Original Research Article

National Association for Music Education

Journal of Research in Music Education
1–16

DOI: 10.1177/0022429419885931
jrme.sagepub.com
SAGE

The Relationships Among Interval Identification, Pitch Error Detection, and Stimulus Timbre by Preservice Teachers

Laura A. Stambaugh[1] and Bryan E. Nichols[2]

Abstract

We examined the relationship between interval identification skill and error detection skill in preservice teachers, accounting for timbral differences by including piano and vocal stimuli. The interval identification test was comprised of 33 items spanning from C2 to B5. Fifteen error detection items were monophonic melodies, two measures long, in 4/4 meter, and included one pitch error. Music education majors ($N = 50$) completed both tests in vocal and piano timbres during one individual study session. Interval identification performance was significantly correlated with error detection performance, $r = .75$. Additionally, interval identification score was a significant predictor for error detection when also accounting for variance from numbers of semesters of enrollment and theory/aural skills courses. Response times for correct responses of interval identification were faster than for incorrect responses. We found no main effects or interactions between primary performance area and timbre of test item. The results suggest interval identification skill generally can be used to predict error detection, reinforcing the importance of developing interval identification as a basic musical skill.

Keywords

error detection, aural skills, interval, preservice teacher, musicianship

[1]Georgia Southern University, Statesboro, GA, USA
[2]The Pennsylvania State University, State College, PA, USA

Corresponding Author:
Laura A. Stambaugh, Department of Music, Georgia Southern University, Box 8052, Foy Building #1004, Statesboro, GA 30460, USA.
Email: lstambaugh@georgiasouthern.edu

Preservice teachers are expected to develop basic aural skills that can be applied to more complex musical tasks (Shatzkin, 1984). Error detection (a teacher's or conductor's ability to identify errors in performance) is one of those critical, complex tasks (Sheldon, 2004). Waggoner (2011) suggested that understanding how specific factors affect error detection skill may lead to the development of more effective curricula. Interval perception and error detection have long histories of study in music education and music cognition research, yet researchers have only included them within the same empirical design in general ways. For example, Brand and Burnsed (1981) examined the predictive validity of grades in theory and aural skills classes on performance on an error detection test and determined it was difficult to generalize about the relationship between course grading and a singular outcome such as error detection because courses will vary in terms of content, topical frequency, and grading criteria. Here, we examined whether the ability to identify intervals was related to error detection skill and whether both skills were bound by timbral constraints.

Listeners must process intervals in various types of presentation: melodic ascending, melodic descending, or harmonic. College students identified melodic intervals more easily than harmonic (Balzano & Liesch, 1982; Buttram, 1969; Wuthrich & Tunks, 1989), ascending intervals more successfully than descending (Russo & Thompson, 2005a, 2005b), and consonant intervals more successfully than dissonant intervals (Kuusi, 2007; Seror, 2011). The range of pitches used in most studies has been one octave within E3 and G4 (e.g., Shatzkin, 1981, 1984). Smaller ranges have been used (Siegel & Siegel, 1977), and intervals larger than an octave were tested by Miyazaki (1988, C3–C6) and Russo and Thompson (2005b, F2–F5). Russo and Thompson found trained listeners were better at estimating the magnitude of melodic intervals an octave or smaller, but trained listeners showed no advantage for intervals larger than an octave. They suggested all listeners struggled to identify larger intervals because those are not commonly found in melodies or because the listeners segregated the low and high pitches into two different auditory streams (e.g., Bregman, 1990). This finding is particularly interesting because ensemble conductors regularly assess vertically aligned pitches spanning over an octave.

Several researchers have ranked the difficulty level of specific intervals, generally finding the P8 is the most correctly identified interval, while the m6, m7, and TT (tritone) are most likely to be identified incorrectly. Listeners showed the m2 has been found to be both easy (Balzano & Liesch, 1982; Poland, 1960; Ponsati, Miranda, Amador, & Godall, 2016; Samplaski, 2005) and difficult (Plomp, Wagenaar, & Mimpen, 1973). Despite the trends found in these studies, Buttram (1969) examined harmonic intervals through their quale (the unique sound produced by simultaneous pitches) and asserted that there is no innate quality of intervals that would allow the development of a taxonomy of difficulty.

A ubiquitous aspect of any interval test is timbre. Piano was used in several interval identification studies (Seror, 2011; Shatzkin, 1984; Spohn, 1963), but the most frequently used stimulus timbre has been pure and complex sinusoidal tones (Balzano & Liesch, 1982; Kuusi, 2007; Lou, Masterson, & Wu, 2014; Miyazaki, 1988; Plomp et al., 1973; Rakowski, 1990; Schellenberg & Trehub, 1996; Siegel & Siegel, 1977; Zarate, Ritson, &

Poeppel, 2012). Loh (2007) stated that the piano is the most commonly found instructional instrument in college classrooms, likely resulting in students hearing aural skills examples played on piano. He questioned the appropriateness of piano timbre for the majority of music majors, who would not become professional pianists. Timbre has received limited attention as an independent variable. Loh found guitar was more effective for interval training than piano, and Russo and Thompson (2005a) found that presenting melodic intervals using bright and dull timbres affected perception of interval size. Other instrumental timbres used in research are organ (Jeffries, 1967), flute (Zarate et al., 2013), and "pseudo-clarinet" (Samplaski, 2005). A noticeable gap in the literature is vocal timbre, with only one study including a synthesized vocal timbre (Zarate et al., 2013).

