Proceedings of Meetings on Acoustics

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Musical Acoustics

Session 5aMUb: Digital Libraries for Speech and Singing

5aMUb1.   A digital library to advance interdisciplinary research in singing

Annabel J. Cohen*, Coralie Vincent, Donald S. Moses, Ichiro Fujinaga, Nyssim Lefford, Theresa Leonard and George Tzanatakis

 *Corresponding author's address: Psychology, University of Prince Edward Island, 550 University Ave, Charlottetown, C1A3Z6, PE, Canada, acohen@upei.ca

  In 2008, at the ASA/EAA symposium honouring pioneering scientist of singing, Johan Sundberg, the AIRS (Advancing Interdisciplinary Research in Singing) project was introduced as a major collaborative research initiative on singing [Cohen, A.J. (2008) Acoustics 08 Paris, 3177-3182]. Over 70 collaborators around the world were to investigate singing from perspectives of development, education and well-being. A digital library was to facilitate distant team members' work on the same data, such as examples from voice studios around the world, performance stages, playgrounds, public places, solos, groups, classrooms, intergenerational or multicultural choirs, therapeutic settings or new tests of singing skills [Vincent, C., Lane, J. & Cohen, A. (2011) PEVOC9]. Plans also included tools for annotation and analysis along with relevant documents and images. The present progress report on this endeavor describes preliminary prototypes, stages of development, and the current functional implementation. It is noted that although singing is primarily an acoustic and auditory phenomenon, video records of the singer are highly valuable. Their benefit, however, must be weighed against challenges arising from ethical considerations and storage requirements. Issues of ownership and data sharing are also raised as are practical matters of choice of platform, storage, formats, backup, human resources, and long-term preservation.

AIRS RESEARCH IN SINGING AND THE NEED FOR A DIGITAL LIBRARY

Singing is a natural and powerful form of human communication that impacts lives in countless ways from birth to the grave. Yet singing has attracted relatively little research attention, in comparison to speech, the other primary form of human vocal communication (Cohen, 2011). There is of course a sizeable body of research on singing, as represented by pioneers such as Johan Sundberg (1987, 2013), Inge Titze (1994, 2011), Graham Welch (2005) and the researchers they have inspired; however, much remains unknown. In 2008, a proposal was put forward to the Social Sciences and Humanities Research Council of Canada to support a Major Collaborative Research Initiative that aimed to advance interdisciplinary research in singing, leading to "AIRS" as a fitting acronym for the megaproject entitled "Advancing Interdisciplinary Research in Singing" (Cohen, 2008). The research was to focus on three primary questions: How does singing develop in every human being? How should singing be taught and used to teach? and How does singing impact well-being? These questions were to be considered from three perspectives: human universals, culture influences, and individual idiosyncracies, all with the aim of developing a grand theory of how singing works and to use it as the basis for mobilizing knowledge about singing for the betterment of society (see Figure 1). The international project would involve researchers in 16 countries representing a range of disciplines (e.g., music, psychology, psychoacoustics, acoustics, education, sociology, computer science).
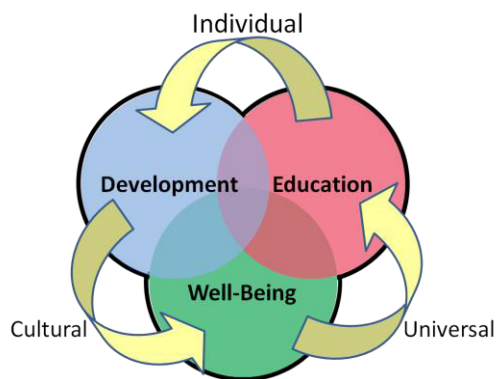

FIGURE 1. Framework for the three themes of AIRS, Development, Education, and Well-being) and the factors that influence them (Cultural, Universal, and Individual) [image courtesy Bing-Yi Pan].

Success for the AIRS project required collecting data in different locations where the researchers worked or travelled and sharing the data across the network, freeing from geographic restriction analysis by researchers who had the capability or interest, and then subsequently sharing the results. By 2008, many aspects of information technology had developed to the extent that a digital library could in theory meet this need. Hence, a key component for the proposed research was the creation of digital library to support the AIRS research on singing.

