JRME SPRING 1997, VOLUME 45, NUMBER 1, PAGES 51-66

The purpose of this study was to examine graduate and undergraduate music majors' ability to detect pitch and rhythm errors in one-, two-, and three-part settings of texturally contrasting musical excerpts. A stimulus audiotape consisting of 12 excerpts resulted from the following arrangement by texture and number of parts: 4 one-part excerpts, 2 two-part and 2 three-part homorhythmic excerpts, and 2 two-part and 2 three-part polyrhythmic excerpts. Subjects (N = 150) listened to purposefully marred recorded performances of these excerpts and attempted to identify pitch and rhythm errors by circling appropriate places on correctly notated scores. Results of analyses of variance with repeated measures on correct responses indicated significant main effects of degree status (graduate students were more accurate than were undergraduates); number of parts (subjects were most accurate in one-part settings, less accurate in two-part settings, and least accurate in three-part settings); texture (subjects were more accurate in homorhythmic texture than in polyrhythmic texture); and error type (subjects were more accurate detecting rhythm errors than pitch errors). Significant interactions indicated that these variables did not function independently of each other.

James L. Byo, *Louisiana State University*

The Effects of Texture and Number of Parts on the Ability of Music Majors to Detect Performance Errors

Relatively little information of an empirically derived and specific nature is available concerning the background and experience necessary for one to acquire advanced-level skill in performance error detection. Brand and Burnsed (1981) examined several factors bearing on the preparation of instrumental music education students—factors that might be expected to affect one's ability to detect performance errors. Somewhat surprisingly, achievement in music theory and sight-singing, years of applied music instruction and ensemble experience, and number of instruments played were not significantly related to skill in error detection. Similarly, Sidnell (1971) reported no relationships among visual score reading, musical memory, and aural harmony achievement.

This research was supported by a grant from the Louisiana State University Council on Research. James L. Byo is an associate professor of music in the School of Music at Louisiana State University, Baton Rouge, LA 70803-2504.

These findings seem to be substantiated by research in which graduate and undergraduate music majors responded alike on a test of their ability to detect performance errors (Byo, 1993); more "experience" did not give graduate students an advantage in the perception of pitch and rhythm errors.

In contrast, Gonzo (1971) found that experienced choral music educators performed better than undergraduate music majors on a pitch error detection test. Furthermore, superior performance in music theory and experience in choral arranging were indicators of success in the detection of pitch errors. An investigation by Dolbeer (1969) yielded similar results—undergraduate music majors who excelled in music theory and/or had more extensive music performance experience improved more in their ability to detect pitch and rhythm errors. Larson (1977) reported significant relationships among pitch error detection, melodic dictation, and sight-singing tasks.

Given a close relationship between error detection and performance adjudication (the former being one of several processes involved in the more comprehensive latter), information culled from research in adjudication and relative to the background and experience of the evaluator would seem to apply here. There is evidence suggesting that, in the context of overall or "global" evaluation, the ability to reliably adjudicate instrumental performance is not dependent on adjudicator level of instrumental performance skill (Fiske, 1977). Less accomplished university-level performers evaluated a series of trumpet performances with much the same outcome as more accomplished performers, a result that seems to hold up also in comparisons of more and less experienced adjudicators of band performance, again in the context of overall ratings (Byo & Brooks, 1994; Winter, 1990). Another study indicated that it may not be necessary to have brass instrument expertise to reliably judge high school trumpet playing; musicians with nonbrass and nonwind specialty areas, when compared to brass and wind specialists respectively, gave trumpet performances similar overall and categorical ratings (Fiske, 1975).

According to the research cited thus far, different levels of achievement and experience in ear training, applied music, and ensemble participation are not always reflected in similar differences in performance scores on various measures of ability in error detection. Accomplished sight-singers, advanced performers, and those with well-developed skills in harmonic dictation do not necessarily possess more highly developed skills in performance error detection than those with average marks in these areas. This seems to point to the need for context-specific strategies in the development of what appears to be "a highly synthesized aural/visual skill [that is] more than an amalgamation of factors thought to be relevant" (Sidnell, 1971, p. 90). The difficulties inherent in transferring knowledge or skills developed in one arena to those required in another are well-documented in general cognitive domains (see Gardner, 1991, and Perkins & Salomon, 1989) and less well-documented in music teaching (Price, 1992) and music performance (Duke & Pierce, 1991; Grutzmacher, 1987; McKnight, 1975;

Price, Blanton, & Parrish, 1993).

