Melodic Dictation Scoring Methods: An Exploratory Study

Jeffrey L. Gillespie

The task of melodic dictation has long been an established component of the aural skills curriculum, having earned its place as a valuable and multi-faceted learning tool (see Rogers [1984], Karpinski [1990]). An entire chapter is devoted to melodic dictation in Karpinski (2000), where the author calls the task "an important focal point of undergraduate music training" and states: "Working with melodic dictation can help to develop some very important musical skills such as musical attention, extractive listening, short-term musical memory, musical understanding, and notation" (p. 62).

Along with the rich resources in melodic dictation for developing a variety of musical skills, instructors are faced with the complex task of determining how best to evaluate dictation work. Karpinsky (2000) offers three main principles for guiding evaluation of dictation: "First, there should be some absolute standard against which the work is judged--reported, for example, in the form of a percentage or letter grade. Second, feedback should be given with regard to precisely what was correct and what was incorrect in each dictation. Third, when appropriate, suggestions should be made about what to do to correct systemic problems, whenever possible" (p. 103). Karpinsky (2000) emphasizes the crucial need "to offer meaningful, usable feedback for listeners' future development" (p. 110). This author wholeheartedly agrees with the pedagogically-sound principles outlined by Karpinsky. As instructors, we face the constant challenge of finding ways to incorporate these principles into the practical classroom.

To date, there are very few published methodologies for evaluating melodic dictation. Among them are the guidelines for

the standardized Graduate Record Examination (GRE) in Music and the Advanced Placement (AP) Examination in Music Theory. Both exams include only tonal melodies for dictation and evaluate both pitch and rhythm of solutions. The somewhat more elaborate system of the GRE does not accept enharmonic equivalents as being correct, but it does give "points back" for segments that are correctly notated but metrically shifted to an improper position. The AP exam does not give points back for such shifted segments. Neither the GRE nor the AP exams provide any feedback other than point scores.[1] Blombach's MacGamut 2000 computer software program improves upon these evaluation methods not only by accounting for temporal shifts within a response, but by providing markings on the music indicating the locations of errors, as well as messages describing the types of errors that are made. As an added pedagogical "bonus," the program checks for segments in which contour is maintained, despite the presence of pitch errors. Such contour-matching segments are marked for the respondent to see, even though point credit is not given in such cases. The scoring methodologies for all three of these sources are useful to study. However, the GRE and AP exams serve as evaluative tools, while MacGamut serves a pedagogical function. It is refreshing to see MacGamut reporting more to the students than simply a basic percentage score on a dictation exercise. Of Karpinsky's principles discussed earlier, all are represented in MacGamut except for suggestions on how to correct mistakes.

For the present study, several questions were considered in exploring how decisions are made by aural skills instructors in evaluating melodic dictation solutions. What factors should be involved in an instructor's diagnosis of error? Should evaluation methodology be affected by a melody's characteristics, such as its degree of tonal stability, the presence of melodic sequence, or other distinctive features? Should certain types of errors be weighed differently than other types? In this study, these questions, as well as others, are addressed through the results of two surveys that examine scoring methods and preferences among college instructors with interest

[1]See Phillips (1998) for more detailed discussion of AP and GRE scoring guidelines, as well as his own application of those models in the classroom, particularly with sight singing.

and experience in aural skills. The first survey, which is the more exploratory of the two, was designed to gather information on how instructors score dictation solutions. Information gathered from Survey I was influential in designing the more controlled Survey II, in which instructors were asked to rank several scoring methods as applied to specific melodies and dictation solutions. Of particular interest were melodies that contained varying degrees of tonal stability, including melodies that could be considered "non-tonal." Because of the complexity of the test material, rhythm was not included as a variable for study, although all melodies did include rhythm. Since rhythm is an integral and significant component of melody, it certainly deserves attention in a later study.

SURVEY I

In the initial survey (Survey I), six contrasting melodies were presented as hypothetical melodic dictation stimuli. The six melodies were comprised of two diatonic melodies (one major key and one minor key), two tonal melodies with some chromatics (one major key and one minor key), and two melodies that could be considered "non-tonal" or "tonally ambiguous." Presented with each melodic stimulus were two sample dictation "solutions" that were specially constructed by the researcher to contain a variety of different pitch errors. Errors included incorrect or missing accidentals, inappropriate enharmonic spellings, incorrect intervals resulting in transposed segments, incorrect structural pitches, missing segments (portions left "blank"), and in some cases, complete loss of tonal center. For reasons of simplicity, variables of contour and rhythm were not included in Survey I; thus, all solutions contained the correct contour and correct rhythm.