Finally, several authors included measures of response time as a dependent variable in interval identification research. Miyazaki (1988) suggested response time "reflects accessibility to the internal representation of music pitch" (p. 510). He found a consistent relationship between response times and standard deviation of responses. Overall, shorter response times were found for tasks considered easier than other tasks, with "ease" indicated by correct/incorrect answer rates (Balzano & Liesch, 1982; Miyazaki, 1998; Seror, 2011; Zarate et al., 2012).

Error detection can be defined as the ability to evaluate aural excerpts whether live or recorded, isolated or in context, compared to notation or aural concept (Davis, 2010). Doerksen (1999) referred to it as a type of diagnostic aural skill. The goal of previous studies has been to describe abilities in specific populations regarding type and placement of errors, including how students respond to such errors in front of actual ensembles (Cavitt, 2003; Waggoner, 2011). Generally, studies present a specific instrument for monophonic excerpts or instrumental ensembles playing multiple parts. One study used string instruments to play stimuli (Stuart, 1979), but vocal or choral timbres are underrepresented in error detection research.

Musicianship tasks can tax cognitive load in various ways. First, error detection assessments can be comprised of rhythm or pitch components or both. Collegiate participants performed more accurately at detecting instrumental rhythm errors than pitch errors (Byo, 1993, 1997), and one researcher chose to isolate pitch by using all quarter note rhythms with junior high participants (Killian, 1991). Second, error detection performance decreased with additional parts in the texture (Byo, 1997; Crowe, 1996; Sheldon, 1998), suggesting additional parts increase complexity of the task and divided attention is a factor in aural skill development. Introducing certain score study strategies like singing has not increased performance in multipart excerpts (Byo & Sheldon, 2000). Another facet of divided attention is when choral directors utilize the piano to support singers. When junior- and senior-level music education majors played piano while simultaneously listening to an ensemble, their error detection ability decreased (Napoles, Babb, Bowers, Hankle, & Zrust, 2016). In addition, conductors have shown superior divided attention ability compared to pianists for timing tasks but not for pitch tasks—and experts to a greater extent than students (Wöllner & Halpern, 2016). This suggests timing elements prevail over pitch elements and lends support to the recent suggestion that pitch elements should be considered independently of timing elements (Nichols, Wöllner, & Halpern, 2018).

Error detection is a skill shown to improve with age, experience, or training. For example, graduate students were more accurate than undergraduate students (Byo, 1997) and upperclassmen more accurate than underclassmen (Gonzo, 1971). Performance on specific error detection tasks improved with instruction, and comparisons of preservice teachers to experienced teachers revealed developmental differences. It is unclear what implications exist for preservice training.

Since an initial study of self-instructional materials for student instrumental conductors (Sidnell, 1971), modern textbooks that focus on error detection are now available for instrumental music (e.g., Spradling, 2010) and choral music students (e.g., Hondorp, 2015), although it is unknown how widespread the use of error detection and other diagnostic materials are in aural skills, methods, or conducting courses. Such training materials are specific to instrumental or choral domains, and research indicates that preservice teachers are more accurate at detecting errors within their domain (Stambaugh, 2016). For those participants, theory course grades were not significantly correlated to error detection performance, and aural skills course grades were significantly correlated to error detection outside the participant's main area (choral vs. instrumental). Although a player's main instrument did not influence ear training proficiency (Wolf & Kopiez, 2018), there remain other possible correlates to error detection skills (e.g., piano experience; Lehmann, 2014).

Researchers have suggested that aural skills do not necessarily transfer to error detection skills required for conducting ensembles and that the specific relationship between interval identification and pitch error detection has not been carefully explored (Byo, 1993). Examining factors related to these skills, such as grades earned from theory course work, has yielded inconsistent results (Brand & Burnsed, 1981; Gonzo, 1971). Evidence points to the role of piano experience for music majors studying musical scores prior to error detection (de Stwolinski, Faulconer, & Schwarzkopf, 1988). It is unclear whether instrument study, lesson history, or class piano, as well as other variables such as interval identification skill, influence error detection skill in preservice teachers. Therefore, the purpose of this study was to explore the relationships among these variables, plus between interval identification and error detection, as well as the potential effect of timbre match/mismatch conditions. We developed two tests in each of two timbre conditions to assess preservice teachers' skills and then analyzed their responses for accuracy and response time. The research questions were (1) What is the relationship between interval identification ability and error detection ability by preservice teacher candidates? (2) Do interval identification and error detection skills show timbral context constraints (e.g., Are instrumentalists better at listening to an instrumental timbre?)? and (3) What is the relationship between response time and accuracy for interval identification and error detection?

