Cohen, A. J., Pan, B.-Y., Da Silva, E., & Dutton, K. (2020). Performance of Canadians on the automated AIRS Test Battery of Singing Skills: music training an age. In F. A. Russo, B. Ilari, & A. J. Cohen (Eds.), *The Routledge companion to interdisciplinary studies in singing. Volume I: Development* (pp. 357-371). Routledge. doi.org/10.4324/9781315163734

Performance of Canadians on the automated AIRS Test Battery of Singing Skills: music training and age

Annabel J. Cohen, Bing Yi Pan, Eric da Silva, & Kyle Dutton

University of Prince Edward Island

Introduction

The AIRS Test Battery of Singing Skills (ATBSS) provides the opportunity to collect data on a wide range of singing abilities. It includes such tasks as singing a familiar song and a favorite song, learning a new song, creating a song, producing one's highest and lowest note as an estimate of vocal range, and imitating basic musical elements.[1] It explores a wider variety of singing abilities than more specifically focused test collections.[2] The ATBSS also gathers data on several speech tasks including verbal fluency, imitation of familiar and unfamiliar phonetic material, and creation of a story. Its broad scope addresses an ambitious aim of providing data for understanding how singing develops across the lifespan and across cultures. This goal was associated with the AIRS project (Advancing Interdisciplinary Research in Singing) as described in the Preface of this volume. Between 2008 and 2012, the ATBSS was designed for both suitability to a wide age range of individuals and adaptation to countries and cultures throughout the world. After the English language version was piloted in Canada (Cohen, Armstrong, Lannan, & Coady, 2009), the test was administered in England, Brazil, and Estonia, translated into Portuguese and Estonian in the latter two cases. Some of these studies were represented in a special issue of *Musicae Scientiae* (Gudmundsdottir & Cohen, 2015), for example, work of Ilari and Habibi (2015) with children of Latino culture, and of Raju, Välja, and Ross (2015) with Estonian children.

Test sessions were recorded for purposes of later analysis, and video recordings provided useful information about participants behavioral responses, but also revealed subtle variability in procedure between testers or even within an individual tester (e.g., offering positive or negative feedback, allowing more than one try on a task, or offering hints or cues. Different treatment of participants, at the very least, adds unwanted variability, and, in the worst case, can bias results (Greenberg & Folger, 1988). A more standardized procedure was needed to reduce differential treatment of participants. This could be accomplished by replacing the human tester with a video of a guide who appears on a computer screen in the same way for every participant. The development of such an automated version took place between 2012 and 2015 (Pan, Liu, & Cohen, 2012). As well as offering a standardized procedure, the automated test in theory opened testing to anywhere in the world with a stable Internet connection.

Music training is often associated with superior or unique performance on musical as well as other tasks (Schellenberg, 2019). Persons with music training often perform differently than those without such training. Other variables are associated with music training (e.g., innate musical ability, higher socioeconomic status). Without a true experiment in which participants are randomly assigned to different training conditions, it is not possible to say that musical training causes a behavior (Jakobson & Cuddy, 2019; Schellenberg, 2019). Nevertheless, comparisons of extreme groups, that is, participants with a great deal of music training versus those with no music training, can be informative and provide a useful framework for considering other factors besides music training (e.g., age and culture) that impact singing development. Another research design considers participants representing a wide range of music training, and measuring the degree of correlation between the music training variable and the other outcome measures of interest (e.g., pitch accuracy, voice quality, creativity). With such a design it is possible only to report the strength of the relationship between music training and the outcome variable; again it is not possible to say that the music training caused the effect.

The present chapter reviews performance of components of the automated version of the ATBSS by several different groups of Canadians. First performance is compared between Canadian university students who are either highly experienced with music or have no musical training. Next, the focus is on student athletes, because athletes and musicians share a life of disciplined focus, and the comparison helps to isolate the impact of music training from simply disciplined practice, be it sports or music. Like music, sports can be both individual and collective, thus, the comparison between those with training in music and athletes can be informative. Adding further perspective, an example of a lifespan cross-sectional approach is described along with an example of a longitudinal study of an older adult over a three-year period.

The challenges of measurement

Each of the tasks of the ATBSS provides a wealth of information, and more than one measurement could be useful to capture it all. For example, in the task in which the participant is asked to create a song, the produced melodies can be subjected to consensual assessment by experts for musicality or creativity (Amabile, 1982), as well as analysis with respect to adherence to the rules of music theory or some other world music descriptive system (Ellis et al., 2018). Video recordings of the participant can supplement the acoustical information, for example revealing the level of stress a participant may experience for a particular task (for example, when asked to make up a song), or the extent of emotion that a certain task might initiate.

Most people experience an emotional response to music (Juslin, 2019; Sloboda, 1986), and music that is linked to a personal memory is more likely to induce an emotional response (Juslin, Liljeström, Laukka, Västfjäll, & Lundqvist, 2011). Facial expression provides information regarding emotional experience (Ekman, Friesen, & Ancoli, 2001; Kaiser, 2017; Weth, Raab, & Carbon, 2015), while information about melodic intervals is also conveyed by the mouth (Russo, Sandstrom, & Maksimowski, 2011. The focus of the present chapter is however on the sounds produced in singing.

It would be ideal to analyze every pitch produced by the participant for the entire 11 components of the test battery, however, the time to accomplish this can be prohibitive. To give an example, the 32 notes of the song *Brother John* are sung on four occasions during the ATBSS, and on one other occasion (a training phase), the participant echoes back each of the four phrases of the song, thus singing one half of the notes (i.e., 16 notes). Thus, just this one test component produces a total of 144 notes (considering the 4.5 repetitions). To analyze the pitch of these notes would take about three hours for a single practiced individual, and additional time to obtain inter-rater reliability. We have found that focusing analysis on salient notes of the melody provides useful information at a reduced time requirement.