The survey was distributed to a group of 13 college-level music theory instructors with experience in teaching aural skills. Respondents were asked to score each solution on a 16-point scale and to provide detailed explanations of the scoring methodology they chose to use.

Respondents to Survey I revealed preferences for an assortment of scoring methods. For the two diatonic and two tonal-chromatic

melodies, the most popular scoring method was to count pitch errors, with no consideration given for any transposed segments of the melody.[2] There were several scorers in Survey I, however, who did choose to give half credit back for transposed segments. Another scorer chose to score solely by interval instead of pitch, regardless of the types of errors present in the solutions. Another scored by pitch with extra points "added back" for correct contour. Roughly half the respondents gave no credit for any pitch spelled as an enharmonic equivalent to a given pitch. For the two non-tonal melodies, there was a fairly even mixture of preferences for scoring by interval and scoring by pitch.

Despite the small pool of respondents, general scoring preferences lean toward scoring "by pitch" for the four tonal melodies, with or without credit for transposed segments. There was not a strong preference toward scoring by pitch or by interval for the two melodies with atonal characteristics, which suggests that there may be considerable difference of opinion on scoring methods for such melodies.

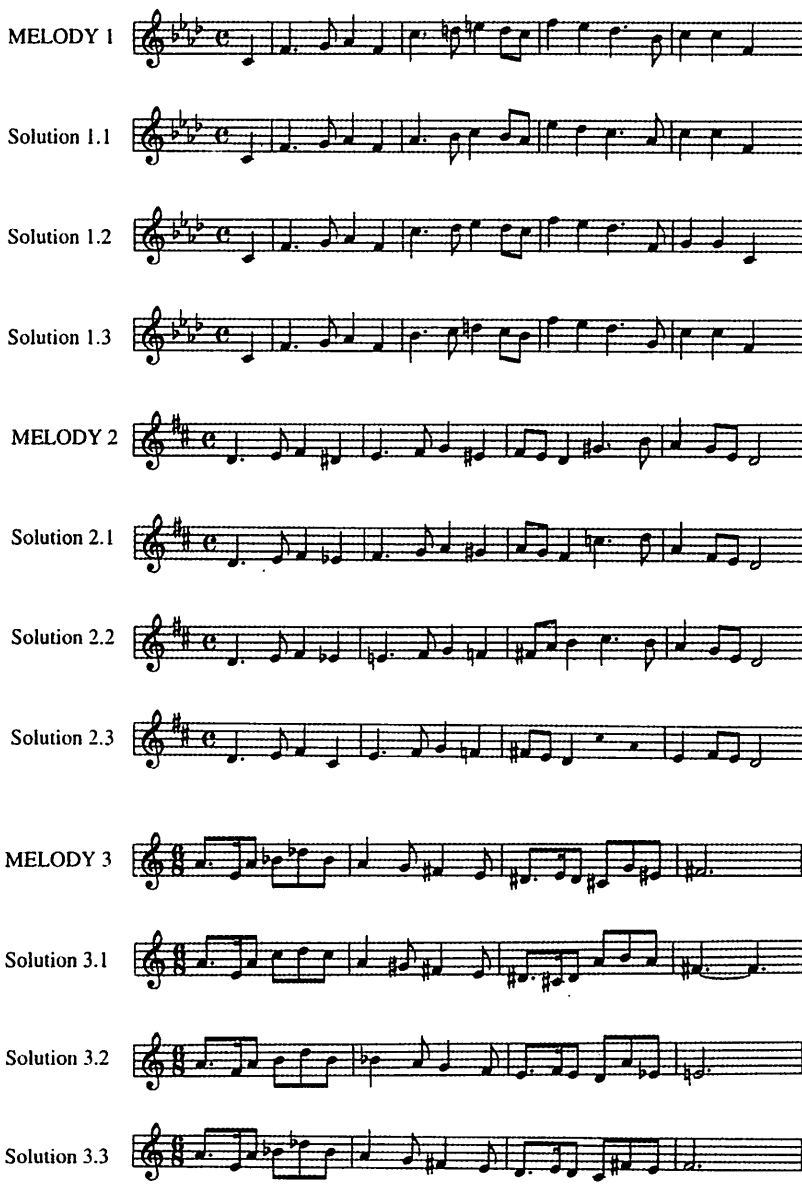
SURVEY II

Methodology

The more controlled Survey II featured three of the test melodies from Survey I (one diatonic, one tonal-chromatic, and one non-tonal), and each melody was paired with three different sample dictation solutions containing pitch errors. Unlike Survey I, in which dictation errors were designed by the researcher, Survey II featured dictation solutions selected from an exam administered to 51 students enrolled in the final semester of aural skills at Butler University. The solutions chosen for Survey II represent a variety of some of the most common errors from the student exam (see Example 1). As in Survey I,

[2]A "transposed segment" may be defined as a segment of the melody, generally three or more pitches in length, that maintains correct consecutive intervals but is "off-track" because of an earlier intervallic error. Thus, that portion of the melody is "transposed" to a different pitch level.

Example 1: The three dictation test melodies of Survey II, each paired with its three solutions.



rhythm was not included as a variable in solutions. Unlike Survey I, some errors in contour were included (see Solutions 2.2, 2.3, and 3.1 from Example 1).

