 JRME 1986, VOLUME 34, NUMBER 4, PAGES 238-261

One hundred and thirty-six theory students participated in a melodic dictation experiment. Subjects were randomly assigned to one of six dictation strategy groups contrasted by simultaneous writing, concentration before notation, and singing before writing. The task within the experiment consisted of notating 12 melodies presented aurally. Strategies were combined across single and dual melodic presentations to see if any of three methods were significantly more effective regardless of the number of presentations. Results of the analysis revealed no significant differences (.05) among the strategies. When the six groups were analyzed individually, significantly higher scores were revealed for Groups 4 (writing while hearing the melody twice) and 5 (writing after hearing the melody twice). Overall, written response accuracy for the six groups was 48%. Subjects in groups that sang reproduced the melodies at a 43% level of correctness. Their written responses matched their vocal responses 61% of the time.

Randall G. Pembrook, *Center for Music Research, Florida State University*

Interference of the Transcription Process and Other Selected Variables on Perception and Memory During Melodic Dictation

The processing and storage abilities of human beings seemingly have always been the objects of investigation. Perception of sounds in a melodic context and the memory patterns for those sounds have elicited much interest. Theories of musical memory and perception reflecting various degrees of complexity and development have been attempted by Connor (1970), Larsen (1973), Tallarico (1974), Deutsch (1975), Dowling (1978), Cuddy (1982), and Moog (1982). Still others have attempted to measure memory patterns in various melodic contexts (Deutsch, 1982; Dowling, 1978; Holly, 1980; Long, 1977; Madsen & Staum, 1983; Pembrook, 1983; Taylor, 1976; Williams, 1975). Several aspects of

For reprints of this article, contact Randall G. Pembrook, Center for Music Research, Florida State University, Tallahassee 32306.

memory specific to music have been noted. Among them are the following:

1. The maintenance of newly arrived information in memory requires the use of a limited capacity system.

2. Such stored information decays more rapidly if exposed to subsequent aural musical events.

3. Increasing interpolated tones during the retention interval increases memory loss (or decay rate).

One task in music that requires the use of perception and memory is that of melodic dictation. Melodic dictation and its inverse, sight singing, are usually thought to be essential to musicians. Dictation is not seen as an end in itself but rather a means toward achieving a more important skill for the musician, that of aural acuity. McGaughey (1961) stated that "qualified musicians must develop reading, singing, *and* [italics added] notational skills in order to achieve acuity of aural perception and to make effective use of that acuity" (p. i).

Otto Ortmann, a pioneer in music research, investigated melodic perception and the problems specifically encountered when dealing with melodic dictation. His set of studies entitled *Research studies in music: Problems in the elements of ear dictation* (1934) dealt with melodic, harmonic, and rhythmic dictation, and are particularly insightful regarding areas of difficulty for music students taking dictation. However, one area in which Ortmann's theories seem dubious includes the initial perception of the aural events. Ortmann claimed that of the dictation mistakes made by his students, only 13% were a result of improper perception of the stimuli. He suggested that most mistakes were caused by improper interpretation or manipulation of the notational symbols (e.g., notating a perfect fourth when actually thinking of the sound of a major third).

Murdock (1974), however, stated that incorrect reports are a function of interference, which accounts for 85–98% of variance, with memory decay being responsible for the remainder. This interference may be caused by previously heard melodies and also parts of the present melody interfering with memory for other parts.

In reference to musical errors of recognition, Cuddy (1982) has gone so far as to state that "recognition mistakes may in fact not be an error of memory. It may reflect the richness or complexity of the cognitive structures that are being evaluated at that time" (p. 27). Without specific inquiries (such as having the student sing back the "perceived" melody), any categorization of inaccurate reports as a function of improper perception or misuse of the notational system is mere speculation.

Although noting other approaches, Ortmann (1934) found that most students in his study employed a dictation strategy wherein they tried to listen *and* write simultaneously. The efficacy of this approach, however, includes the question of multidimensional task abilities discussed by those such as Neisser (1976) and Neisser, Hirst, and Spelke (1981). They have suggested that the simultaneous performance of two or more tasks is jeopardized if one task suddenly increases in difficulty. If processing of incoming material ceases when the student cannot instantaneously notate difficult intervals, the progressive method of dictation probably should be forbidden as an option for melodic dictation students even though Ortmann cited it as the most preferred method.

Perhaps because of the seeming complexity of simultaneous listening and writing, and problems caused by large skips, most ear training texts do not endorse this approach. Dictation texts such as those by Benward (1961), Edlund (1963), Herder (1973), and Trubitt and Hines (1979) have embraced various approaches with most calling for audio reinforcement through singing or clapping after the melody has ceased.

One problem that may be encountered as a result of listening to the entire melody before attempting to write, as the previously mentioned texts suggest, is that of memory capacity. Miller (1956) found the short-term memory capacity to be approximately seven bits. That these bits are somewhat equivalent to notes can be substantiated by studies of Tallarico (1974), Long (1977), and Pembrook (1983), which indicated the short-term memory limit to be somewhere between 7 and 11 notes. Thus students may not be capable of storing very long melodies.

The combination of the perceptual problems associated with immediate writing and the memory limits encountered as a result of delayed transcription produces a dilemma for the melodic dictation instructor. What type of approach should be recommended to students and is one particular approach consistently better for different types of melodies?

It was the purpose of this study to investigate variations in students' dictation scores as a result of different transcription procedures. Because of the paucity of research in the field of melodic dictation, the following hypotheses were stated in the traditional null form even though the majority of texts suggested the superiority of the concentration or singing approaches:

1. There will be no differences in the accuracy of student responses using three different dictation techniques (progressive, concentration and singing) on selected melodies.
2. There will be no difference in notational accuracy by subjects as a result of single presentations and repeated presentations.
3. There will be no difference in the accuracy of responses between freshmen and sophomore theory students either vocally or in written form.
4. There will be no difference in the accuracy of student responses on the second half of melodies containing certain selected difficult intervals on the second half of melodies containing only conjunct motion.
5. During warm-up exercises (prior to experimental group assignment) there will be no difference in the number of students already using each of the three dictation methods.

This study also attempted to address the question of whether dictation errors are a result of perceptual problems or represent a misuse of the notational system.

METHODS AND PROCEDURES

The major research question in this study focused on the approaches (strategies) used during the stimulus presentation period in both single and multiple presentation settings. A second area of interest concerned possible differences in the efficacy of these strategies in relation to certain melodic characteristics. The melodic factors that were selected for study included tonality (tonal and atonal), melodic length (6, 10, and 16 notes), and motion. Two types of motion were used. One contained only seconds and thirds while the other consisted of seconds, thirds, and a difficult skip near the center of the melody. The tritone, minor sixth, and minor seventh were the difficult skips selected to initiate a processing problem.

Many additional factors such as contour, tempo, and rhythm were not included for experimental investigation in this study because of the extremely long testing period that would have been required. Therefore, attempts were made to simplify these elements as much as possible according to related literature in each of the areas. Contour was restricted to the A form (see Long, 1977; Pembrook, 1983), which begins in an ascending fashion and changes direction only once thus ending in a descending manner. All melodies were played at a tempo of ♩ = 90, which has been recommended as an optimum tempo by Ortmann (1934) and Pembrook.