Required Functionality of the AIRS Digital Library

The AIRS digital library would have the following functionality:

1. store (unlimited number of) vocal recordings at potentially high resolution and for the long term—including examples derived from laboratory studies, solo (voice studio) and choral pedagogy, choirs, and singing in natural (school playgrounds) and therapeutic (homes for seniors) settings.
2. allow researchers to add information (annotation) or link to their data (e.g., acoustic) analyses so that researchers could build on each other's work and so that knowledge would accumulate rather than duplicate. A single example of singing can contain more information than any one researcher can analyze completely from all perspectives. By depositing new knowledge associated with examples, a multidisciplinary team can more effectively build understanding of the nature and significance of singing.
3. provide examples for educators and others;

(a) not only of singing and songs but also of the cultural context which gives meaning to those songs or through which the songs bring cultural understanding
(b) pedagogical examples of models (e.g., singing lesson within a voice studio) that can be used for serious voice study or voice pedagogy study (i.e., how to teach)
(c) materials for use in an intergenerational singing curriculum (e.g., in homes for seniors), or for developing international choirs
(d) singing exercises for various health disorders, e.g., lung diseases, Parkinson's disease

4. provide a place for rare, potentially otherwise perishable, cultural archives (e.g., native songs that exist only as oral tradition) that can be used for studies of song and cultural evolution (although there may be more appropriate ethnomusicological repositories)

5. Tools for analysis. Song and singing can be analyzed in many ways: acoustical, musical structure, linguistic, anthropological, psychological (audiovisual measurement of an individual or the group interaction). Specific procedures developed by one researcher or research group (e.g., Ness, Tzanetakis, & Biro, 2010) can be made available in one place for all researchers

6. Search features which depend on an appropriate well-designed metadata scheme and leading edge technology

DIGITAL LIBRARIES FOR SOUND, SPEECH AND MUSIC

There are increasing numbers of digital repositories for sound, speech, and music. The functionality depends on initial design issues, defining the kind of content and the way of describing it (metadata). Any library is useful only to the extent that its content is accessible. There are at least three relevant digital libraries that are remarkable for the accessibility of their data (to at least a significant group of users) and consequent ability to revolutionize research and scholarship in their specific fields.

CHILDES - Children`s Language Date Exchange System

The understanding of the acquisition of speech was greatly advanced as a result of Brian MacWhinney and Catherine Snow's co-founding of the Children's Language Data Exchange System (CHILDES) (MacWhinney, 2000). Approximately 100 researchers transcribed discourse with children in accordance with codes developed for CHILDES. This coding system for words and utterances, as well as for higher-level descriptors of data sets (age of speaker, gender, context) was designed to allow flexible searching of samples. The database has led to over 3000 published research papers.

VARIATIONS for Music

The University of Indiana embarked on a project in the late 1990's to provide online access to sound recordings from the library's collection. The aim was to "integrate a database of music information objects (text, images, scores, sound, and a catalogue) with a graphically oriented hypermedia user interface" (Dunn & Mayer, 1999, p. 12). Thus music students and scholars would have access to the variations of information that pertain to a particular musical object. The project proved enormously successful and continued and continues to expand with Variations 2 in 2005, and Variations 3 in 2009, that enabled other institutions to take advantage of University of Indiana's Digital Library software for music. In 2011 a project Variations on Video was funded as a multi-institutional effort to update the original Variations technology and add video capabilities. Proposed improvements of metadata processes aim to increase discovery capabilities of users and to experiment with providing end-users with the potential for enriching basic metadata.

Cornell Lab of Ornithology Macaulay Library

The Macaulay Library claims to be "the world's largest and oldest scientific archive of biodiversity audio and video recordings" (macaulaylibrary.org). Its original focus on birdsong eventually expanded to the sounds of all living species. Originating in the late 1920's, the first recordings were audiovisual and were made with new film technology. Eventually various audiotape recording techniques provided better archiving potential leading to an entirely audio collection fo the rest of the 20th century. In the 21st century, recognizing that audio is but one aspect of behavior, video regained interest at the same time that digitized audiovisual recording became feasible. The

library for sound now welcomes video recordings. It is noted that this largest sound repository has only 16 recordings of humans (all audio) compared to 8000 audio and 1500 video examples of sparrows.

The development and maintenance of each of these three repositories has entailed enormous commitments of human expertise and time; steady financial support from major granting agencies, foundations, and industry; application of technologies, in some cases state-of-the-art infrastructure; and dedicated space. Yet for all their strengths, none of these outstanding digital libraries for sound could readily serve all the functions required by a digital library that would support the AIRS international, multimodal, interdisciplinary research on singing. Yet the task is within the capability of current technology.