A number of researchers have tested the effects of training that are directly related to the development of skill in error detection. From these efforts, we know that subjects' ability to detect performance errors improves with practice (Costanza, 1971; Deal, 1985; DeCarbo, 1982; Doane, 1989; Grunow, 1980; Malone, 1985/1986; Ramsey, 1979; Sidnell, 1971; Stuart, 1979). Training was frequently modeled after aspects of the programmed instruction format (Costanza, 1971; Deal, 1985; Malone, 1985/1986; Ramsey, 1979; Sidnell, 1971), and longer formats were more effective than shorter packages (Ramsey, 1979). Training involving four-voice reduced scores and electronically produced sounds in an interactive, computerized program of error detection was as effective as training using conventional means, that is, recordings of a 30-piece wind ensemble and full scores (Deal, 1985; Ramsey, 1979). Timbre was a variable in a study by Byo (1993) in which music majors perceived errors with greater success in the context of a synthesized piano timbre condition than in a synthesized and heterogeneous wind instrument condition.

Podium-based instruction and experience while conducting a live ensemble facilitated development of skill in error detection and, in view of potential problems in transferring skills developed from tape recordings or synthesized sounds to live conditions, might be the preferred means of improving performance in this area if the live, rehearsal room format is feasible (DeCarbo, 1982; Doane, 1989). Podium-based experiences beg the question of whether the physical act of conducting inhibits one's ability to discriminate performance errors. An answer to this question is uncertain due to conflicting results (Blocher, 1990; Doane, 1989; Forsythe & Woods, 1983). Finally, the effect of one's knowledge of the score on the ability to detect errors is also uncertain due to so few studies in this area (Grunow, 1980). Grunow's research was an outgrowth of the commercially available MLR Score Reading Program (Froseth & Grunow, 1979).

The construction of a program in error detection, particularly one intended to be sequential in the manner suggested by programmed formats, requires that one make several choices relative to music selection (texture and number of parts) and unique to the task (error type, insertion of errors, and timbre). Lack of a sufficient empirical base and otherwise conflicting findings have forced investigators to make assumptions, resulting perhaps in a certain degree of arbitrariness in extant programs. To illustrate, it could be argued that development of valid training programs in error detection should be based on knowledge of the mental processes that are or perhaps should be operating in the detection of performance errors. In this regard, it has been popular to promote strategies that focus attention on a specific line or lines in the musical texture while ignoring other lines. This most certainly has been motivated by a desire to make the aural task less daunting for novice conductors, while supposedly providing a means for conductors to avoid being "fooled" by, for example, inner parts, which are commonly considered to be inherently more difficult to hear than outer voices. Research, however, does not overwhelmingly support either the focused approach or the inner-part contention, though the number of related studies is scant.

Hayslett (1992) investigated subjects' ability to simultaneously focus on one voice and detect performance errors in a concurrent voice within three-voice musical textures. Subjects were significantly more accurate in identifying voice parts that had been deleted when attention was not focused on a particular voice. In other words, attending to multiple auditory streams (as it was assumed subjects were doing in an unfocused listening condition) seemed to promote one aspect of "peripheral" hearing, that of deleted parts. Subjects' ability to detect deleted parts was most accurate when that part was the top line (melody), less accurate when that part was the bass, and least accurate when that part was the inner line. Huron (1989), like Hayslett, found subjects to be more discerning of outer voice manipulations. Although Byo (1993) found no differences in subjects' responses with regard to the main effect of location of errors in four-voice textures, this result was obfuscated by an interaction of error location, error type, and timbre.

Determining a hierarchy, by difficulty, of error types has been a major concern of investigators given its fundamental relationship to any type of systematic approach to training. Considered collectively, however, the findings of the following studies are inconclusive. Geringer and Madsen (1989) speculated that musicians, when listening to music, attend to musical elements in the following order: rhythm, pitch, tone quality, and loudness. In an attempt to assess the accuracy of part of this proposed hierarchy, subjects listened to pairs of excerpts that were differentiated by various manipulations of pitch and tone quality. Musicians responded with greater accuracy to pitch alterations than to tone-quality changes, lending some support to the proposal. Blocher (1990), in an attempt to determine an error hierarchy, found no significant differences in difficulty among six error types. Subjects listened to 10 four-measure excerpts recorded by a brass trio (two trumpets, one trombone) and identified errors in articulation, dynamics, intonation, note accuracy, phrasing, and rhythm. In a study that comprised 20 six-measure excerpts, isolated two error types and two timbre conditions (single and multiple), and controlled for placement of errors across soprano, alto, tenor, and bass voices, Byo (1993) found music majors to be significantly more discerning of rhythm errors than pitch errors.

Results of research by Doane (1989) and Hayslett (1992) indicated the opposite with regard to pitch and rhythm errors. Doane's music major subjects were involved in instruction in error detection that spanned from 6 to 12 weeks and entailed practice using the MLR Instrumental Score Reading Program (Froseth & Grunow, 1979) in addition to audiotape and videotape examples of instrumental ensemble performances that were marred by purposefully inserted errors. Excerpts were up to 37 measures long and comprised four to six voices. Subjects' postinstruction responses to error types were as follows (least to most difficult to detect): pitch, balance, precision, and rhythm.