Sink (1983) found that simultaneous presentation of rhythm and melody reduced attention to rhythmic structures. To completely remove rhythm would have been artificial so an attempt was made to simplify rhythm according to suggestions by Hofstetter (1981). He reported that the greatest dictation accuracy occurred in simple time signatures and in melodies where non-dotted or dotted rhythms were used exclusively rather than combined. Therefore, only simple ($\frac{2}{4}$, $\frac{3}{4}$, and $\frac{4}{4}$) meters were used. Rhythms within this context of meter were limited to whole, half, quarter, and eighth notes. Two dotted half notes also were included to avoid the use of rests in $\frac{3}{4}$ examples.

The 12 melodies used in the study are illustrated in Figure 1 along with their randomly selected presentation order.

Because subjects from two groups were asked to sing the melodies, the melodies required a restricted range. The largest melodic range used was an octave and a third (minor tenth). While this was acknowledged to be a larger range than desired, a 16-note melody with an A contour and a skip of at least a tritone necessitated it. To restrict the range on the long melodies, seconds were used almost exclusively. Thus the longer melodies contain very few thirds. However, because seconds and thirds were considered to represent conjunct motion, it was assumed that this would not seriously interfere with the research design. To facilitate singing, all melodies were transposed from the key in which they were listed in the various dictation texts (Edlund, 1963; Herder, 1973; McHose & Tibbs, 1945) so that the lowest pitch never fell below Great B(♭) (A(♯)) for men and small B(♭) for women. The highest note in any of the melodies was C(♯)^1 for men and C(♯)^2 for women. Different beginning tones for each melody were used to limit any sense of key orientation and therefore possible interference between trials.

The melodies selected for this study were recorded from a Bösendorfer piano. All melodies were originally recorded on Scotch 250 tape using a Technics 1500 and a pair of Newmann KM84 microphones. A DBX Model 155 noise reduction system was used to create the cleanest possible sound. From this tape, directions and melody numbers were

Figure 1. Warm-up melodies, experimental melodies, and presentation order. T = tonal, A = atonal, C = conjunct, and D = disjunct.

inserted to create the stimuli cassette tapes. An Aiwa 5D-L5OU cassette recorder was used for presentation in conjunction with a set of KEF Model 15 speakers, located approximately 5 ft (1.52 m) from the subject.

Subjects for this study consisted of freshmen and sophomore music theory students at Florida State University. All subjects were randomly assigned to one of the six treatment groups and tested individually. Before beginning the experiment, each subject was asked if he or she had perfect pitch. Those responding affirmatively were asked to identify the note names of several pitches played on the piano by the experimenter. Data for subjects found to have perfect pitch were not eliminated from the study but were evaluated separately.

Because sung responses were essential to this study, it was necessary to ensure that subjects could vocalize pitches that they were thinking. Therefore two tests were administered to subjects in those groups required to sing (Groups 3 and 6). The first simply consisted of having female subjects sing from small B(♭) to D^2 and male subjects sing from Great B(♭) to D_1 to insure that each subject could cover the range needed for the vocal section of the test. In addition the Taylor Music Perception Ability Test (TMPAT), which specifically tests for the ability to sing what one is thinking, was administered. (For a discussion of the TMPAT, see Taylor, 1976.) Data from subjects who failed to achieve a 90% rating on the TMPAT test (9 of the 10 sung tones being not more than 50 cents higher or lower than the frequency of the correct tone) and also data from those unable to sing the required range of tones were disregarded in comparing vocal and written accuracy and determining perception accuracy.

After completing the TMPAT, students were given an answer sheet that presented them with a staff, beginning note, time signature, and key signature for 15 examples. Three of these examples were warm-up exercises and the other 12 were the experimental melodies. The object of the first two warm-up melodies (aside from allowing the student to become comfortable with the experimental environment) was to examine the student's dictation method and compare actual behaviors to Ortmann's statement that most students use the progressive method of dictation. The third warm-up melody was presented after the student had received directions on how to take dictation to ensure that the directions were understood.

Because six different strategies were employed in the study, there were six different sets of directions used. The specific tasks for the first three groups are compared in Figure 2. Tasks for Groups 4 to 6 were identical to Groups 1 to 3 except that the entire process was repeated for each melody. So that the total response time was identical in all strategies, subjects in Groups 4 to 6 received only 22, 45, and 60 seconds of silence, respectively, after each hearing.

Before each melody, the student heard only the first pitch. In some classrooms a harmonic orientation was used (e.g., I-IV-V-I) to prepare the student for the melodic dictation example. This was avoided in the experiment because key orientation was impossible for the atonal melodies used in this study. Therefore, any such key orientations before

Group 1	First tone in melody sounded	2 sec silence	2 measures metronome clicks	melodic presentation; subject simultaneously notates		45, 90, or 120 sec of silence depending on the melodic length for written response		announcement of next melody
Group 2	First tone in melody sounded	2 sec silence	2 measures metronome clicks	melodic presentation		45, 90, or 120 sec of silence depending on the melodic length for written response		announcement of next melody
Group 3	First tone in melody sounded	2 sec silence	2 measures metronome clicks	melodic presentation	2 sec silence	beginning tone	subject sings	45, 90, 94 120 sec of silence depending on the melodic length for written response	announcement of next melody

Figure 2. Task analysis by strategy (Groups 1 to 3). Groups 4 to 6 executed these steps twice.

only tonal melodies might have biased the subject toward more correct responses in these melodies. All melodies began on the downbeat and were preceded by two measures of the ensuing melodic tempo as indicated by the clicking of a metronome. This clicking ceased when the melody began.

On the two warm-up melodies subjects were allowed to respond in any fashion they preferred. On the third warm-up melody and the ensuing 12 experimental melodies they used the assigned strategy. Before attempting to sing each melody, all subjects employing the singing strategy received the first tone, which had been recorded utilizing the piano to provide them with a reference.

During the experiment, the students' vocal responses were recorded using a Technics 1520 reel-to-reel recorder and a pair of Shure 545S Unidyne III microphones. A model 6T5 Strobo Conn Tuner was used to test the calibration of the Technics recorder to ensure that the subjects' vocal responses were recorded in order and reproduced at the frequency at which they were originally sung. All sound production was controlled by a Crown DL2 Distinction Stereo Controller (Preamplifier) and a PSA-2 Professional Self-Analyzing Amplifier.

Subjects in Groups 1 and 2 took approximately 25 minutes to complete the melodies while those in Groups 3 through 6 usually required 30 to 40 minutes to finish. Answer sheets were collected by the test administrator upon completion of the task.

Determining correct responses is problematic when interpreting transcribed pitches. For the purposes of this experiment, a response was considered correct if it represented the correct note *or* if the second note of each pair constituted a correct interval from the previous note (even if the preceding note was missed, thus creating an absolute pitch error for the second note). This approach to data analysis was thought to be particularly important in assessing memory after large skips. If student responses were evaluated only upon absolute pitch, large skip errors seemingly would guarantee other errors. This represented an unjustifiable approach.