As shown by Wolfe, Garnier, Henrich and Smith (Chapter 5) and Devaney (Chapter 10) and as discussed by Stadler Elmer (Chapter 2) and Pfordresher and Demorest, (Chapter 24), analysis of pitch of the voice is a complex task, particularly for notes sung by singers who are less accurate.[3] Pitch analysis software, such as Praat (Boersma & Weenink, 2013) may introduce octave errors, which, being 12 semitones “out” can have dramatic effects on the calculation of a group average pitch, and human overview is needed whenever these tools are implemented. When notes are not separated by silence, human intervention is also required to provide information about the start of a note. While our earliest studies did in fact measure each pitch of the song *Brother John* (Cohen et al., 2015), our later efforts focused on salient pitches within a song. In particular we have focused on the key-note or tonic (*do* in the *do re mi* system) of the piece. Of the 32 notes of *Brother John*, 10 are tonic notes as seen in Figure 27.1.

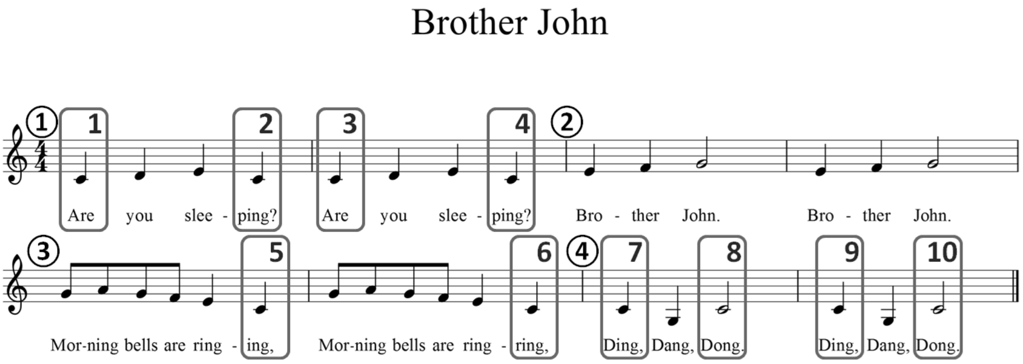

Figure 1. Ten tonic notes of the melody of *Brother John* (*Frère Jacques*).

The four phrases of the melody are identified by encircled numbers.ab

Accurate singing would ideally produce identical pitches for these 10 notes. Much can be learned from focusing on how participants handle these structurally significant tones. Measurements of the 10 frequencies (f_o's) can be analysed for variability. If all 10 f_o's are identical, the variability will be 0, but if the singer misrepresents the melody and 10 f_o's are dramatically different, then the variability will be high. The variability of the 10 measures of f_o about the mean f_o informs us about the stability of the singing voice. In other work (independent of the ATBSS), Beck, Rieser, and Erdemir (2017) reported the standard deviation of just three tonics of the *Happy Birthday* song as a measure of the influence of different types of concurrent feedback on singing accuracy. Depending on the problem asked, different tones can be targeted.

The reader has probably experienced attending a birthday celebration when the singing of *Happy Birthday* led to a cacophonous sound reflecting that different singers start the song on different notes. Everyone may be singing the melody correctly, but in different transpositions. Eventually it all works out and most people settle in on the same key. But what leads to key choice in the first place? There can be several factors. The ATBSS provides an opportunity to consider this question, and to obtain information beyond questions of pitch accuracy from the *Brother John* data, having important information about cognition and attention during singing

Data sets

Since the development of the online version of the ATBSS, several types of data collection have taken place. Here we focus on the data collected at the University of Prince Edward Island between the years 2012 and 2015 with native speakers of English. We examine primarily three sets of data. The first two entailed 20 musicians and 20 persons without musical training respectively, and the third entailed 40 élite university athletes. We also report on a cross-sectional and a longitudinal study that involved older participants.

University music majors and non-majors

The life of a student enrolled in a music degree program differs from that of a student majoring in most other disciplines. The music student focuses many hours a day on music. Some of these hours are spent in practice and performance; while others demand musical imagery for studies of harmony, and counterpoint. Yet other courses such as music history require listening. Music students typically engage in choral singing as part of their degree program requirements, regardless of whether they are voice majors. Students who are enrolled in majors other than music, may also listen to a lot of music every day, but usually it is popular music.[4] Generally their attention is directed to a broader range of extra-musical domains of interest (television, sports events, part-time jobs) as compared to the music-focused attention of music students.

The comparison of persons with and without music training provides the opportunity to explore what singing ability is common to all young adults and what is associated with a life in music. We have studied the performance of 20 young adults who have an average of seven years of training on a musical instrument and two years of voice lessons. We have also studied a contrasting group of 20 young adults who have never played a musical instrument nor taken voice lessons. Participants in the *untrained* group have had no more than 1 year of training on a single instrument, while participants in the *trained* group have had no less than 2 years on their major instrument and met at least two of the three following criteria: enrolled as a university music major, had at least 5 years of private lessons on musical instrument or voice, obtained at score of at least 9 out of 10 in the music reading and theory test that accompanied the ATBSS. For the non-musician group, the average score on this music test was 0.25.