Method

Participants

Participants (N = 50; females, n = 23; males, n = 27) were undergraduate music education majors (instrumentalists, n = 27; vocalists, n = 23) from two universities with enrollments of 25,000 to 27,000 students. The number of participants from each site was equal, and students ranged from freshmen to seniors. A background questionnaire was used to establish how many semesters each participant was enrolled in college; how many courses were completed in music education, music theory, aural skills, group piano, and conducting; and document private lessons in the primary performance area and piano. At one site, the degree sequence included four semesters of combined theory and ear-training courses (four credits each), two semesters of class piano, and two semesters of keyboard harmony. At the other site, the degree sequence included four semesters of theory (three credits each), four semesters of a separate sight-singing/ear-training skills (one credit each), and four semesters of group piano.

Stimuli

Interval identification test. We used a range from C2 through B5 for the test items because it is similar to the ranges teachers encounter when working with high school bands, choirs, or orchestras. We devised a test using melodic and harmonic interval presentations. The starting pitch for melodic intervals and the bottom (low) pitch for harmonic intervals were equally distributed across four octaves. An equal number of ascending, descending, and harmonic items occurred in the two octaves above and below middle C. The template for the melodic interval test items was to hear 1,500 ms of white noise, Pitch 1 for 750 ms, and then Pitch 2 for 750 ms (Balzano & Liesch, 1982; Kuusi, 2007; Samplaski, 2005; at 80 beats per minute [bpm], two beats of white noise, followed by Pitch 1 for one beat, and then Pitch 2 for one beat). We included white noise to reduce potential carryover effects from the previous item during testing. The only difference for the harmonic interval presentation was that the pair of pitches sounded simultaneously for 750 ms (one beat). When listening to melodic intervals, participants had to wait to hear the beginning of the second pitch before they could respond with their answer. However, they could respond to the harmonic interval stimuli immediately upon hearing the initial presentation of both pitches. This potential difference of .75 s in minimal response time was not a practical concern because the mean response times for the three types of intervals were later found to be 8.09 to 10.61 s. In addition, the slowest response times were for descending intervals and not harmonic intervals.

Consistent with previous research (e.g., Byo, 1993; Thornton, 2008), synthesized sounds in prerecorded files were used to ensure identical presentations of the stimuli. The test items were created in Finale using the Grand Piano timbre and the Vocal Oohs timbre, which allowed us to control all parameters of the sound test files. We constructed separate tests for the two timbres, with 33 test items unique to each timbre. The 66 Finale files were exported as .mp3 files, for use with Qualtrics survey software.

We included the interval of a perfect octave only for practice items because previous research has established it is the easiest interval to identify. The intervals from m2 to M7 were presented within the bass and treble ranges in three presentation modes:

melodic ascending, melodic descending, and harmonic. In the compiled tests, the presentation orders were random for mode and range. The administration time for each test was less than 10 min.

Error detection test. In developing a test comparing piano and vocal timbres in isolated monophonic excerpts, we excluded rhythm due to the primacy of rhythm effects in previous error detection tests. We modeled this pitch-only test (see Figure S1 in the online version of the article) after two-bar phrases used by Killian (1991), containing eight quarter notes. After conducting a pilot test (N = 9) using Killian's eight melodies, we made several changes to the stimuli. Because there was a ceiling effect, we created additional items of increased difficulty by including notated chromatic pitches (chromatic pitches in the notated stimuli) and chromatic pitch error foils (chromatic foils for diatonic notated pitch excerpts). This pilot process resulted in a revised test built of 15 items. Tonal or atonal context is shown to affect performance (Groulx, 2013). Therefore, the two-bar phrases were each based in one key center of F major. Each excerpt began on the first through sixth scale degrees, and excerpts did not always end on the key center. The phrases contained eight quarter notes and were presented in the range C3 to F#4. They were prepared in Finale using the Grand Piano and Vocal Ooohs timbres at 60 bpm. Each phrase contained only one error, and the participant was asked to mark the inaccurately played pitch or select "no errors." We anticipated the resulting 15 items would present easy, moderate, and difficult test items for the target population.

Procedure

We tested participants in individual study sessions that lasted 45 to 60 min (demographic survey, two interval tests, and two error detection tests at 10 min each) in a quiet room with a researcher present. Qualtrics survey software was used to administer the tests on MacBook Pro computers. First, participants completed the background questions. Next, they completed the four tests in an order counterbalanced across participants. Each test began with two practice questions presented in the same format as the test items. During the practice items, participants could adjust the volume of the computer speakers to their comfort level. We told participants they could play each sound file one time, and the researcher was able to monitor participants' activity to ensure they adhered to this protocol. Participants were also told that their answers were being timed but accuracy was more important than speed.

In the interval tests, the computer screen showed an audio file, the question, "Which interval do you hear?" and radio buttons for all intervals from m2 to M7 to select their answer. Participants clicked on the sound file, listened to the item, and chose their answer as soon as they wanted to. It was possible to select an answer and then change that answer before proceeding to the next item, after which a response was recorded and could not be changed. The software recorded response time as the onset of clicking the audio file until the last answer button was selected. After

Table 1. Means and Standard Deviations for Interval Identification and Error Detection Tests.