In assessing vocal accuracy, all tones were transcribed to the nearest pitch (e.g., a vocalized pitch determined to represent 449 Hz was considered as an A because it was less than 50 cents from the given tone of A 440 and thus closer to it than to any other note). This transcription was accomplished by using a pitch extractor which translated recorded analog input into digital output and then classified the frequency as one of the 12 notes of the music alphabet. Transcribed vocal tunes were evaluated in an identical manner as notated responses.

RESULTS

The *N* for this study was 136 subjects (76 college freshmen and 60 sophomores) with 22 assigned to each of the six treatments and 4 extras, 3 of which were in Group 1 and the remaining subject assigned to Group 2. Data collected represented subject responses in both written and sung form. Only 1 of the 44 subjects in the singing groups scored below the acceptable level in the TMPAT. Furthermore, two subjects failed to sing at a level loud enough to be recorded. Therefore, vocal responses were based on an n of 41. The written data from the student who had scored below the acceptable TMPAT level were included for data analysis.

Hypothesis 1: There will be no differences in the accuracy of student responses using three different dictation techniques (progressive, concentration, and singing).

Because the three strategies were the important elements to be evaluated in this study—regardless of the different number of presentations—data from Groups 1 and 4 (listening while simultaneously writing) were combined. Similar action was taken with data from Groups 2 and 5 (listening then writing), and Groups 3 and 6 (listening, singing, then writing). Data were based on the total number of correct responses on the 12 experimental melodies and the third warm-up melody (where subjects practiced the assigned strategy). This totaled 125 responses.

Results from the one-way analysis of variance are shown in Table 1. The value of $F(2,133) = 1.14$ was determined to be less than the value required to reject the null hypotheses at alpha = .05. (The actual associated probability was .32). Therefore the decision was to fail to reject null hypothesis 1. The order of means was Strategy 1 (writing while hearing, $M = 70.96$), Strategy 2 (listening then writing, $M = 68.96$), and Strategy 3 (listening, singing, then writing, $M = 65.39$).

To see if there were significant differences among the six groups when analyzed separately, a second analysis was carried out. In addition, each of the six strategies was retained rather than condensed as before to see if there were significant interactions between the various levels of group and the selected melodic factors. These factors included length (6, 10, and 16 notes), motion (conjunct and disjunct), and tonality (tonal and atonal).

The BMDP8V—Analysis of Variance Computer Program with the four factors of group (G), length (L), tonality (T), and motion (M), and subjects nested within groups was used to evaluate this question. Data were limited to subjects' responses on each of the 12 experimental melodies.

The results are shown in Table 2. All four factors (main effects) were found to be significant, $p < .05$. The group factor was significant with $F(5, 126) = 6.24$, $p < .001$. The means for groups (ordered highest to lowest) were as follows (scores represent percentage correct): Group 4 (writing while listening, two presentations), $M = 59.20$, Group 5 (listening then writing, two presentations), $M = 56.17$, Group 6 (listen-

Table 1
ANOVA Summary Table for Differences in Accuracy for Three Dictation Strategies

Source of variance	Sum of squares	*df*	Mean squares	*F*	*p*
Strategy	721.52	2	360.76	1.14	0.323
Within	42075.88	133	316.36		

Table 2
ANOVA Summary Table for the Factors of Group (G), Length (L), Tonality (T), and Motion (M)

Source of Variance	Sum or squares	*df*	Mean squares	*F*
Mean	3945114.02	1	3945114.02	2266.71
Group (G)	54266.40	5	10853.28	6.24*
Length	224370.80	2	112185.40	272.31*
Tonality	43753.65	1	43753.55	133.75*
Motion	13399.37	1	13399.27	42.68*
Subjects within Groups	219297.96	126	1740.46	
G-L	2594.50	10	259.45	.63
G-T	2358.15	5	471.63	1.44
L-T	35666.00	2	17833.00	60.92*
G-M	2653.15	5	530.61	1.69
L-M	13788.72	2	6894.36	23.56*
T-M	4633.01	1	4633.01	17.04*
S-L(G)	103818.96	252	411.98	
S-T(G)	41217.12	126	327.12	
S-M(G)	39561.48	126	313.98	
G-L-T	4259.20	10	425.92	1.45
G-L-M	3282.50	10	328.25	1.12
G-T-M	1911.25	5	382.25	1.41
L-T-M	42723.10	2	21361.55	71.80*
S-L-T(G)	73770.48	252	292.74	
S-L-M(G)	73745.28	252	292.64	
S-T-M(G)	34260.66	126	271.91	
G-L-T-M	6888.70	10	688.87	2.32
S-L-T-M(G)	74975.04	252	297.52	

*n <. 0 5
*p <. 0 5

ing, singing, then writing, two presentations), M = 49.85, Group 3 (listening, singing, then writing, single presentation), M = 44.81, Group 1 (writing while listening, single presentation), M = 44.72, and Group 2 (listening then writing, single presentation), M = 44.70. Scheffe's Test for Multiple Contrasts (Hicks, 1973) was used to determine where significant differences (.05) in means occurred. As can be seen in Figure 3, Group 4 differed significantly from Groups 6, 3, 1, and 2; and Group 5 differed significantly from Groups 3, 1, and 2. All other mean comparisons were not found to be significantly different. The factor of groups accounted for 2% of the sample variance.

For the factor of length, also found to be significant (F = 272.31, df = 2, 252, p < .0001), Scheffe's Test was again used. Levels were found to differ significantly (.05) between the 6- and 10-note melodies and also between 10-and 16-note melodies as well as between 6-and 16-note lengths with means decreasing as length increased (6-note melody = 62.16, 10-note melody = 53.77, and 16-note melody = 33.78). The factor of length accounted for 13% of the sample variance.

The factor of tonality also was significant (F = 133.75, df = 1, 126, p < .0001). The scores on tonal melodies (M = 55.16) were significantly higher than those on atonal melodies (M = 44.65). Tonality accounted for 3% of the total variance.

Finally, motion was found to be significant (F = 42.68, df = 1, 126, p < .0001). Accuracy on conjunct melodies (M = 52.81) was significantly higher than accuracy on the disjunct ones (M = 47.00). This factor, however, accounted for only 1% of the variance.

There were three two-way interactions significant at p<.01. These included length-tonality (F[2, 252] = 60.92), length-motion (F[2, 252] = 23.56), and tonality-motion (F[1, 126] = 17.04). The graphs of these two-way interactions are shown in Figures 4, 5, and 6. In the length-tonality interaction, mean differences decreased as melodic length increased such that as the melody reached the 16-note length, the means were nearly equal (tonal = 34.42, atonal = 33.15). The length-motion interaction graph illustrates a diverse pattern with the means being nearly equal at the shortest (6-note conjunct = 62.80; 6-note disjunct = 61.52) and longest (16-note conjunct = 34.79; 16-note disjunct = 32.78) lengths, but quite different at the 10-note length (conjunct = 60.85, disjunct = 46.70). No notable patterns were evident in the tonality-motion interaction, with the two lines being nearly parallel.

