HARMONY IN ISOLATION: A REAL-TIME VOCAL HARMONIZER

IS BIOMETRIC VOICEPRINT COLLECTION AND USE SECURE?

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Electrical Engineering

By
Laura Gustad

November 2, 2020

Team Members:
Nate Hunter, Noah Mills

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS
Catherine D. Baritaud, Department of Engineering and Society
Harry Powell, Department of Electrical and Computer Engineering

A voice is one of the most important tools we have access to. We use our voices to talk, to express ourselves, and to sing. The technical project at hand is to create a tool to enhance a person's singing voice and create harmonies in real time. However, voices can also be used as identifiable personal footprints. While some devices that utilize voice inputs seem entertaining and harmless, the frequency identification methods are similar to those in devices that can isolate and store each person's unique voiceprint. The *Gale Encyclopedia of E-Commerce* defines voiceprint as a behavioral biometric consisting of the unique pattern of pitch, dynamics, and underlying characteristics that can be obtained from a voice recording ("Biometrics", 2002). The use of voice recognition and voice-control is continually increasing in the Internet of Things, especially via voice assistants, in-home devices, and authentication measures. Devices that are continually listening in or that use linguistics recognition to identify voice commands have access to the owner's voiceprint, which is a unique, measurable identifier. Due to the adoption of voiceprint as an authentication method, a breach of this information creates a major risk of identity theft.

As a singer, musician, and electrical engineer, I was drawn to technical and STS projects that incorporate music and sound with signal processing. The vocal harmonizer proposes the creation of a device to assist musicians in times of distance learning by applying hardware and software to make one voice sound like many. On the other end of the spectrum, the STS focus explores the data privacy legislation that regulates the collection and use of an individual's voiceprint in technologies ranging from banking to voice assistants. The technical project is tightly coupled with the STS topic because the vocal harmonizer creates, but does not save, voiceprint identifiable data.

The timeline for the work on both the STS and technical portions of the thesis are shown in Figure 1, with STS tasks highlighted in blue and technical tasks highlighted in red. The STS tasks will be completed during STS 4500 and STS 4600 under the guidance of Professor Catherine Baritaud in the Department of Engineering and Society. The technical project work will be designed, built, and tested in ECE 4440 with Professor Harry Powell and Professor Adam Barnes in the Department of Electrical and Computer Engineering.

Semester 1	08/25-11/02	11/02-11/24	11/24-12/4
Design, build, and test vocal harmonizer			
Write IEEE conference paper			
Research and write Prospectus			
Revise Prospectus			
Semester 2	02/1-03/25	03/25-05/04	
Research a write STS scholarly article			
Make any necessary changes to thesis before it is bound			

Figure 1: Timeline of STS and Technical Projects. The table divides the STS tasks (blue) and technical tasks (red) by semester, highlighting the dates for each task to be completed in yellow (Gustad, 2020).

HARMONY IN ISOLATION: A REAL-TIME VOCAL HARMONIZER

Our project to create a vocal harmonizer came from the realization that during the COVID-19 pandemic, musical groups are no longer able to safely practice or perform together. The merits for this restriction were further verified through research by Helding et al. (2020) which revealed that singing produces aerosolized particles too small to be stopped by masks (para. 10). Our proposed device synthesizes vocal audio from a microphone with chord input from a keyboard. The user will be able to play the desired chord on the keyboard while singing any note into the microphone, and the speaker will output the voice reharmonized into the chosen chord. This device may be assistive in musical education and will introduce a sense of togetherness to musical groups, even in this time when singers are forced to perform their music alone. As the user can sing any note, not necessarily one of the notes in the desired chord, the vocal harmonizer device would eliminate the limitations of individual vocal range and assist in the vocal experimentation needed to learn to recognize harmony by ear (Spies, 2015, p. 168).