		Combined Score 66 Items	Piano Timbre 33 Items	Vocal Timbre 33 Items
Interval test score	All students	38.29 (13.71)	21.35 (6.72)	16.57 (7.59)
	Instrumentalists (n = 27)	37.19 (12.06)	20.96 (6.01)	16.22 (6.70)
	Vocalists (n = 23)	39.71 (15.78)	21.82 (7.62)	17.00 (8.70)
Mean response time per item	All students	9.33 (2.64)	8.89 (2.56)	9.78 (3.16)
	Instrumentalists	9.18 (2.69)	8.91 (2.79)	9.44 (2.93)
	Vocalists	9.52 (2.64)	8.86 (2.32)	10.18 (3.43)
Error detection test score	All students	20.19 (6.98)	9.92 (3.93)	10.04 (3.64)
	Instrumentalists (n = 27)	20.48 (6.47)	10.11 (3.64)	10.37 (3.24)
	Vocalists (n = 23)	19.81 (7.74)	9.70 (4.32)	9.62 (4.25)
Mean response time per item	All students	10.07 (3.84)	10.13 (4.89)	10.04 (3.48)
	Instrumentalists	9.65 (2.07)	9.73 (2.38)	9.57 (2.29)
	Vocalists	10.57 (5.23)	10.60 (6.79)	10.61 (4.54)

Note. Response time is in seconds.

completing an item, they clicked on a next-page arrow to move to the next item. In the error detection tests, the computer screen showed the notation for the melody, an audio button to click to start the sound file, the question, "Which pitch (if any) is performed incorrectly?" and then the numbers 0 to 8 or "no error" with radio buttons. Response time was recorded from the onset of the sound file until the participant selected their final answer.

Preliminary Analysis

The mean age for all participants was 20.44 years (SD = 1.28). Instrumentalists (n = 27; M_{age} = 20.85, SD = .95) were significantly older than vocalists (n = 23; M_{age} = 19.96, SD = 1.46), $t(48) = -2.606$, $p = .012$, $d = .72$, and instrumentalists had completed significantly more courses in theory/aural skills (M = 4.59, SD = 0.97) than vocalists (M = 3.57, SD = 1.62), $t(48) = -2.767$, p =.008, d = .76. Males and females had similar levels of piano experience (see Table S1 in the online version of the article), $F(1, 48) = 1.204$, $p = .278$, $\eta_p^2 = .024$, although more females than males had private piano experience. Age was significantly correlated with theory/aural skills class ($r = .667$, $p < .001$). Therefore, subsequent analyses comparing vocalists to instrumentalists used the number of theory/aural skills courses as a covariate. Table 1 presents the means and standard deviations for total and subscores.

We scored interval identification answers at the item level dichotomously (0 = incorrect, 1 = correct), and a composite score was calculated. Error detection answers were also summed, reflecting the number of correct responses per timbre condition. We screened all data for missing values, and response times were also screened for outliers. One participant failed to complete the vocal error detection test due to

Table 2. Regression of Error Detection Scores on Interval Identification Scores and Theory Course Level.

	b	*SE*	β	*t*	95% Confidence Interval	*p*
Intercept	2.98	2.43	—	1.23	−1.93, 7.88	.228
Semesters enrolled	.78	.46	.26	1.72	−1.14, 1.70	.093
Theory enrollment	.11	.84	.02	.13	−1.60, 1.81	.900
Interval identification	.34	.05	.69	6.55	0.23, 0.44	<.001

protocol error. We included this person's data for the other three tests completed. After reviewing our response time data, we were concerned about excluding answers with response times more than 3 *SD* from the corresponding participant's mean because it has already been established that some intervals are more difficult to identify (e.g., Buttram, 1969). Also, we allowed participants to use their own strategies, such as singing, to help determine their answers (Miyazaki, 1988). We set 60 s maximum as the cutoff point for outliers because that length of time could indicate a study session was interrupted or there was a technological problem. All data fit within this threshold. Visual inspection of histograms indicated positive skew for responses times. Therefore, we logarithmically transformed all response times.

The theory and aural skills course sequences varied between the two universities, so one researcher recoded these responses to conform them in one variable. The difficulty level of the interval test items, as measured by rate of correct responses, ranged from .16 to .96 for piano and from .18 to .78 for vocal intervals. Item difficulty for piano error detection ranged from .36 to .88 and from .47 to .82 for the vocal form. We calculated internal consistency and reliability using Cronbach's alpha and Spearman-Brown split-halves method. The Cronbach's alpha values were $\alpha = .81$ to .91, and $r_{SB} = .75$ to .93. These values were considered to be within acceptable ranges (Allen & Yen, 2001).

Results

The first purpose of this study was to examine the relationship between interval identification skill and error detection skill on a test of pitch errors in monophonic excerpts. Results indicated these preservice music teachers' abilities to identify intervals was very strongly related to their abilities to detect errors in this test, $r(46) = .75, p < .001$. While this relationship was significant, our main interest was whether interval identification performance could be used to predict error detection performance even when accounting for age or experience. Thus, we tested a linear regression model including number of semesters of enrollment, theory/aural skills class level, and interval identification score. The level of collinearity was considered acceptable because all variance inflation factors were below 2.9. The resulting model, $F(3, 42) = 23.720$, $p < .001$, explained 63% of the variance in error detection score. The standardized beta coefficient shows that interval identification was the strongest predictor and semesters of enrollment was a secondary source of variation, as shown in Table 2.