Hayslett's (1992) primary question involving focus of attention and peripheral hearing as measured by detection of deleted parts is addressed above. As one means of focusing subjects' attention, pitch and rhythm errors were inserted in parts to which attention was to be drawn. In completing the experimental task, subjects were asked to indicate departures from the score involving deleted parts, rhythm errors, and pitch errors. Though detection of deleted parts alone was the dependent variable in this study, post facto analysis of the detectability of the three error conditions yielded the following order (least to most difficult): pitch, rhythm, and deleted parts.

The present study extends the body of research that investigates performance error detection in the context of various textural factors. Specifically, it examines subjects' ability to detect pitch and rhythm errors in one-, two-, and three-part settings of homorhythmic and polyrhythmic musical examples.

METHOD

To avoid overriding possible effects, the musical stimuli selected for use in a performance error detection task must present an optimum amount of interference (neither too much nor too little) from musical elements other than those targeted for investigation. Thus, it was determined a priori that (a) the number of concurrent parts in any music excerpt would not exceed three; (b) parts within each excerpt would maintain separate identities in the musical texture, that is, involving minimal if any voice crossing and sufficient variety to function as separate lines; and (c) the level of melodic, rhythmic, and harmonic complexity among the excerpts would be quite modest. The design of this study necessitated musical stimuli entailing two categories of texture—that in which voice relationships are primarily homorhythmic and that in which the voice relationships are primarily polyrhythmic.

The following excerpts were suitable relative to these criteria and thus were selected for inclusion on the stimulus audiotape: measures (mm.) 9–19 of "Passepied" from *Suite in C* by J. S. Bach (Voxman, 1953), mm. 9–16 of "Gavotte" by Kranz (Voxman, 1952), mm. 1–14 of "Ritornello" from *Symphony and Two Ritornellos* by Johann Kindermann (Istvan, 1973), and mm. 11–18 of "Dance" from *Three Dance Movements* by Paul Peuerl (Istvan, 1973). The Kranz and Peuerl excerpts are homorhythmic; the Bach and Kindermann excerpts are polyrhythmic.

Each excerpt represented a complete musical idea ranging in length from 8 to 14 measures. Melodically, the excerpts are predominantly diatonic, nonsequential, and stepwise in contour. Rhythms involve note and rest values through sixteenth notes and little use of syncopation. Harmonic language is tonal. Tempos are moderate. The two primarily homorhythmic excerpts consist of a melody in the top voice that is supported by vertically oriented accompanying parts. The lines are decorated by the inclusion of several nonharmonic tones. The two primarily polyrhythmic excerpts feature rhythmic variety in simultaneous parts, not cross rhythms as is implied by an alternative meaning for the label.

One of these excerpts consists of melody, countermelody, and bass line; the other derives its independence of line and rhythmic activity from canonic devices.

To further isolate texture, each homorhythmic and polyrhythmic excerpt was presented in one-part, two-part, and three-part settings. A total of 12 excerpts resulted from the following arrangement by texture and number of concurrent parts: 4 one-part (monophonic) excerpts, 2 two-part and 2 three-part homorhythmic excerpts, and 2 two-part and 2 three-part polyrhythmic excerpts. In the assignment of original excerpt parts to the one-part settings, only alto and bass parts were used, as a means of establishing some degree of balance across textures. Thus the 4 one-part settings were represented as follows: an alto line from one of the original three-part homorhythmic excerpts, the bass line from the other original homorhythmic excerpt, an alto line from one of the original three-part polyrhythmic excerpts, and the bass line from the other original polyrhythmic excerpt.

In the development of the stimulus audiotape for error detection, FINALE version 2.6, an Apple Macintosh LC computer, a Sentech Mini Mac 100 MIDI interface, a Yamaha SY-77 digital synthesizer, and a Sony HX Pro stereo cassette deck were used to record and perform the four original excerpts and the various settings of each. Furthermore, MIDI playback files were edited to include purposeful errors of pitch and rhythm that were considered reflective of "typical" erroneous performance, not contrived or out of context (see Byo, 1993, and Ramsey, 1979, for descriptions of typical errors). A total of 36 inserted performance errors (18 pitch and 18 rhythm) are categorized and described in Table 1.

Twelve errors (six of each error type) were inserted into the four excerpts constituting each of the one-, two-, and three-part conditions. An effort was made to distribute inserted errors as evenly as possible across parts, although complete control in this regard, as might be achieved through random assignment, was eschewed in favor of a process in which errors were inserted at points where student musicians, for example, are prone to err. This procedure is supported by research in which the main effect of error location within four-part textures did not affect subjects' ability to detect performance errors (Byo, 1993). Each performance error was inserted into the excerpts only once. The number of inserted errors per excerpt ranged from two to four. The first measure of each excerpt remained error-free in order to allow subjects one measure to acclimate to a tonal center.