PROTOTYPES FOR THE AIRS DIGITAL LIBRARY

The AIRS project assumes that every research question associated with singing can find a place within its tri-partite research framework that focuses on development, education, and well-being with three further primary subcategories (see Table 1). All digital library prototypes have been based on the hierarchical structure of the AIRS Research Framework as below.

TABLE 1. Structure of AIRS Research into three primary themes each with three sub-themes.
A fourth primary research theme is that of the Digital Library.

Development	Education	Well-being
Perception and Production of Singing	Learning to Sing Naturally	Singing and Cross-cultural Understanding
Multimodal (audiovisual) aspects of Singing	Formal Teaching of Singing	Singing and Intergenerational Understanding
AIRS Test Battery of Singing Skills	Using Singing to Teach	Singing and Direct Benefits to Mental and Physical Health

Prototype AIRS DL 0.0

The first prototype of the AIRS Digital Library provided for a further breakdown within each of these nine above categories in terms of Experiments, Studies, Demonstrations, and Miscellaneous. Navigating to Experiments led to a further drop-down menu item that listed the particular experiments for which data existed. Navigation to an experiment with the AIRS Test Battery of Singing Skills (Cohen et al., 2009; Vincent, Lane & Cohen, 2011), for example, produced a further menu that revealed the code name for a participant, one of five sessions, and for each session of 11 different components. Each of the components revealed an audiovisual clip of a participant in the study performing a particular singing skill, such as creating a song, or singing a familiar song accurately, or learning a new song, or carrying out several language tasks. An ingest system was devised by the UPEI Robertson Library that enabled uploading each segment and the provision of searchable metadata. While the process of ingestion was slow, the primary problem of the data base was that it was not built for scale. Its great value was in the establishment of a working prototype, an example whereby audiovisual information from a study could be uploaded for sharing across a network of individuals.

Prototype AIRS DL 1.0

During the early years of the project, there was general discussion among potential depositors to and users of the digital library, however, a subsequent concerted effort on the part of Nyssim Lefford in 2012 aimed to determine the types of data that might be expected from the AIRS research project. To this end, Lefford developed a formal survey of needs and used assessments of researchers regarding data types, quantity, and format. With this information at hand, a new prototype was developed which represented a wider variety of data. The model was developed completely within the Drupal open-source-software content-management environment. This prototype primarily held text information, including descriptors of data sets and links to their owners (i.e., individual AIRS researchers). While AIRS DL1.0 represented a wide variety of singing research examples, the search features were limited. The prototype itself was highly accessible, available as part of the basic web-site of AIRS to those with a password. While still possible to search within the nine categories, the new model aimed to promote cross-linking and searching across categories.