Survey II was distributed to a broad sampling of college and university theory instructors from both large and small institutions across the United States. Of the 41 instructors who responded, 22% were professors, 29.3% associate professors, 19.5% assistant professors, 9.8% lecturers, and 19.5% "other" (primarily graduate students). Years' experience teaching music theory was distributed as follows: 1-5 years = 17.1%, 6-10 years = 24.4%, 11-15 years = 14.6%, 16-20 years = 17.1%, 21-25 years = 9.8%, 26-30 years = 7.3%, and more than 30 years = 9.8%. Years' experience teaching aural skills was distributed as follows: 1-5 years = 26.8%, 6-10 years = 17.1%, 11-15 years = 19.5%, 16-20 years = 22.0%, 21-25 years = 4.9%, 26-30 years = 2.4%, and more than 30 years = 7.3%. Those with experience teaching advanced-level aural skills (specifically post-second-year college undergraduate level or advanced graduate level) represented 37% of the subject pool. Each participant was also asked if his or her institution utilized any standardized scoring methodology for melodic dictation. Surprisingly, only 17.1% of the participants reported that their institution followed any sort of standardized methodology.

The 41 respondents were asked to rank five specific scoring methods, according to their preferences, for each of the sample solutions provided in the survey. The five scoring methods to be ranked were scoring by:

1) interval,
2) pitch,
3) pitch with transposition credit,
4) interval or pitch (choosing the method or combination of methods that causes the least point reduction),
5) interval and pitch (combining all errors of pitch and interval and dividing by two).

All five of these methods were among those chosen by participants in Survey I, and they were selected for Survey II either for their popularity in Survey I or simply to provide more variety among the methods in Survey II. Optional scoring methods were also available

to be ranked, at the respondents' discretion. The optional methods included opportunities to subtract points for such errors as missed structural scale degrees (such as tonic or dominant occurring on downbeats), incorrect contour, enharmonic spelling errors, or loss of tonal center. Example 2 presents the instruction page given to each participant, outlining all the scoring methods as applied to a sample melody and dictation solution. An actual scoring page for "Melody 2 – Solution 2.1" from Survey II is shown in Example 3, exactly as it would have appeared to participants as they completed their responses.

Results and Discussion

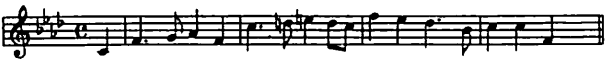
Statistical measurements of Survey II data addressed the following questions: 1) How do the five scoring methods rank according to popularity among scorers? 2) Does preference for a particular scoring method vary significantly with the style or character of the three test melodies? 3) Does preference for a particular scoring method vary significantly with the nine dictation solutions (three for each melody), or with the types of errors contained within those solutions? 4) Does each subject (each scorer) exhibit significant variation in scoring preferences?