Parts were entered into the computer such that they would be played back in perfect time, at a constant velocity, and at the tempo of quarter note equals 92. The piano timbre on the synthesizer was used for all excerpts. Previous research had shown that subjects were better able to detect performance errors in a piano timbre condition than in a multiple timbre condition (Byo, 1993). Each excerpt was dubbed onto a cassette tape two times in succession as subjects were to be given two opportunities to detect performance errors. Five seconds of silence separated each trial, and 20 seconds of silence separated each set of two

Table 1
Performance Error Descriptions and Number

Error description	Item (Error) number
Pitch error	
Major second high	1, 5, 6
Major second low	3
Minor second high	9, 15, 18, 26
Minor second low	11, 12, 14, 20, 24, 27, 29, 32, 34, 35
Rhythm error	
Two-note figure performed in reverse	8, 17, 19, 36
Sustained note not given full value	13, 23
Note held through a quarter rest	25
Three eighth notes performed as an eighth-note triplet	7
Dotted rhythm performed as if double-dotted	4, 22
Two notes of equal duration performed unevenly	2, 10, 33
Three-note figure performed at half value	28
Two quarter notes performed as if tied	21
Two-note dotted figure performed as two notes of equal duration	16, 30
Dotted-quarter-note, three-eighth-note figure in 3/4 time performed as a quarter note, two eighth notes, and a quarter note	31

trials. The latter interval served to allow subjects time to turn to the next score page and acclimate themselves to the visual representation of the next excerpt. To test for order effect, two excerpt orders were determined at random.

Correctly notated full scores in concert key were created without tempo, dynamic, and interpretation indicators in order to reduce or eliminate possible interference with the task at hand, namely detection of pitch and rhythm errors. Scores to each of the 12 excerpts and printed instructions to subjects were compiled to form a packet. The two excerpt orders necessitated the preparation of two audiotapes and two sets of score packets.

The error detection task was pilot tested by four graduate students. A pitch error was defined as any note that deviated from the printed score (specifically, "wrong notes," not "out-of-tune notes"). Rhythm errors were any rhythmic deviation from the printed score. Subjects circled the specific location (part, measure, and beat) of each error and described each by writing an R (rhythm error) or a P (pitch error) next to each circle. Subjects were instructed to mark nothing in the score if they heard no errors. As a means of verifying that inserted errors were perceivable, any error that was not identified correctly by three of four graduate students was determined to have been too difficult to detect and was subsequently changed by editing the intensity level of individual parts or by inserting a different performance error, as was the case in several instances. The stimulus audiotape was rerecorded to reflect these changes.

In the second phase of this study, graduate ($n = 45$) and undergraduate ($n = 105$, 69 of whom were second-semester freshmen) music majors at two large southern universities listened to the stimulus tape while watching correctly notated scores and indicated what they perceived to be pitch and rhythm discrepancies between printed score and aural stimulus. Professional-quality audio equipment was used during the administration of the error detection task.

RESULTS

For statistical analysis, subjects were labeled according to degree status (graduate or undergraduate), and correct responses were categorized by error type, number of parts, and texture. Due to the repetition of excerpts in various settings (e.g., an excerpt occurring in its two-part setting, followed at some point by the three-part setting, and finally the one-part setting—an order resulting from random assignment of settings), several means were used to examine the effects of order of excerpt presentation on subjects' responses.

First, a two-way analysis of variance (ANOVA) comparing degree status across the two excerpt orders was calculated on total correct response scores. As is shown in Table 2, the main effect of order was not significant [$F(1, 146) = .84$, $p = .36$]; there was, however, a significant difference attributable to degree status [$F(1, 146) = 29.89$, $p = .0001$]. Graduate students responded to the task with greater overall accuracy ($M = 17.78$ out of 36 possible correct responses) than did undergraduate students ($M = 13.64$, freshmen $M = 12.83$). There was no significant interaction.

Next, correct responses by excerpt were examined to determine whether scores evidenced any trend across time in the error detection task. Pearson product-moment correlations were computed in order to compare percentages of correct responses on identical excerpts in each order of presentation for both the undergraduate and graduate groups. Undergraduate students' scores were consistent regardless of excerpt order ($r = .96$). This high correlation was due in part to a preponderance of inserted errors that went undetected by many undergraduate subjects. A correlation of .68 in the graduate student group provided evidence that excerpt order affected graduate students to a greater degree than undergraduates, though the moderate nature of this relationship would seem to indicate that order did not exert greater influence than location of performance errors within the musical texture or error type.

