Methodological Perspectives on Singing Accuracy: An Introduction to the Special Issue on Singing Accuracy (Part 2)

Steven M. Demorest
Northwestern University

Peter Q. Pfordresher
University at Buffalo, State University of New York

Simone Dalla Bella
University of Montpellier 1, Montpellier, France

Sean Hutchins
The Royal Conservatory of Music, Toronto, Canada

Psyche Loui
Weslyan Unversity

Joanne Rutkowski
The Pennsylvania State University

Graham F. Welch
University of London, London, United Kingdom

The problem of poor pitch singing has been a topic of both practical and empirical concern for music educators for many years. Earlier efforts focused either on interventions that might help students develop the skills (Joyner, 1969; Yank Porter, 1977) or age-related changes in singing accuracy and proposed models for how such skills might develop (Welch, 1985; 1986). More recently, music educators have explored the influences of training, maturation, and task difficulty on children's singing accuracy (Demorest & Clements, 2007; Nichols, 2013; Welch et al., 2009) and use of singing voice (Rutkowski & Miller, 2003). Researchers in psychology and cognitive neuroscience have become interested in poor pitch singing in adults as a kind of cognitive deficit, and have begun to explore the various conditions under which people have difficulty singing accurately (Dalla Bella & Berkowska, 2009; Hutchins & Peretz, 2012; Loui, Guenther, Mathys, & Schlaug, 2008; Pfordresher & Brown, 2007). One crucial piece of information lacking in these efforts is a shared definition of what constitutes accurate singing, as well as a shared approach to measuring this skill (see Dalla Bella and Demorest & Pfordresher this volume). While individual studies have proposed various assessments and scoring systems, the lack of a core set of tasks has made it extremely difficult to compare findings across studies, or to develop a sense of how prevalent poor pitch singing is in the general population across the lifespan.

While several groups have developed or are developing batteries of tasks related to singing skills (cf. Berkowska & DallaBella, 2013; Cohen et al., 2012), one goal of the Seattle International Singing Research Symposium was to attempt to design a minimum set of tasks focused on measuring pitch accuracy in singing and associated skills that could be easily administered and scored. The ideal battery would be comprehensive enough to yield a meaningful set of data regarding singing accuracy performance, but short enough that various research groups could include it as a part of any larger battery of tasks that might be of particular interest to their research questions. Several papers in this issue involve measuring singing accuracy and related skills on multiple tasks (Dalla Bella, Demorest & Pfordresher, Hutchins) or the same task scored multiple ways (Dalla Bella, Demorest & Pfordresher, Rutkowski), but none use the exact set of procedures proposed here. The proposed protocol was designed to include information needed by the various investigators that are exploring poor pitch singing. What follows is a set of tasks that create a basic "singing accuracy protocol" for any group interested in contributing to this larger research effort.

The following procedures were proposed during the Seattle International Singing Research Symposium (October 2013). We refer to the battery on the whole as the Seattle Singing Accuracy Protocol, or SSAP. Tasks are designed to provide a baseline for any study of singing that could be used to compare the performance of one study population directly to the performance of populations from other studies across different ages and levels of training. The tasks are designed to be short, easy to administer, and able to be scored acoustically through (relatively) automatic means or through a simple rating method. In order to standardize the administration of the SSAP, the practice examples, stimuli, and instructions for each task will eventually be prerecorded

Music Perception, volume 32, issue 3, pp. 266–271, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2015.32.3.266

either in online form or minimally as a downloadable recording. Comparing the performance of subjects across research studies offers a number of advantages:

1. We can begin to develop a database of average singing performance across ages and levels of experience that could be used eventually to assess individual performance against a consistent standard.
2. Researchers can compare the performance of their subjects on the tasks unique to their study with the more standard SSAP tasks.
3. Researchers can begin to identify the prevalence of various causes of poor pitch singing, as well as the participant's current level of singing ability and development, by including several procedures: a perceptual psychophysical staircase measure, various stimulus timbres for matching, and singing a familiar song on text and neutral syllable (doo) from long-term memory.