Example 4 lists the means and standard deviations for the three melodies and their five scoring methods. The means alone indicate that two scoring methods, "pitch" (scoring method "B") and "pitch with transposition credit" (scoring method "C"), are the most consistently popular, compared to the other three methods. Also, as the melodies become more chromatic (and perhaps more "tonally ambiguous"), the means for the "interval" method increase, while the means for the "pitch" method decrease. Results indicate that for one of the scoring methods, the "interval or pitch" method (scoring method "D"), there is a statistically significant degree of variance across the three test melodies. With "interval or pitch," its mean ranking is significantly lower for Melody 1 than for Melody 2, and significantly lower for Melody 1 than for Melody 3. It may be recalled that with the "interval or pitch" method, the evaluator selects whichever method, or combination of methods, produces the

Example 2: SURVEY II - Instructions and Sample of Scoring Method

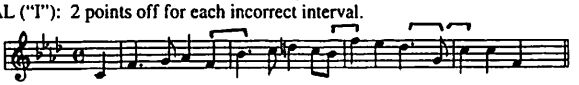
For this survey, you are asked to rank several scoring methods as applied to melodic dictation solutions for a given melody. There are three different dictation melodies, each with three different solutions. For your convenience, the scoring has already been done. Your task is simply to rank the appropriateness of each scoring method as applied to each solution.

For each solution, as many as seven different scoring methods may be ranked. Below is a sample melody, followed by the seven scoring options used in this survey, with clear illustrations of each scoring method. (The sample is Melody 1/Solution 3 from the actual survey.) **Since space does not permit detailed illustrations of each scoring method within the actual survey, it is important to study this page carefully before proceeding on to the survey!** Also, it should be mentioned here that RHYTHM IS NOT INCLUDED AS A VARIABLE in this study, for purposes of simplicity. Consequently, points are not allotted for rhythm!

SAMPLE TEST MELODY: 

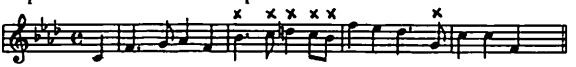
SCORING METHODS (out of a total of 32 possible points):

a) INTERVAL ("I"): 2 points off for each incorrect interval.



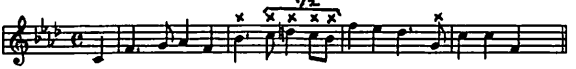
-8 = 24

b) PITCH ("P"): 2 points off for each incorrect pitch.



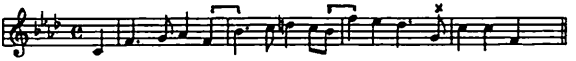
-12 = 20

c) PITCH with TRANSPOSITION CREDIT ("PwT"): 2 points off for each incorrect pitch. Then add back one point per pitch if part of a transposed unit of at least 2 consecutive pitches.



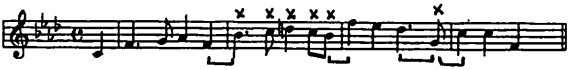
-8 = 24

d) INTERVAL or PITCH ("IorP"): 2 points off for incorrect intervals or pitches. Error types are selected according to whichever method, or combination of methods, arrives at the highest overall score. The score can be based on an all-interval approach, an all-pitch approach, or combination of both.



-6 = 26

e) INTERVAL and PITCH ("I&P"): 1 point off for each incorrect interval, and one point off for each incorrect pitch.



-10 = 22

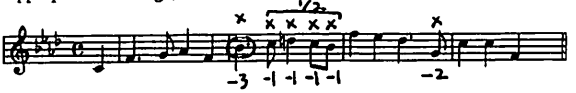
OPTIONAL SCORING METHODS (complete these as needed):

f) One of the above 5 methods (circle one: I P (PwT) IorP I&P), with additional points off for (check all that apply):

___ incorrect contour ___✓___ missed structural scale degrees, such as tonic and dominant

___ losing tonic ___ other (specify):_____

Add appropriate markings, and indicate numerical score:

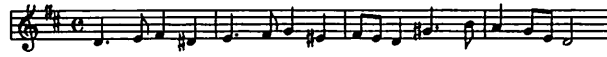


SCORE = -9=23

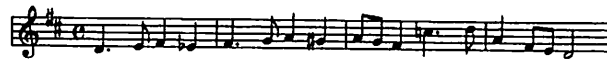
g) OTHER METHOD – Describe the method, and mark the notated example accordingly:
(no other method is presented here for this sample)

Example 3: Sample scoring page from Survey II, to be completed by respondents.

