and for Listener 4 where an I tone is presented. Much less difference between presentation conditions is noted for Listener 1, but, given that performance for this listener is essentially perfect under all conditions at  $t = 0.0$  sec, slower forgetting under fixed conditions may be inferred from the finding that fixed performance exceeds roving on 8 of the remaining 10 possible comparisons and is equivalent on one (binomial  $p < .02$ ). The data for Listener 4, blank interval, are equivocal regarding the difference between fixed and roving conditions. Accuracy of judgment was very high (producing a possible ceiling effect). Superior performance under fixed conditions was noted for this listener for only

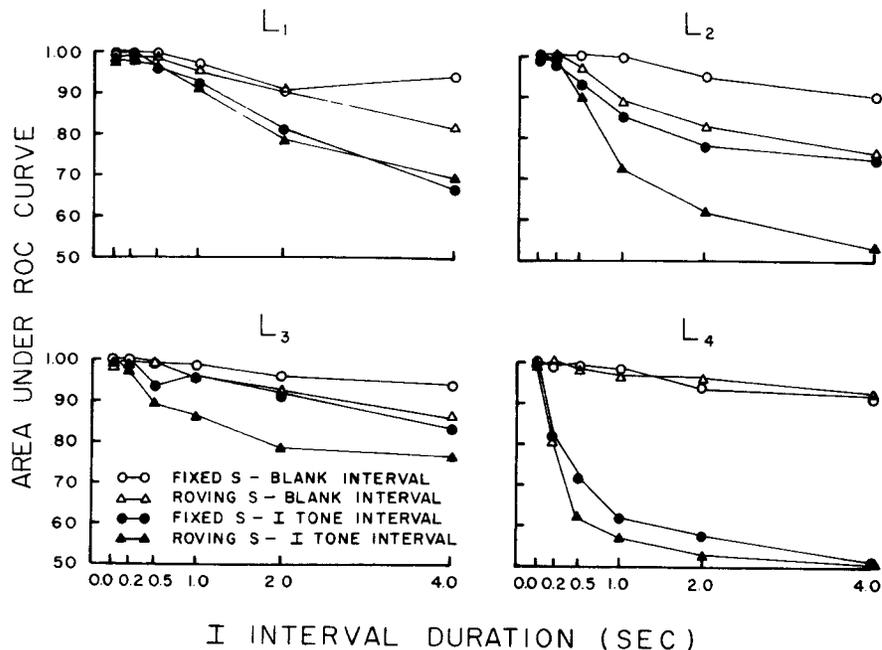


Fig. 3. Individual forgetting functions (Experiment 1).

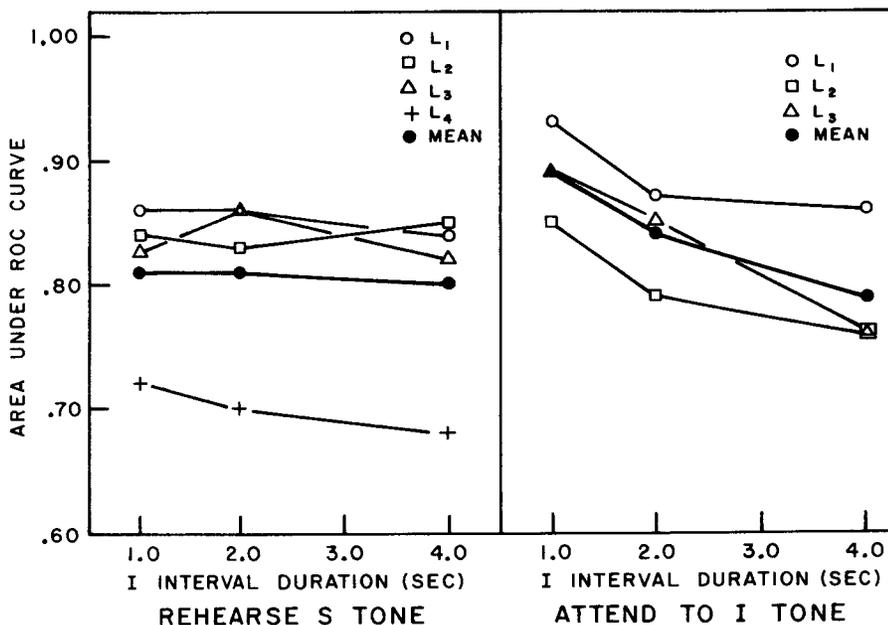


Fig. 4. Forgetting functions obtained under rehearsal and nonrehearsal instructional conditions (Experiments 2A and 2B).

two of the five comparisons after  $t = 0.0$  sec. Had the difficulty of the pitch discrimination task been increased for Listener 4, however, performance under roving-S conditions would likely have approached chance before  $t = 2.0$  sec. Listener 4 was the most sensitive to auditory interference.

In summary, three listeners showed slower forgetting under fixed-S than under roving-S conditions; for a fourth

listener, the conclusion holds for the filled I-tone interval. There were considerable individual differences in forgetting rate under fixed- and roving-S conditions and in size of decrement attributable to auditory interference.