A three-way ANOVA with repeated measures was calculated on correct response scores in order to test the effect of number of parts on graduate and undergraduate subjects' ability to detect pitch and rhythm errors. As shown in Table 3, the main effects of degree status (graduate $M = 2.95$; undergraduate $M = 2.27$ out of a possible 6), num-

Table 2
Summary Table for Two-Way Analysis of Variance

Source	*df*	Sum of squares	Mean square	*F*	*p*
Order (A)	1	15.22	15.22	.84	.36
Degree (B)	1	536.69	536.69	29.89	.0001
A × B	1	43.58	43.58	2.43	.12
Residual	146	2621.23	17.95		

ber of parts (one-part $M = 3.77$; two-part $M = 2.52$; three-part $M = 1.13$), and error type (rhythm $M = 3.35$; pitch $M = 1.60$) were significant at the .0001 level. There were also two significant ($p = .0001$) two-way interactions (error type by degree status and number of parts by error type). These results are most appropriately represented, however, by a significant three-way interaction among the main effects [$F(2, 296) = 7.21$, $p = .0009$]. Concerning pitch errors, subjects' scores were markedly higher on one-part excerpts than on either the two- or three-part excerpts, with graduate students' scores clearly exceeding those of undergraduates. For rhythm errors, a different pattern of correct responses surfaced. Graduate and undergraduate subjects' scores, which were similar across all settings, evidenced a notable drop on the three-part setting. With one exception, pitch scores were lower than rhythm scores.

To examine the influence of homorhythmic and polyrhythmic textures on subjects' responses, a four-way ANOVA with repeated measures (degree status, number of parts, texture, and error type) was calculated. In this analysis, scores for one-part settings were removed due

Table 3
Summary Table for Three-Way Analysis of Variance

Source	*df*	Sum of squares	Mean square	*F*	*p*
Degree (A)	1	87.91	87.91	28.70	.0001
Residual	148	453.33	3.06		
Parts (B)	2	861.49	430.75	537.59	.0001
B × A	2	.47	.23	.29	.75
Residual	296	237.17	.80		
Error (C)	1	450.82	450.82	260.07	.0001
C × A	1	47.25	47.25	27.26	.0001
Residual	148	256.56	1.73		
B × C	2	273.29	136.64	162.36	.0001
B × C × A	2	12.14	6.07	7.21	.0009
Residual	296	249.11	.84		

Table 4
Summary Table for Four-Way Analysis of Variance

Source	*df*	Sum of squares	Mean square	*F*	*p*
Degree (A)	1	31.36	31.36	27.04	.0001
Residual	148	171.67	1.16		
Part (B)	1	119.73	119.73	377.46	.0001
B × A	1	.13	.13	.41	.52
Residual	148	46.95	.32		
Texture (C)	1	1.90	1.90	6.43	.01
C × A	1	.04	.04	.15	.70
Residual	148	43.83	.29		
Error (D)	1	187.72	187.72	313.47	.0001
D × A	1	6.70	6.70	11.19	.001
Residual	148	88.63	.60		
B × C	1	.10	.10	.29	.59
B × C × A	1	1.90	1.90	5.26	.02
Residual	148	53.56	.36		
B × D	1	131.59	131.59	363.71	.0001
B × D × A	1	.03	.03	.08	.78
Residual	148	53.55	.36		
C × D	1	9.65	9.65	22.77	.0001
C × D × A	1	.01	.01	.03	.87
Residual	148	60.06	.41		
B × C × D	1	24.08	24.08	64.50	.0001
B × C × D × A	1	4.61	4.61	12.36	.0006
Residual	141	55.25	.37		

to their irrelevance to textures involving multiple parts. As shown in Table 4, main effects of degree status (graduate $M = 1.16$; undergraduate $M = .81$ out of a possible 3), number of parts (two-part $M = 1.26$; three-part $M = .57$), texture (homorhythmic $M = .96$; polyrhythmic $M = .87$), and error type (rhythm $M = 1.38$; pitch $M = .50$) were significant at or beyond the .02 level. In addition, there were numerous significant ($a = .05$) two- and three-way interactions; however, these are most appropriately represented by a significant four-way interaction [$F(1, 148) = 12.36$, $p = .0006$]. In each condition (degree status, number of parts, and texture), graduate students were slightly more accurate than were undergraduate students. The effect of texture on undergraduate students was more clear than it was for graduate students, with homorhythmic excerpts evidencing slightly higher scores than polyrhythmic excerpts. Subjects in both groups scored highest on two-part homorhythmic excerpts and lowest on two-part polyrhythmic excerpts.

A different pattern of responses is evident for rhythm errors. Scores for two-part excerpts were higher than three-part excerpts for both groups; the highest scores involved the polyrhythmic texture. The three-part settings resulted in notably lower scores, with the effects of texture being unclear.

DISCUSSION

The primary goal of this study was to gather data relevant to performance error detection that might ultimately aid in establishing a solid empirical base from which to proceed in the development of instructive materials to be used in the preparation of prospective music teachers. As such, it is an extension of previous research (Byo, 1993) and is an attempt to answer some of the questions raised by that and other related research.