Seattle Singing Accuracy Protocol (SSAP)

PROCEDURES

1. *Comfortable Range*. This is a series of short tasks that allow the researcher to reliably identify the singer's comfortable singing range. By identifying the comfort pitch from the average of these three tasks, the SSAP can be presented at a pitch level that will insure that accuracy problems are not a result of being asked to sing outside of a comfortable range.
    a. Count backwards from ten quickly. Previous studies (Harries, Griffin, Walker, & Hawkins, 1996; Welch et al., 2010) have used this procedure to identify a person's mean comfortable speaking pitch that is predictive of their singing pitch.
    b. Sing a familiar song from the list shown below (with no starting pitch given). The mean pitch of the song will be compared to the pitch identified previously from the counting task.
    c. Sustain a single comfortable pitch on "oo" for several seconds.
2. *Singing Imitation Tasks*. Once a comfortable pitch has been identified, subjects will be asked to perform three singing imitation tasks. Target pitches will span the range of a fifth, with the mean centered around the comfort pitch. By keeping the stimuli within the range of a fifth we further insure that singers will not be hampered by register limitations. Participants will be given a practice example before each block of echo tasks.
    a. Task 1: Single pitch human voice. Imitate a single pitch sung by a prerecorded human model that represents the vocal range and timbral category (male /female) of the participant (younger children will imitate a female model) for 10 trials (5 pitches x 2 trials per pitch). Research has demonstrated that poor pitch singers and inexperienced singers do better when the timbre of the stimulus matches their own voice (Green, 1990; Hutchins & Peretz, 2012).
    b. Task 2: Imitate a single pitch played in a piano timbre for 10 trials – 5 pitches x 2 trials per pitch). This task helps identify those singers who might be accurate when hearing vocal stimuli but have difficulty generalizing to a non-vocal timbre (cf. Hutchins & Peretz, 2012).
    c. Task 3: Pattern task. Imitate six four-pitch patterns in the range of a 5th based on the participant's comfort pitch (3 ascending and 3 descending) presented in the same vocal timbre as in Task 1 (see Appendix A for patterns). Previous research has found performance differences between single pitch matching, and matching longer patterns, which introduce memory demands, but also provide more of a tonal context (e.g., Pfordresher & Brown, 2007; Wise & Sloboda, 2008).
3. *Two Song-Singing Tasks*. These tasks will explore participants' ability to sing accurately from their long-term memory under two conditions – text and neutral syllable. Participants will not be given a starting pitch or starting tempo.
    a. Task 1: Sing a familiar song from the list on text a cappella
    b. Task 2: Sing the same song on a neutral syllable e.g. "doo"

    The list of songs from which the familiar song is chosen should be graded for difficulty and may need to be altered depending on the culture in which testing is taking place. Below is a short preliminary list of well-known songs. These are in order of difficulty from easy to hard based on the content of the songs and the experience of our symposium participants:[1]

[1] These songs were found by our participants to be known in some form across many Western cultures. A database of such songs, categorized by complexity and familiarity does not, to our knowledge exist, but would be very valuable.

1) Jingle Bells
2) Frere Jacques - (or Brother John or Are you Sleeping?)
3) Twinkle, Twinkle Little Star (or ABC song)
4) Happy Birthday

4. *Adaptive Pitch Discrimination Task (2-4 minutes).* This is a three-up one-down staircase procedure (Cornsweet, 1962) for pitch discrimination around the center frequency of 500 Hz. The task allows the researcher to identify quickly whether or not their participant has a perceptual deficit. Inability to distinguish variations in tones smaller than a semitone has been shown to be predictive of poor performance on the Montreal Battery for the Evaluation of Amusia (MBEA) (Loui, Alsop, & Schlaug, 2009; Loui et al., 2008).[2]
5. *Musical Background Questionnaire.* The goal of the SSAP background questionnaire (see a draft in Appendix B) is to quickly determine basic information about a participant's musical background. Individual studies would likely want more specific information in some areas, but including these items at a minimum allows the outcomes between studies to be compared while factoring in musical experience. One of the challenges here is to adopt terminology that works internationally (e.g., "band, choir, or orchestra" or "music class" might be too U.S. focused)