This project will mostly serve as a learning exercise for our group, but we believe that the idea could be put to positive use in today's difficult environment. This project involves collaboration, technical skills, and expertise with specific tools. From this project, we all hope to gain a better understanding of digital signal processing, printed circuit board (PCB) design and construction, embedded systems analysis, and programming microcontrollers.

HARDWARE AND SOFTWARE DESIGN

To implement our harmonizer, our hardware design approach begins with an audio signal input via the microphone. This signal travels through an anti-aliasing filter on the PCB to minimize noise in the signal. The anti-aliasing filter was designed to have a breakpoint frequency of 22 kHz, half the sampling frequency of 44.1 kHz, according to filter design rules (Engineer Ambitiously, 2019, para. 1). Then, the filtered audio is inputted into the codec, which converts it to a 16-bit digital signal. The digitized voice signal is then fed from the codec on the PCB into the myRIO embedded device via serial communication. The keyboard audio input is digitized and communicated to the myRIO by the same process.

Digital signal processing occurs on the myRIO as follows. The keyboard signal is translated into a set of frequencies representing the chord that is played, using a Fast Fourier Transform (FFT). The FFT resolution will be 4096 bins with 10.8 Hz separation, which minimizes the time needed to identify the pitch using Gaussian interpolation, while keeping the frequency resolution at an acceptable level (Gasior & Gonzalez, 2004, pp. 5-6). The voice signal is translated from the time domain to the frequency domain using an FFT. This frequency representation is used to identify the fundamental frequency of the voice signal. The frequency-domain voice signal is then shifted from its fundamental frequency center to each of the keyboard frequencies in the set. Then, the individual signals for each keyboard note are added together into a unified, frequency-domain, chordal output signal. Then, the summed signal is converted back to the time domain with an Inverse Fast Fourier Transform (IFFT).

The digital output signal is sent from the myRIO back to the codec via serial communication. The codec converts the signal back to analog. The analog signal may need to be buffered on the PCB before going to the piano amp. Finally, the amplified output signal is fed into a speaker, which plays the desired output audio. The block diagram representing the flow of the audio signal within the system is shown in Figure 2.