Table 3. Descriptive Statistics for Correct and Incorrect Response Times per Item in Seconds.

		N	Minimum	Maximum	Mean	*SD*
Piano error detection	Correct	49	4.59	39.83	10.04	4.97
	Incorrect	45	3.58	40.33	10.86	6.18
Vocal error detection	Correct	49	4.87	34.18	9.67	4.39
	Incorrect	46	6.47	36.74	11.80	6.24
Piano interval identification	Correct	50	4.40	15.22	8.07	2.28
	Incorrect	50	4.60	28.09	10.84	4.55
Vocal interval identification	Correct	50	4.76	17.19	8.54	2.74
	Incorrect	41	5.79	21.81	11.34	4.22

Note. Missing data is due to skipped answers and one participant failing to complete one test.

The second research question was whether vocal and instrumental music majors show timbral constraints in interval identification and error detection tests. We employed two-way repeated analysis of covariance (ANCOVA) measures to examine this question. For interval identification, we found no main effect of instrumental or vocal emphasis on overall ability to identify intervals in both timbres, $F(1, 45) = 3.013, p = .089$, $\eta_p^2 = .060$, and no interaction between piano/vocal interval timbre and instrumental/vocal emphasis, $F(1, 45) = .002, p = .962$, $\eta_p^2 = .000$. For error detection, we also found no main effect of students' area of emphasis on overall error detection ability, $F(1, 45) = 0.91, p = .764$, $\eta_p^2 = .002$, and no interaction between piano/vocal error timbre and instrumental/vocal emphasis, $F(1, 45) = 1.627, p = .209$, $\eta_p^2 = .035$.

Finally, our third research question asked what the relationship was between response time per item and accuracy. Participants were significantly faster at making correct responses than incorrect responses for piano intervals, $F(1, 48) = 3.961, p = .05$, $\eta_p^2 = .076$. All other comparisons were not significant ($p > .05$), as shown in Table 3. Based on the recent results of Wolf and Kopiez (2018), who found males significantly outperformed females on an ear-training assessment, we used ANCOVA to examine the impact of gender on our aural skills tests. We found no significant differences by gender for accuracy scores. However, females ($M = 7.89$ s, $SD = 2.14$) were significantly faster than males ($M = 9.64$ s, $SD = 2.69$) at identifying piano intervals, $F(1, 47) = 8.046, p = .007$, $\eta_p^2 = .146$, and vocal intervals (females, $M = 8.98$, $SD = 3.72$; males, $M = 10.51$, $SD = 2.72$), $F(1, 47) = 4.688, p = .035$, $\eta_p^2 = .091$, with a trend for also being faster at detecting piano errors (females, $M = 8.95$, $SD = 1.66$; males, $M = 10.03$, $SD = 2.58$), $F(1, 47) = 3.786, p = .058$, $\eta_p^2 = .075$.

Discussion

We examined the extent to which interval identification skill accounted for monophonic pitch error detection skill in preservice teachers. We presented results based on interval presentations in two timbres across a four-octave range to vocalists and instrumentalists. The ability to identify intervals in all ranges was strongly correlated with the ability to identify errors in brief monophonic melodies. In addition, a regression analysis accounting for semesters of enrollment and number of theory/aural skills classes suggested that performance on a test of interval identification was a significant predictor of pitch error detection skill. While primary performance area did not interact with the timbre of presented test items, we did find an unexpected gender difference of females responding significantly faster than males for interval identification items.

The development of error detection is an important skill for preservice teachers. Because we found evidence suggesting interval identification skill can generally be used to predict pitch error detection in simple contexts, we suggest students may benefit from starting error detection work early in their aural skills training. It is unknown if a similar relationship exists between interval identification and more complex music, which may be perceived in larger vertical and horizontal units. For every point scored higher on this test of interval identification, a significant gain of .69 points was earned on pitch error detection. This finding addresses a long-standing question about error detection: Is error detection a separate skill unto itself (Brand & Burnsed, 1981; Byo, 1993), or is error detection a product or synthesis of other, more elemental aural skills (Gonzo, 1971; Sidnell, 1971)? In addition, our regression model indicated semesters of enrollment made a unique contribution, beyond theory/aural skills class, to error detection ability, consistent with graduate students being more accurate than undergraduate students (Byo, 1997) and upperclassmen more accurate than underclassmen (Gonzo, 1971). Future research could investigate additional sources of variance that have not previously been examined, such as working memory capacity.

We selected a range of four octaves for the intervals in this study because music teachers conduct ensembles that perform across many octaves and previous literature has tested a more limited range of intervals. Several participants commented that the lowest octave of vocal intervals, C2 to B2, was difficult to hear. However, the difficulty ratings for the lowest intervals encompassed a wide range (.18–.72), similar to scores in all other octaves, and no difficulty rating was lower than other octaves. Therefore, the students' comments may reflect a lack of confidence at this low octave. Russo and Thompson (2005b) suggested identical intervals played in different registers can be perceived as being different sizes. In addition, it is possible the participants had not heard intervals in such low ranges during their theory/aural skills classes.