The three-way interaction of LTM (F = 71.880, df = 2, 252, p < .0001) also was significant. Because in practice the factors of tonality, length, and motion are interwoven in melodies, the three-way interaction's effect on memory is of interest even though interpretation is somewhat difficult. Three interesting phenomena can be seen in Figure 7. First, the tonality-conjunct motion level was greatest for facilitating memory at the 6-note length, second at the 10-note length, but was third at the 16-note length. Conversely, the atonal-conjunct motion level moved from poorest for facilitating memory at the 6-note length to the best at the 16 note length. The atonal-disjunct condition resulted in poor scores across all lengths. In this analysis, there were significant interactions but no significant strategy-melodic construction factor (G-L, G-T, or G-M) two-way interactions.

Hypothesis 2: There will be no differences in notational accuracy by subjects as a result of single presentations and repeated presentations.

As in Null Hypothesis 1, an attempt was made through data analysis to isolate a factor (in this case number of presentations) apart from the different strategies. Therefore, data from the first three groups were combined as were data from Groups 4 to 6. Thus data reflected accuracy from the single and multiple (two) presentations with the null hypothesis stating no differences between the two groups. For this test, as in the first ANOVA, data consisted of the correct response totals on experimental melodies 1–12 and Warm-up Melody 3. A t test for independent samples

Group 4	Group 5	Group 6	Group 1	Group 2	Group 3
59.20	56.17	49.85	44.81	44.72	44.70

Figure 3. Scheffe's Test for significant differences between levels for the factor of Group.

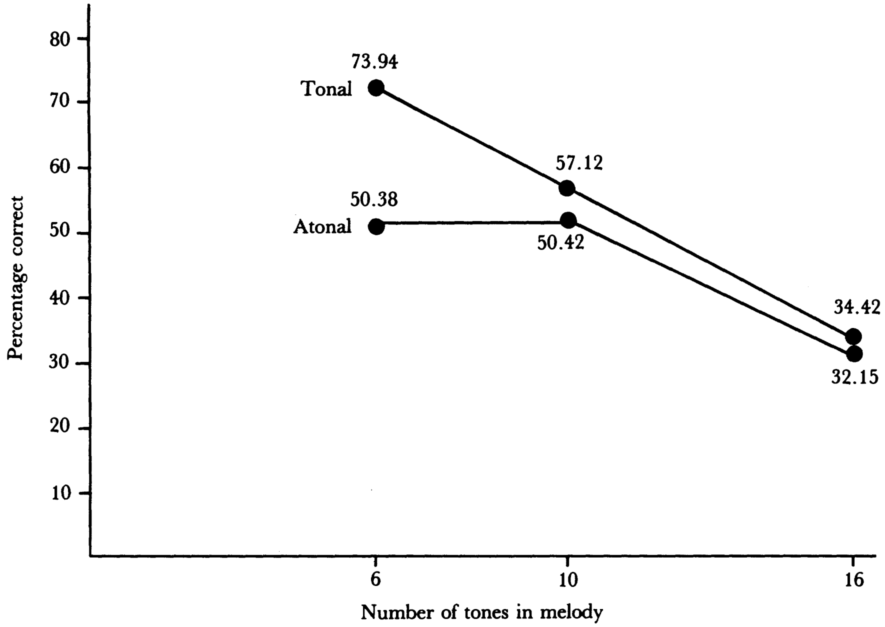

Figure 4. Length-Tonality two-way interaction.

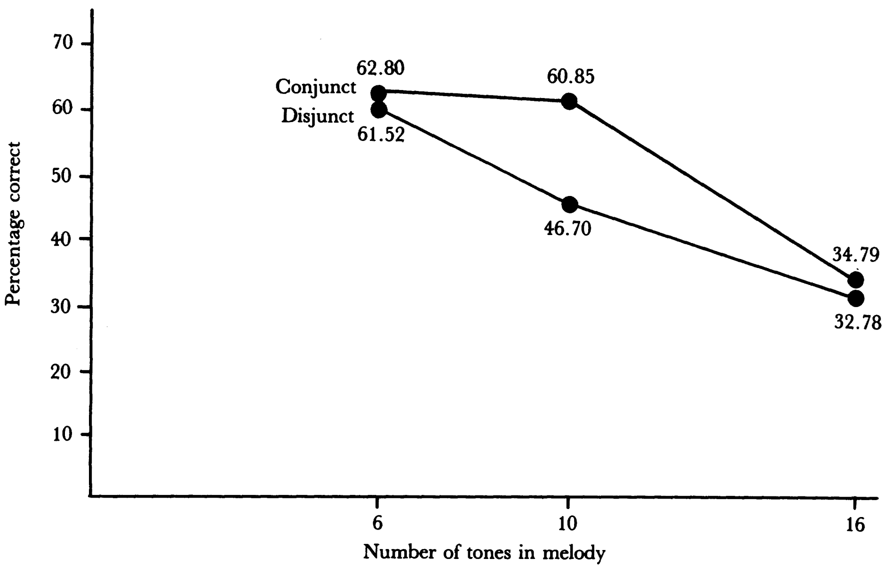

Figure 5. Length-Motion two-way interaction.

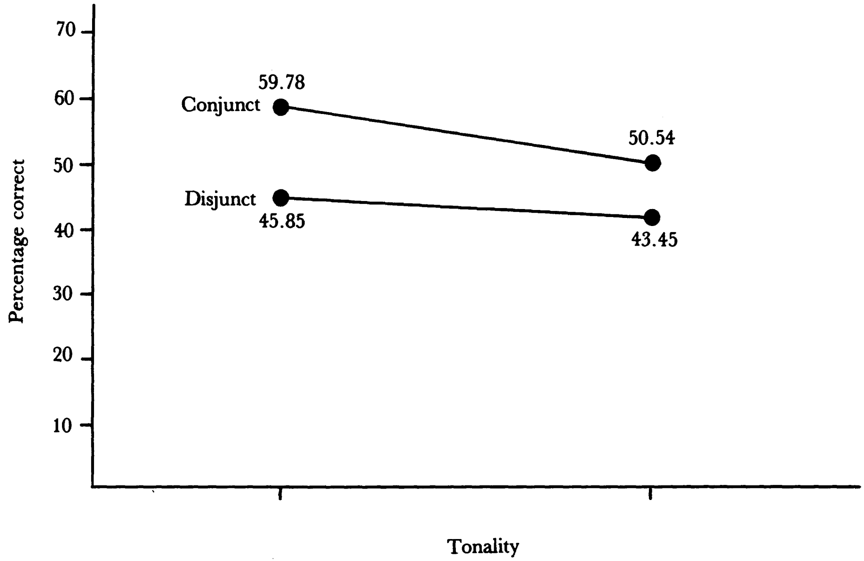

Figure 6. Tonality-Motion two-way interaction.

