Defining Poor-Pitch Singing: A Problem of Measurement and Sensitivity

Simone Dalla Bella
University of Montpellier 1, Montpellier, France

Important individual differences characterize singing in adult nonmusicians. In spite of the fact that the majority can carry a tune, some occasional singers are particularly inaccurate or imprecise in producing or imitating pitch (poor-pitch singers). Poor-pitch singing can be defined via acoustical analyses of vocal performances. In spite of the objective nature of this method, however, to date there is not a standard strategy for identifying and describing poor-pitch singers. Different tasks (e.g., singing from memory vs. imitation), cut-offs (50 cents, 100 cents, vs. variable criteria), and metrics (e.g., accuracy vs. precision) are typically used for assessing singing proficiency. Here the pros and cons of different methods and measurements are discussed. The boundary between poor-pitch singing and good singing depends on these factors, which should be carefully taken into account when assessing singing abilities. An approach based on multiple tasks sharing a common set of measures of singing accuracy and precision, with a relative cut-off (i.e., 2 *SD* from the average of a normative group) is favored for identifying poor-pitch singers.

Received: June 5, 2014, accepted November 1, 2014.

Key words: singing, pitch, disorders, poor-pitch singing, tone deafness

Singing in tune is not the exclusive privilege of professional musicians. Most of us, regardless of the fact that we have not received dedicated vocal training or musical tutoring, can carry a tune (Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2007). The majority, referred to as "occasional singers" throughout this paper, displays accurate singing (although rather imprecise, see Pfordresher, Brown, Meier, Belyk, & Liotti, 2010; Berkowska & Dalla Bella, 2013) on both pitch and time dimensions (Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007). Proficient singing is as natural as speaking, is widespread and universal (Mithen, 2006). Humans are likely to be hard-wired for singing. The first vocal productions emerge spontaneously by imitation of maternal singing (e.g., Trehub & Trainor, 1999), as soon as during the first months of life (Papoušek, 1996). Starting from these first examples of singing skills, vocal performance slowly develops over time thanks to exposure, spontaneous practice, and in some cases early musical tutoring (e.g., see Ostwald, 1973; Welch, 2006, for a review) until the accuracy and precision characteristic of adult performance are achieved (Berkowska & Dalla Bella, 2013; Pfordresher et al., 2010). Finally, singing is typically a participatory activity. People sing in group during religious ceremonies, at parties, in the military, an activity which is thought to foster group bonding (Mithen, 2006; Wallin, Merker, & Brown, 2000).

In spite of increasing evidence that singing is natural, widespread, and probably deeply rooted in human biology, we tend to underestimate our ability to sing in tune. When more than 100 university students were asked to assess whether they can imitate melodies, about 60% of them answered negatively (Pfordresher & Brown, 2007). Systematic studies of singing skills in adults reveal a scenario that is less defeatist, though. Only a minority of occasional singers show difficulties in singing in tune (about 10-15% according to previous estimates based on singing accuracy; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007; Pfordresher & Brown, 2007). They are particularly inaccurate in producing pitch, by singing quite far from the target pitches in experiments requiring production of familiar melodies or imitation of single pitches, intervals, or simple melodies. Note, however, that the number of poor-pitch singers can vary as a function of the measure of singing proficiency that is considered and the cut-off score. These issues will be discussed in detail below. For example, when considering pitch precision (i.e., the consistency of repeated attempts to produce a pitch) instead of accuracy (i.e., the distance between the produced pitch and the target pitch), this estimate rises to about 55% (Pfordresher et al., 2010; see also, Berkowska & Dalla Bella, 2013).

The study of poor-pitch singing is an important source of information for advancing our knowledge

Music Perception, volume 32, issue 3, pp. 272–282, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2015.32.3.272

on the song system. Since poor-pitch singing can selectively result from the malfunctioning of a component of the song system (e.g., perceptual, motor, or sensorimotor processes) (Pfordresher & Brown, 2007; Hutchins & Peretz, 2012), the study of this condition can help uncover the functional and neuronal architecture of the mechanisms underlying proficient singing. In addition, a better knowledge of the song system and of the circuitries malfunctioning in poor-pitch singing will foster the development of theory-driven remediation strategies for reducing inaccurate singing (e.g., during childhood). In the present article, I will first provide a short introduction to a model we proposed that illustrates the mechanisms underlying proficient singing (i.e., the Vocal Sensorimotor Loop - VSL; Berkowska & Dalla Bella, 2009a; Dalla Bella, Berkowska, & Sowiński, 2011). It will be shown that poor-pitch singing can originate from the malfunctioning of different components within the VSL, thus leading to a variety of phenotypes of singing disorders. In the second part of the article, the question of detecting poor-pitch singers and of characterizing their performance will be raised. Four critical questions will be addressed, relative to the tasks needed to uncover poor-pitch singing, the analysis method, the metrics for defining inaccurate/imprecise singing, and the cut-off-criteria.

The Song System

The song system is a complex set of structures and processes that involves both physical mechanisms and a dedicated neuronal network. Physical mechanisms include respiratory, laryngeal (i.e., the vocal folds), and articulatory structures (i.e., the vocal tract). The air supply needed for vocalization is provided by the lungs. The airstream coming from the lungs is modulated by the vocal folds in a process referred to as "phonation." It is the vocal tract that conveys to sound the spectral and temporal features typical of sung voice (e.g., Sundberg, 1999; Titze, 1994). The quality of the vocal output, which allows the listener to distinguish a professional singer from an amateur, depends on the coordination of these mechanisms (Sundberg, 1987, 1999).