University athletes

Singers are sometimes referred to as “vocal athletes” (LeBorgne & Rosenberg, 2014). The concept highlights the fact that singing entails sophisticated motor co-ordination as a result of dedicated practice and training. Performance on any musical instrument, not just the voice, entails a high level of motor control and integration of multi-sensory information. The required extensive practice leads to adaptive neuroplasticity in sensory and motor systems and enhanced processing of bodily information (Schirmer-Mokwa et al., 2015). Inter-university competitive sport also entails extensive training, and in some ways resembling that of music students. It is possible, on the one hand, that the focus on motor control and disciplined attention that both music students and student athletes share might lead to some similarities in their singing test performance. On the other hand, the lives of athletes also resemble those of students who are not musicians, especially for athletes who have had little music training. Thus, we direct attention to a group of 40 university athletes who had been dedicated to training in one of a wide range of varsity sports. Some of the athletes had had some musical training, and the group of 40 was divided into two groups based on a median split of their level of music training. The level of music training was determined by performance on a music theory test and by the ranking of three experienced musicians who were given background information on a file card for each athlete. The judges were asked to rank order the athletes on this basis. The musician-athlete group obtained a score of 3.75 on the music questionnaire as compared to a score of 1.80 for the non-musician-athlete group. The musician athlete group had played an average of 1.14 musical instruments while the non-musician group had played an average 0.1 musical instruments.

Lifespan Study

The automated ATBSS was developed for administration to persons representing a wide age range. Hence one initiative aimed to test persons representing every decade through to the ninth. Although all decades up to the age of the 80's were eventually represented, numbers of persons in each decade was relatively small (for some decades, only 2 persons), and did not provide a representative sample. The initiative however revealed that it was possible to obtain data from young children (age 4 to 6) with an experimenter sitting with the child. The youngest child to take the test independently was 8 years old. Further, persons in their 80's were able to manage the test independently.

Longitudinal case study of one octogenarian

In the earliest exploration of the ATBSS, participants were tested monthly for 5 months (Cohen, 2015; Cohen et al., 2009). Subsequently, most studies with the automated ATBSS typically tested participants only once. However, a case study was conducted with a gentleman who was tested on seven occasions between the ages of 84 to 87 years. He had expressed an interest in taking voice lessons over this period of time, and the researchers were interested in his general ability at this age and its sustainability across the sessions. It is noted that Bugos (Chapter 32) collected data from the on-line ATBSS before and after several weeks of group singing lessons for a group of older adults in the United States.

Method: Application of the automated ATBSS

Above we have described four different sets of data collected with the automated ATBSS: musicians, athletes, persons ranging in age across eight decades, and a longitudinal focus on one octogenarian. Whereas we will report results of these four different Canadian studies, because the testing procedure is automated and was conducted in the same location, the following description of the procedure for testing is general to all four studies.

Procedure

The procedure was approved by the University Research Ethics Board in accordance with the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans. The test was presented using the online AIRS-Test system (Pan, et al., 2012) via the computer. At the participant's keyboard in the double-walled testing room, the experimenter called up the Test Battery website and entered the appropriate (female or male) invitation code. The experimenter then left the booth and monitored the behavior of the participant from the adjacent single-walled booth where the output of video recorder was displayed. The experimenter could speak to the participant through a microphone if necessary. Audiovisual responses of the participant were recorded by a webcam and sent to a Mac Pro computer (housed in the booth where the experimenter sat). When the online test was completed, the audiovisual data were sent for storage in the remote (cloud) ATBSS data base. As a backup, a video camera in the testing booth recorded the session and picking-up the audio and visual image of the participant.

The full script of the ATBSS is provided in the companion website [CW27.1] at www.airsplace.ca. In brief, an audio-visually depicted pleasant female of university age (a professional vocalist and actress) invites the participant to carry out a series of singing and speaking tasks. Acting as a guide, she introduces each of the tasks. Written instructions also appear on the screen, sometimes in abbreviated form. On completion of each task, the participant initiates the next task with a mouse click. Depending on the task, the computer program saves audiovisual data of speaking and singing as recorded by the web-cam, typed text to open-ended questions, or a mouse-click indicative of a choice among several alternatives (e.g., when four pictures are presented and the participant must select one).

Test Components of the ATBSS

The complete sequence of 11 test types were as follows (see also Figure 22.1):

1) opening conversation: respond verbally to one of three questions about musical interests;
2) sing the familiar song *Brother John* (*Are you sleeping/ Frère Jacques*), also after components 5, 10, and 11. [prompted by two lines in English from the song];
3) repeat (a) an English sentence that contained most of the English speech sounds, and (b) four Mandarin words, each representing a different linguistic tone;
4) repeat a 5-note easy melody *sol-mi-mi-sol-mi*
5) sing your highest and lowest note;
6) sing your favorite song;
7) repeat several musical elements (e.g., single notes, intervals, triads, scales);
8) continue an 8-note song (melody completion/improvisation);
9) create a song inspired by a choice of four pictures, with and without words (i.e., 2 songs in total);
10) sing a new song (presented in the key of E≅);

11) make up a story inspired by a picture.

After component #5, there was a request to sing the entire song, and then to imitate each of the four phrases of the *Brother John* song. The request to sing the *Brother John* song also appeared after component #10 (after learning a new song in the key of E≅) and after component #11(creating a story). The repeated testing of the *Brother John* song allows for testing of the participants’ ability to sing accurately, but also determining the influences of various contexts (such as the influence of a prior musical key, C vs E≅).