The detection of pitch and rhythm errors certainly is not a skill at which music majors excel, as evidenced by overall correct response rates of less than 50% in both the present study and in previous research involving four-part excerpts (Byo, 1993). However, these data resulted from conditions in which subjects were initially unfamiliar with the music, were given minimal time to study the scores, and did not hear an aural representation of the music before commencing with the task—conditions that need not prevail in authentic rehearsal environments. Additionally, the approach taken in the development of the musical stimuli, one that sought optimum levels of interference, by necessity precluded a preponderance of high individual subject scores. (Overall rates of 80% correct and above certainly would have indicated the task was too easy, thus obfuscating results.) Last, these unimpressive correct response percentages should be considered in light of the fact that this research endeavors to investigate aural/visual characteristics, unaided by prior aural models. Accordingly, the specific nature of subjects' correct responses seems to show some tendencies and patterns—perhaps clues—that might lead to empirically based and effective pedagogy in performance error detection.

In the context of the performance errors selected for this study, music majors were more accurate in the detection of rhythm errors than pitch errors. This result substantiates the rhythm-pitch hierarchy, by increasing difficulty, of error detectability (Byo, 1993), one that is to some extent validated by the variety of musical contexts within which errors have been inserted in the musical stimuli of these studies. It should be noted, however, that all rhythm errors are not easier to detect than all pitch errors, as is evident in an item analysis of inserted errors (compare Table 4 to Tables 5 and 6). Despite the fact that all rhythm errors have duration in common, they differ in detectability due to the musical context in which they occur. The four-way interaction of the variables in the present study substantiates the context-specific nature of inserted errors. The hierarchy question will remain difficult to answer definitively given an inherent lack of equality in detectability among error conditions.

Graduate students responded more accurately to the error detection task than did undergraduates, a result that is consistent with the findings of Gonzo (1971) and Dolbeer (1969), but in conflict with results

Table 5
Item Difficulty and Discrimination Values for Pitch Error Conditions

	Difficulty		Discrimination power		
Item	Graduate	Undergraduate	Graduate	Undergraduate	Error condition
3	.82	.77	.30	.39	1 part
11	.82	.34	.50	.54	1 part
18	.76	.70	.50	.61	1 part
20	.74	.53	.80	.46	1 part
12	.63	.26	.70	.57	1 part
24	.58	.33	.70	.61	3 parts; top voice; homo
27	.45	.26	.80	.54	2 parts; top voice; homo
26	.42	.07	.60	.21	2 parts; top voice; homo
29	.39	.16	.80	.32	2 parts; top voice; homo
1	.37	.08	.30	.18	3 parts; top voice; poly
14	.32	.07	.60	.14	3 parts; top voice; poly
15	.29	.09	.70	.21	3 parts; bottom voice; poly
5	.24	.06	.50	.14	2 parts; top voice; poly
32	.21	.03	.60	.04	1 part
35	.16	.03	.40	.14	3 parts; top voice; homo
6	.13	.06	.40	.11	2 parts; bottom voice; poly
34	.05	.04	.00	.00	3 parts; bottom voice; homo
9	.00	.01	.00	.00	2 parts; top voice; poly

Note. Homo = homorhythmic; poly = polyrhythmic.

of an investigation involving similar comparisons (Byo, 1993) as well as data provided by Brand and Burnsed (1981), Sidnell (1971), and Fiske (1975, 1977). In the present study, data indicate that a substantial portion of the difference in overall correct response scores between graduate and undergraduate students involved pitch rather than rhythm. In other words, graduate students scored higher overall because they demonstrated greater ability to detect pitch errors; rhythm error responses were similar across groups.

These findings might enter into the development of instructive material in several ways. Perhaps one's ability to detect pitch errors is something that can be expected to improve with an increase in the number and intensity of opportunities to perform and listen to music, opportunities that often mark the experiences of graduate students and certainly differentiate graduate students from undergraduates. Despite the apparent advantages that graduate students have in this regard, undergraduate scores for rhythm errors closely rival those of the graduate students. One approach to the teaching of performance error detection, then, might be to begin instruction with experiences in detecting rhythm errors only, as a means of introducing students to a more focused or specific form of listening within the fabric of con-

Table 6
Item Difficulty and Discrimination Values for Rhythm Error Conditions

Item	Difficulty		Discrimination power		Error condition
	Graduate	Undergraduate	Graduate	Undergraduate	
4	.97	.95	.00	.04	1 part
31	.95	.90	.10	.04	1 part
7	.89	.95	.30	.29	2 parts; top voice; poly
28	.89	.90	.10	.11	2 parts; top voice; homo
10	.89	.87	.20	.39	2 parts; bottom voice; poly
33	.89	.87	.10	.14	1 part
21	.87	.91	.10	.18	1 part
30	.82	.77	.30	.50	2 parts; bottom voice; homo
22	.66	.28	.10	.18	3 parts; middle voice; homo
19	.63	.73	.10	.14	1 part
8	.58	.49	.80	.68	2 parts; bottom voice; poly
16	.29	.26	.00	.32	3 parts; middle voice; poly
25	.18	.22	.60	.00	2 parts; bottom voice; homo
2	.18	.21	.10	.46	3 parts; middle voice; poly
17	.18	.07	.40	.07	3 parts; top voice; poly
36	.13	.17	.00	.07	3 parts; bottom voice; homo
23	.13	.10	.30	.11	3 parts; bottom voice; homo
13	.08	.24	.40	.04	1 part