Individual differences were not related to amount of prior musical training, nor with information transmission scores later obtained from an experiment in absolute judgment of pitch. All listeners reported that they followed the instructions not to rehearse the standard tone or tones.

A further analysis, conducted on percent correct responses for each of the 18 experimental sessions, agreed with the above findings. It was also found that performance across sessions was stable, with no practice effects or Practice by Presentation Condition interactions discernible. This finding was not unexpected, since the listeners were highly trained before the experimental data were collected.

#### EXPERIMENT 2

The data of Experiment 1 were collected under instructions not to rehearse the S tone, and presumably the results are dependent upon this instructional condition. Experiment 2 was conducted to determine whether instructions to rehearse do affect the shape or regularity of the memory function.

#### Method

Test procedure was similar to that for the roving-S/I-tone-interval condition described in Experiment 1,

except that tones were recorded and reproduced through Sharpe HA-10 headphones by a Roberts tape recorder, Model 990. The roving-S frequencies were 725, 755, 785, and 815 Hz, the I-tone frequency was 930 Hz, and the S-C difference was either 0 or +10 Hz.

Four test tapes were recorded, each consisting of 48 trials. For each trial, the duration of the I interval was selected randomly from one of three alternatives, 1.0, 2.0, or 4.0 sec, with the restriction that each duration occur an equal number of times. The value of the S-C difference was also determined randomly, with the restriction that each value occur equally often.

For Experiment 2A, listeners received a practice session of 192 trials, then each of the four test tapes at each of five experimental sessions, with the order of tapes randomized across listeners and sessions. Each listener made 960 judgments in all, 320 per I duration. Experimental instructions permitted listeners to "attend to or rehearse the S tone during the I interval, and to obliterate the I tone," but forbade singing aloud. Listeners were four students, one male and three female, ranging in age from 17 to 25.

Experiment 2B was conducted to ensure that listeners tested in Experiment 2A would produce regular forgetting functions under nonrehearsal instructions (cf. Experiment 1). After a rest break of several days, each of three listeners (the fourth was unfortunately no longer available) made an additional 960 judgments as above. Instructions directed the listener to attend to the I tone during the I interval and not rehearse the S tone.

### Results

The results of Experiments 2A and 2B are presented in Fig. 4, where area under the ROC curve is plotted against I-interval duration in seconds. For Experiment 2A (left-hand panel), where instructions emphasized rehearsal of the S tone, the mean effect of I interval is negligible; two listeners show a reversal in trend in their individual forgetting functions. However, for the three listeners who subsequently judged under instructions to attend to the I tone (Experiment 2B, right-hand panel), a regular decrease in performance is obtained as a function of I-interval duration.

The results indicate that instructional conditions do affect the shape of the forgetting function. Whether or not rehearsal has a facilitative or interfering effect is, as Wickelgren (1969) has suggested,

dependent upon the I-interval duration. What is clear is that even under roving conditions, regular forgetting functions are most reliably obtained when listeners are instructed to attend to the I tone. Nonrehearsal instructions were employed by Wickelgren and by Massaro, but many other investigators do not report instructional conditions. Without such information, it is difficult to interpret previous failures to find forgetting in the delayed comparison of pitch. In pilot work, it was noted that in the absence of instructions many listeners would later report that they adopted a rehearsal strategy.

### DISCUSSION

The main conclusion is that under nonrehearsal instructions the differences between fixed- and roving-S conditions attributable to time or to interference are quantitative, not qualitative. All four listeners in Experiment 1 showed that memory for both fixed and roving standards deteriorates over short retention intervals and is adversely affected by auditory interference. The data may be interpreted in terms of a model for intensity resolution proposed by Durlach and Braid (1969). According to Durlach and Braid, two modes of operation are available to the listener in the delayed comparison task—a sensory-trace mode where the amount of noise or trace diffusion is a function of time and amount or type of extraneous interference, and a context-coding mode where noise results from imprecise representation or abstraction of the stimulus range and is assumed to increase with increasing width of range. Where information is available from both sources, the listener forms his decision using information from the least noisy source or some optimal combination of both sources. In Experiment 1, sensory-trace diffusion is evident in the data of all listeners, listeners differing only in rate of trace diffusion and degree of susceptibility to interference. Experiment 2A suggests that trace diffusion may be prevented or retarded by maintaining the trace in a rehearsal loop. Operation of the context-coding mode is reflected in Experiment 1 where fixed-S performance exceeds roving-S performance. A plausible prediction is that the difference between fixed-S and roving-S conditions will depend upon the stimulus range of the roving standards. The individual parameter here may well be the value of a frequency bandwidth which must be exceeded before the context-coding information for roving standards is sufficiently degraded to produce experimental differences. It may be

noted that this prediction would not follow from a model which simply postulated a long-term or different type of trace under fixed-S conditions.