The aforementioned physical mechanisms are under the control of functional and neuronal structures that have made the object of a separate line of research. In this article I will focus primarily on an account we recently proposed for describing the main mechanisms underlying proficient singing, in tasks such as imitation (e.g., pitch or melody matching tasks), and singing a familiar melody from memory (VSL; Berkowska & Dalla Bella, 2009a; Dalla Bella et al., 2011; see also Zarate, 2013). Note that the concepts behind the VSL have been further elaborated in a model proposed by Pfordresher and collaborators in this volume. In this new proposal, the role played by hierarchical pitch control on real time pitch production (i.e., planning and monitoring of pitch intervals and higher-level tonal structures) is considered.

The VSL outlines the functional components of the song system underlying unimpaired pitch production, including perceptual, motor, auditory-motor mapping, and memory components (see Figure 1). When singing a familiar melody, pitch and temporal information stored in long-term memory is retrieved, and sent to motor planning/implementation mechanisms whereby articulatory and phonation instructions are executed. Since the VLS is built as a closed-loop mechanism, the song output is fed back to the system (i.e., to perceptual processing), and can influence further processes in the VLS. Feedback analysis and in particular auditory-motor mapping is a crucial aspect for accurate vocal performance (for more discussion see Pfordresher et al.'s contribution in this volume). Indeed, error correction during performance is possible because the percept of the produced melody is compared to the memory representation of the intended melody, a discrepancy is detected, and thereby the motor plan is subsequently modified. A similar loop is thought to subserve imitation of single pitches or of more complex pitch sequences. In this case, however, working memory rather than long-term memory will be engaged. For example, in a single-pitch matching task the target pitch to be reproduced is perceptually analyzed, stored in working memory, and mapped to the corresponding motor plan and motor gestures. As with singing from memory, the analysis of the auditory feedback allows singers to self-monitor their own vocal performance, and to make adjustments to the motor plan to correct errors or improve accuracy or precision when needed. It is noteworthy that two pathways for pitch perception are identified in the VSL. The overt pathway is engaged in those tasks where explicit judgments of pitch differences are produced, such as in pitch discrimination tasks, or same-different tasks with melodic material. The covert pathway is supposed to account for cases in which singers, in spite of being very inaccurate in judging pitch differences, do still exhibit some singing abilities, such as producing the correct pitch direction (Griffiths, 2008; Loui, Guenther, Mathys, & Schlaug, 2008), or relatively spared singing abilities, at least when singing with lyrics (Dalla Bella, Giguère, & Peretz, 2009). This condition is possible only under the assumption that the difference between pitches can be computed covertly by an implicit mechanism.

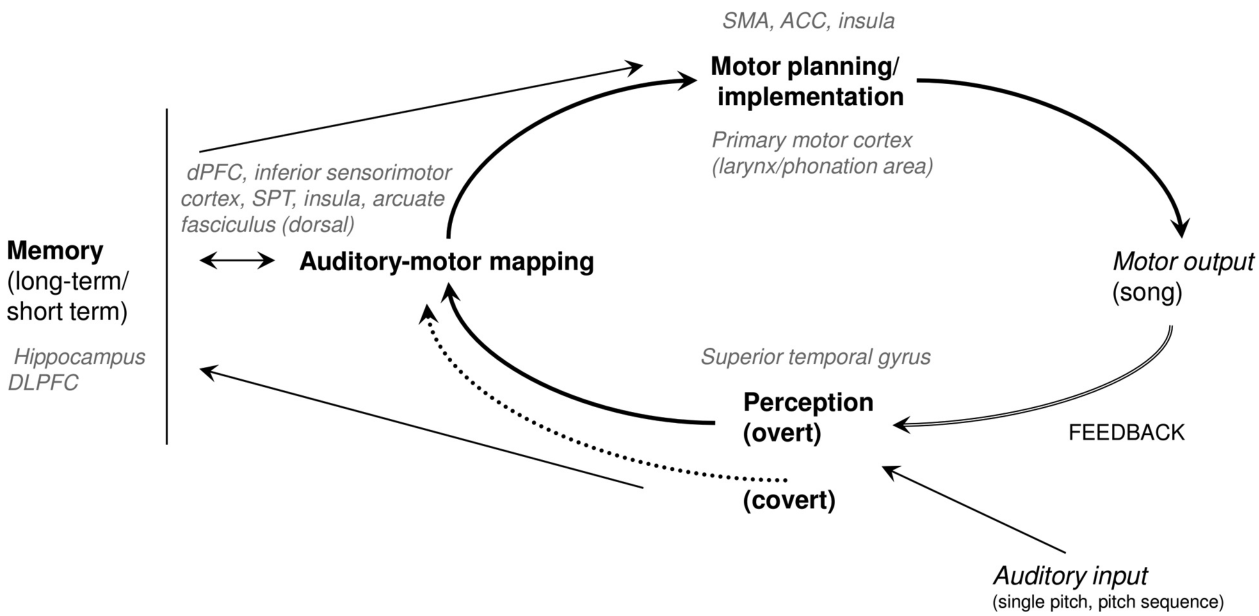

FIGURE 1. Schema of the Vocal Sensorimotor Loop (VSL) (Berkowska & Dalla Bella, 2009a; Dalla Bella, Berkowska, & Sowinski, 2011). SMA = supplementary motor area; ACC = anterior cingulate cortex; dPFC = dorsal prefrontal cortex; SPT = cortex of the dorsal Sylvian fissure at the parietal-temporal junction; DLPFC = dorsolateral prefrontal cortex.

The functional components of the VSL (i.e., sensory, motor, and auditory-motor integration) are sustained by a complex neuronal network (Dalla Bella et al., 2011), as uncovered by neuroimaging studies (see Zarate, 2013; Loui in this volume; see also Brown, Martinez, Hodges, Fox, & Parsons, 2004). Some of the involved areas have been indicated in Figure 1, and described in detail elsewhere (Dalla Bella et al., 2011). In keeping with the functional separation of overt and covert pathways highlighted in the VSL, there is increasing evidence that the two pathways possess dedicated neuronal substrates (Zarate, 2013; Zarate & Zatorre, 2008; Zarate, Wood, & Zatorre, 2010). The covert pathway is associated with the activity of the auditory cortex and of the inferior frontal gyrus, while the overt pathway involves the auditory cortex, the intraparietal sulcus, the anterior insula, the anterior cingulate cortex, and dorsal premotor cortex. In sum, neuroimaging evidence points to a complex neuronal circuitry that affords proficient singing in the majority, and which, when disrupted, can bring about poor-pitch singing.