Prototype AIRS DL 2.0

AIRS DL 2.0, developed later in 2012, improved permission features by enabling a group of researchers to share otherwise confidential data. DL 2.0 also added advanced search features, based on metadata options thought to reflect concepts important to users. While adhering to a particular metadata standard was discussed, the team chose to develop a number of custom content types based on needs described by the AIRS user community. These were: Bibliography, Data, Group (a collection of contextual content that can have a group of users associated with it), Interview, Literature, Popular media coverage, Presentation, Study, and Teaching material. Each content type was associated with a potentially unique set of metadata fields.[1] As an example, fields for the Bibliography content type are: Title, Digital Library collections, AIRS Theme, Researchers/authors, Publication date, Keywords, Description, Format, Technical description, and Permissions. Other content types share many field labels, but some like Data and Interview include extensive Subject/Country, and Authority lists. Figure 2 below shows an example of Teaching material content type which adds Format (e.g., video recordings), Associated Files (for file names), and Contexts: Performances and Contexts: Education settings.
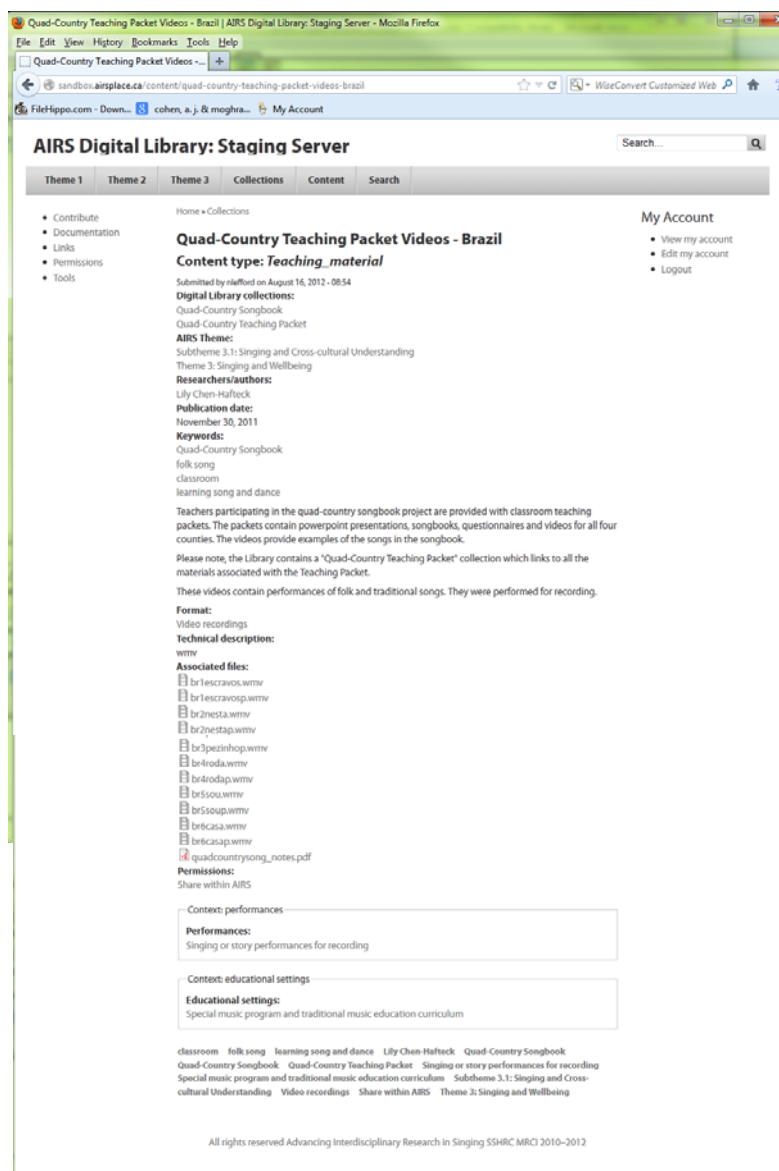

FIGURE 2. Metadata fields and metadata for one example of teaching materials including 12 videofiles and a pdf of explanatory notes. Not all possible metadata fields have been completed. [The image concatenates two overlapping screen shots]

Each of the content types have the same basic fields so that with some additional work, the current content could be crosswalked/migrated to a standardized format like Dublin Core should it at some time seem prudent to do so.

The particular page of metadata can be accessed from several routes. For example, a search on "Brazil" and "folk song" would lead to this collection "Quad-Country Teaching Packet Videos – Brazil". Figure 3 shows a screen shot from each of the last two video files (*br6escravos.wmv* and *br6escravosp.wmv*) represented in this above example. The left panel shows a screen shot from a video clip of children in Brazil singing the song *Casa de Farinha* (The Flour House) and on the right, one of the children pronouncing the words (without singing) for the first line "Mandai fazer uma casa de farinha", providing a model to enable teachers from outside Brazil to teach pupils this song from t the Brazilian culture. This video is part of a larger corpus of children's songs from three other countries (Canada, China, and Kenya) plus Brazil, for a study led by Lily Chen-Hafteck. For each of six songs from each country, a video of the children singing the song is matched by a second video provided by one child modeling the pronunciation of each word of the song. The microphone is visible in this example from Brazil.

FIGURE 3. A screen shot from an audiovisual example of Brazilian children singing (left panel) and an individual child modeling the pronunciation of the first line of the song [with kind permission of Alda de Jesus Oliveira, from the Quadcultural Songbook study under direction of Lily Chen-Hafteck].

A parallel to the Brazilian example above is shown below (see Figure 4) from the Kenyan collection where here the children are singing (and teaching) the song *Jamba*. Notably the original purpose for these materials goes beyond teaching the songs, and is rather to study how learning the songs of another culture may improve attitudes toward the other culture, as has been found in a social psychological study in Portugal by Sousa, Neto, and Mullet (2005) and in a qualitative music education study of Chen-Hafteck (2007). However, beyond the immediate purposes of this multinational study, these rich materials can provide useful information for studies of musical development, pitch accuracy, and rhythmic ability, for example. As mentioned the complete set of materials includes songs of two additional cultures, Canada and China, which are not shown.

FIGURE 4. A screen shot from an audiovisual example of children from Brazil singing (left panel) and an individual child modeling the pronunciation of the first line of the song [with kind permission of Elizabeth Andango, from the Quadcultural Songbook study under direction of Lily Chen-Hafteck].