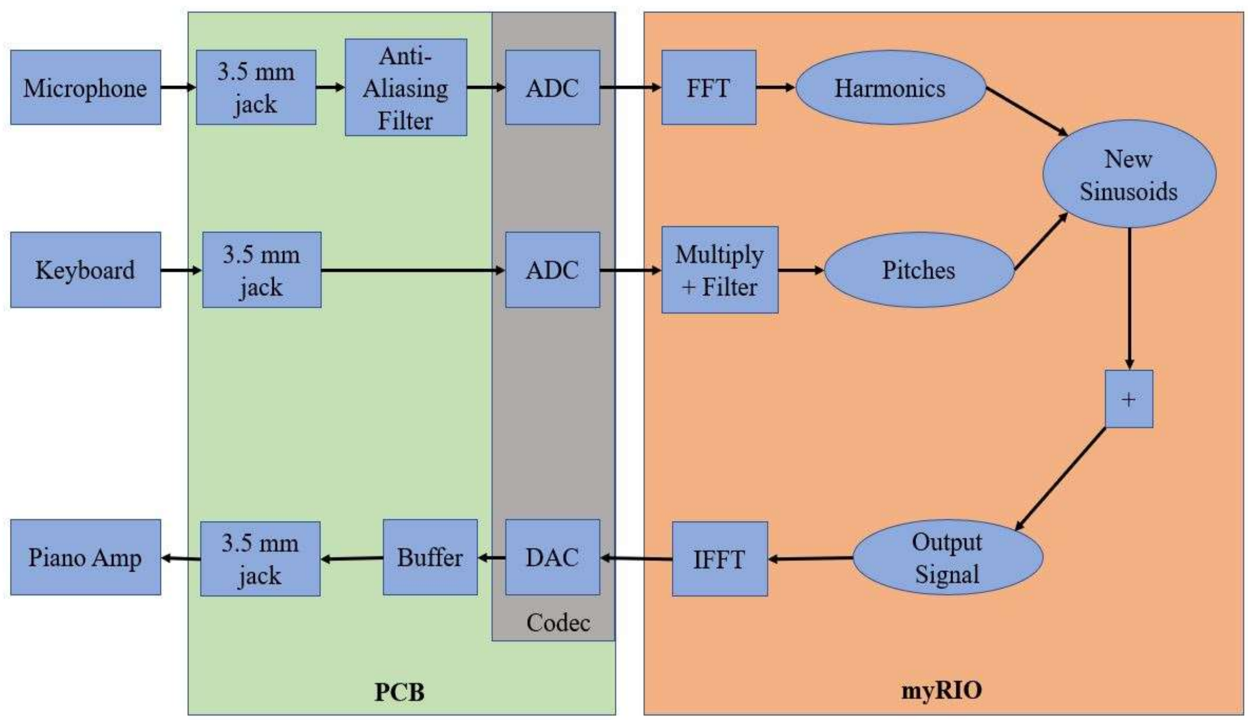

Figure 2: Block Diagram of Vocal Harmonizer. Full block diagram illustrating the signal path from the inputs through the PCB and myRIO to the piano amp (Mills, 2020).

As of November 2020, the overall hardware structure is fairly well developed. The first iteration of the PCB has been printed and tested, and the final draft of the PCB will be ordered on November 6, 2020. The software development in LabView on the myRIO is underway (National Instruments, 2019). The exact functions that will be used in the digital signal processing on the myRIO are yet to be determined.

EXPECTED OUTCOMES, TIMELINE, AND BUDGET

To complete this project, we have designed circuitry to connect our user inputs and outputs to our myRIO processing unit. Using Multisim, we were able to design the system and perform tests to confirm that it should work as we planned. The $500 budget for the project was provided by the School of Engineering and Applied Sciences. After we assemble the system, we will do the majority of testing in the National Instruments (NI) laboratory in the C-Wing of Thornton Hall. The NI laboratory provides the soldering equipment for connecting components to the PCB and the VirtualBench equipment needed to test signals and voltage readings. Testing the vocal harmonizer together in the lab will allow us to collaborate and measure the effectiveness of the system with different voice ranges in a controlled environment.

Due to this semester's unique schedule, we plan to have this project completed no later than the end of November 2020 and the corresponding conference paper written by the exam date in December. The technical project will be completed together with team members Nate Hunter and Noah Mills, and our technical advisors are Professor Harry Powell and Professor Adam Barnes in the Electrical and Computer Engineering Department.

IS BIOMETRIC VOICEPRINT COLLECTION AND USE SECURE?

With the rise of the Internet of Things (IoT) and remote access, voice control and biometric authentication have become the buzz terms for new home technologies and devices. In fact, a Visa (2020) survey about consumer opinion on biometric authentication revealed 72% interest in fingerprint identification, 45% in eye scan, and 32% in voice recognition (slide 4). The graphic in Figure 3 illustrates the different types of biometric identifiers currently being used in identification technology.

Figure 3: Types of Biometric Identifiers. The main biometric identifiers, from facial recognition to vein patterns, are divided into physiological or behavioral categories, based on their consistency level (Thales Group, 2020, Biometrics: trends section).

As voiceprint is the only biometric authentication method that can be used via a remote channel, it is the preferred method of confirming identity in security for applications from remote banking to connected devices for the home (Wood, 2020, para. 2). Regarding the popularity of home devices that utilize voiceprint identification, Ammari, Kaye, Tsai, and Bentley (2019) found that "45% of Americans use digital assistants," whether through free-standing IoT devices or the voice assistants on their phones (17:2). A complication with increased use of connected devices and biometric authentication is the privacy concern that arises with the simultaneous, increased access to biometric identifiers. In 2017, research by Fairhurst, Li, and Da Costa-Abreu illuminated the increased vulnerability of identifiable information due to the sale of consumer data to allow companies to predict a potential customer's gender, age, and more based on biometrics (pp. 3-5). As the collection of voice data grows, the risk that our identifiable, biometric data could be vulnerable to unlawful collection, use, storage, or leak also increases.