We chose the piano as an instrumental timbre for comparison to the synthetic vocal timbre because none of our participants were piano majors and to provide continuity with previous research. It is unknown how participants may have responded to test intervals or errors played on an instrument similar to or different from their own instrument family. Our results are consistent with Ponsati and colleagues (2016), who did not find an advantage for match conditions between primary performance area and test timbre. However, Stambaugh (2016) did find match/mismatch differences for vocal and instrumental preservice teachers listening to recordings of actual choruses and bands when rhythm was included in error detection. Our use of one instrumental timbre and the Finale Vocal Oohs timbre as a vocal timbre are limitations of the present study. Future research should include recordings of vocalists, and instrument representation should be accounted for carefully among participants in studies of these skills.

It remains possible that participants' main instrument may mediate the effect of timbre on interval identification or error detection and that range may serve as a moderator for the effect of range on interval identification or error detection.

Previous studies of musicianship skills have made comparisons between male and female genders. For example, Wolf and Kopiez (2018) recently validated an assessment of analytical hearing, reporting males significantly outperformed females on the Musical Ear Training Assessment. The items in that assessment included melodic and harmonic identification and rhythm production. However, Ponsati and colleagues (2016) did not find gender differences in their test of harmonic intervals with adolescents in Spain, ages 11 and older. Adding to this inconsistency, females in our study were significantly faster than males for both piano and vocal intervals and (not significantly) for piano errors, with medium to large effect sizes. Notable differences among the three studies are in the characteristics of the samples. In Wolf and Kopiez, 59% of participants played a harmonic instrument, such as piano, while none of our participants indicated a harmonic instrument as their primary performance area. Neither Wolf and Kopiez nor Ponsati and colleagues included vocalists, while we did. Also, the majority of participants in Ponsati and colleagues' study were under the age of 14. It is difficult to draw conclusions about the role of gender in aural skills testing given the disparities in these samples and test items.

When teachers are on the podium, they need to identify and address errors in a timely manner. Our design included response time measurement not only because it is ecologically valid but also because it may represent fluency in cognitive processing (Miyazaki, 1988). Furthermore, Ranger and Kuhn (2013) noted, "It can turn out beneficial to measure response speed in addition to ability in the case that both characteristics of an individual contribute to the prediction of some external criteria like professional success" (p. 75). We found partial consistency with existing research in music (Miyazaki, 1988; Seror, 2011): Response times for correct identification of piano intervals were shorter than response times for incorrect answers, but this difference was not as strong for the error detection test. It is likely that when preservice teachers were confident in knowing the answer, they answered quickly. When they were unsure of the answer, they employed strategies such as quiet singing or mentally stepping through intervals (e.g., for a major third, singing *do–re–mi*). Future research could employ a limited response time window wherein participants must respond within a specified amount of time. This paradigm would simulate better actual teaching conditions with the pressure of students waiting for directions from the teacher.

Brand and Burnsed (1981) reported the test-retest reliability of the 10-day test window for their junior high band error detection test instrument (r = .59). Conflicting reports about the relationships between component skills lead us to recommend using previously employed measures and reporting item difficulty and internal consistency in newly designed tests. Our test included items that were easy, medium, and difficult in difficulty, according to standard principles of test design (Allen & Yen, 2001), and future designs should include measures and reporting of this information. In addition, our pitch error detection test was designed for efficiency and was limited to a small treble range to allow time for piano and vocal presentations. This is a limitation of the present study, and future work in error detection ranges can be undertaken to explore timbral and range differences among preservice teachers. Also, it is unclear whether the ability to detect errors in monophonic melodies is highly related to multipart ensemble recordings, and future comparisons of the predictive validity of monophonic error detection is required.

Implications for Teaching and Learning

The results of the present study highlight that the advanced teacher skill of error detection is strongly related to the more basic aural skill of interval identification. The findings further suggest the need to use a variety of pitch ranges and timbres to build students' confidence as they develop aural skills. Our test was highly reliable, and future work could be done to develop shorter versions of this test for research or instructional use. It is not uncommon for preservice teachers to use key phrases from familiar songs as surrogates for specific intervals. Smith, Nelson, Grohskopf, and Appleton (1994) suggested taking this strategy one step further: Early in their training, let students respond to intervals with the name of the surrogate instead of the corresponding interval name. Using the surrogate allows students to draw on long-term memory and their existing musical schemata, and it also builds their confidence and attitude toward aural skills training. This may be particularly useful for developing accuracy and confidence in the range of C2 to C3 and for intervals larger than an octave (Russo & Thompson, 2005b).

Finally, this study focused on monophonic error detection. Because error detection performance has been shown to decrease with additional parts (Byo, 1997; Crowe, 1996; Sheldon, 1998), this suggests that the ability to divide one's attention may be relevant for more complex error detection contexts. Students may benefit from using commercially available self-instructional materials for error detection that include multipart excerpts. The process of marking the error while listening versus marking after listening is similar to the conflict in dictation pedagogy: to write while listening or after listening (Paney & Buonviri, 2017). Buonviri (2015) found these two strategies were equally effective for undergraduates taking short tonal melodic dictation. Likewise, preservice teachers should explore a variety of approaches to using practice materials for error detection.