The above examples of children singing their well known songs show how data of singing can serve a variety of research purposes. The acquisition of data on singing takes financial resources, human effort and cooperation of many people. Ingestion into a digital library can also be time and resource intensive. In the long term though, data in a functional digital library may ultimately serve hundreds of research or educational purposes adding value to all the resources spent to acquire the data.

THE FUTURE FOR THE AIRS DIGITAL LIBRARY FOR SINGING RESEARCH

Developing a digital repository for a singing has special challenges. One of them that may come as a surprise is that of the special issue of privacy. In Western culture, singing in front of another person can be a highly embarrassing activity. This means that individuals who sing in studies may not be comfortable in allowing their video, or even audio data, to be broadly shared. We can compare this situation to that of a mega-study being conducted in Australia called AusTalk (Burnham et al., 2011). The Big Australian Speech Corpus aims to comprise 1000 Australian English speakers representing the geographical and social diversity within Australia. Each speaker is to produce 3 1-hour sessions for a total of 3,000 hours of high quality audiovisual data. The project resembles that of AIRS in its emphasis on audiovisual data and high quality recording, however, all of the AusTalk data appear to represent a single protocol, so in this sense the metadata problem is much simpler. It is likely that permissions to share the data on speaking are less of a concern than for data on singing.

A second issue that arises in connection with rare vocal materials or performances is that of ownership and copyright. This problem is not unique to the AIRS project and is faced in ethnomusicological work. A third issue concerns quality control. At the recent AIRS 4th Annual meeting, Coralie Vincent, who has a background in sound and video recording, delivered a welcomed presentation on sound recording. She described the general problems encountered when recording singing, and provided guidelines for achieving audio standards (sampling rate of 48,000 Hz and bit-depth of 24 bits). It is noted that the standard sampling rate for the Cornell sound library is twice this recommendation. Additional technical issues concern issues of storage, as video data places far greater demands than audio, and audio is more demanding than text. Creating a digital library, such as AIRS, that is part of a larger institutional system, entails adopting the standards of that system and working within its constraints. Fortunately for AIRS, the opportunity to develop the DL prototypes 1.0 and 2.0 has been quite unconstrained through the provision by the university library of a standalone 'sandbox'.

 A promising direction for the future is that of adopting a new institutional repository system in development since 2009 by the Robertson Library of the University of Prince Edward Island. The scheme links Drupal (with which AIRS is highly familiar) and the powerful Fedora (or Flexible Extensible Digital Object Repository Architecture) digital asset management architecture. The combined resource is referred to as Islandora (islandora.ca; Leggott, 2009). The search capability of Fedora will have advantages over Drupal, and it can be seen that given the richness of data of singing, the many associations to a song and the many levels and kinds of analysis of a song, the optimization of search and discovery is a priority for a future AIRS digital library. As one example, in the AIRS Test Battery of Singing Skills, participants are asked to sing a favorite song. Assuming names of favourite songs are identified by the participant in the on-line query in the automated version of the test (Pan & Cohen, 2012) appropriate metadata and search would allow probing the entire AIRS data base for other examples of the same song title, examples that could arise in teaching situations, or situations of learning songs naturally in a playground. Consider, for example, if Jamba were the favourite song of several participants in the test battery, these could be found along with the example produced by the Kenyan children in an entirely different research project. Appropriate music analysis tools could complement the search (i.e., when audio records lacked an associated song title). For its potentially superior search capabilities and its association with leading edge digital library technology, the Islandora platform is attractive. The Islandora scenario would entail migration of the existing assets of AIRS DL 2.0 to Islandora, translating former ingesting procedures, and developing newer ones. The entire enterprise would entail considerable dedicated human resources to start.

The development of the ideal digital library to support research in singing would require the resources of the kind that have gone into other major repositories mentioned earlier (VARIATIONS, CHILDES, the Cornell/Macaulay sound collection).   Because of the richness of data of singing, a small but highly searchable library with annotation capacity, far exceeds having a large library of content that is inaccessible and fixed. AIRS looks to continuing progress in this very exciting age of rapid technological development in the areas of psychoacoustics, cognitive musicology, bioacoustics, speech perception and production, media technologies, and digital libraries. Sharing

experiences with other research groups that are developing voice and music repositories will be invaluable to making the best choices and decisions regarding the next rendition of the AIRS DL 3.0 to support singing research.

