SIMPLY ACAPELLA

Xinyue Ma

Acapella, a music format characterized by multiple vocal parts without instrumental accompaniments, has gained popularity on college and university campuses across the nation. Acapella ensembles typically rely on musically talented members to teach and guide the learning of new songs. In situations where sheet music is unavailable, leaders must manually distinguish each vocal part by ear, transcribe it, and record it for other members to learn. This process poses challenges, as it requires expertise in music notation and can limit individuals without formal musical training from exploring their interest in acapella.

The Journey Begins. My journey with acapella traces back to 2018 when I, an amateur vocalist in a singing club, received an invitation from an acapella ensemble to join them as their alto singer. It was a new semester, and they was to rehearse Pentatonix's "Royals" for their upcoming final concert. These ensemble members were close friends from my singing club days, so I eagerly accepted the opportunity, taking my first step into the captivating world of acapella.

The Beauty of Acapella. However, my involvement in acapella music challenged me to shift my focus towards the voices of my fellow singers and appreciate the profound significance of harmony in music. The success of our performance of "Royals" solidified our team, and we continued to rehearse and explore various acapella songs in different styles. The impact of this acapella experience is truly unforgettable. Nonetheless, an issue that arose during our practice sessions caught my attention.

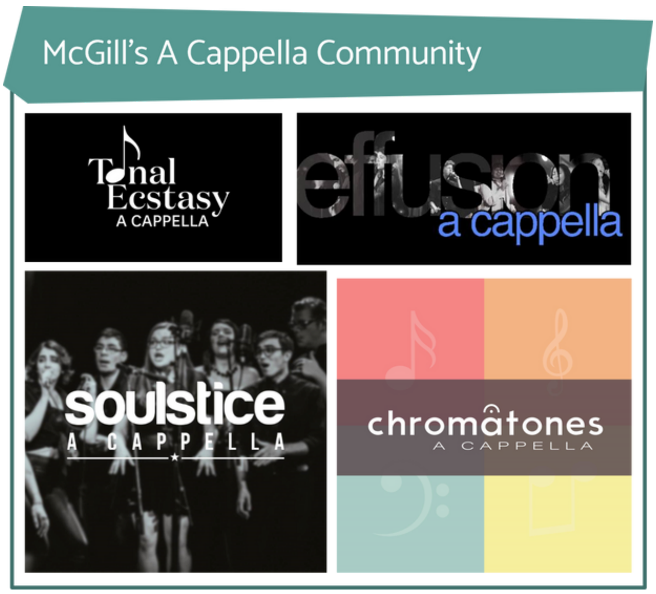

Figure 1. Four acapella clubs at McGill

The Problem that Sparked Curiosity. Not all members in every acapella ensemble, including myself, have received formal music training. Particularly in collegiate acapella groups, participants often rely on a knowledgeable and skilled member to act as the leader, guiding rehearsals and teaching new songs. In the past, we utilized a mobile app called MuseScore, which could play individual vocal parts if the sheet music was available.

However, it was a common occurrence to be captivated by a vocal track or YouTube video of an acapella song for which no corresponding sheet music existed. In these situations, the leader had the arduous task of discerning each vocal part by ear, transcribing it, and recording it for the rest of the ensemble to learn. This process proved to be challenging and burdensome. As we understand, music is comprised of sound waves that transform energy, and each wave's frequency corresponds to its pitch. Acapella songs, being a combination of various pitches over time, add complexity to the transcription process (see Figure 2). Traditionally, transcribing music involved transforming it into the frequency domain to identify the pitches, a task that can be simplified through computation. Every time we had to abandon an intriguing piece due to the lack of sheet music, I pondered the existence of a toolkit that could automatically separate and transcribe each vocal part of any acapella song. Despite conducting a thorough search, I found no such solution available. This motivated me to embark on the development of my own toolkit, leading me to undertake the project at Building 21.