The development of error detection skills is, of course, not simply a function of having excellent interval detection ability. The manipulations present in previous research spotlight the many factors involved in identifying incorrect pitches, rhythms, dynamics, and so on in music performance. This study confirmed the assumption that a specific aural skill was important for a more advanced aural skill. Further research should continue to elucidate other constituent factors of error detection to better prepare preservice teachers for their future conducting experiences.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Laura A. Stambaugh https://orcid.org/0000-0002-7459-5890

Supplemental Material

Figure S1 and Table S1 are available in the online version of the article at https://doi.org/10.1177/0022429419885931

References

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Balzano, G. J., & Liesch, B. W. (1982). The role of chroma and scalestep in the recognition of musical intervals in and out of context. *Psychomusicology*, *2*, 3–31. Retrieved from http://psycnet.apa.org/psycinfo/1985-16421-001

Brand, M., & Burnsed, V. (1981). Music abilities and experiences as predictors of error-detection skill. *Journal of Research in Music Education*, *29*, 91–96. doi:10.2307/3345017

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Buonviri, N. (2015). Effects of a preparatory singing pattern on melodic dictation success. *Journal of Research in Music Education*, *63*, 102–113. doi:10.1177/0022429415570754

Buttram, J. B. (1969). The influence of selected factors on interval identification. *Journal of Research in Music Education*, *17*, 309–315. doi:10.2307/3343879

Byo, J. (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. doi:10.2307/3345405

Byo, J. (1997). The effects of texture and number of parts on the ability of music majors to detect performance errors. *Journal of Research in Music Education*, *45*, 51–66. doi:10.2307/3345465

Byo, J., & Sheldon, D. (2000). The effect of singing while listening on undergraduate music majors' ability to detect pitch and rhythm errors. *Journal of Band Research*, *36*, 26–46.

Cavitt, M. E. (2003). A descriptive analysis of error correction in instrumental music rehearsals. *Journal of Research in Music Education*, *51*, 218–230. doi:10.2307/3345375

Crowe, D. (1996). Effects of score study style on beginning conductors' error-detection abilities. *Journal of Research in Music Education*, *44*, 160–171. doi:10.2307/3345668

Davis, S. (2010). Error detection in the aural skills class: Research and pedagogy. *Journal of Music Theory Pedagogy*, *24*, 37–67.

de Stwolinski, G., Faulconer, J., & Schwarzkopf, A. B. (1988). A comparison of two approaches to learning to detect harmonic alterations. *Journal of Research in Music Education*, *36*, 84–94. doi:10.2307/3345242

Deal, J. (1985). Computer-assisted instruction in pitch and rhythm error detection. *Journal of Research in Music Education*, *33*, 159–166. doi:10.2307/3344803

Doerksen, P. (1999). Aural-diagnostic and prescriptive skills of preservice and expert instrumental music teachers. *Journal of Research in Music Education*, *47*, 78–88. doi:10.2307/3345830

Gonzo, C. (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. doi:10.2307/3343762

Groulx, T. (2013). The influence of tonal and atonal contexts on error detection accuracy. *Journal of Research in Music Education*, *61*, 233–243. doi:10.1177/0022429413485602

Hondorp, P. (2015). *Choral error detection: Exercises for developing musicianship*. New York, NY: Carl Fischer.

Jeffries, T. B. (1967). The effects of order of presentation and knowledge of results on the aural recognition of melodic intervals. *Journal of Research in Music Education*, *15*, 170–190. doi:10.2307/3343858

Killian, J. (1991). The relationship between sightsinging accuracy and error detection in junior high singers. *Journal of Research in Music Education*, *39*, 216–224. doi:10.2307/3344721

Kuusi, T. (2007). Interval-class and order of presentation affect interval discrimination. *Journal of New Music Research*, *36*, 95–104. doi:10.1080/09298210701755164

Lehmann, A. C. (2014). Using admission assessments to predict final grades in a college music program. *Journal of Research in Music Education*, *62*, 245–258. doi:10.1177/0022429414542654

Loh, C. S. (2007). Choice and effects of instrument sound in aural training. *Music Education Research*, *9*, 129–143. doi:10.1080/14613800601127619

Luo, X., Masterson, M. E., & Wu, C.-C. (2014). Melodic interval perception by normal-hearing listeners and cochlear implant users. *The Journal of the Acoustical Society of America*, *136*, 1831–1844. doi:10.1121/1.4894738

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception & Psychophysics*, *44*, 501–512. doi:10/3758/bf03207484

Napoles, J., Babb, S., Bowers, J., Hankle, S., & Zrust, A. (2016). The effect of piano playing on preservice teachers' ability to detect errors in a choral score. *Journal of Music Teacher Education*, *27*, 149–162. doi:10.1177/1057083716639724

Nichols, B., Wöllner, C., & Halpern, A. (2018). Score one for jazz: Working memory in jazz and classical musicians. *Psychomusicology: Music, Mind, and Brain*, *28*, 101–107. doi:10.1037/pmu0000211

Paney, A. S., & Buonviri, N. O. (2017). Developing melodic dictation pedagogy. *Update: Applications of Research in Music Education*, *36*(1), 51–58. doi:10/1177/875512331668615

Plomp, R., Wagenaar, W. A., & Mimpen, A. M. (1973). Musical interval recognition with simultaneous tones. *Acustica*, *29*, 101–109.