Contexts for singing of *Brother John / Frère Jacques*

Component #2 of the ATBSS refers to singing the song *Brother John* / Frère Jacques without any given musical context. Two verbal cues were provided on the screen - (the first line *Are you sleeping*, and the third line *Morning Bells are Ringing*) to enable the participant to reveal his or her melodic and remaining verbal knowledge of the song. No melodic cues were provided. The training condition, which began after Component #5, presented a vocal model of the entire song of *Brother John* (female voice for female participants, male voice for male participants) in the key of C. The guide asks the participant first to listen to the entire song, and then to sing it back. Immediately after this, the first bar of each of the four phrases of *Brother John* is presented, and the participant is asked to repeat this phrase. Thus, the voice of the model and the echo of the participant together render the entire 32 notes of the song again. Whereas the participant is not told to sing in the same key as the model, most participants regard the demand as implicit. The exercise assured that each participant had an opportunity to hear and sing each part of the song. The next request to sing the song follows Component #10, in which the participant has just listened to two presentations of an unfamiliar song in the key of E≅ and subsequently was asked to sing back the song as accurately as possible using “la” (without words). The participant hears the E≅ song four times and tries to sing it twice. A research interest was to determine whether the E≅ context would influence the subsequent choice of key when asked to next sing *Brother John*. The final request to sing the *Brother John* again follows Component #11 which invites the participant to create and tell a verbal story based on a choice of two pictures.

After singing a final version of *Brother John*, the participant clicked a button to submit the results to the remote server. The on-line music background questionnaire followed (approximately 10 minutes for those with considerable music training, and much less if there was no training to report). The questionnaire asked about basic music theory (naming of notes and chords and identification of chord progressions), and the ability to recognize a familiar melody from music sight-reading, with a maximum score of 10. The total test time was approximately 40 minutes.

The vocal model

The melodic materials had been recorded using a Sennheiser e855 microphone and the Audacity program (16bit, 44.1kHz). The Melodyne v3.2 program was used to adjust the pitch to conform to the equal-tempered scale (A_4 = 440.00 Hz) and note durations were also exact. There was one “perfect” female recording of the vocal model of *Brother John* in the key of C and the similarly perfect male recording was created by transforming the female down one octave using Melodyne.

Observations and discussion

Each component of the ATBSS provides a wealth of information and is like a mini-experiment unto itself. Researchers typically focused on only a few components, or even just one, but all of the data remain available for eventual analysis. We report a selection of observations on the components from the musicians and non-musicians and then discuss the observations from the remaining data sets.

Musicians and non-musicians

Speaking voice analysis – Components 1 and 3a

The average f_o was calculated for verbal and melodic materials. Measurement of the average f_o of the opening conversation revealed a lower frequency for the non-musicians. The f_o of the first singing of *Brother John* was also lower for non-musicians and showed greater spread across the chromatic tones. A correlation between this frequency and that of the choice of tonic in Hz was significant for the non-musicians but not for the musicians, supporting the view of a stronger relation between speaking voice and the choice of tonic for decreasing levels of music experience. The mean f_o of the repeated English sentence was lower for the non-musicians, and the musicians' f_o came closer to the f_o of the model on the repetition. Details of the analysis are found in the extended report at [**CW27.2**]

Voice range – Component 5

The highest and lowest notes reflected a smaller range for the non-musicians, not surprisingly. Details of the analysis are found in the companion website at [**CW27.2**]

Brother John – Choice and variability of tonic - Component 2

The mean tonic of *Brother John* after the presentation of the model (in C) showed perfect compliance of key with the model for musicians but less so for the non-musicians, although the modal value (C) was the same for both groups. After the song in E≅, the distribution of keys for the musicians and non-musicians differed, with the musicians showing more choices in and near E≅. After creating the story, the tonic of *Brother John* was lower for the non-musicians with the musicians staying close to E≅ or moving to C (the key of most of the ATBSS components). The overall standard deviation of the tonics of *Brother John* was greater for the non-musicians. [**CW27.2**]

Melody completion – Component 8

The improvised endings (including the repetition of the original 8-note melody *do re re mi re do re mi*) were notated by two musicians independently and then compared and reconciled with the help of the Melodyne program. There were 34 participants (19 musicians), who completed the task. Musicians repeated the 8-note model significantly more correctly than did the non-musician and had a significantly large range (8 vs 5 semitones). Musicians began the improvised (new) part of the melody most frequently (45%) on the dominant note (G), whereas non-musicians chose the tonic (C) most frequently (40%). All musicians ended on the tonic (C, 100%), while only 47% of non-musicians did. These results suggest that music training (and its correlates) shapes melody expectancy when improvising and expands the materials one can comfortably use in instant melody creation. [**CW27.3**]

Melody improvisation and story creation – Components 9 and 11 [**CW27.4**]

The creativity of the songs (without words) of 24 of the participants was assessed by 6 experts (Ellis et al., 2018), and their mean ratings correlated with the number of words produced in story and the duration of the story. As well, the music test score showed the same pattern of correlation. These results were consistent with the idea that musical experience is associated with domain-general divergent thinking that applies to non-musical tasks. It is likely that domain-specific convergent thinking constrained the duration of the created melodies, but had no influence on narrative creation, and current research by Erin Hannah is examining the melodic structure of all melodies (with and without words) to examine this idea further.