Poor-Pitch Singing: Four Critical Questions

Understanding the functional and neuronal architecture of the song system paves the way to understanding poor-pitch singing. This condition can manifest due to the malfunctioning of any of the components of the VSL. Brain damage and neurogenetic (i.e., congenital) disorders can disrupt the song system. Here I will focus on poor-pitch singing in otherwise healthy participants who are occasional singers (for reviews on poor-pitch singing resulting from brain damage, see Ackermann, Wildgruber, & Riecker, 2006; Berkowska & Dalla Bella, 2009a; Gordon, Racette, & Schön, 2006; Marin & Perry, 1999; Stewart et al., 2009). Particular attention will be paid to four questions that are critical for defining poor-pitch singing and for uncovering cases of individuals affected by this condition among occasional singers. First, the tasks needed to examine singing proficiency in the general population will be reviewed. Second, the analysis methods (peer judgments vs. acoustical analysis) will be compared in order to provide raw measures of singing proficiency. Third, different metrics for poor-pitch singing computed based on these raw measures will be presented. Finally, various criteria to decide whether an individual sings proficiently or s/he rather exhibits poor-pitch singing will be discussed.

Different sources can bring about singing disorders, such as poor perception, poor motor control, sensorimotor mapping errors, or poor memory retrieval (Dalla Bella et al., 2011; Dalla Bella, Tremblay-Champoux, Berkowska, & Peretz 2012; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). Poor-pitch singing is often treated as the outcome of perceptual deficits (e.g., in congenital amusia or tone-deafness; Ayotte et al., 2002; Peretz, 2001; Peretz et al., 2002; Peretz & Hyde, 2003), namely erroneous extraction of pitch information from the auditory input. Indeed, poor-pitch singing and perceptual deficits are in general associated in individuals with congenital amusia when they are asked to sing a familiar melody from memory (Ayotte et al., 2002; Dalla Bella et al., 2009), or in pitch-matching tasks (Hutchins, Zarate, Zatorre, & Peretz, 2010). That poor-pitch singing may stem from poor perception is consistent with the VSL. Indeed, poor representation of auditory feedback impinges on auditory-motor mapping and eventually hinders accurate/precise singing. Yet, poor-pitch singing can occur with spared pitch perception (Bradshaw & McHenry, 2005; Dalla Bella et al., 2007; Pfordresher & Brown, 2007; Wise & Sloboda, 2008). This finding is consistent with alternative explanations of poor-pitch singing positing that auditory-motor mapping is not carried out properly, or ascribing inaccurate singing to poor motor control (Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). Indeed, an accurate auditory representation of the vocal performance can be inaccurately mapped to motor representations for phonation. Finally, poor-pitch singing can result from (or may be linked to) a faulty memory system (e.g., Dalla Bella et al., 2012). In the presence of spared perception, functional motor control, and auditory-motor mapping, difficulties with encoding, storage of novel melodies, or retrieval of pitch information from long-term memory during reproduction (e.g., in imitation tasks or in singing from memory) may cause poor-pitch singing. For example, individuals suffering from congenital amusia are typically poor-pitch singers (Ayotte et al., 2002; Dalla Bella et al., 2009), but also show poor memory (Ayotte et al., 2002; Gosselin, Jolicoeur, & Peretz, 2009; Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010). In this condition, memory traces of stored melodies are typically less stable and decay more rapidly, thus they are likely to negatively affect sung performance in singing from memory and in pitch- or melody-matching tasks (Tremblay-Champoux, Dalla Bella, Phillips-Silver, Lebrun, & Peretz, 2010).

Because poor-pitch singing is the consequence of different causes, important variability and individual differences among poor-pitch singers are foreseen. Accordingly, there are indications that poor-pitch singing is far from a monolithic disorder. The studies mentioned so far provide compelling evidence that there exist at least two categories of poor-pitch singers, namely those with or without concurrent perceptual deficits. In addition, recent data reveal finer distinctions within these two categories, leading to a variety of singing "phenotypes." Dissociations have been shown when singing proficiency was considered in terms of absolute pitch vs. relative pitch and when the ability to sing in tune was measured in terms of accuracy vs. precision. Note that the term "absolute pitch" is used throughout the paper to refer to the production of absolute pitch height, as compared to "relative pitch," referring to the production of pitch intervals. For example, poor imitation of target pitches is not necessarily accompanied by poor repetition of pitch intervals. Moreover, imprecise (or inconsistent) singing does not always lead to inaccurate singing. In sum, a growing body of evidence points toward a diversity in performance patterns in poor-pitch singing. This raises questions about the very definition of poor-pitch singing, the procedure to be followed when analyzing vocal renditions, its measurement, and the sensitivity of the various criteria available.

WHICH TASKS?