Note. Homo = homorhythmic; poly = polyrhythmic.

current voices, and initially avoiding the pitch error. This might address a need for development of the ability to purposefully ignore certain aural stimuli in order to isolate a targeted stimulus. (It is conjectured that "ignoring" in this context is easier said than done, given the preponderance of possible aural and visual distractors present in the performance of music, especially apparent under live conditions.)

Results relevant to the context within which performance errors might occur, namely the number of concurrent parts, seem to suggest a systematic, successive approximations-type approach to early training. This might entail the introduction of one error type (preferably rhythm), in one part only, with initial two-part experiences implementing rhythm errors only. Subsequently, another error type (e.g. pitch) might be introduced by itself in a series of one-part experiences prior to being combined with a second error type (e.g. rhythm, as it presents fewer challenges to students) in two parts. Generalizing to the live rehearsal setting, one might decide a priori to listen for rhythm errors in the context of a single line of music, and then to listen for rhythm errors in two specified parts. In time, a focus on pitch errors in one part might precede listening for a combination of pitch and rhythm errors in two parts. Pedagogy in performance error detection might entail sectional rehearsal practica that initially involve only one line of music and subsequently progress to sectional rehearsing involving two and three parts. Given the common rehearsal technique of isolating large ensemble performance problems by reducing the number of concurrent instrumental or vocal sections (e.g. from full band to woodwinds alone to clarinets alone to second clarinets alone), a well-developed ability to detect errors in one-, two-, and three-parts would serve a novice conductor well.

This focused approach to teaching and learning skills in error detection seems at first glance to conflict with the "fuzzy" focus or nonspecific error detection strategies espoused by Hayslett (1992). The latter procedure for listening is likened to a visual approach in which a nonspecific gaze might allow one to see more than would be the case if one had focused on specific elements. Specific and nonspecific strategies for performance error detection may not be incompatible. In fact, the issue may not be one of whether to use either strategy; instead it may be an issue of when and for whom each is most appropriate. Future research is needed to sort out this issue. What might be the effects of training that proceeds from the general (nonfocused) to the specific (focused) before returning to the general? Do the different learning styles of students preclude certain approaches and likewise invite others? Is it possible, perhaps even likely, that initial preservice teacher experiences in error detection involving nonspecific strategies could be quite daunting, especially considering the large number of possible performance errors and error contexts that confront young teachers in elementary and secondary school performing ensembles?

The specific effects of texture on error detection responses are unclear. Results of the four-factor ANOVA do not show consistent patterns in subjects' correct responses. It is difficult, for example, to interpret a result that shows graduate students performing better on three-part polyrhythmic excerpts than on two-part polyrhythmic excerpts involving pitch errors. Factors that might contribute to this lack of clarity include the possibility of an excerpt-specific variable in detection of performance errors that precludes neatly packaged interpretations. One possible explanation for the difficulty subjects experienced with the two-part polyrhythmic excerpts is that the insertion of performance errors in these excerpts did no readily apparent harm to the sound of the music. Nothing sounded wrong, although what was notated was not what was performed. However, comparison with inserted errors in the three-part polyrhythmic excerpts, for which subjects were more accurate, does not support this argument, as no "alarming" sounds resulted from inserted errors under this textural condition either. One implication of these data is that when texture is considered as a factor, an increase in the number of parts and in the independence of these parts does not necessarily increase the complexity of the error detection task.

It seems safe to assume that, in order to detect performance errors with a high degree of accuracy, one must at some point formulate accurate expectations of the sound of the music. The error detection task, in this case, becomes one of determining whether musical realization is consistent with expectation. Of course, development of accurate aural images is one benefit of score preparation. Results of this study indicate that accurate expectations of rhythm performance are formulated with greater ease and accuracy than are those for pitch performance. Subjects were given only 20 seconds to acclimate to each new musical example. Given the brevity of this period, it is likely that expectations involving rhythm performance were being formulated simultaneously with the performance of the excerpts. Future research might note subjects' responses to conditions in which aural expectations, particularly those involving pitch, are established through various forms of score preparation.

REFERENCES

Blocher, L. R. (1990. February). *An analysis of college band instrumentalists' detection of common performance errors.* Paper presented at the Southern Division CBDNA/NBA Conference, Tallahassee, 1990.

Brand, M., & Burnsed, V. (1981). Music abilities and experiences as predictors of error-detection skill. *Journal of Research in Music Education, 29,* 91–96.