CORRECT MELODY 2



SOLUTION 2.1



SCALE FOR RANKING the Scoring Methods Listed:
1= Strongly Disagree 2=Disagree 3= Neutral 4=Agree 5=Strongly Agree

SCORING METHODS FOR SOLUTION 2.1:

a) INTERVAL:
6 incorrect intervals / -12 = 20 points 1 2 3 4 5

b) PITCH:
11 incorrect pitches (enharmonic spelling ok) / -22 = 10 points 1 2 3 4 5

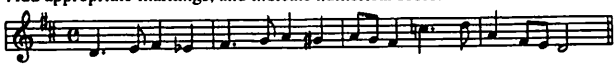
c) PITCH WITH TRANSPOSITION CREDIT:
10 incorrect pitches with one transposed segment / -16 = 16 1 2 3 4 5

d) INTERVAL OR PITCH (choose method with least point reduction):
3 interval errors and 1 pitch (F# in m.4) / -8 = 24 1 2 3 4 5

e) INTERVAL AND PITCH:
Combine all interval and pitch errors, at 1 point each / -17 = 15 1 2 3 4 5

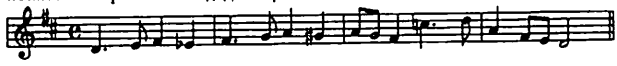
OPTIONAL SCORING METHODS FOR SOLUTION 2.1 (complete these as needed):

f) ONE OF THE ABOVE 5 METHODS (circle one: I P PwT IorP I&P),
WITH ADDITIONAL POINTS OFF FOR (check all that apply): 1 2 3 4 5
___enharmonic spelling errors ___missed tonic or dominant
___missed step-progression downbeats of m.2 and m.3 ___other (specify):_____
Add appropriate markings, and indicate numerical score:



SCORE =_____

g) OTHER METHOD NOT LISTED - Describe the method, and mark the notated example accordingly, and provide a numerical score: 1 2 3 4 5



SCORE =_____

lowest point reduction for that solution. In other words, it is a flexible scoring method that is not based upon the style of the melody, but on the types of errors found in the solution. The significantly lower mean score for this method with Melody 1 suggests that instructors would not likely be so flexible with scoring method for such a strongly diatonic melody, and would instead select a method that would weigh pitch errors more heavily than interval errors, because of the tonal strength of the melody itself.

In addressing the consistency of scorers in their ratings across the three melodies, a correlation analysis reveals that scorers are

consistent with two scoring methods, "pitch with transposition credit" (method "C") and "interval and pitch" (method "E"), across all melodies and solutions. For the other three methods, there is a strong correlation between scoring of Melodies 1 and 2, but less so between Melodies 1 and 3, or between Melodies 2 and 3. These results suggest that instructors favor methods incorporating some degree of compromise between scoring by pitch and scoring by interval, with a tendency to weigh scoring by pitch a bit more heavily. This remains consistent for diatonic melodies, even those with some degree of chromaticism included (Melodies 1 and 2 from this test). Instructors who do change their scoring methodology are most likely to do so when evaluating a melody that could be considered non-tonal (Melody 3 from this test).

The results discussed from Example 4 and from the correlation analysis also suggest that scoring by "pitch" and "pitch with trans-

Example 4: Means and standard deviations fo all five scoring methods for te three test melodies from Survey I. Scoring Methods: A = Interval, B = Pitch, C = Pitch with Transposition Credit, D = Interval or Pitch, E = Interval and Pitch. N = 41.

Melody	Scoring Method	Mean	Std. Dev.
1	A	2.40	1.18
	B	3.21	1.38
	C	3.50	1.14
	D	2.11	0.91
	E	2.75	1.17
2	A	2.50	1.15
	B	3.14	1.34
	C	3.44	0.83
	D	2.62	0.84
	E	2.56	0.93
3	A	2.85	1.11
	B	2.98	1.16
	C	3.33	0.98
	D	2.60	1.00
	E	2.74	1.21

position credit" are the preferred methods overall. The popularity of these two methods reflects a desire by many of the instructors surveyed to choose scoring methods that reinforce "tonal" listening strategies (hearing and processing "in a key") when the melody itself is considered tonal. However, as was stated earlier, there was noticeable shift to scoring by "interval" for the tonally ambiguous (non-tonal) melody. (It should be mentioned that test results reveal wide minimum and maximum ranking preferences across the sample, which indicates that the test itself was sensitive in picking up differences.)