The first decision to be made before assessing whether an occasional singer is a poor-pitch singer is to choose the appropriate task(s). Singing proficiency is tested with a variety of tasks. A common task that can be easily performed by nonmusicians is to sing a well-known song (e.g., "Happy birthday" or "Brother John") with lyrics from memory (e.g., Berkowska & Dalla Bella, 2013; Dalla Bella et al., 2007, 2009). The song can be produced with lyrics or the melody can be sung on a syllable (e.g., /la/). The latter condition typically favors enhanced performance in occasional singers, probably because of the reduced memory load (Berkowska & Dalla Bella, 2009b, 2013). Nevertheless, singing on a syllable is often disrupted in congenital amusia, whereas singing with lyrics can be unimpaired (Dalla Bella et al., 2009). Imitation tasks can also be used. A model stimulus (e.g., a single pitch, an interval, or a short novel melody) is presented, which is imitated by participants using a vowel or a syllable (pitch-matching). Single-pitch matching is a very common task (Goetze, Cooper, & Brown, 1990; Hutchins et al., 2010; Pfordresher & Brown, 2007), and is treated as a core task for assessing musical talent (Watts, Murphy, & Barnes-Burroughs, 2003). Imitation of more complex patterns (e.g., melodies sung with lyrics or on a syllable), as opposed to simpler patterns or individual pitches, is also a quite common task (e.g., Pfordresher & Brown, 2007; Berkowska & Dalla Bella, 2013; Rutkowski, 1990; Rutkowski & Miller, 2003a). Note that in all aforementioned tasks natural feedback (i.e., auditory and proprioceptive) is provided. A variant of the task, however, consists of presenting additional or altered auditory feedback. In this condition, participants are asked to imitate a pitch sequence while the correct pitch or melody are delivered as one sings (Pfordresher & Brown, 2007; Tremblay-Champoux et al., 2010), or when the produced pitch is altered (e.g., Hafke, 2008; Zarate & Zatorre, 2008).

To summarize, there is not a single way to test singing proficiency. In a perfect world, however, we may aspire to use a single task that provides a reliable and consistent estimate of singing proficiency. Even though this would be desirable for a variety of reasons (e.g., time effectiveness), an assessment of singing proficiency based on a single task is problematic. Different tasks are better suited to reflect the activity of partly independent components of the VSL. For example, melody imitation in the presence of augmented feedback, or singing on a syllable of a well-known song, puts less demands on working memory or long-term memory as compared to imitation of novel pitch sequences. Thus, in order to obtain a thorough profile of singing abilities I advise researchers and music educators to use a rich set of tasks. This approach has been adopted in a recent study in which we proposed a battery referred to as "the Sung Performance Battery," which includes tasks ranging from single pitch-matching to imitation and production from memory of complex pitch sequences (Berkowska & Dalla Bella, 2013). A smaller set of core tasks (the Seattle Singing Accuracy Protocol – SSAP) is also proposed in the present volume by Demorest and collaborators. Similar tools test a wide spectrum of processes and mechanisms underlying proficient singing, and offer the possibility to uncover specific disorders characterizing a particular phenotype of poor-pitch singing. In addition, if these tools are shared across different contexts (e.g., research laboratories and music schools) they can allow researchers in music cognition and music educators to compare their results across studies.

ANALYSIS: WHICH METHOD TO EXTRACT RAW DATA FROM SUNG RECORDINGS?

Once the performance in a given vocal task has been recorded, there are different methods available for analyzing sung renditions. The easiest procedure is to ask expert musicians, singers, or music educators to provide subjective ratings of the recordings (e.g., Hébert, Racette, Gagnon, & Peretz, 2003; Racette, Bard, & Peretz, 2006; Rutkowski, 1990; Schön, Lorber, Spacal, & Semenza, 2004; Wise & Sloboda, 2008). This method is well suited to obtain a global and fast assessment of singing proficiency. Subjective ratings can be quite reliable, as shown by high inter-rater reliability (e.g., Guerrini, 2006; Rutkowski & Miller, 2003b); yet, at times they are inevitably associated with inconsistencies among raters (Kinsella, Prior, & Murray, 1988; Prior, Kinsella, & Giese, 1990). Moreover, the method lacks reliability when specific music dimensions (e.g., pitch vs. rhythm) or peculiar musical features (e.g., the exact deviation from a target pitch or interval, the discrepancy from a given tonality, etc.) have to be assessed.

For more precise and reliable estimates of singing proficiency, methods relying on objective measurements are typically preferred (for a comparison between subjective and objective methods, see Larrouy-Maestri, Lévêque, Schön, Giovanni, & Morsomme, 2013). Acoustic methods thus appear a viable alternative (Dalla Bella et al., 2007, 2009; Murayama, Kashiwagi, Kashiwagi, & Mimura, 2004; Pfordresher et al., 2010; Terao et al., 2006). These methods, requiring substantial off-line processing of the auditory signal, consist of computing note pitch onsets and pitch height after acoustic segmentation of the sung renditions. Note pitch onsets and pitch height relative to a model performance (e.g., an interval or a melody) are sufficient to obtain a variety of measures of accuracy and precision in vocal performance (e.g., Berkowska & Dalla Bella, 2013; Pfordresher et al., 2010). Objective acoustically-based measures of singing accuracy and precision have a few advantages. They yield a quantitative and thorough profile of singing proficiency for occasional singers, thus providing substantial information to qualify their singing abilities and to uncover cases of poor-pitch singing. In addition, the method makes the criteria for teasing apart good from poor-pitch singers more explicit (see below for further discussion on this point). Finally, acoustical analyses provide a common set of methods for the analysis of the performance across tasks and across studies. This solution offers the possibility to compare the results obtained in different laboratories. The main drawback of the method is that at the present stage it cannot be performed in a totally automated fashion, especially when dealing with complex pitch sequences. Verification of major steps of the analysis (e.g., phoneme segmentation of the performance, pitch extraction, artifact removal) has to be performed off-line and is very time-consuming. Yet, it is not unlikely that further technological development will improve existing algorithms, limiting human contribution to the analysis, and leading to an optimal solution in the near future.

ANALYSIS: WHICH METRICS FOR QUALIFYING POOR-PITCH SINGING?