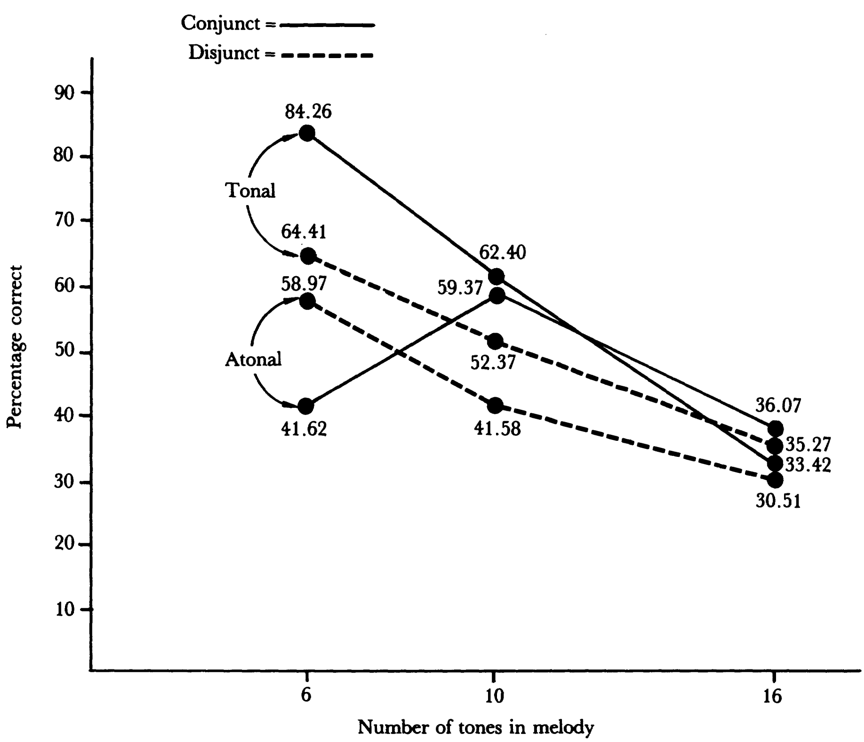

Figure 7. Length-Tonality-Motion three-way interaction.

was used. The $t(134)$ value of 4.906 was significant ($p < .05$), with those subjects in the "two hearings" group attaining a mean number of notes correct of 75.62 (out of 125) and the single hearing group achieving a mean number of 61.77. This factor accounted for 15.2% of the sample variance. Therefore, Null Hypothesis 2 was rejected.

Hypothesis 3: There will be no difference in the accuracy of responses between freshmen and sophomore theory students either vocally or in written form.

To assess the hypothesis regarding differences between freshmen and sophomores on written and vocal accuracy, two separate t tests for independent samples were used. Data for the first consisted of accurate vocal response totals on experimental melodies 1–12. The total number of possible correct tones was 116. A calculated $t(39)$ of 1.795 was not significant at .05. (The actual probability was .08). The total n for this test was 41, consisting of 26 freshmen ($M = 47.31$) and 15 sophomores ($M = 54.13$). Only data from the subject's first try at reproducing the melody were used. (It was believed that the number of subjects singing twice $n = 21$ represented too small a group to be used exclusively for meaningful interpretation of the data, while totals from the first and second singing attempts were not combined because of the potential confounding effects due to different numbers of freshmen and sophomores. It was found that 7.6% of the variance in singing scores was due to class level (freshmen and sophomores.)

A t test for independent groups again was used to assess differences in writing scores between freshmen and sophomores. The probability associated with the calculated $t(134)$ value of 1.914 was greater than the selected alpha of .05. (The calculated probability was .0577.) Sophomores, however, scored higher in total correct responses, averaging 71.75 versus 65.92 for freshmen. Class level was calculated to account for 2.7% of the sample variance. Therefore, Null Hypothesis 3 not was rejected for singing or written responses.

Hypothesis 4: There will be no difference in the accuracy of student responses on the second half of melodies containing certain selected difficult intervals versus the second half of melodies containing only conjunct motion.

The accuracy of each subject on the tones following intervals of tritones, minor sixths, and minor sevenths (melodies 2, 4, 6, 8, 9, and 11) and accuracy on tones in the same serial positions that were preceded by the more conjunct intervals of seconds and thirds (melodies 1, 3, 5, 7, 10, and 12) were compared. On melodies 6 notes in length this meant analyzing accuracy on positions 5 and 6 since tones 3 and 4 created the large interval in disjunct melodies. Similarly in 10-note melodies, it meant comparing tones 7 through 10 as positions 5 and 6 were the large skip tones. Finally in 16-note melodies, tones 10 through 16 were examined because positions 8 and 9 represented large skips. Using this formula the total possible number of correct tones for each of the two groups (following disjunct and following conjunct motion) was 26. A paired t test ($dt = 135$, $t = 4.622$) was significant ($p<.05$) and the means for the two groups revealed that accuracy was greater for conjunct melodies. Therefore Null Hypothesis 4 was rejected. Motion accounted for 13.7% of the sample variance.

To determine if any of the three basic dictation strategies were superior in promoting accuracy following difficult intervals, a one-way analysis of variance was used to evaluate second-half accuracy by group. Each number in the paired *t* test (described in the previous paragraph) that was associated with disjunct melodies was codified according to group. Results (show) in Table 3) indicated that there were no significant differences (.05) between groups in scores on second-half accuracy following a large skip. (The actual probability was .157).

Hypothesis 5: During warm-up exercises (prior to experimental group assignment) there will be no difference in the number of students already using each of the three dictation methods.

To evaluate Null Hypothesis 5, a chi-square test was used to differentiate between the number of students selecting each strategy—progressive, concentration, and singing on the two initial warm-up melodies. Of the 136 subjects, 97 employed the progressive approach: writing while the melody was sounding. Only 34 listened during the playing of the melody (concentration strategy) and then wrote. Just five sang during the melodic presentation before attempting to notate it. The chi-square value of 97.61 confirmed these differences as significant beyond the .001 level; therefore Null Hypothesis 5 was rejected.

The final area of investigation addressed by this study involved the question of whether dictation errors are the result of improper perception and memory or the misnotating of properly perceived sounds. The first method of determining an answer to this question examined the number of pitches with which subjects responded. The mean numbers of pitches subjects *attempted* vocally per melody are shown in Table 4. These figures are based on tones attempted (both correct and incorrect responses), not merely correct responses. It can be seen clearly that accuracy in assessing the number of notes sounded abated as length

Table 3
ANOVA Summary Table for Second-half Accuracy for Three Dictation Strategies

Source of variance	Sum of squares	*df*	Mean squares	*F*	*p*
Group	62.7	2	31.37	1.18	.157
Within	2223.8	133	16.72		

Table 4
Number of Notes Attempted for Written and Sung Responses (6-, 10-, and 16-Note Melodies)

	6-note melodies		10-note melodies		16-note melodies	
	M	*sd*	*M*	*sd*	*M*	*sd*
Written	6.00	.46	9.57	1.62	11.84	3.65
First hearing (vocalized)	5.86	.83	9.16	1.82	10.23	3.75
Second hearing (vocalized)	5.99	.44	9.06	2.42	11.50	3.35

increased. This observation is reinforced by the standard deviation figures which indicate a diversity of responses on the 10- and 16-note melodies. Means for the 6- and 10- note melodies were fairly close to the actual number of pitches sounded but the means for the 16-note melodies were well below the actual number of pitches. Subjects who sang attained an overall written accuracy rate of 43%. Subjects actually wrote what they sang 61% of the time.