Emotional response as indicated by facial expressions – Components 2 and 6

For a subset of 12 participants cross-classified as musician/ non-musician and male/female gender (3 in each group), an analysis was conducted comparing the facial expression of the favorite song versus a previous and later singing of the familiar (but not favorite) *Brother John* song. An 18-item facial movement scale was developed.[5] Four judges rated performance of the three songs of the 12 singers using this scale. The scale represented a simplified version of the Ekman Facial Action Coding scheme (Ekman, Friesen, & Hager, 2002). Ratings on the majority of the items were influenced by one or more of the three variables (meaningfulness of the song, gender, and level of music training). It was predicted that self-selected song, females, and musicians would show greater movement. Inter-rater reliabilities between the four judges for the 19 movement scales for each of the 3 song conditions led to Cronbach's Alpha > .70 for 15 scales, and these scales were examined in further analysis. Results of an analysis of variance suggested that females and males differed in the extent to which they looked down during the familiar song: males looked down the most during the familiar song (a nursery rhyme) as compared to females who looked down relatively little, but looked down relatively more for their favorite song. Other measures showing influences were left head tilt, lip corner puller, mouth stretch, eyes up, eyes turn left, eyes turn right, and eyes close (MacDonald & Cohen, 2013.

Summary

As summarized in Table 27.1 Row 1 (Musicians/Non-Musicians), the components offer a wealth of data that can be brought to bear on the question of how singing develops. A number of aspects of performance were associated with music training such as accuracy/variability, quality of improvised melodies, number of ideas in created stories, well-formedness of improvised song endings, sensitivity to prior musical context in choice of context; wider voice range, and greater independence of speaking voice. As well evidence of a relation between emotional action of the face and singing contexts was also shown.

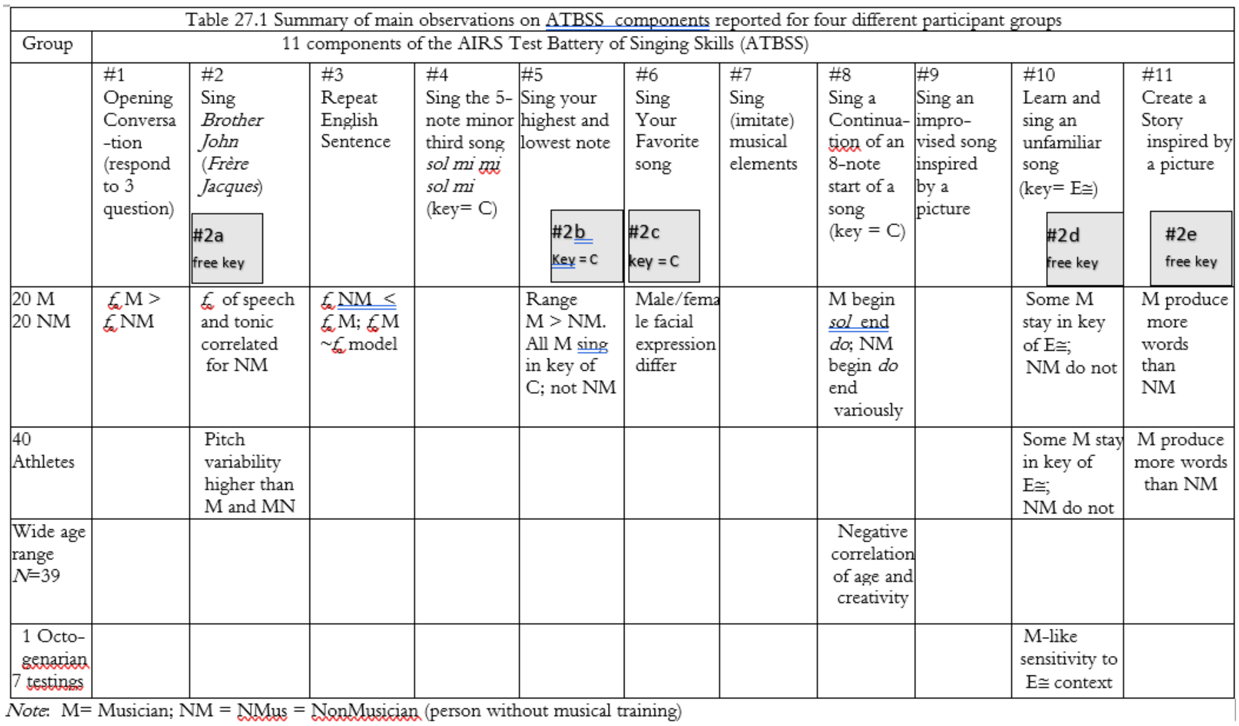

Table 27.1 Summary of main observations on ATBSS components reported for four different participant groups

Group	11 components of the AIRS Test Battery of Singing Skills (ATBSS)										
	#1 Opening Conversa-tion (respond to 3 question)	#2 Sing *Brother John* (*Frère Jacques*) #2a free key	#3 Repeat English Sentence	#4 Sing the 5-note minor third song *sol mi mi sol mi* (key= C)	#5 Sing your highest and lowest note #2b Key = C	#6 Sing Your Favorite song #2c key = C	#7 Sing (imitate) musical elements	#8 Sing a Continua-tion of an 8-note start of a song (key = C)	#9 Sing an impro-vised song inspired by a picture	#10 Learn and sing an unfamiliar song (key= E≅) #2d free key	#11 Create a Story inspired by a picture #2e free key
20 M 20 NM	f_0 M > f_0 NM	f_0 of speech and tonic correlated for NM	f_0 NM ≤ f_0 M; f_0 M ~f_0 model		Range M > NM. All M sing in key of C; not NM	Male/female facial expression differ		M begin *sol* end *do*; NM begin *do* end variously		Some M stay in key of E≅; NM do not	M produce more words than NM
40 Athletes		Pitch variability higher than M and MN								Some M stay in key of E≅; NM do not	M produce more words than NM
Wide age range *N*=39								Negative correlation of age and creativity			
1 Octo-genarian 7 testings										M-like sensitivity to E≅ context	