One may believe that after choosing the appropriate and most sensitive method for analyzing sung recordings — namely acoustical methods — the task of telling whether someone is a poor-pitch singer or a proficient singer is already completed. Unfortunately, this is far from the truth. First, once raw data is obtained from acoustical analysis (i.e., F0 trajectories and timing), the segment of F0 corresponding to each note has to be chosen. For example, the analysis can focus on the entire segment associated with a given note, including the initial vocal scoop toward the beginning pitch of a note (e.g., Pfordresher & Brown 2007), on the voiced part in the production of a note (e.g., steady-state pitch of a note, Hutchins et al., 2010; Pfordresher et al., 2010), or on a segment defined based on verbal segmentation of the performance (i.e., vowel group in a syllable; Dalla Bella et al., 2007, 2009). Second, after the F0 segment for a given note is chosen, its pitch height is obtained either by computing the mean F0 or median F0. The last solution is preferable, because the median is less affected by extreme values of F0, resulting from artifacts in pitch extraction or extreme variation of F0.

Based on the obtained measure of pitch height, a wealth of different metrics can be calculated. Some metrics focus on the deviation of each produced note from a given target pitch (referred to as *absolute pitch metrics*, for simplicity), expressed in cents or semitones, such as the mean signed note error or the mean absolute note error (Berkowska & Dalla Bella, 2013; Hutchins et al., 2010; Pfordresher & Brown, 2007; Pfordresher et al., 2010; an alternative measure is initial pitch deviation, which is the absolute error of the first note only, see Berkowska & Dalla Bella, 2009b; Dalla Bella & Berkowska, 2009). In the case of signed note error, the direction of the error (i.e., whether the produced note is higher or lower than the model) is included when computing the metric. In the calculation of absolute note error, the absolute value is used in the analysis of the errors to avoid sharp and flat errors canceling one another out. Other measures can be used to reflect internal note stability (e.g., Dalla Bella et al., 2007; Hutchins, Larrouy-Maestri, & Peretz, 2014); for example, by computing the *SD* of the extracted fundamental frequency within the steady-state portion of a note, or the stability or the reproduction of the same segment in a melody (pitch stability, Dalla Bella et al., 2007, 2009). *Relative pitch metrics* are similarly computed considering the difference between two subsequent pitches, or interval, expressed in cents or semitones, relative to the interval in the notation, such as the absolute pitch interval error (e.g., Pfordresher & Brown, 2009; Pfordresher et al., 2010; see also pitch interval deviation; Berkowska & Dalla Bella, 2009b; Dalla Bella & Berkowska, 2009; Dalla Bella et al., 2007, 2009). Finally, the number of interval errors, defined based on a given criterion (e.g., interval departing from the target interval by more than 1 semitone) and the number of contour deviations are also computed (e.g., Dalla Bella et al., 2007, 2009).

A particularly useful classification of metrics based on the distinction between *accuracy* and *precision* has been proposed by Pfordresher and collaborators (2010) — and recently extended by Berkowska and Dalla Bella (2013) — to a variety of vocal tasks in the Sung Performance Battery. This distinction and the associated tasks are illustrated in Figure 2. Separate measures of accuracy and consistency can be obtained for absolute and relative pitch (for equations, see Berkowska & Dalla Bella, 2013; Pfordresher et al., 2010). Accuracy in terms of absolute pitch refers to the average difference between sung and target pitches. This difference is typically measured irrespective of the direction of pitch change (i.e., ascending or descending) (Pfordresher et al., 2010; but see Pfordresher & Brown, 2007). Precision of absolute pitch is defined as the singer's consistency in repeating the same target pitch in pitch-matching tasks or across more complex pitch sequences (Berkowska & Dalla Bella, 2013; Pfordresher et al, 2010). Accuracy and precision can be computed in terms of relative pitch. Here, accuracy indicates the average difference between sung pitch intervals and model intervals; for example, based on the melody score. Precision of relative pitch (i.e., consistency in repeating the same target interval) has been examined, for example, in two recent studies (Berkowska & Dalla Bella, 2013; Pfordresher et al., 2010).

Because of the diversity of metrics that are indicators of singing proficiency, there is little agreement on the measure(s) that allow one to characterize an individual as a poor-pitch singer. Accuracy in producing or imitating pitches — in terms of absolute and relative pitch — is often considered as critical for identifying poor-pitch singers (Dalla Bella et al., 2007; Pfordresher & Brown, 2007; Pfordresher et al., 2010). Nevertheless, treating low pitch accuracy as the sole indicator of poor-pitch singing may underestimate the occurrence of this condition among occasional singers, and thus may be misleading. The prevalence of poor-pitch singing is higher when considering precision instead of accuracy (Pfordresher et al., 2010). Thus, different metrics (i.e., accuracy vs. precision, in terms of absolute or relative pitch) can result in discordant estimates of the prevalence of poor-pitch singing. This fact obviously hinders the comparison of results from different studies that do not share the same metrics of singing proficiency. Here I propose adopting a set of metrics based on accuracy and precision, estimated in terms of both absolute and relative pitch, which is an approach we adopted in a recent study (Berkowska & Dalla Bella, 2013). These metrics have the advantage that they can