SCORING

The goal of the SSAP will be to provide a system of scoring that is as automated as possible, sensitive to meaningful variations in singing accuracy, and reflective of human judgments regarding what constitutes accurate singing. While any of these tasks could be scored by expert human raters, it would be preferable to automate the scoring as much as possible. The perceptual procedure outlined above exists online now and provides a result in terms of performance related to the larger population (http://www.musicianbrain.com/pitchtest/) but would be modified to produce a discrimination threshold for each participant. Our plan is that the range finding procedure will also be automated by analyzing the fundamental speaking or singing frequency via the YIN algorithm (de Cheveigné & Kawahara, 2002) in real-time in the Max/MSP platform (Zicarelli, 1998), but of course the tasks could be used to guide a researcher *in situ* regarding the best pitch range for testing. The three imitation singing tasks will be scored acoustically providing both mean deviation scores for all items in a block and absolute note error scores. See Demorest and Pfordresher (this volume) for data comparing this acoustic scoring procedure to human judgment. At present, the easiest way to score singing a song reliably is through the use of a rating procedure (Demorest & Pfordresher, this volume; Welch, et al., 2010; Wise & Sloboda, 2008). These global rating procedures, while reliable, do not provide very sensitive measures regarding the nature or location of the difficulty a singer might have within a melody. However, acoustic scoring of song singing currently requires that each tone of the melody be isolated manually before analysis. Previous investigators have explored scoring song-singing by using an acoustic analysis of the interval relationships between adjacent pitches (Berkowska & Dalla Bella, 2009, 2013). The challenge is to parse the individual pitches of a song accurately so that the interval analysis could be done automatically. The reason for including task two under song singing is that singing on a neutral syllable is easier to parse automatically into separate syllables that could then be analyzed acoustically (note that stop consonants (e.g. /d/ are significantly easier to analyze than liquids /l/; thus we prefer "doo" to something like "la"). Also, research has demonstrated that adults sing more accurately on a neutral syllable (Berkowska & Dalla Bella, 2009; 2013) while research with children has yielded mixed results (Goetze, Cooper, & Brown, 1990).

Conclusion

The goal of the SSAP will be to provide a measure that can be: a) standard in administration and scoring across many studies allowing for a direct comparison of results, b) short and easy to administer, and c) comprehensive enough to provide some diagnostic information regarding the possible cause of an individual's poor pitch singing. We are in the process of creating and piloting an automated version of the protocol that will be freely available to any scholars interested in using it. Common usage of such a measure would dramatically increase our knowledge base regarding the prevalence and possible causes of poor pitch singing across a variety of populations. This information should help us to better identify the most common sources and types of singing difficulty and how they vary by age or training. This in turn should provide better information to guide research into the possible neurological underpinnings of such difficulties and how they might relate to other deficits in audio-motor skill. In the future, the SSAP may also be usable as a standard for measuring the efficacy of possible interventions for improving the fundamental musical skill of singing.

[2] This procedure was used with adults and would need to be piloted with children to insure it could function in the same way.

Author Note

Correspondence concerning this article should be addressed to Steven M. Demorest, Henry and Leigh Bienen School of Music, Northwestern University, 711 Elgin Road, Evanston, IL 60208-1200. E-mail: sdemorest@northwestern.edu

References

Berkowska, M., & Dalla Bella, S. (2009). Reducing linguistic information enhances singing proficiency in occasional singers. *Annals of the New York Academy of Sciences*, *1169*, 108-111. doi:10.1111/j.1749-6632.2009.04774.x

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The Sung Performance Battery (SPB). *Frontiers in Psychology*, *4,* 714. doi: 10.3389/fpsyg.2013.00714

Cohen and the AIRS Team (2012). *Midterm report (2009-2012): Advancing interdisciplinary research in singing*. Downloaded from: http://www.airsplace.ca/sites/discoveryspace.upei.ca.airs2010/files/AIRS%20Midterm%20Report%20-%202012-07-09-Cohen_AJ.pdf

Cornsweet, T. N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology, 75*(3), 485-491.

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority. *Annals of the New York Academy of Sciences*, *1169*, 99-107. doi:10.1111/j.1749-6632.2009.04558.x

Demorest, S. M., & Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *Journal of Research in Music Education*, *55*, 190-203. doi:10.1177/002242940705500302.

de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America, 111*, 1917-1930.

Goetze, M., Cooper, N., & Brown, C. J. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education, 104*, 16-37.

Green, G. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education, 38,* 225-231.