ENDNOTE

[1]The ingest process aims to capture a significant amount of information about each catalogue entry. This information pertains to authorship, the format of the content, the process of collection, date of creation and permissions, as well as, detailed descriptions of content itself— both the author's own descriptions and utilizing terminology familiar across the AIRS project. AIRS is comprised of researchers from many different disciplines, each with their own specialized vocabulary, the multiple descriptions facilitate data sharing and re-use (across areas of study). The ingest process is (presently) manual, and managed via a set of web forms. From a user standpoint, it is very easy to use. However, to date, to ensure consistency and security, ingest is performed by specialized library administrators who are familiar with the AIRS project, its research and its researchers. Whenever there is uncertainty about labels, categorizations, permissions, etc. the contributing researcher is contacted directly. Despite the extra care required, the system offers some excellent flexibility. A very wide range of file types may be ingested. Different types of files can be grouped and associated within a single catalogue entry; for example, an audio file, video file and tabular data may be bundled. However, the more granular the data, the more labor intensive to ingest. If those same three elements, the audio and video files and the tabular data file, are ingested as individual catalogue entries, the labor increases exponentially. Plus, the library administrator must make sure that the information for each entry is significantly distinct to facilitate searching. But this is possible and such content is currently available in the library.

ACKNOWLEDGMENTS

REFERENCES

Burnham, D., Estival, D., Fazio, S., Cox, F., Dale, R., Viethen, J., Cassidy, S., Epps, J., Togneri, R., Kinoshita, Y., Göcke, R., Arciuli, J., Onslow, M., Lewis, T., Butcher, A. & Hajek, J. (2011). Building an audio-visual corpus of Australian English: large corpus collection with an economical portable and replicable Black Box, *Proceedings of the 12th Annual Conference of the International Speech Communication Association Interspeech*, 2011, Florence, 28-31.

Chen-Hafteck, L. (2007). Contextual analyses of children's responses to an integrated Chinese music and culture experience. *Music Education Research, 9*, 337-353.

Cohen, A. J. (2008). Advancing interdisciplinary research in singing through a shared digital repository. Acoustics '08 Paris (2nd joint conference of the Acoustical Society of America and the Association of European Acoustics). Paris, June 29 – July 4. 3177-3181 (CD-ROM)

Cohen, A. J. (2011). Introduction to the special volume on the Psychology of Singing. *Psychomusicology: Music, Mind & Brain, 20*.

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.

Dunn, J. W., & Mayer, C. A. (1999). VARIATIONS: A digital music library system at Indiana University. DL '99: Proceedings of the Fourth ACM Conference on Digital Libraries, Berkeley, CA. 12 – 19.

Leggott, M. (2009). Islandora: a Drupal/Fedora Repository System https://smartech.gatech.edu/bitstream/handle/1853/28495/117-631-1-PB.pdf?sequence=2

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analysing Talk, Vol. 1: Transcription Format and Programs* (3rd ed). Mahweh, NJ: Erlbaum, *Vol. 2: The Database* (3rd ed.). Mahweh, NJ: Erlbaum.

Ness, S., G. Tzanetakis, D.P. Biro "Computer-assisted cantillation and chant research using content-aware web visualization tools" *Multimedia Tools and Applications, Springer Verlag* , 48(1), 2010

Pan, B-Y., & Cohen, A. J (2012). The AIRS On-line Test Battery of Singing Skills. [abstract of presentation at CSBBCS Annual Meeting]. *Canadian Journal of Experimental Psychology, 66*, 307-307.

Proutskova, P. (2008). Data infrastructure for ethnomusicological archives - current situation and future perspectives. *Journal of International Society of Sound- and Audiovisual Archives*, (31):45–54.

Sousa, M. D. R., Neto, F., & Mullet, E. (2005). Can music change ethnic attitudes among children? Psychology of Music, 33, 304-316.

Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica, 22*, 28-48.

Sundberg, J. (2013). Perception of singing. In D. Deutsch (Ed.) *The psychology of music*. Third Edition. (pp. 69-105). New York: Elsevier.

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

Titze, I. (2011). A tessitura-gram for singers. *Journal of the Acoustical Society of America, 129*, 2581-2581.

Vincent, C., Lane, J., & Cohen, A. J. (2011). Le projet AIRSde recherche sur l'acquisition de la voix. Chantée Research on the Acquisition of Singing : the AIRS Test Battery and Digital Library, 9th Pan European Voice Conference PEVOC9, Marseille.

Welch, G. F. (2005). Singing as communication. In D. Miell, R. MacDonald, & D. J. Hargreaves, (Eds). *Musical communication* (pp. 239-259). Oxford, UK: Oxford University Press.