PRIVACY REGULATION CHANGES WITH THE ADOPTION OF BIOMETRIC AUTHENTICATION

At a glance, the outdatedness of biometric data protections in the United States is on display as federal law specifically protects only data related to healthcare information, children under the age of thirteen, and financial or credit information (Tschider, 2018, pp. 122-126). Tschider (2018) defines the three factors of the consent myth as “(1) individuals have meaningful choice with respect to the privacy notice, (2) an individual reasonably should and can invest time in reviewing a privacy notice as part of a contractual bargain, and (3) individuals can understand what privacy notices mean in terms of real-life impact” (p. 112). The reliance of United States data protections on the factors of the consent myth without regulations that require companies’ adherence in practice often leaves consumers without a full understanding of the consent being given.

However, some states have created more comprehensive data privacy provisions that better enforce consent standards for data collection. As outlined in Prescott’s (2020) article in *The National Law Review,* Illinois enacted the Biometric Information Privacy Act (BIPA) in 2008, followed by Texas’ Capture or Use of Biometric Identifier Act, Washington’s HB 1493, and California’s Consumer Privacy Act in 2009, 2017, and 2020 respectively (paras. 4-11). Each of these statutes, protects consumer consent requirements and sets permitted durations for storage of consumer biometric data.

CLASS ACTION AND THE ACTOR NETWORK ANALYSIS OF BIOMETRICS

The Actor-Network Theory framework, first implemented by Callon (1984) and Latour (2005), is often used for analyses seeking to find a balance of power between actors surrounding a technology. Latour (2005), enumerates the requirements for a system analysis to be Actor-Network Theory as: agency assigned to non-human actors, a social environment that continues to change, and the goal of reconstructing the social landscape (pp. 10-11). In the STS analysis, voiceprint biometric technology and government regulation will both be assigned agency as actors, and the social network perspectives change with each new device or data breach. As Callon (1984) explains, the actor-network framework allows for analysis of not only each actor's interaction with the technology, but also the "interessement," identities imposed by actors on each other, which will help to identify the unique web of social forces between the actors in the network of voiceprint biometrics technology (p.207).

When the Actor-Network framework is applied to voiceprint biometrics identification, the main actors can be defined as businesses, engineers, hackers, users, and government and their interactions can be mapped as shown in Figure 4. The hackers find new spoofing methods, which forces the engineers to advance encryption defense and data storage techniques. Engineers provide the biometric identification technology and algorithms for businesses to create new connected devices, either through industry-independent research or company-funded design. Businesses use predictive algorithms to target consumers, and the consumer response to products, which product lines receive more attention, determines direction of market trends. Finally, government regulations protect specific consumer data and require businesses to notify consumers of a breach of their information. But in this Actor-Network model, the user impact on government regulation requires further investigation.

In a review of states' consumer data privacy legislation, Rosenthal and Oberly (2020) pinpoint as the strictest state biometric privacy law the Illinois BIPA, the only law that includes a private right of action for consumers (para. 23). The private right of action allows individuals to file a civil suit when their biometric privacy has been violated, when an entity is collecting biometric data without written consent or is storing that data for a nondisclosed amount of time. Through such legal action, individuals can influence data practices and how regulations are changed. Without the opportunity for citizens to be involved in pushing for regulation or amendment to existing legislation, the Actor-Network theory model of biometric identification has a missing interaction, as shown in Figure 4.