Certain factors did contribute to more accurate vocal responses. A *t* test comparison of accuracy between group 6 subjects' first attempts (39% accuracy) and second attempts (52% accuracy) revealed that stimulus repetition made a significant difference in eliciting correct responses (t = 11.50, df = 20, $p<.05$). Number of presentations accounted for 87% of the sample variance. Tonal melodies elicited higher scores in first and second singing responses—46 and 54% respectively—than did atonal melodies (39 and 50 %). Likewise conjunct melodies resulted in greater accuracy in the first and second attempts (45 and 56% respectively) than did disjunct ones (40 and 47%). However, the 16 note-length greatly inhibited accuracy as none of the 44 singers sang any of the 16-note melodies absolutely correctly on the first attempt and only 2 of 21 were able to sing all 16 notes correctly after two hearings.

A phenomenon that the experimenter had not anticipated but observed during the execution of this study was that rhythm seemed to be an entity that was preserved regardless of whether the subject perceived and remembered the proper tones and intervalic relationships. This was confirmed by a paired *t* test comparing each student's total of correct pitch responses and correct rhythmic responses. The $t(135)$ value of 24.45 was significant (.05) and accounted for 81.6% of the variance between scores. While it may seem an unfair comparison because of the greater number of possibilities from which to choose each interval, the actual number of presented intervals and rhythms in the two tasks was identical with pitches consisting of four items (major and minor seconds and thirds if the occasional large interval is disregarded) and rhythms also consisting of four items—eighth, quarter, half, and whole notes.

DISCUSSION AND RECOMMENDATIONS

In both single and combined hearings the three methods used in this study were not found to differ significantly in producing correct dictation responses. Significant differences were found only when comparing the three methods for the condition of two melodic presentations. In this condition writing while hearing and listening before writing proved superior to listening, singing, and writing. This is a surprising result when considering the recommendations and warnings in most melodic dictation texts. The superiority of the progressive and concentration approaches may be the result of singers attempting more notes than they can remember. Such attempts may create incorrect aural models from which to notate. If such is the case, it would behoove aural skills teachers to refrain from compelling students to sing beyond their memory limits.

The significantly dramatic increase in scores when subjects were given a second presentation raises questions as to why this occurs. It is possible that a first hearing provides a schemata or expectancy that allows more meaningful interpretations of the events during a second playing. Another explanation is that information that already has been written can be disregarded during the second playing, thus eliminating the capacity problem discussed by Miller (1956).

In observing the melody components of variance factors, the largest percentage of variance was attributable to length (13%). This may indicate simply that no matter what type of melody is presented (tonal/atonal, conjunct/disjunct, etc.), subjects cannot remember more than a certain number of tones. In this study there was a noticeable decrease in accuracy from the 6-note (62.16%) to 10-note (53.77%) length, but the 10-to-16 note drop was dramatic, falling from 53.77 to 33.79%. Therefore, it may be speculated that the short-term memory limit falls somewhere between 10 and 16 tones.

The tonality-length interaction was worth noting. In short melodies, a large difference was seen between accuracy on tonal and atonal melodies but as the length increased (perhaps approaching the memory capacity limit), tonality made less difference. As previously reported, this may indicate that as length increases, the ability to remember tones decreases regardless of divergent melodic characteristics.

Surprisingly, there were no significant interactions between group and the melody factors of length, tonality, and motion. Apparently no strategy used in this study greatly altered subjects' abilities to remember and notate accurately, regardless of the melody's length, tonality, or motion.

Before explanations are given for why there were no significant differences in either vocal or written response accuracy between freshmen and sophomores, it should be noted that differences would have been significant in both written and sung categories had an alpha of .10 been selected. This may offer little consolation, however, to those expecting a dramatic difference as a result of drilling on these skills (sophomores used in this study had spent a year doing melodic dictation utilizing a computer program designed specifically for that task). Again, conjecture and possible explanations for no significant differences can be attempted. First and perhaps most likely, the task was too difficult, falling beyond that area where training discrepancies between freshmen and sophomores would be reflected. Certainly the 16-note length exceeded the skills of all but a few. For most, the 10-note melodies also seemed to be too long.

Another explanation for the lack of significant differences might be attributed to a dearth of practice. During the semester in which the experiment took place, sophomores were primarily involved in harmonic dictation with no in-class melodic dictation work; and because of a shortage of computer terminals, they did not have access to the computer-based melodic dictation program previously mentioned.

Comparing second-half accuracy following large skips (disjunct) to the first half of the melodies (consisting of steps) indicated that large skips do seem to cause interference in processing tones that follow such a skip.

Unfortunately, there was not a simple solution to this problem through use of one particular melodic dictation strategy. If unfamiliar intervals (those that cannot be instantaneously recognized) cause processing problems then one of the following solutions may be necessary: (a) intense drill on the problematic interval(s) including the "skip" interval, (b) implementation of strategies that encourage a temporary disregard of the interval and a concentration on the relationships of ensuing tones to previous ones, or (c) isolation of the large intervals upon one hearing for identification and practice.

Students exhibited an overwhelming predilection toward the progressive approach (writing while listening). Seventy-one percent of the subjects used this method when given a choice. Of the students who were not assigned to the progressive approach, many said (at the completion of the test) they felt their performance would have been better had they been allowed to write while the tones were sounding.

The fact that students sang melodies back on one hearing at a 43% rate of correctness establishes a reasonable doubt regarding Ortmann's 1934 assertion that dictation errors are 87% notational in nature and only 13% perceptual. If subjects are trained to listen more intently and sing before taking dictation, results may be different, but results from the present study seemingly would indicate that many students are not perceiving the proper pitches.

Ortmann's contention that contour is a fundamental perceptual phenomenon seemed to be supported by this study. The basic A contour was altered only 10% of the time by writers, 5% being one directional substitutions and 5% reflecting more than one change in direction. For singers, changes in contour occurred 12% of the time with 10% being only one directional substitution and 2% containing more than one direction change.

As a result of this study, the experimenter is inclined to disagree or at least clarify Ortmann's (and others) statement that seconds are the easiest interval to identify. When assorted minor and major seconds of no particular scale pattern were presented consecutively, subjects seemed to employ seconds haphazardly such that their patterns reflected little resemblance to the original melody. They knew that several notes were close to each other but could not remember the correct order of the major and minor seconds. In seeming frustration, many jumped to the highest pitch in the melody after attempting a small number of consecutive seconds and descended from that point. It may be that randomly presented major and minor seconds represent the most difficult memory task as they defy codification or condensing and must be remembered separately. One evidence of the confusion or lack of memory caused by multiple minor and major seconds presented consecutively is people who wrote the same sequence of tones for two or three melodies (specifically numbers 5 and 6 (16 AC and 16 AD, see Figure 1) in a row as if to say, "What is the difference?"

Regarding perceptual limitations, two melodies seemed to act as indicators of perceptual ability. Warm-up Melody 3 (tonal and 10 notes in length) and Experimental Melody 12 (tonal and 10 notes in length) were the longest melodies to which subjects responded with total accuracy with any regularity. Those who correctly answered Warm-up Melody 3 and experimental melody 12 usually performed well overall on the test. Those who did not do well on these seemed to have difficulty even on the 6-note melodies. As a result of this study, it is believed that perception and/or memory for an unfamiliar melody will not be accurate for anything longer than a 10-note melody. Furthermore, it still may not be accurate unless the stimulus is of a nature that can be easily codified (e.g., tonal setting).