Harries, M. L. L., Griffin, M., Walker, J. & Hawkins, S. (1996). Changes in the male voice during puberty: Speaking and singing voice parameters. *Logopedics Phoniatrics Vocology, 21*(2), 95-100.

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General, 141*, 76-97.

Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education, 17,* 115-124.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, *29*(33), 10215-10220. doi:10.1523/JNEUROSCI.1701-09.2009

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology*, *18*(8), R331-R332.

Nichols, B. E. (2013). *Task-based variability in children's singing accuracy* (Unpublished doctoral dissertation). University of Washington, Seattle, WA.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception*, *25*, 95-115. doi:10.1525/mp.2007.25.2.95

Rutkowski, J., & Miller, M. S. (2003). The effect of teacher feedback and modeling on first graders' use of singing voice and developmental music aptitude. *Bulletin of the Council for Research in Music Education*, *156*, 1-10.

Welch, G. F. (1985). A schema theory of how children learn to sing in-tune. *Psychology of Music, 13,* 3-18.

Welch, G. F. (1986). A developmental view of children's singing. *British Journal of Music Education. 5,* 295-303.

Welch, G. F., Himonides, E., Papageorgi, I., Saunders, J., Rinta, T., Stewart, C., et al. (2009). The National Singing Programme for primary schools in England: An initial baseline study. *Music Education Research, 11*(1). 1-22.

Welch, G. F., Himonides, E., Saunders, J., Papageorgi, I., Preti, C., Rinta, T., et al. (2010). *Researching the impact of the National Singing Programme 'Sing Up' in England: Main findings from the first three years (2007-2010). Children's singing development, self-concept and sense of social inclusion.* London, UK: International Music Education Research Centre, Institute of Education. ISBN13: 978-1-905351-13-8

Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined "tone deafness": Perception, singing performance and self-assessment. *Musicae Scientiae, 12*(1), 3-26. doi:10.1177/102986490801200102

Yank Porter, S. (1977). The effect of multiple discrimination training on pitch-matching behavior of uncertain singers. *Journal of Research in Music Education, 25,* 68-81.

Zicarelli, D. (1998). *An extensible real-time signal processing environment for Max.* Paper presented at the Proceedings of the International Computer Music Conference, University of Michigan, Ann Arbor, MI.

Appendix A:

Sample four-note pitch patterns written in moveable solfège (from Pfordresher & Brown, 2007).
1. D-R-M-S
2. S-F-M-D
3. D-S-F-M
4. S-D-R-M
5. D-M-R-S
6. S-M-F-D

Appendix B

SSAP Musical Background Questionnaire (NOTE: This will likely be done in a tree format where prior answers determine next question)

1. Age in years and months ____________
2. I am __right __left-handed. Sex: M F
3. Have you ever been diagnosed with a hearing impairment?? Yes No
4. Any personal history of neurological or psychiatric disorders? Yes No
   (your responses are completely confidential)

****

	Entirely Disagree	Mostly Disagree	Somewhat Disagree	Neither Agree or Disagree	Somewhat Agree	Mostly Agree	Entirely Agree
	1	2	3	4	5	6	7
5. I enjoy singing.	1	2	3	4	5	6	7
6. People think I am a good singer.	1	2	3	4	5	6	7
7. I am musically talented.	1	2	3	4	5	6	7

****

8. What was the last year in school where you had a music class and how old were you?
   Year______________________ Age______
9. Have you had any private lessons in music?
   - IF NO Go to #10
   - IF YES then "Was it on an __ instrument or __ voice (singing) or __both?
     - If "instrument" then "What instrument did you study longest?" ________________
       - How many years? __________________
       - Any other instruments studied including voice? ______________________
     - If "voice" then "How long did you take lessons?"
     - If "both" then go through instrument tree.
10. Have you ever played or sung in a music group? YES NO
    - IF NO then go to #11
    - If yes – "Please check all that apply" –
      - Band
      - Choir
      - Orchestra

    - Percussion ensemble
    - Jazz Band/Combo
    - Mariachi Band
    - Rock Band
    - Folk group
    - Other
        - If Other ask them to list group(s)
    - How many years have you played in your primary group?

11. What is your ethnicity? ________________
12. What is your first language? ______________
13. Any other fluent languages? YES NO
    - If yes please list in order of fluency
      ____
      ____________________.