Note: M= Musician; NM = NMus = NonMusician (person without musical training)

Athletes

Choice and variability of tonic- Component 2

A detailed analysis of the 10 tonics of the *Brother John* data examined performance on each of the 10 tonics occurring across the piece. The athletes with less musical experience had trouble with the third phrase of the piece while those with musical training were more stable throughout the piece. The standard deviation of the tonics of *Brother John* were greater than that of both non-musicians and musicians previously described, although the athletes had more musical training than the non-musicians in that group. As the pitch analysis was conducted by different coders using a slightly different algorithm in this study, this result needs to be taken with caution. [**CW27.6.1**] As well, even though the average training of the athletes who were more musical was much lower than those in the musically trained group previously described, they showed a similar influence of the E≅ context that was not found for the athletes who lacked musical training. It was also found, not surprisingly that the athletes who spent more time on music spent less on their sport. [**CW27.6.2**]

Song and story creation - Components 9 and 11

Once again, music test scores predicted story duration and number of words in the story, but not duration of the song. It was interesting to note that the time to create the stories was less than that for the song, but the story lyrics were much longer than that of the song. (See Table 27.1 Row 2). Song and story length and number of words in the song were significantly inter-correlated, suggesting domain generality of a creativity component. The results are consistent with separate ideational and control networks proposed by Beaty, Benedek, Silvia, and Schacter (2016) underlying creative cognition in general with application to vocal improvisation. [**CW27.4**]

Lifespan cross-sectional study

Create an ending– Component 8

Completions (improvised endings) of the 8-note song were analyzed. There were initially 39 participants ranging in age from 4 to 87 years, however oldest and youngest were eliminated because they could not create an ending.[6] The improvisations were analyzed in three ways. The first examined the visual representation of the acoustic waveform produced by the Musical Micro Analysis program designed to represent the singing voice, with the *y*-axis depicting notes in semitone steps (Elmer, 2006; Stadler Elmer & Elmer, 2000). This depiction provided information about the number of notes in the melody, 7.1 (SD= 4.7), and the contour of the melody produced by the participant. Participant's produced on average 5.8 contour changes (including repeating the model which entailed 2 contours changes) with a mean vocal range of 7 semitones. Sixty six percent of participants were able to correctly repeat the model melody and continue to create an improvised ending. Seventy six percent of participants ended their melodies on the tonic note.

The second analysis conducted correlations on several different measures. Correctly repeating the model melody was significantly correlated with ending the improvisation on the tonic note, $r = .61$, $p < .001$), with the number of contour changes found in participant's entire melody, $r = .39$, $p < .032$), and with the number of notes in the melody, $r\ (29) = .38$, $p < .03$. Being able to correctly repeat the model shows a competency and understanding of the musical structure and rules governing the model melody. This musical knowledge may offer confidence in the participants' abilities for composition and improvisation. A foundational knowledge allows one to push those boundaries to create something new. A correlation related to age was conducted using data from all 39 participants. A negative correlation ($r = -0.327$, $p < .02$) was found between age and ability to make up a melody, indicating that as participants became older they were less likely to engage in the improvisation component of this task. (See Table 27.1 Row 3). This leads one to wonder if their ability to perform this task declines with age, or if older participants might be more shy. The third analysis focused on the musical notation of the melodies and revealed a large variety of completions from the simple beginning 8-note pattern. [CW27.7]

Singing octogenarian

Previous explorations with four persons in their 80's tested twice revealed evidence of the basic singing skills.[7] The question addressed here is: can repeated vocal tests track general cognitive and motoric resilience associated with singing skills, and to what extent can an older person benefit from voice lessons?

Data obtained from the online ATBSS were compared to group data of the 20 musicians and 20 non-musicians. The focus was the singing of the familiar song under four different contexts over the course of the entire test. Our specific attention was on the choice of initial tonic under the different context conditions, and pitch variability of all 10 tonics in the song. As previously described, young musicians (e.g., enrolled in a university music program) and non-musicians show different patterns of tonic choices. Young musicians chose higher pitches for the key-note and show greater sensitivity to the previous musical context (greater flexibility). But what would a much older person do over a period of 3 years while taking voice lessons?

Regarding choice of key, following the C-model, all but one of his key-note choices were C (the other 1 semitone away), and following the E≅ context, there was a strong tendency to sing on or near the key of E≅, (i.e., 3 E≅,’s, 2 D’s -1 semitone away, and 2 other key choices). This pattern is more musician-like than non-musician-like. Consistency of the key-notes within a piece suggest stable cognitive-motoric processes over the more than 3.0 years during which performance data were collected (i.e., between the ages of 84 to 87 years). (See Table 27.1 Row 4).

The change of key following the change of context (particularly the C-model and the E≅-melody) suggests cognitive flexibility. Other components within ATBSS (e.g., favorite song, pitch accuracy in singing back the melodic elements, making up an ending to the song, creating a new song and a story) remain to be analyzed and may provide information about other cognitive, motor, and emotional processes in this older gentleman. It is recognized that this is a case study of an unusually capable participant, a former clinical psychologist, independent living and travelling, who claimed to have little music training but who greatly enjoyed classical music. Of interest is that he behaved much like the young musically trained group, for the key choice, routinely selecting the E≅ key for *Brother John*, following the E≅ context. He had said he was not musically trained, however, we learned that his brother had engaged in singing as an enthusiastic amateur. [CW27.8]

Conclusion

The ATBSS leads to an almost overwhelming amount of information that can further our understanding of the acquisition of singing across the lifespan and across cultures. The analysis of the data already obtained deserves continuing extension. Every note and every musical interval is informative on its own and in conjunction with other measures.[8]