T-tests were calculated to check for differences in ranking preferences based upon: years' experience in teaching theory, years' experience in teaching aural skills, and aural skills teaching level. Of those three categories, "aural skills teaching level" produced results that might warrant further study. The 41 subjects were divided into the more advanced teaching level (N=15) and the less advanced teaching level (N=26). The more advanced level included post-second-year college undergraduate level, plus graduate-level advanced aural skills. The less advanced teaching level included pre-college level, the basic two-year theory sequence, and graduate-level aural skills review. The two groups exhibited considerable differences regarding the "interval or pitch" scoring method for all three melodies. For that method, the more advanced instructors produced quite lower mean rankings than the less advanced. The same holds true with the "interval" method, but only for Melodies 1 and 2. This lack of enthusiasm for the "interval or pitch" method, plus the low degree of interest in the "interval" method for Melodies 1 and 2, suggests that more experienced aural skills instructors believe more strongly in a "pitch" (or "tonal") approach to listening and evaluating tonal melodies than do less experienced instructors. Further study to compare these two groups could prove fruitful.

Example 5 presents mean rankings of the five scoring methods for all nine dictation solutions (three solutions for each of the three test melodies). Two scoring methods, "pitch" (method "B") and "pitch with transposition credit" (method "C"), received consistently high means for nearly all of the nine solutions. The other three methods (interval, interval or pitch, interval and pitch) received consistently lower means in most cases. This further reinforces the

Example 5: Means of the rankings of scoring methods for all nine dictation solutions from Survey II. Variable 1.1A refers to "Scoring Methods A for Melody 1/Solution 1,"variable 1.2B refers to "Scoring Method B for Melody 1/Solution 2," etc.; Scoring Method: A = Interval, B = Pitch, C = Pitch with transposition, D = Interval or Pitch, E = Interval and Pitch; "Rank" indicates the five rankings from the Survey: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree; N = the entire subject pool of 41 participants. Standard deviations are not shown here, but all are between 0.91 and 1.47.

	MELODY 1		MELODY 2		MELODY 3	
	Variable	Mean	Variable	Mean	Variable	Mean
SOLUTION 1	1.1A	2.34	2.1A	2.49	3.1A	2.56
	1.1B	3.15	2.1B	2.93	3.1B	3.46
	1.1C	3.63	2.1C	3.46	3.1C	3.32
	1.1D	2.12	2.1D	2.05	3.1D	2.63
	1.1E	2.80	2.1E	2.80	3.1E	2.73
SOLUTION 2	1.2A	2.39	2.2A	2.49	3.2A	2.95
	1.2B	3.29	2.2B	3.15	3.2B	2.44
	1.2C	3.41	2.2C	3.51	3.2C	3.22
	1.2D	2.02	2.2D	3.29	3.2D	2.46
	1.2E	2.68	2.2E	2.24	3.2E	2.63
SOLUTION 3	1.3A	2.46	2.3A	2.51	3.3A	3.05
	1.3B	3.20	2.3B	3.34	3.3B	3.05
	1.3C	3.44	2.3C	3.34	3.3C	3.44
	1.3D	2.20	2.3D	2.51	3.3D	2.68
	1.3E	2.76	2.3E	2.63	3.3E	2.85

popularity of both the "pitch" and "pitch with transposition credit" methods.