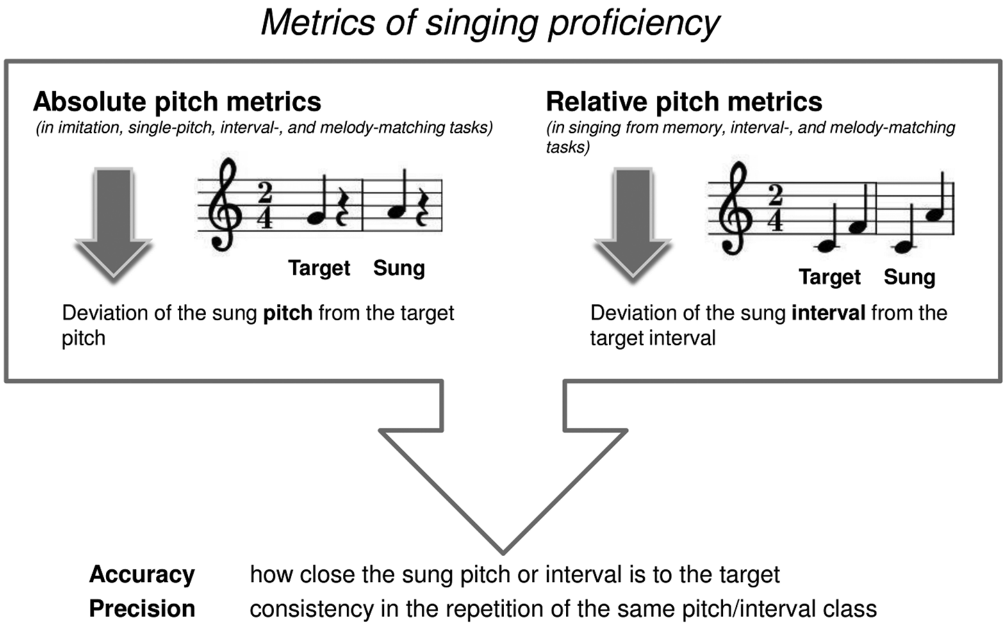

FIGURE 2. Metrics of singing proficiency based on absolute pitch and on relative pitch differences.

be applied to different tasks ranging from single-pitch matching to imitation of complex melodies, thus allowing between-task comparisons. Moreover, they provide a profile of singing abilities that is thorough and particularly well-suited for identifying different profiles of poor-pitch singers. This approach revealed important individual differences among poor-pitch singers, showing profiles of impairment involving specifically absolute or relative pitch, and accuracy or consistency (Berkowska & Dalla Bella, 2013; Dalla Bella & Berkowska, 2009).

In spite of the fact that metrics based on objective acoustic measures are preferable to peer judgments for the aforementioned reasons, some limitations are worth mentioning. If these metrics are supposed to replace, at least in part, human judgment of pitch accuracy and precision, it would be important to identify which of them more closely reflect perceived pitch. This is presently unknown. However, a psychophysical study may be in order to validate the match between the metrics commonly used in the field and perceived pitch. Another issue concerns the inevitable fluctuations in F0 trajectory corresponding to a given note. These fluctuations enter in the computation of pitch height (i.e., when calculating mean or median F0) and are qualified as "inaccuracies" or variability by metrics of internal note stability (e.g., Dalla Bella et al., 2007; Hutchins et al., 2014). However, such variability may occasionally be conflated with expressive variations, which are typical of professional singing. Thus, further research should be devoted to finding measures of F0 fluctuations that tease apart random variability from expressive variations (e.g., vibrato).

WHICH CRITERIA?

After choosing one or a set of metrics of singing proficiency, a cut-off score has to be determined so that a performance below its value will be considered indicative of poor-pitch singing. Unfortunately, again, there is no consensus about the final criterion used to tease apart poor-pitch singers from proficient singers (Berkowska & Dalla Bella, 2013; Dalla Bella et al., 2011; Hutchins & Peretz, 2012). Individuals can be qualified as poor-pitch singers based on an absolute criterion, namely if their productions deviate from a target pitch (e.g., in a pitch-matching task) by more than a fixed value (i.e., accuracy). The cut-off value ranges from one semitone (100 cents = 1/12 of an octave; e.g., Berkowska & Dalla Bella, 2013; Pfordresher & Brown, 2007; Pfordresher et al., 2010), to a quarter of a tone (50 cents; Berkowska & Dalla Bella, 2013; Demorest & Clements, 2007; Hutchins & Peretz, 2012; Hutchins, Roquet, & Peretz, 2012). The semitone cut-off appears *prima facie* as a reasonable choice, given that this interval is the smallest difference between two neighboring notes in Western music. However, its drawback is that since a pitch or an interval may be both smaller or larger than a semitone, the acceptable range of performance around the target pitch is indeed 200 cents (e.g., see Hutchins, Roquet, & Peretz, 2012). Thus, this criterion, by being not stringent enough, may lead a poor-pitch singer to be categorized as a good singer. Based on perceptual factors, the more stringent ±50-cent cut-off is proposed. It is assumed that if a note deviates from a target tone by more than half a semitone it will be likely perceived as the other note instead of the target. Note that 50 cents corresponds to the threshold for sung tones to be considered in tune, at least for listeners with musical experience (Hutchins et al., 2012).

An alternative to the fixed cut-off is to adopt a variable criterion. This possibility is justified by the fact that norms for singing accuracy and precision are likely to depend on the tested group (e.g., musicians vs. nonmusicians; individuals with vocal training vs. without training). Accordingly, the criterion to identify poor-pitch singers should vary for different population groups (e.g., it will differ for adults and children). In addition, average accuracy and precision are affected by task factors, with typically worse performance in imitation tasks compared to singing from memory tasks (Berkowska & Dalla Bella, 2013). With a variable cut-off, poor-pitch singers can be identified relative to the performance of a control/comparison group. This method is often used to uncover impaired performances in patients with brain damage (e.g., Schön et al., 2004). An example of variable cut-off is to consider poor-pitch singers those individuals who are outliers in a group on one or a few measures of singing proficiency. Outliers typically deviate from the average of the group by more than 2 *SD* (Berkowska & Dalla Bella, 2013; Dalla Bella & Berkowska, 2009). If the assumption of a normal distribution for the measures of singing accuracy and precision is met, this procedure limits the incidence of poor-pitch singing to a constant percent of individuals located in the tail of the distribution. A similar variable cut-off served in previous studies to determine if an individual suffers from congenital amusia (tested with the Montreal Battery of Evaluation of Amusia; Peretz, Champod, & Hyde, 2003).