Byo, J. L. (1993). The influence of textural and timbral factors on the ability of music majors to detect performance errors. *Journal of Research in Music Education, 41,* 156–167.

Byo, J. L., & Brooks, R. (1994). Comparison of junior high musicians' and music educators' performance evaluations of instrumental music. *Contributions to Music Education, 21,* 26–38.

Costanza, A. P. (1971). Programed instruction in score reading skills. *Journal of Research in Music Education, 19,* 453–459.

Deal, J. J. (1985). Computer-assisted instruction in pitch and rhythm error detection. *Journal of Research in Music Education, 33,* 159–166.

DeCarbo, N. J. (1982). The effects of conducting experience and programmed materials on error-detection scores of college conducting students. *Journal of Research in Music Education, 30,* 187–200.

Doane, C. (1989). The development and testing of a program for the development of conductor aural error detection skills. *CBDNA Journal, 6,* 11–14.

Dolbeer, R. C. (1969). Programmed instruction in detecting aural-notational errors in musical performance (Doctoral dissertation, Ohio State University, Columbus). *Dissertation Abstracts International, 30* (10), 4304A.

Duke, R. A., & Pierce, M. A. (1991). Effects of tempo and context on transfer of performance skills. *Journal of Research in Music Education, 39,* 93–100.

Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education, 23,* 186–196.

Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication. *Journal of Research in Music Education, 25,* 256–263.

Forsythe, J. L., & Woods, J. R. (1983). The effects of conducting on the error detection ability of undergraduate and graduate instrumental conductors. *Contributions to Music Education, 10,* 27–31.

Froseth, J. O., & Grunow, R. A. (1979). *MLR instrumental score reading program.* Chicago: G.I.A. Publications.

Gardner, H. (1991). *The unschooled mind.* New York: Basic Books.

Geringer, J., & Madsen, C. (1989). Pitch and tone quality discrimination and preference: Evidence for a hierarchical model of musical elements. *Canadian Music Educator, 30,* 29–38.

Gonzo, C. L. (1971). An analysis of factors related to choral teachers' ability to detect pitch errors while reading the score. *Journal of Research in Music*

Education, 19, 259–271.
Grunow, R. F. (1980). An investigation of the relative effectiveness of four modes of score preparation on visual-aural discrimination skills development (Doctoral dissertation, University of Michigan, Ann Arbor). *Dissertation Abstracts International, 41,* 580A.
Grutzmacher, P. A. (1987). The effect of tonal pattern training on the aural perception, reading recognition, and melodic sight-reading achievement of first-year instrumental music students. *Journal of Research in Music Education, 35,* 171–181.
Hayslett, D. J. (1992). *The effect of directed focus on the peripheral hearing of undergraduate instrumental music majors.* Paper presented at the national convention of the Music Educators National Conference, New Orleans.
Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception, 6* (4), 361–382.
Istvan, M. (Ed.). (1973). *Chamber music for beginners.* New York: Boosey & Hawkes.
Larson, R. C. (1977). Relationships between melodic error detection, melodic dictation, and melodic sightsinging. *Journal of Research in Music Education, 25,* 264–271.
McKnight, C. B. (1975). Music reading ability of beginning wind instrumentalists after melodic instruction. *Journal of Research in Music Education, 23,* 23–34.
Malone, P. R. (1986). Development and evaluation of an approach to increasing pitch error detection skills in choral music education students (Doctoral dissertation, Florida State University, Tallahassee, 1985) *Dissertation Abstracts International, 46,* 3644A.
Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18* (1), 16–25.
Price, H. E. (1992). Transferring teaching concepts from methods course to studio instruction. *Contributions to Music Education, 19,* 75–86.
Price, H. E., Blanton, F., & Parrish, R. T. (1993, July). *Effect of two instructional methods on high school band students' sight-reading proficiency, music performance, and attitude.* Paper presented at the Sixth International Conference of the World Association for Symphonic Bands and Ensembles, Valencia, Spain.
Ramsey, D. S. (1979). Programmed instruction using band literature to teach pitch and rhythm error detection to music education students. *Journal of Research in Music Education, 27,* 149–162.
Sidnell, R. G. (1971). Self-instructional drill materials for student conductors. *Journal of Research in Music Education, 19,* 85–91.
Stuart, M. (1979). The use of videotape recordings to increase teacher trainees' error detection skills. *Journal of Research in Music Education, 27,* 14–19.
Voxman, H. (1952). *Chamber music for three clarinets, volume II.* Chicago: Rubank.
Voxman, H. (1953). *Chamber music for three woodwinds, volume II.* Chicago: Rubank.
Winter, C. M. (1990). The effects of primacy/recency and excerpted versus full-length performances on large instrumental group adjudication (Doctoral dissertation, University of Alabama, 1989). *Dissertation Abstracts International, 51,* 780A.

Submitted June 30, 1994; accepted July 12, 1995.