Regarding consistency of rankings of the five scoring methods for individual scorers across all nine dictation solutions, data indicate that approximately one-half of the 41 participants remained very consistent with rankings. Those particular participants received low standard deviation measurements for most, if not all, of their rankings for the five scoring methods. In fact, three of the "consistent" participants did not vary their rankings at all for four of the five scoring methods (standard deviation = zero), regardless of the style of the test melody or the types of errors found in the solutions. Only two participants earned what could be considered high standard de-

viation measurements (indicating a great degree of variance), and in both cases, the high value was found in relation to only one scoring method. When each of the five scoring methods is viewed according to the degree of variance among rankings by the 41 subjects, it is clear that the greatest consistency is found with three scoring methods: "interval," "pitch," and "interval and pitch." Considerably less consistent are the rankings for the "pitch with transposition credit" and "interval or pitch" scoring methods. The fact that rankings for "pitch with transposition credit," one of the most popular methods, are less consistent, may be related to the types of errors included in the solutions. Some solutions contained errors that were transposed segments of the given melody, while other solutions did not contain transposed segments. Thus, the ranking for that method tended to vary according to the types of errors present in the solutions. The same might be said for the "interval or pitch" method, since that method is also quite connected to the types of errors found in each solution.

In addition to the statistical results, some observations should be made regarding other aspects of the survey data. As mentioned earlier, scorers not only ranked the five main scoring methods, but they had the option of preferring one of the five methods while subtracting additional points for errors in contour, structural pitches, enharmonic spelling, and other error types. These extra options allowed scorers to be more detailed with scoring, in order to account for specific types of errors found in the solutions (see "Optional Scoring Methods" in Example 2). Nearly two-thirds of the respondents (25 out of 41) chose to utilize these optional scoring methods for one or more of the solutions. Several of those scorers went into great detail with their scoring, adding hand-written commentary in the margin, sometimes modifying their decisions, other times expressing uncertainty about how to approach a particular error. At the other extreme, some scorers chose simplicity of method over complexity, perhaps for reasons of practicality or ease of scoring. The extreme in simplicity is best represented by the scorer who chose one simple scoring method for the entire survey, without varying the method for any of the test melodies or solutions, and without making any additional comments on the answer sheets. In sum, there was an interesting mix of preferences, within a spectrum of methods ranging

from very simple to quite complex. These preferences may reflect one's own personal style of communication, or they may be necessitated by one's own teaching situation. A teacher who has many students and limited time for grading, or who supervises several teaching assistants who do their own grading, may opt for a less complex methodology that could be accomplished more quickly and with greater uniformity.

CONCLUSION

What do these results show, and how can they be applied to the aural skills classroom? The data reveal both general tendencies as well as striking individual preferences in scoring approach. While some scorers worked with painstaking detail, arriving at quite complex scoring systems, a few of those same scorers also acknowledged the impracticality of such complex methods when faced with grading multiple papers or when having to supervise several instructors. In the actual classroom, then, some instructors may end up sacrificing detail out of necessity.

Of the 41 participants in Survey II, only seven reported that their institution follows any standardized guidelines for scoring melodic dictation exercises. Of those seven, one uses standardized scoring only in the freshman year. Possibly the lack of standardized guidelines for so many of the represented schools stems from the difficulty in arriving at a foolproof scoring method. With schools that have multiple aural skills sections taught by several instructors, perhaps it would be beneficial to work toward a common scoring methodology that is not too complex and time-consuming, yet incorporates sound pedagogical principles according to the characteristics of the melodies being tested. More in-depth exploration of how institutions deal with scoring methodologies, how and why such decisions are made, and how successful their system works, would make for an interesting study.

In my first few years of teaching aural skills, I chose to utilize a scoring methodology that gave students the highest possible score on any given dictation solution. My method, then, was the same as the "interval or pitch" method used in Survey II of the present study. I

was quite sympathetic towards students and wanted to give them as much help as possible in conquering a very difficult task. Since the "interval or pitch" method is governed by "high score potential" and has nothing to do with preferred listening strategies, I soon became dissatisfied with it as a pedagogical tool. To illustrate, suppose a student notates a solution for a clearly tonal melody and makes a single interval error very early in the melody. All subsequent intervals are maintained, which means that the rest of the melody is "off-track" because of that single error. Even though all the remaining pitches are incorrect and the tonal center is completely lost, my sympathetic scoring method would give that student a quite positive score of only a single error. Such a score would communicate falsely to the student that his or her listening skills are in good shape, when in fact, the student needs to seriously evaluate the significance of the error and re-examine what the best strategies are for listening to tonal melodies. At the very least, I would need to add commentary with the score, so that the student would clearly know that the error was a serious one.