This Chapter complements the others on the ATBSS that follow in this final section of the volume. While much data have been collected and are reported, they represent the tip of the iceberg. The CHILDES data base for the study of child discourse led to over 3000 publications initially obtained from as few as 100 researchers (MacWhinney, 2000, 2014). We can imagine collections of data ongoing at different sites for different age groups and levels of musical training, and for persons representing a variety of individual differences. It is hoped that the reader will appreciate the scope of this endeavor for obtaining the data needed as the foundation of theories of the development of singing and for advocacy regarding the benefits and applications of singing for society. Already our work in Canada has begun to suggest the effects of musical training on vocal accuracy, voice range, and musical key choice, but also on creativity in the non-musical domain of story creation. We have also seen resilient vocal performance over a period of three years in a non-musician in his late mid-to-late 80's. As we overcome measurement challenges and define ways of measuring performance, our cross-sectional and longitudinal studies recommend the importance of continuing to pursue studies with the automated ATBSS as a basis for developing theories of singing acquisition that can define the vocal potential across cultures and over the lifespan.

Acknowledgements

Acknowledgement is offered to the Social Sciences and Humanities Research Council, the AIRS research facility at UPEI, and all collaborators and students involved in research on the ATBSS.

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43,*997–1013.

Amabile, T. M. (1996). *Creativity in context.* Boulder, CO: Westview.

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends in Cognitive Science, 20*, 87-95.

Beck, S. L., Rieser, J. J., & Erdemir, A. (2017). Singing without hearing: A comparative study of children and adults singing a familiar tune. *Psychomusicology: Music, Mind, & Brain, 27*, 122-131.

Belfi, A. M., & Kacirek, K. (2019). The famous melodies stimulus set: Development and normative data. Poster presented at the Biennial meeting of the Society for Music Perception and Cognition, New York.

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

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.47, retrieved 23 April 2013 from http://www.praat.org/

Bugos, J. A., DeMarie, D., Torres, M., Gbadamos, A., & Anderson, S. (2019). Children’s facial affect on singing tasks: Results of imitated and improvised vocal responses. Poster presented at the Biennial Meeting of the Society for Music Perception and Cognition (SMPC), New York.

Cohen, A. J. (2015). AIRS Test Battery of Singing Skills (ATBSS): Rationale and Scope. *Musicae Scientiae, 19*, 238-264.

Cohen, A. J. (2019). Singing. In P. J. Rentfrow & D. J. Levitin (Eds.). *Foundations in music psychology* (pp. 685-750). Cambridge, MA: MIT Press.

Cohen, A. J., Armstrong, V., Lannan, M., & Coady, J. (2009). A protocol for cross-cultural research on acquisition of singing. Neurosciences and Music III-Disorders and Plasticity. *Annals of the New York Academy of Science, 1169,* 112-115.

Ekman, P., Friesen, W. V., & Ancoli, S. (2001). Facial signs of emotional experience [Abstract]. In W. G. Parrott (Ed.), *Emotions in Social Psychology: Essential Readings* (pp. 255-264). New York, NY US: Psychology Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System*. Salt Lake City, UT: A Human Face

Ellis, B. K., Hwang, H., Savage, P. E., Pan, B.-Y., Cohen, A. J., & Brown, S. (2018). Identifying style-types in a sample of musical improvisations using dimensional reduction and cluster analysis. *Psychology of Aesthetics, Creativity, and the Arts, 12*(1), 110–122.

Elmer, F. J. (2006). Musical Micro Analysis Tools: pitch analyser. [Computer Software]. Available from http://mmatools.sourceforge.net/.

Greenberg, J., & Folger, R. (1988). Experimenter bias. In *Controversial issues in social research methods. Springer series in social psychology* (pp. 121-138). New York, NY: Springer. doi.org/10.1007/978-1-4612-3756-3_7

Gudmundsdottir, H., & Cohen, A. J. (Eds.)]. (2015). The AIRS Test Battery of Singing Skills. Special issue. *Musicae Scientiae, 19,* 233-324.

Ilari, B., & Habibi, A. (2015). Favorite songs, melodic elements and a familiar tune: Performance of children from Brazil and the United States on components of the ATBSS. *Musicae Scientiae, 19*, 265–281. doi:10.1177/1029864915597566

Juslin, P. N. (2019). *Musical emotions explained.* Oxford, UK: Oxford University Press.

Juslin, P. N., Liljeström, S., Laukka, P., Västfjäll, D., & Lundqvist, L. (2011). Emotional reactions to music in a nationally representative sample of Swedish adults: Prevalence and causal influences. *Musicae Scientiae*, *15*, 174-207.

Kaiser, D. (2017). Using facial expressions of emotion as a means for studying music-induced emotions. *Psychomusicology: Music, Mind, & Brain*, *27*, 219-222.

Krumhansl, C. L. (2017) Listening niches across a century of popular music. *Frontiers in Psychology,* 8:431, 1-18, doi: 10.3389/psyg.2017.00431

LeBorgne, W. D., & Rosenberg, M. D. (2014). *The vocal athlete*. San Diego, CA: Plural Press.

MacDonald, G. R., & Cohen, A. J. (2013). The effects of personal song meaning, singer gender and musical training on spontaneous facial and head movement while singing. An exploratory investigation. Poster presented at the 23rd annual meeting of the Canadian Society for Brain, Behavior, & Cognitive Science, Calgary, AB.