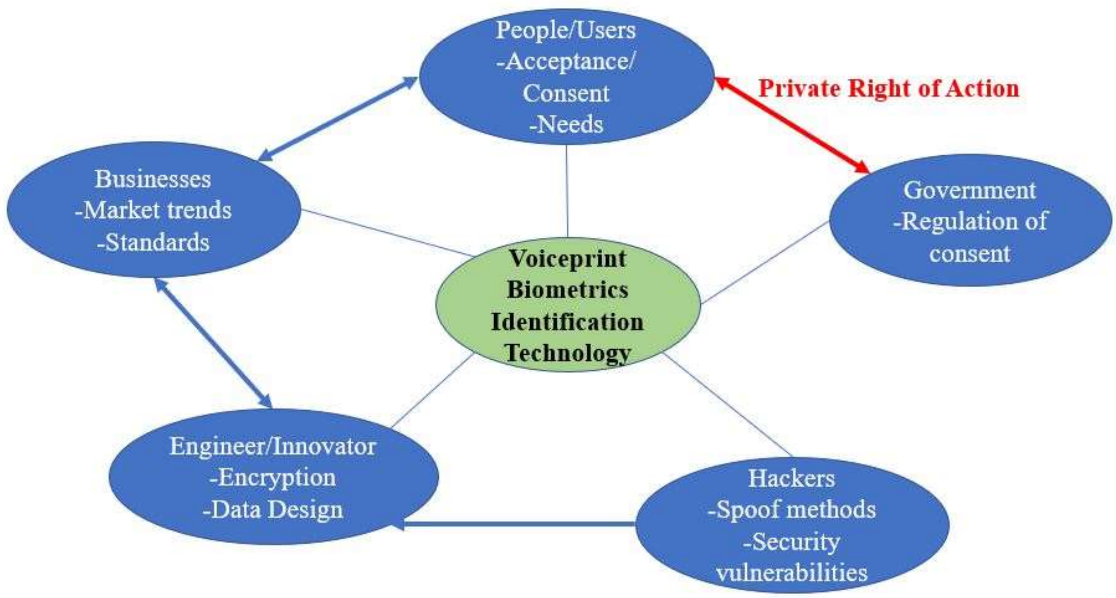

Figure 4: Actor-Network of Voiceprint Biometric Identification Technologies. The actors in the network of biometric identification are shown to interact outside of their main connection to biometrics; the red arrow highlights the often-missing connection between user citizens and protective legislation development (Adapted by Gustad, 2020, from Carlson, 2007).

Individual participation in government regulation of biometric identification technology has already led to important precedent in business standards in Illinois. For example, the case of *Rosenbach v. Six Flags Entertainment Corp.* held that “plaintiffs can pursue Illinois Biometric Information Privacy Act claims even in the absence of an actual harm” (Rosenthal & Oberly, 2020, para. 8). The judicial implication of BIPA has already impacted the platform for individuals to speak out against company data practices in court. Without a meaningful opt-out choice besides not using certain devices, individual users need to be involved in the network that decides the restrictions on the use of their biometric identifiers, whether through private law-suits or other means.

The STS research portion of this thesis project will be focused on the interactions between user perceptions of biometric privacy, business use of biometric data, and the regulation of data privacy, specifically exploring the views on inclusion of the private right of action. The STS discussion will explore the existing regulations and current practices of voice data collection, consent, and use practices and consider how to improve the ethical considerations in the development and regulation of devices capable of collecting, storing, and using voice and biometric data. The STS paper will be completed under the guidance of Catherine Baritaud in the Department of Engineering and Society.

UNDERSTANDING ETHICAL IMPLICATIONS OF VOICEPRINT AUTHENTICATION

In efforts to fully understand a technology's impact, the ethical implication of its adoption is as important as the technical merits of its design. Creating a real-time harmonizer from research and design through to implementation will not only give me experience in the technical façade of research, but will also provide insight into the larger field and a better understanding of the STS topic. Developing a device with a simple level of manipulation of a voice input, the vocal harmonizer, will help me appreciate the depth of knowledge companies are able to gain with more sophisticated devices and AI algorithms, which is critical for considering the ethical concerns of those devices. The technical and STS projects are tightly coupled within the context of voiceprint authentication and will be used in tandem to view the full spectrum of voice data collection.