The possibility of different cut-offs for defining poor-pitch singing is problematic because it underscores a certain degree of uncertainty in the definition of this condition. The performance of some occasional singers will inevitably score "in the normal range" with a tolerant criterion (e.g., a variable cut-off), while it will fall beyond a more stringent cut-off (e.g., 50 cents) and accordingly will be treated as "poor-pitch singing." Unfortunately, at present there is no unequivocal reason to choose one criterion over the others. Nevertheless, to shed light on the incidence of poor-pitch singing based on different cut-offs, namely fixed (100 and 50 cents) versus variable criteria, we recently examined the performance of a group of 50 occasional singers in a variety of tasks against the three aforementioned criteria. The participants were submitted to the Sung Performance Battery (SPB; Berkowska & Dalla Bella, 2013), formed by a comprehensive set of tasks and analysis methods yielding a thorough profile of someone's singing abilities, and capable of detecting poor-pitch singing and its phenotypes. The SPB includes single-pitch matching, pitch-interval matching, novel-melody matching, singing from memory of familiar melodies (with lyrics and on a syllable), and imitation of familiar melodies (with lyrics and on a syllable) at a slow tempo indicated by a metronome. The slow controlled-tempo condition was implemented in the imitation task, since singing at a fast tempo is typically associated with lower pitch accuracy (Dalla Bella et al., 2007). Sung performances were analyzed using acoustical methods, providing comparable measures of accuracy and precision in terms of both absolute pitch and relative pitch across all tasks. We found, in keeping with previous studies (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Pfordresher et al., 2010), that occasional singers were more accurate and precise when imitating or producing from memory well-known songs than when they imitated single pitches, intervals, and short novel melodies. Moreover, their performance was more accurate and precise when they sung on a syllable than with lyrics (Berkowska & Dalla Bella, 2013). A comparison of the three cut-offs across the tasks reveals that the 50-cent criterion is definitely too stringent, leading to extremely high estimates of poor-pitch singing (up to 84% and 64% of the participants, based on absolute and relative pitch, respectively, for interval and melody matching). Analysis with the 100-cent cut-off obviously yields lower numbers of poor-pitch singers; this criterion, however, is too lax and allows identifying very few poor-pitch singers based on relative pitch. Both fixed cut-off criteria lead to different estimates of the occurrence of poor-pitch singing depending on the task, and measure (based on absolute pitch vs. relative pitch, accuracy vs. precision). The variable cut-off provides a more reasonable estimate of poor-pitch singing at the same time controlling for the differences across tasks and measures. With this criterion, applied across all tasks and metrics, 28% of the tested sample revealed poor-pitch singing in at least one task and in one metric. This criterion, in addition, is sensitive enough to individual differences among poor-pitch singers, in terms of absolute pitch vs. relative, and accuracy vs. precision (for details, see Berkowska & Dalla Bella, 2013).

In sum, various cut-offs have been proposed as criteria to detect poor-pitch singing. Although fixed cut-offs have been often used, they are variably sensitive to poor-pitch singing depending on the task and on the metrics. This hinders the use of fixed criteria for obtaining a thorough profile of singing abilities, where singing is assessed with a set of core tasks and on a variety of measures. This situation could be remediated by applying some form of calibration of the fixed cut-off depending on task difficulty, a possibility that deserves further enquiry. In contrast, a variable cut-off, by definition, is less biased by task factors and by the type of measure, while retaining high sensitivity to the different phenotypes of poor-pitch singing. Thus, a variable criterion should be preferred over a fixed criterion to assess poor-pitch singing with different tasks and measures.

Conclusions

Identifying and characterizing poor-pitch singing in occasional singers is essential to understanding which components of the song system are not fully functional (e.g., based on the VSL) in adults with singing disorders. This knowledge plays a critical role for devising remediation strategies that are theory-driven, and targeted to a particular process or mechanism within the song system. Yet, defining poor-pitch singing, as illustrated in this article, may be quite challenging. Four critical questions have been addressed that impinge upon the definition of poor-pitch singing, related to the choice of the task, analysis method, measures of accuracy and precision, and cut-off criteria. In a nutshell, the figure emerging from this discussion is that poor-pitch singing is a rich and multi-faceted phenomenon, involving a considerable number of individual differences. To capture such complexity, the assessment of poor-pitch singing cannot be confined to using a single task, and based on a single measure or criterion of singing proficiency. Indeed, depending on these factors, the boundary between good and poor singing may shift. Rather, testing should strive to provide a profile of singing abilities based on a variety of tasks targeting the different components of the song system, and, ideally, singing proficiency should be assessed with the same set of measures (e.g., accuracy and precision) across all tasks (see the SPB for an example, Berkowska & Dalla Bella, 2013). The choice of the cut-off criterion is critical and, although we have a preference for a relative criterion (e.g., 2 *SD* from the average of a normative group), the results with different criteria can still be compared before making a final decision about poor-pitch singing. This multi-task and multi-measure approach is likely to provide thorough and valuable information paving the ground to successful remediation strategies.

Author Note

We are grateful to Peter Pfordresher, Sean Hutchins, Joanne Rutkowski, and to one anonymous reviewer who offered valuable comments on a previous version of the manuscript. This research was supported by a research grant from the European Commission (ITN-EBRAMUS, n. 238157).

Correspondence concerning this article should be addressed to Simone Dalla Bella, Movement to Health Laboratory (EuroMov), University of Montpellier 1, Montpellier, France. E-mail: simone.dalla-bella@univ-montp1.fr

REFERENCES

Ackermann, H., Wildgruber, D., & Riecker, A. (2006). "Singing in the (b)rain": Cerebral correlates. In E. Altenmüller, M. Wiesendanger, & J. Kesselring (Eds.), *Music, motor control and the brain* (pp. 205-222). Oxford, UK: Oxford University Press.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238-251.

Berkowska, M., & Dalla Bella, S. (2009a). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69-83.