What about the stable memory for pitch presumably underlying "absolute pitch" (Harris, 1948)? Such highly accurate judgment of pitch is undoubtedly present among musicians, but experiments demonstrating successful training of absolute judgment by training recognition of a reference tone (Cuddy 1968, 1970, 1971) differ from Experiment 1 in at least two ways: (1) rehearsal, including vocalizing, of the reference standard is encouraged, and (2) the reference standard is embedded in an experimentally defined structural set of tones. Improvement in judgment for the reference standard was always accompanied by improvement in judgment of the entire set. It was argued in these earlier studies that training establishes structural decision rules organized about the reference standard; without such rules, mere repetition of a standard tone is ineffective in producing improvement in absolute judgment.

Siegel (1971) has found that musicians who possess absolute pitch do not necessarily show superior pitch-discrimination ability in a delayed comparison task, and argues against the "superdiscrimination" hypothesis of absolute pitch. Thus, the trace component of a standard tone may be subject to temporal decay, but its context code may be developed through musical or other training so that task performance dependent on the precision of the context code may be greatly improved. In more general situations, if experiments are designed so that the context code provides sufficient information for accurate judgment, or if the sensory trace may be preserved through a rehearsal loop, it would not be surprising to find evidence for a stable memory for a standard tone.

We may ask whether or not listeners in Experiment 1 would have shown decay under fixed-S conditions if they had not also been tested under roving-S conditions. The within-Ss within-sessions design was chosen so as to measure performance under as nearly equivalent motivational sets or strategies as possible. Extensive pilot work indicated that decay of fixed-S functions would be obtained even if it were the only condition tested at an experimental session (always provided the listener reported he refrained from rehearsal). This procedure was not continued, however, because pilot listeners reported large differences in motivation and fatigue between sessions and we did not want to achieve a "constant set" by discarding

data on the basis of introspective evidence. Another point worth noting is that Listeners 2 and 4 show fixed-S forgetting even though roving-S/I-tone-interval functions approach the chance level. If memory for the roving standards dissipates over 4 sec during the roving block, it is difficult to argue that interference in memory from these roving standards produced the fixed-S decay.

"Peaking" in the shape of ROC curves (Fig. 1) has also been noted by other investigators, particularly for ROC functions derived from response latencies (Moss et al, 1970; Norman & Wickelgren, 1969). The ROC curves for the confidence rating data reported by Moss et al (1970) show peaking for O3, and for all other Os where performance is good ( $d' > 1.00$ ). A number of theoretical possibilities might be advanced; one plausible explanation (cf. Wickelgren, 1968) is that listeners first make a binary judgment with low criterion variance and then categorize the initial judgment with somewhat greater criterion variance. The criterion for the binary judgment is usually, though not necessarily, the "same-different" point on the rating scale, and the lower variance yields an elevated point on the ROC curve. Further investigation of criterion variability in

rating procedures is warranted. In the meantime, area under the ROC curve provides a useful nonparametric description of performance.

#### REFERENCES

- AIKEN, E. G., & LAU, A. W. Memory for the pitch of a tone. *Perception & Psychophysics*, 1966, 1, 231-233.
- CUDDY, L. L. Practice effects in the absolute judgment of pitch. *Journal of the Acoustical Society of America*, 1968, 43, 1069-1076.
- CUDDY, L. L. Training the absolute identification of pitch. *Perception & Psychophysics*, 1970, 8, 265-269.
- CUDDY, L. L. The absolute judgment of musically-related pure tones. *Canadian Journal of Psychology*, 1971, 25, 42-55.
- DURLACH, N. I., & BRAIDA, L. D. Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 1969, 46, 372-383.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- HARRIS, J. D. Discriminability of pitch: Suggestions toward method and procedure. *American Journal of Psychology*, 1948, 61, 309-322.
- HARRIS, J. D. The decline of pitch discrimination with time. *Journal of Experimental Psychology*, 1952, 43, 96-99.
- MASSARO, D. W. Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1970a, 83, 32-39.
- MASSARO, D. W. Forgetting: Interference or decay? *Journal of Experimental Psychology*, 1970b, 83, 238-243.
- MASSARO, D. W. Consolidation and interference in the perceptual memory

- system. *Perception & Psychophysics*, 1970c, 7, 153-156.
- MOSS, S. M., MYERS, J. L., & FILMORE, T. Short-term recognition memory of tones. *Perception & Psychophysics*, 1970, 7, 369-373.
- MYERS, J. L. *Fundamentals of experimental design*. Boston: Allyn & Bacon, 1966.
- NORMAN, D. A., & WICKELGREN, W. A. Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology*, 1969, 6, 192-208.
- SIEGEL, J. A. The nature of absolute pitch. Unpublished doctoral dissertation, University of Michigan, 1971.
- SORKIN, R. D. Extension of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America*, 1962, 34, 1745-1751.
- THOMAS, I. B., DINICOLA, P. D., ELIA, M. P., & PASIERBIAK, C. S. Auditory memory in a pitch-discrimination task. *Journal of the Acoustical Society of America*, 1970, 48, 1383-1385.
- WICKELGREN, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1966, 72, 250-259.
- WICKELGREN, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, 1968, 5, 102-122.
- WICKELGREN, W. A. Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, 1969, 6, 13-61.

(Received for publication September 17, 1970.)