To determine if any of the three basic dictation strategies were superior in promoting accuracy following difficult intervals, a one-way analysis of variance was used to evaluate second-half accuracy by group. Each number in the paired *t* test (described in the previous paragraph) that was associated with disjunct melodies was codified according to group. Results (show) in Table 3) indicated that there were no significant differences (.05) between groups in scores on second-half accuracy following a large skip. (The actual probability was .157).

Hypothesis 5: During warm-up exercises (prior to experimental group assignment) there will be no difference in the number of students already using each of the three dictation methods.

To evaluate Null Hypothesis 5 a chi-square test was used to differentiate between the number of students selecting each strategy—progressive, concentration, and singing on the two initial warm-up melodies. Of the 136 subjects, 97 employed the progressive approach: writing while the melody was sounding. Only 34 listened during the playing of the melody (concentration strategy) and then wrote. Just five sang during the melodic presentation before attempting to notate it. The chi-square value of 97.61 confirmed these differences as significant beyond the .001 level; therefore Null Hypothesis 5 was rejected.

The final area of investigation addressed by this study involved the question of whether dictation errors are the result of improper perception and memory or the misnotating of properly perceived sounds. The first method of determining an answer to this question examined the number of pitches with which subjects responded. The mean numbers of pitches subjects *attempted* vocally per melody are shown in Table 4. These figures are based on tones attempted (both correct and incorrect responses), not merely correct responses. It can be seen clearly that accuracy in assessing the number of notes sounded abated as length increased. This observation is reinforced by the standard deviation figures which indicate a diversity of responses on the 10-and 16-note melodies. Means for the 6-and-10 note melodies were fairly close to the actual number of pitches sounded but the means for the 16-note melodies were well below the actual number of pitches. Subjects who sang attained an overall written accuracy rate of 43%. Subjects actually wrote what they sang 61% of the time.

Certain factors did contribute to more accurate vocal responses. A *t* test comparison of accuracy between group 6 subjects' first attempts (39% accuracy) and second attempts (52% accuracy) revealed that stimulus repetition made a significant difference in eliciting correct responses (t = 11.50, df = 20, p<.05). Number of presentations accounted for 87% of the sample variance. Tonal melodies elicited higher scores in first and second singing responses—46 and 54% respectively—than did atonal melodies (39 and 50 %). Likewise conjunct melodies resulted in greater accuracy in the first and second attempts (45 and 56% respectively) than did disjunct ones (40 and 47%). However, the 16 note length greatly inhibited accuracy as none of the 44 singers sang any of the 16 note melodies absolutely correctly on the first attempt and only 2 of 21 were able to sing all 16 notes correctly after two hearings.

A phenomenon that the experimenter had not anticipated but observed during the execution of this study was that rhythm seemed to be an entity that was preserved regardless of whether the subject perceived and remembered the proper tones and intervalic relationships. This was confirmed by a paired t test comparing each student's total of correct pitch responses and correct rhythmic responses. The $t(135)$ value of 24.45 was significant (.05) and accounted for 81.6% of the variance between scores. While it may seem an unfair comparison because of the greater number of possibilities from which to choose each interval, the actual number of presented intervals and rhythms in the two tasks was identical with pitches consisting of four items (major and minor seconds and thirds if the occasional large interval is disregarded) and rhythms also consisting of four items—eighth, quarter, half, and whole notes.

DISCUSSION AND RECOMMENDATIONS

In both single and combined hearings the three methods used in this study were not found to differ significantly in producing correct dictation responses. Significant differences were found only when comparing the three methods for the condition of two melodic presentations. In this condition writing while hearing and listening before writing proved superior to listening, singing, and writing. This is a surprising result when considering the recommendations and warnings in most melodic dictation texts. The superiority of the progressive and concentration approaches may be the result of singers attempting more notes than they can remember. Such attempts may create incorrect aural models from which to notate. If such is the case, it would behoove aural skills teachers to refrain from compelling students to sing beyond their memory limits.

The significantly dramatic increase in scores when subjects were given a second presentation raises questions as to why this occurs. It is possible that a first hearing provides a schemata or expectancy that allows more meaningful interpretations of the events during a second playing. Another explanation is that information that already has been written can be disregarded during the second playing, thus eliminating the capacity problem discussed by Miller (1956).

In observing the melody components of variance factors, the largest percentage of variance was attributable to length (13%). This may indicate simply that no matter what type of melody is presented (tonal/atonal, conjunct/disjunct, etc.), subjects cannot remember more than a certain number of tones. In this study there was a noticeable decrease in accuracy from the 6- note (62.16%) to 10-note (53.77%) length, but the

10-to-16 note drop was dramatic, falling from 53.77 to 33.79%. Therefore, it may be speculated that the short-term memory limit falls somewhere between 10 and 16 tones.

The tonality-length interaction was worth noting. In short melodies, a large difference was seen between accuracy on tonal and atonal melodies but as the length increased (perhaps approaching the memory capacity limit), tonality made less difference. As previously reported, this may indicate that as length increases, the ability to remember tones decreases regardless of divergent melodic characteristics.

Surprisingly, there were no significant interactions between group and the melody factors of length, tonality, and motion. Apparently no strategy used in this study greatly altered subjects' abilities to remember and notate accurately, regardless of the melody's length, tonality, or motion.

Before explanations are given for why there were no significant differences in either vocal or written response accuracy between freshmen and sophomores, it should be noted that differences would have been significant in both written and sung categories had an alpha of .10 been selected. This may offer little consolation, however, to those expecting a dramatic difference as a result of drilling on these skills (sophomores used in this study had spent a year doing melodic dictation utilizing a computer program designed specifically for that task). Again, conjecture and possible explanations for no significant differences can be attempted. First and perhaps most likely, the task was too difficult, falling beyond that area where training discrepancies between freshmen and sophomores would be reflected. Certainly the 16-note length exceeded the skills of all but a few. For most, the 10-note melodies also seemed to be too long.

Another explanation for the lack of significant differences might be attributed to a dearth of practice. During the semester in which the experiment took place, sophomores were primarily involved in harmonic dictation with no in-class melodic dictation work; and because of a shortage of computer terminals, they did not have access to the computer-based melodic dictation program previously mentioned.

Comparing second-half accuracy following large skips (disjunct) to the first half of the melodies (consisting of steps) indicated that large skips do seem to cause interference in processing tones that follow such a skip. Unfortunately, there was not a simple solution to this problem through use of one particular melodic dictation strategy. If unfamiliar intervals (those that cannot be instantaneously recognized) cause processing problems then one of the following solutions may be necessary: (a) intense drill on the problematic interval(s) including the "skip" interval, (b) implementation of strategies that encourage a temporary disregard of the interval and a concentration on the relationships of ensuing tones to previous ones, or (c) isolation of the large intervals upon one hearing for identification and practice.

Students exhibited an overwhelming predilection toward the progressive approach (writing while listening). Seventy-one percent of the subjects used this method when given a choice. Of the students who were not assigned to the progressive approach, many said (at the completion of the test) they felt their performance would have been better had they been allowed to write while the tones were sounding.

The fact that students sang melodies back on one hearing at a 43% rate of correctness establishes a reasonable doubt regarding Ortmann's 1934 assertion that dictation errors are 87% notational in nature and only 13% perceptual. If subjects are trained to listen more intently and sing before taking dictation, results may be different, but results from the present study seemingly would indicate that many students are not perceiving the proper pitches.