A Solution Takes Shape. The objective of this project is to develop an audio source separation toolkit capable of accurately transcribing and separating each vocal part from any acapella song. This toolkit caters to both music enthusiasts and experts, while also holding potential interest for researchers in the field. The project implementation is divided into two main components: polyphonic separation and audio transcription.

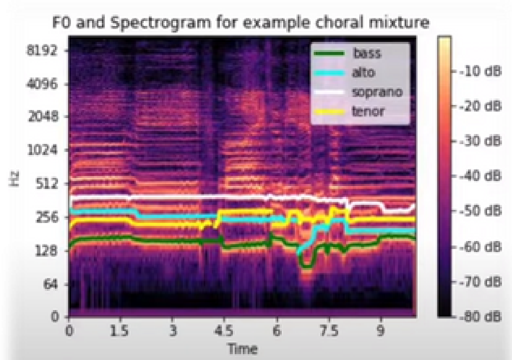

Figure 2. Spectrogram of four-part acapella song with fundamental pitches of is constituent parts

Polyphonic separation involves extracting individual vocal signals from an acapella song. This task is accomplished using Asteroid, a PyTorch-based audio source separation toolkit. Asteroid offers pre-defined neural network models for speech separation, allowing researchers to utilize their pre-trained models or train their own models using custom datasets.

In my project, I adapted two specific models: the fully convolutional time-domain audio separation network (ConvTasNet) and the dual-path transformer network (DPTNet) (see Figure 3). These models were trained using publicly available datasets comprising 26 songs from Bach chorales and 22 songs from Barbershop Quartets. Each song contains recordings of the four vocal parts: Soprano, Alto, Tenor, and Bass (SATB). Barbershop Quartets feature all-male voices, while Bach chorales comprise two female and two male voices. The output of the models provides separated vocal parts in the "wav" format. While the separated recordings serve as a valuable learning tool for those who learn music by ear, it is important to consider individuals who rely on sheet music. To address this, I took another step to transcribe the individual vocal recordings into MIDI format, a standardized file format for music transmission and storage. The audio transcription was accomplished using BasicPitch, a lightweight neural network library for automatic music transcription developed by Spotify's Audio Intelligence Lab. BasicPitch automatically transcribes the music into MIDI files, further enhancing accessibility and enabling learners to follow along with sheet music.

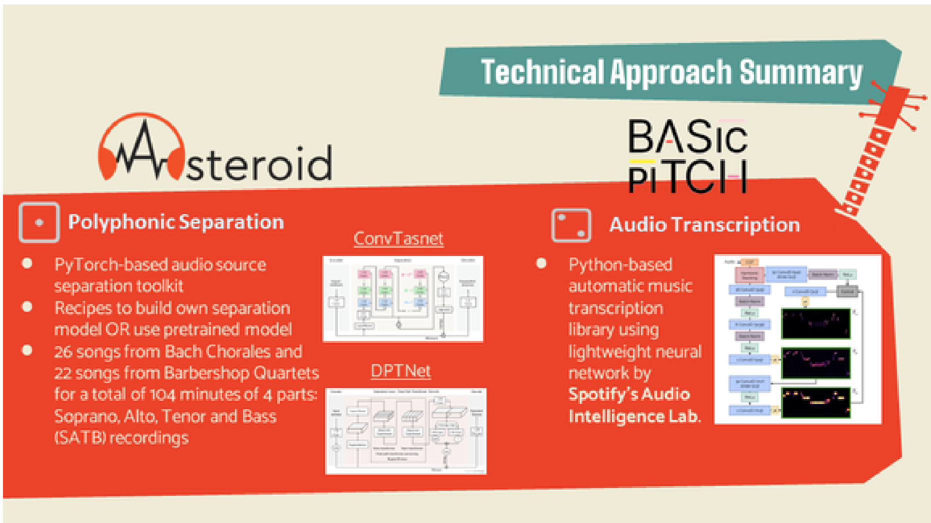

Figure 3. Descriptions of the tasks and the corresponding python-based library.