MacLean, B. (2018). Learning and memory for popular music in two generations. Unpublished honours thesis. Department of Psychology, University of Prince Edward Island. http://www.islandscholar.ca/islandora/object/ir%3A22507?solr_nav%5Bid%5D=e8739f00ef1b505942f2&solr_nav%5Bpage%5D=0&solr_nav%5Boffset%5D=0

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Erlbaum.

MacWhinney, B. (2014). What we have learned. *Journal of Child Language, 41*, Supplement S1, 124-131.

Pan, B. Y., Liu, D., & Cohen, A. J. (2012). A system for developing a series of interactive tests of vocal production requiring on-line audiovisual recording. *Journal of the Acoustical Society of America*, *131(4)*, 3377-3377.

Raju, M., Välja, L., & Ross, J. (2015). Estonian children's improvisational songs, the nature of performance and songs' coherence with the Western tonal musical canon. *Musicae Scientiae, 19,* 282–300.

Ross, E. (2012). Singing across the lifespan: AIRS Test Battery of Singing Skills. An exploratory investigation of creativity through improvisation. Unpublished Honours theis, Department of Psychology, University of Prince Edward Island.

Russo, F. A., Sandstrom, G. M., & Maksimowski, M. (2011). Mouth versus eyes: Gaze fixation during perception of sung interval size. *Psychomusicology: Music,*

Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly Journal of Teaching and Learning, 1*, 81-85.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15,* 511–514. http://dx.doi.org/10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2019, May 20). Correlation = Causation? Music Training, Psychology, and Neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication.http://dx.doi.org/10.1037/aca0000263

Schirmer-Mokwa, K. L., Fard, P. R., Zamorano, A. M., Finkel, S., Birbaumer, N., & Kleber, B. A. (2015). Evidence for enhanced interoceptive accuracy in professional musicians. *Frontiers of Behavioral Neuroscience, 9*, 349. doi:10.3389/fnbeh.2015.00349

Sloboda, J. (1986). *The musical mind*. Oxford, UK: Oxford University Press.

Stadler Elmer, S., & Elmer, F. J. (2000). A new method for analyzing and representing singing. *Psychology of Music, 28*, 23-42.

Welch, G. F. (1986). A developmental view of children's singing. *British Journal of Music Education, 3*, 295-303.

Welch, G. F. (1998). Early childhood musical development. *Research Studies in Music Education, 11*, 27-41.

Weth, K., Raab, M., & Carbon, C.-C. (2015). Investigating emotional responses to self-selected sad music via self-report and automated facial analysis. *Musicae Scientiae, 19,* 412–432.

Wise, K. J. (2009). Understanding "tone deafness": A multi-componential analysis of perception, cognition, singing and self-perceptions in adults reporting musical difficulties. Doctoral Dissertation, Keele University. http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.502980.

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*, 3-26.

---

[1] The design of the ATBSS benefitted from input and feedback from researchers involved in the AIRS project during the initial years of the test's development between 2009 to 2012. For example, workshops focusing on the ATBSS were held at the AIRS Annual meetings in Charlottetown (2009), Seattle (2010), and St. John's Newfoundland (2011), consistent with the naming of the test battery "AIRS". The song *Brother John /Frère Jacques* figures prominently in the test as the example of a familiar song. The familiarity is justified by recent research of Belfi and Kacirek (2019) in their development of a famous melodies stimulus set. Its structural organization also recommended it.

[2] For example, the Sung Performance Battery (SPB) of Berkowska and Dalla Bella (2013) assesses the participant's vocal range, single-pitch matching, pitch-interval matching, novel-melody matching, singing familiar melodies with and without lyrics but does not explore singing a favorite song or improvisation. In the context of defining tone deafness, Wise (2009; Wise & Sloboda, 2008) developed measures for

---

adult singing proficiency and included a perceptual rating scale for use by expert judges, reflecting a similar scale developed for children by Welch (1986, 1998). The tests of Welch, along with that of Rutkowski (1990) assume an early dominance of speech in singing development. Consequently, Wise's perspective considers speech-like components.

[3] Measuring pitch is more complex than measuring the results of a dichotomous perceptual task involving a yes/no judgment. The sampling rate of the acoustic signal of the voice typically exceeds 20,000 times per second, and while computer algorithms such as Praat (Boersma & Weenink, 2013) can estimate pitch, it is still necessary to provide the algorithm with a starting and ending point for each note of interest, and look out for octave errors. The musically trained human ear is also a viable source for identification of tones that fit the prevailing musical scale framework, however, analysis by ear (with comparison to a fixed pitch reference scale such as a piano keyboard) is tedious and subjective.

[4] In an online questionnaire regarding knowledge of popular music of the last five decades, MacLean (2018) university students, unselected for music experience, reported listening to an average of 20.50 (*SD* = 24.28) hours per week (i.e., over 3 hours per day). Krumhansl (2017) collected data in an online survey regarding patterns of listening to music across the lifespan. Over 1900 respondents, regardless of age, pop and rock music was listened to most (as compared to classical, country and folk, jazz, among others).

[5] Conducted by Gillian MacDonald for her Honours thesis and presented at the Canadian Society for Brain Behavior and Cognitive Science, Calgary, 2015. See further [**CW27.5**] and for children by Bugos, DeMarie, Torres, Gbadamos, & Anderson (2019).

---

[6] Conducted by Erica Ross for her Honors thesis. See further [**CW27.6**].

[7] Data collected by Emily Gallant (2009) for her honours thesis entitled: Singing ability of persons with dementia.

[8] Due to the changes made at the site of the server company, the version of the ATBSS became non-functional in August 2017. The studies and data reported here encourage this redevelopment with new technologies.