Ortmann's contention that contour is a fundamental perceptual phenomenon seemed to be supported by this study. The basic A contour was altered only 10% of the time by writers, 5% being one directional substitutions and 5% reflecting more than one change in direction. For singers, changes in contour occurred 12% of the time with 10% being only one directional substitution and 2% containing more than one direction change.

As a result of this study, the experimenter is inclined to disagree or at least clarify Ortmann's (and others) statement that seconds are the easiest interval to identify. When assorted minor and major seconds of no particular scale pattern were presented consecutively, subjects seemed to employ seconds haphazardly such that their patterns reflected little resemblance to the original melody. They knew that several notes were close to each other but could not remember the correct order of the major and minor seconds. In seeming frustration, many jumped to the highest pitch in the melody after attempting a small number of consecutive seconds and descended from that point. It may be that randomly presented major and minor seconds represent the most difficult memory task as they defy codification or condensing and must be remembered separately. One evidence of the confusion or lack of memory caused by multiple minor and major seconds presented consecutively is people who wrote the same sequence of tones for two or three melodies (specifically numbers 5 and 6 16 AC and 16 AD, see Figure 1) in a row as if to say, "What is the difference?"

Regarding perceptual limitations, two melodies seemed to act as indicators of perceptual ability. Warm-up Melody 3 (tonal and 10 notes in length) and Experimental Melody 12 (tonal and 10 notes in length) were the longest melodies to which subjects responded with total accuracy with any regularity. Those who correctly answered Warm-up Melody 3 and experimental melody 12 usually performed well overall on the test. Those who did not do well on these seemed to have difficulty even on the 6-note melodies. As a result of this study, it is believed that perception and/or memory for an unfamiliar melody will not be accurate for anything longer than a 10-note melody. Furthermore, it still may not be accurate unless the stimulus is of a nature that can be easily codified (e.g., tonal setting).

SUMMARY

As a result of this study it appears that melodic dictation strategy makes little difference in response accuracy. However, two important trends did emerge. First, the best way to improve dictation response accuracy appears to be a repetition of the melodic stimuli. Those hearing the stimuli twice consistently did better than those hearing them only once. A second phenomenon worthy of noting is that accuracy scores decreased dramatically as the melodic length increased from 10 to 16 notes, even after two hearings. Errors in dictation for longer melodies more likely reflect an inability to remember all of the tones rather than deficiencies in the process of notating them. In this instance, emphasis should thus be placed on careful listening or coding such that the problem of limited memory can be overcome. Endless drill on isolated intervals alone will not result in great accuracy on longer melodies, regardless of how proficient the student becomes at such a skill.

The value of singing before notating also was found questionable in this study. Although there were no significant differences among the three strategies when the melody was presented once, the singing method was found to be the worst of the strategies in the condition with two presentations of the melody and when scores on single and dual presentations were combined. If students cannot sing what they have heard, asking them to sing may provide a "self-composed" inaccurate model from which to dictate, especially for longer melodies. The finding that subjects write what they sing only 61% of the time also raised the question of the reasoning for having subjects sing when they fail to notate what they have just sung.

Finally, the progressive method, which allows the student to write as the melody is sounding, does not appear as poor a technique as Benward suggests (1961). Overall, the progressive technique resulted in the highest scores. If the progressive technique is implemented using only noteheads and these placed particular distances apart to indicate duration, the subject may be able to notate the melody as it is sounding in all but the quickest of examples.

REFERENCES

Benward, B. (1961). *Teacher's dictation manual in ear training*. Dubuque: Wm. C. Brown.

Conner, A. O. (1970). The perception of musical intervals: The theory of property arrays. *Council for Research in Music Education, 20,* 38–42.

Cuddy, L. (1982). From tone to melody to music: Some directions for a theory of musical cognition. *Council for Research in Music Education, 71,* 15–29.

Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Detusch & J. A. Deutsch (Eds.), *Short-term memory* (pp. 108–151). New York: Academic Press.

Deutsch, D. (1982). The influence of melodic context on pitch recognition judgement. *Perception and Pschophysics, 31* (5), 407–410.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85* (4), 341–354.

Edlund, L. (1963). *Modus novus: Studies in reading atonal melodies*. London: J. W. Chester.

Herder, R. (1973). *Tonal/atonal; Progressive ear training, singing, and dictation studies in diatonic, chromatic, and atonal music*. New York: Continuo Music Press.

Hicks, C. R. (1973). *Fundamental concepts in the design of experiments*. New York: Holt, Rinehart, & Winston.

Hofstetter, F. T. (1981). Computer-based recognition of perceptual patterns and learning styles in rhythmic dictation exercises. *Journal of Research in Music Education, 29(4),* 265–277.

Holly, K. M., Jr. (1980). An experimental investigation of the effects of selected factors on the short-term retention of pitch sequences. *Council for Research in Music Education, 63,* 48–51.

Larsen, R. L. (1973). Levels of conceptual development in melodic permutation concepts based on Piaget's theory. *Journal of Research in Music Education, 21(3),* 256–263.

Long, P. A. (1977). Relationships between pitch memory in short melodies and selected factors. *Journal of Research in Music Education, 25(4),* 272–282.

Madsen, C. K., & Staum, M. J. (1983). Discrimination and interference in the recall of melodic stimuli. *Journal of Research in Music Education, 31(1),* 15–31.

McGaughey, J. M. (1961). *Practical ear training.* Boston: Allyn & Bacon.

McHose, A. I., & Tibbs, R. N. (1945). *Sight singing manual.* New York: Appleton-Century-Crofts.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review, 63,* 81–97.

Moog, H. (1982). The foundations of musical learning through hearing: A contribution towards a learning theory of music. *Psychology of Music, Proceedings of the Ninth International Seminar on Research in Music Education.*

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* New York: John Wiley & Sons.

Neisser, U. (1976). *Cognition and reality.* San Francisco: W. H. Freeman.

Neisser, U., Hirst, W., & Spelke, E. S. (1981). Limited capacity theories and the notion of automaticity: Reply to Lucas and Bub. *Journal of Experimental Psychology: General, 110(4),* 499–500.

Ortmann, O. O. (1934). *Problems in the elements of ear dictation.* Baltimore: Peabody Conservatory of Music.

Pembrook, R. G. (1983, February). *The effects of contour, length, and tonality on melodic memory.* Paper presented at the Southeast Regional Meeting of the Music Educators National Conference, Louisville, KY.

Sink, P. E. (1983). Effects of rhythmic and melodic alterations on rhythmic perception. *Journal for Research in Music Education, 31(2),* 101–113.

Tallarico, P. T. (1974). A study of the three phase concept of memory: Its musical implications. *Council for Research in Music Education, 39,* 1–15.

Taylor, J. A. (1976). Perception of tonality in short melodies. *Journal of Research in Music Education, 24(4),* 197–208.

Trubitt, A. R., & Hines, R. S. (1979). *Ear training & sight-singing: An integrated approach.* New York: MacMillan.

Williams, D. B. (1975). Short-term retention of pitch sequence, *Journal of Research in Music Education, 23(1),* 53–66.

November 13, 1985