REFERENCES

Ammari, T., Kaye, J., Tsai, J. & Bentley, F. (2019). *Music, search, and IoT: How people (really) use voice assistants*. ACM Transactions on Computer-Human Interaction. 26. 1-28. doi: https://doi.org/10.1145/3311956

Biometrics. (2002). In J. A. Malonis (Ed.), *Gale Encyclopedia of E-Commerce* (Vol. 1, pp. 64-65). Gale. https://link.gale.com/apps/doc/CX3405300051/UHIC?u=viva_uva&sid=UHIC&xid=1fbd7b9f

Callon, M. (1984). Some elements of a sociology of translation: domestication of the scallops and the fisherman of St Brieuc Bay. *The Sociological Review*, *32*(1), 196-233. doi: https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Engineer Ambitionsly. (2019). *Anti-aliasing filters and their usage explained.* Retrieved from https://www.ni.com/en-us/innovations/white-papers/18/anti-aliasing-filters-and-their-usage-explained.html

Fairhurst, M., Li, C., Da Costa-Abreu, M. (2017, November). Predictive biometrics: a review and analysis of predicting personal characteristics from biometric data. *IET Biometrics,* 6(6), 369-378. doi: 10.1049/iet-bmt.2016.0169

Gasior, M. & Gonzalez, J. (2004). Improving FFT frequency measurement resolution by parabolic and gaussian spectrum interpolation. *AIP Conference Proceedings*, 732 (1). doi: https://doi.org/10.1063/1.1831158

Gustad, L. (2020). *Actor Network of Voiceprint Biometric Identification Technologies* [Figure 4]. *Prospectus* (unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Gustad, L. (2020). *Timeline of STS and Technical* Projects [Figure 1]. *Prospectus* (unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Helding, L., Carroll, T. L., Nix, J., Johns, M., LeBorgne, W., & Meyer, D. (2020). COVID-19 after effects: Concerns for singers. *Journal of Voice*. doi: https://doi.org/10.1016/j.voice.2020.07.032

Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford, NY: Oxford University Press.

Mills, N. (2020) *Block Diagram of Vocal Harmonizer* [Figure 2]. *Prospectus* (unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

National Instruments (2019). *LabView* (19.0) [Computer software]. https://www.ni.com/en-us/shop/labview.html

Prescott, N. (2020, January 15). The anatomy of biometric laws: What U.S. companies need to know in 2020. *The National Law Review.* Retrieved from https://www.natlawreview.com/

Rosenthal, J. & Oberly, D. (2020, February). Biometric privacy in 2020: The current legal landscape. *Law360.* Retrieved from https://www.blankrome.com/

Spies, B. (2015, February 1). Introducing music students to harmony – an alternative method. *Education As Change*, 19(1), 165 - 187.

Thales Group. (2020, September). *Biometrics: definition, trends, use cases, laws and latest*. Thales: Building a future we can all trust. https://www.thalesgroup.com/en/markets /digital-identity-and-security/government/inspired/biometrics

Tschider, C. (2018). Regulating the internet of things: Discrimination, privacy, and cybersecurity in the artificial intelligence age. *Denver Law Review*, *96*(1), 87–143.

VISA (2017). *Goddbye, passwords. Hello, biometrics* [Presentation slides]. Retrieved from https://usa.visa.com/dam/VCOM/global/visa-everywhere/documents/visa-biometrics-payments-study.pdf

Wood, L. (2019, December). *Worldwide voice biometrics markets, forecast to 2024- Emerging lucrative opportunities within the BFSI vertical.* Business Wire. https://www.businesswire.com/