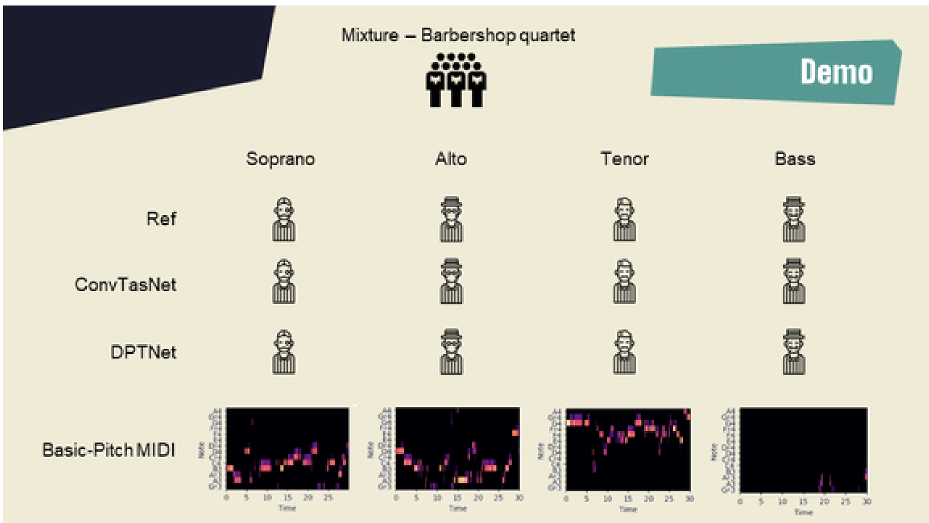

Figure 4. An example implementation of vocal part separation of a barbershop quartet.

Conclusion and Perspectives. This result is promising; however, there is still a long way to go before achieving the ultimate goal of benefiting every acapella community. Firstly, the performance of the polyphonic separation models needs to be tested on acapella songs beyond barbershop quartets and Bach chorales. With additional time and resources, I can train models to accommodate other music formats as well. Moreover, it is necessary to package the code as a mobile application, ensuring intuitive usage and inviting individuals of all expertise levels to explore and learn acapella. This will further expand the reach and accessibility of this captivating musical art form.

One of the memorable experiences I had at Building 21 was participating in the explorative workshops hosted by fellow scholars. These workshops covered a wide range of topics, spanning from archeology to AI chatbot. They provided a valuable opportunity for the benefit of the presenter and their project, and the format is diverse: background surveys for their projects; behavioral testing for self-exploration; cross-cultural and interdisciplinary discussions, and a inclusive platform to express their own viewpoints. During my research on music theory, I stumbled upon an intriguing subject: audio illusions. What fascinated me was the fact that individuals exhibit varying levels of sensitivity and even contrasting responses to the same audio stimuli. Eager to delve deeper into this phenomenon, I decided to involve others and embark on an exploration of these captivating audio illusions.

One of the audio illusions we explored is known as the tritone paradox. We played different tritones in ascending and descending order, and participants were asked to report the order they perceived. The intriguing aspect of this illusion is that certain pairs of notes can elicit different pitch perceptions in individuals. Some people hear an ascending pattern, while others, when listening to the exact same pair of tones, hear a descending pattern instead. This experience can be particularly astonishing for musicians who are confident in their judgments but completely disagree on whether the tones are moving up or down in pitch. Furthermore, these perceptions can vary among individuals from different countries and with different native languages. We collected responses from all participants who played the game. The majority of participants accurately reported the correct tone orders (Figure 5, top), but we were fascinated to find one individual with the opposite perception (Figure 5, middle), and another who provided a completely different and confusing report (Figure 5, bottom).

iv) Creative impression

Figure 5. Three examples of perceived tritone orders from Building 21 members.