Berkowska, M., & Dalla Bella, S. (2009b). Reducing linguistic information enhances singing proficiency in occasional singers. *Annals of the New York Academy of Sciences, 1169*, 108-111.

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

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch-matching abilities of adults who sing inaccurately. *Journal of Voice*, *14*, 431-439.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. R., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research*, *20*, 363-375.

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences*, *1169*, 99-107.

Dalla Bella, S., Berkowska, M., & Sowiński, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology*, *2*, 164.

Dalla Bella, S., Giguère, J-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America*, *121*, 1182-1189.

Dalla Bella, S., Giguère, J-F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America*, *126*, 414-424.

Dalla Bella, S., Tremblay-Champoux, A., Berkowska, M., & Peretz, I. (2012). Memory disorders and vocal performance. *Annals of the New York Academy of Sciences*, *1252*, 338-344.

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

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.

Gordon, R. L., Racette, A., & Schön, D. (2006). Sensory-motor networks in singing and speaking: a comparative approach. In E. Altenmüller, M. Wiesendanger, & J. Kesselring (Eds.), *Music, Motor control and the brain* (pp. 205-222). Oxford, UK: Oxford University Press.

Gosselin, N., Jolicoeur, O., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 270-272.

Griffiths, T. D. (2008). Sensory systems: Auditory action streams? *Current Biology*, *18*, R387-388.

Guerrini, S. C. (2006). The developing singer: Comparing the singing accuracy of elementary students on three selected vocal tasks. *Bulletin of the Council for Research in Music Education*, *167*, 21-31.

Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *Journal of the Acoustical Society of America*, *123*, 273-278.

Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain*, *126*(8), 1838-1850.

Hutchins, S., Larrouy-Maestry, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception and Psychophysics*, *76*, 2522-2530.

Hutchins, S., & 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.

Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, *30*, 147-159.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America*, *127*, 504-512.

Kinsella, G., Prior, M. R., & Murray, G. (1988). Singing ability after right and left sided brain damage. A research note. *Cortex*, *24*, 165-169.

Larrouy-Maestri, P., Lévêque, Y., Schön, D., Giovanni, A., & Morsomme, D. (2013). The evaluation of singing voice accuracy: A comparison between subjective and objective methods. *Journal of Voice*, *27*, 259.e1–259.e5.

Loui, P., Guenther, F., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*, R331-332.

Marin, O. S. M., & Perry, D. W. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *Psychology of music* (pp. 653-724). San Diego, CA: Academic Press.

Mithen, S. (2006). *The singing Neanderthals*. Cambridge, MA: Harvard University Press.

Murayama, J., Kashiwagi, T., Kashiwagi, A., & Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain and Cognition*, *56*, 36-42.

Ostwald, P. F. (1973). Musical behavior in early childhood. *Developmental Medicine and Child Neurology*, *15*, 367-375.

Papoušek, H. (1996). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliège & J. A. Sloboda (Eds.), *Musical beginnings* (pp. 37-55). Oxford, UK: Oxford University Press.

Peretz, I. (2001). Brain specialization for music: New evidence from congenital amusia. *Annals of the New York Academy of Sciences*, *930*, 189-192.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 185-191.

Peretz, I., Champod, S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58-75.

Peretz, I., & Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*, 362-367.

Prior, M., Kinsella, G., & Giese, J. (1990). Assessment of musical processing in brain-damaged patients: implications for laterality of music. *Journal of Clinical and Experimental Neuropsychology*, *12*, 301-12.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of 'tone-deafness.' *Music Perception*, *25*, 95-115.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception and Psychophysics*, *71*, 1385-1398.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America*, *128*, 2182-2190.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*(10), 2571-84.

Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly: Center for Research in Music Learning and Teaching*, *1*(1-2), 81-95.

Rutkowski, J., & Miller, M. S. (2003a). 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.

Rutkowski, J., & Miller, M. S. (2003b). The effectiveness of frequency of instruction and individual/small group singing activities on first graders' use of singing voice and developmental music aptitude. *Contribution to Music Education*, *30*, 23-38.

Schön, D., Lorber, B., Spacal, M., & Semenza, C. (2004). A selective deficit in the production of exact musical intervals following right-hemisphere damage. *Cognitive Neuropsychology*, *21*, 773-784.

Stewart, L., von Kriegstein, K., Dalla Bella, S., Warren, J. D., & Griffiths, T. D. (2009). Disorders of musical cognition. In S. Hallam, I. Cross, & Thaut, M. (Eds.), *Oxford handbook of music psychology* (pp. 184-196). New York: Oxford University Press.

Sundberg, J. (1987). *The science of the singing voice*. DeKalb, IL: Northern Illinois University Press.

Sundberg, J. (1999). The perception of singing. In D. Deutsch (Ed.), *The psychology of music* (pp. 171-214). San Diego, CA: Academic Press.

Terao, Y., Mizuno, T., Shindoh, M, Sakurai, Y., Ugawa, Y., Kobayashi, S., et al. (2006). Vocal amusia in a professional tango singer due to a right superior temporal cortex infarction. *Neuropsychologia*, *44*, 479-88.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Trehub, S. E., & Trainor, L. J. (1999). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43-77.

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M-A., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive Neuropsychology*, *27*, 463-476.

Wallin, N. L., Merker, B., & Brown, S. (2000). *The origins of music*. Cambridge, MA: MIT Press.

Watts, C., Murphy, J., & Barnes-Burroughs, K. (2003). Pitch-matching accuracy of trained singers, untrained participants with talented singing voices, and untrained participants with non-talented singing voices in conditions of varying feedback. *Journal of Voice*, *17*, 187-196.

Welch, G. F. (2006). Singing and vocal development. In G. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 311-329). New York: Oxford University Press.

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology*, *6*, 15-22.

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,* 3-23.

Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience*, *7*, 237.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-related neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, *40*, 1871-1887.

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, *48*, 1871-1887.