Poland, B. W. (1960). *An investigation of some aural and notational elements in music theory* (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.

Ponsati, I., Miranda, J., Amador, M., & Godall, P. (2016). Students' performance when aurally identifying musical harmonic intervals: Experimentation of a teaching innovation proposal. *International Journal of Music Education*, *34*, 445–458. doi:10.1177/0255761415619391

Rakowski, A. (1990). Intonation variants of musical interval in isolation and in musical contexts. *Psychology of Music*, *18*, 60–72. doi:10.1177/0305735690181005

Ranger, J., & Kuhn, J.-T. (2013). Analyzing response times in tests with rank correlation approaches. *Journal of Educational and Behavioral Statistics*, *38*, 61–80. doi:10.3102/1076998611431086

Russo, F. A., & Thompson, W. F. (2005a). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, *67*, 559–568. doi:10.3758/BF03193514

Russo, F. A., & Thompson, W. F. (2005b). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin & Review*, *12*, 1068–1075. doi:10.3758/BF03206445

Samplaski, A. (2005). Interval and interval class similarity: Results of a confusion study. *Psychomusicology*, *19*, 59–74. doi:10.1037/h0094040

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*, 272–277. doi:10.1111/j.1467-9280.1996.tb00373.x

Seror, G. A., III. (2011). *Intonation variants of musical interval in isolation and in musical contexts* (Unpublished master's thesis). State University of New York at Albany.

Shatzkin, M. (1981). Interval and pitch in and out of immediate context. *Journal of Research in Music Education*, *29*, 111–123. doi:10.2307/3345020

Shatzkin, M. (1984). Interval recognition in minimal context. *Journal of Research in Music Education*, *32*, 5–14. doi:10.2307/3345276

Sheldon, D. (1998). Effects of contextual sight-singing and aural skills training on error-detection abilities. *Journal of Research in Music Education*, *46*, 384–395. doi:10.2307/3345550

Sheldon, D. (2004). Effects of multiple listenings on error-detection acuity in multivoice, multitimbral musical examples. *Journal of Research in Music Education*, *52*, 102–115. doi:10.2307/3345433

Sidnell, R. (1971). Self-instructional drill materials for student conductors. *Journal of Research in Music Education*, *19*, 85–91. doi:10.2307/3344117

Siegel, J. A., & Siegel, W. (1977). Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, *21*, 143–152. doi:10.3758/BF03198717

Smith, J. D., Nelson, D. G. K., Grohskopf, L. A., & Appleton, T. (1994). Familiar tunes and music perception novice listeners. *Cognition*, *52*, 23–54.

Spohn, C. L. (1963). Programming the basic materials of music for self-instructional development of aural skills. *Journal of Research in Music Education*, *11*, 91–98. doi:10.2307/3344147

Spradling, R. (2010). *Error detection: Exercises for the instrumental conductor*. New York, NY: Carl Fischer.

Stambaugh, L. (2016). Differences in error detection skills by band and choral preservice teachers. *Journal of Music Teacher Education*, *25*(2), 25–36. doi:10.1177/1057083714558421

Stuart, M. (1979). The use of videotape recordings to increase teacher trainees' error detection skills. *Journal of Research in Music Education*, *27*, 14–19. doi:10.2307/3345115

Thornton, L. (2008). The effect of grade, experience, and listening condition on the melodic error detection of fifth- and sixth-grade woodwind students. *Update: Applications of Research in Music Education*, *26*(2), 4–10. doi:10.1177/8755123308317612

Waggoner, D. (2011). Effects of listening conditions, error types, and ensemble textures on error detection skills. *Journal of Research in Music Education*, *59*, 56–71. doi:10.1177/0022429410396094

Wolf, A., & Kopiez, R. (2018). Development and validation of the musical ear training assessment (META). *Journal of Research in Music Education*, *66*, 53–70. doi:10.1177/0022429418754845

Wöllner, C., & Halpern, A. (2016). Attentional flexibility and memory capacity in conductors and pianists. *Attention, Perception, & Psychophysics*, *78*, 198–208. doi:10.3758/s13414-015-0989-z

Wuthrich, C. E., & Tunks, T. W. (1989). The influence of presentation time asynchrony on music interval perception. *Psychomusicology: A Journal of Research in Music Cognition*, *8*, 31–46. doi:10.1037/h0094233

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *The Journal of the Acoustical Society of America*, *132*, 984–993. doi:10.1121/1.4733535

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2013). The effect of instrumental timbre on interval discrimination. *PLoS ONE*, *8*(9), Article e75410. doi:10.1371/journal.pone.0075410

Author Biographies

Laura A. Stambaugh is an associate professor of music education at Georgia Southern University. Her research interests include error detection skills, music-motor learning, and executive functions.

Bryan E. Nichols is an assistant professor of music at Pennsylvania State University. His research interests include singing accuracy, pitch perception, and the acquisition of music literacy.

Submitted May 22, 2018; accepted December 4, 2018.