In recent years, I have progressed to scoring by pitch for tonal melodies, with half credit back for transposed segments of three or more consecutive pitches. In comparison with my previous method, this method reinforces hearing "in a key," but does compromise a bit by giving a slight reward for transposed segments. For atonal melodies or melodies with no clear tonal center, such as Melody 3 from Survey II, I employ the "interval or pitch" method, which basically gives credit wherever possible. I am hesitant to score atonal melodies solely by interval, because I presume that students may utilize a combination of listening strategies for such melodies, perhaps juggling intervallic listening with tonal listening in various combinations. The "interval or pitch" method seems to be a neutral method to correspond to students' expected listening strategies for such melodies.

At Butler University, I supervise the aural skills program, which includes three or four theory faculty members (including myself) and two teaching assistants who are graduate students in theory/composition. We teach approximately six freshman sections and six sophomore sections of aural skills, and each section contains around twelve students. I construct all lesson plans and exams, and

we all utilize the same scoring methodologies for melodic dictation. The standardized scoring method works well for our institution, particularly as an aid for the teaching assistants, who typically have no prior classroom teaching experience. Students are made aware of the scoring methodologies we use, and on occasion they mark their own work in class, so that we can discuss connections between scoring and hearing. One specific activity I now include at the sophomore level is something similar to Survey II. Students are actually provided with melodies of varying styles and asked to assign a scoring methodology that would be reflective of how they would hear and process that melody. We then collect all the responses and discuss the results.

I believe very strongly that the scoring methods should correlate as closely as possible with the listening strategies advocated in the classroom. I will be honest in that regard and say that I am still not satisfied with my scoring methods or how I communicate results to students. I often find myself in a hurry just to get exams graded and returned, with little time to write many comments on students' papers. My current goal is to further refine how I score solutions, based in part on the results from the present surveys. For example, if a tonal melody has a sequential passage with a clear step progression in effect on downbeats, then I would like for the downbeat pitches to carry more weight in the scoring process. If I refine my scoring methods, however, I must also refine my teaching methods, making sure to emphasize the importance (and the advantage) of hearing such structurally significant pitches within the context of a melody.

Despite the need for scoring methods that are time-efficient and relatively simple, we must not overlook the need for methods to reflect how we wish the students to hear and process melodies. If we encourage different listening and processing strategies for a tonal melody than for an atonal melody, then shouldn't our scoring of such melodies reflect those differences as well? If we stress the importance of structural scale degrees and step-progressions within a tonal melody, then shouldn't our scoring method reflect such important components? If we encourage students to utilize all their listening skills to understand a melody, then perhaps we should evaluate using a combination of approaches (interval and pitch, for

example). The difficulty lies in the fact that despite our teaching approaches, we have no sure method of determining exactly how a student will actually listen and process a melody when completing a dictation exercise.

Perhaps there is a middle ground for which we can strive, one which maintains some degree of simplicity but which also allows for some detailed error assessment according to the structure of the melody and the types of errors found. To affirm the views of Karpinsky (2000), quoted at the beginning of this article, it is also crucial that we take time to provide feedback that can prove beneficial to the listener's development of aural skills. If indeed the task of melodic dictation is an important and multi-faceted learning tool, then we certainly should strive to make the best use of it in the classroom.

REFERENCES

Blombach, Ann. *MacGamut*. Columbus, OH: MacGamut Music Software International, 2000.

Educational Testing Service. *Practicing to Take the GRE Music Test*. 2nd ed. Princeton, NJ: Educational Testing Service, 1997.

Educational Testing Service. *Released Exam: 1998 AP Music Theory*. Princeton, NJ: Educational Testing Service, 1998.

Karpinksi, Gary. *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-Level Musicians*. New York: Oxford University Press, 2000.

Karpinski, Gary. "A Model for Music Perception and Its Implications in Melodic Dictation." *Journal of Music Theory Pedagogy* 4 (1990): 191-229.

Phillips, Joel. "Using Models Derived from Those Used in Nationally Administered Examinations." *Journal of Music Theory Pedagogy* 12 (1998): 65-80.

Rogers, Michael. *Teaching Approaches in Music Theory: An Overview of Pedagogical Philosophies*. Carbondale: Southern Illinois University Press, 1984.