Sounds of Health: Using Personalized Sonification Models to Communicate Health Information

MATTHEW CLARK, University of Virginia, USA
AFSANEH DORYAB, University of Virginia, USA

This paper explores the feasibility of using sonification in delivering and communicating health and wellness status on personal devices. Ambient displays have proven to inform users of their health and wellness and help them to make healthier decisions, yet, little technology provides health assessments through sounds, which can be even more pervasive than visual displays. We developed a method to generate music from user preferences and evaluated it in a two-step user study. In the first step, we acquired general healthiness impressions from each user. In the second step, we generated customized melodies from music preferences in the first step to capture participants' perceived healthiness of those melodies. We deployed our surveys for 55 participants to complete on their own over 31 days. We analyzed the data to understand commonalities and differences in users' perceptions of music as an expression of health. Our findings show the existence of clear associations between perceived healthiness and different music features. We provide useful insights into how different musical features impact the perceived healthiness of music, how perceptions of healthiness vary between users, what trends exist between users' impressions, and what influences (or does not influence) a user's perception of healthiness in a melody. Overall, our results indicate validity in presenting health data through personalized music models. The findings can inform the design of behavior management applications on personal and ubiquitous devices.

CCS Concepts: • **Human-centered computing → Auditory feedback**; **Ubiquitous and mobile computing design and evaluation methods**.

Additional Key Words and Phrases: Sonification, ambient displays, behavior change, personalized healthcare, mobile health, music

ACM Reference Format:
Matthew Clark and Afsaneh Doryab. 2022. Sounds of Health: Using Personalized Sonification Models to Communicate Health Information. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 206 (December 2022), 31 pages. https://doi.org/10.1145/3570346

1 INTRODUCTION

With the advancement of wearable technology in recent years, digitized and personalized healthcare has increased in popularity [26]. As popularity increases, the need for a wider variety of presentation methods to appeal to the increasing user pool also increases. A popular method to portray information is through ambient displays [4, 7, 8, 14, 16, 17, 22, 23, 28, 44]. Ambient displays provide users information through a metaphor they frequently encounter, making the data easily accessible and understandable. Often, these displays provide an analysis of physiological sensor data, offering an easily evaluated depiction of health such as ratings or goal progress [7, 8, 16, 23]. These displays have been proven to successfully alter the behavior of users, encouraging them toward healthier decisions.

Authors' addresses: Matthew Clark, wdd6ts@virginia.edu, University of Virginia, Charlottesville, Virginia, USA; Afsaneh Doryab, ad4ks@virginia.edu, University of Virginia, Charlottesville, Virginia, USA.

2474-9567/2022/12-ART206
https://doi.org/10.1145/3570346

Despite the success of ambient displays in invoking healthier behaviors, there is a noticeable gap in the literature on using automated audio systems to deliver analyzed health information and incite behavior change ambiently. This gap is particularly surprising, as music's ability to spark behavior change in individuals is well-documented [9, 13, 27]. In business settings, marketers frequently use changes in audio to invoke specific emotions, influencing customers into making specific shopping decisions [27]. In a health setting, music therapy is a widely accepted form of psychiatric treatment, where clinicians use music and sounds to influence patients into better psychological well-being. Studies have even found music therapy more effective at treating depression than traditional psychotherapy treatments [13]. As such, creating auditory wellness ambient displays may engage users differently than visual displays, as well as engage users that may be unable to receive information through visual cues, such as the visually impaired. So, the question arises, if music is so effective at invoking behavior change, why has it been largely overlooked in systems that use health data in ambient methods to incite healthy behaviors?

The gap is not due to a lack of technical capability in presenting data as music. The field of sonification has made great strides in encoding various types of data into sound. Sonification has found intuitive uses, like aiding in the navigation of the visually impaired [5, 18, 19, 24] and embedding subtle notifications ambiently within music while accounting for music style [2] to highly complex and unintuitive uses, such as demonstrating crime statistics against women in India [31] and aiding the visually impaired in interacting with visual artwork [39]. Sonification has also investigated making music out of longterm biological data [37]. Still, this data often directly represents physiological changes prioritizing music creation rather than providing distinct health feedback. Even more strongly supporting our claim that the cause of this gap is not due to technical inability to present data is that sonification has already been deployed in behavior change. By encoding physiological processes as sounds, researchers have been able to change participants' mindsets [43], and cause more mindful behavior [6]. Sonification is heavily researched in physical therapy settings, changing participant's behavior during treatment [32–34]. However, in these settings, sonification provides immediate sound mapping based on real-time data, providing no feedback on the user's health. All of these systems encode physiological data for users to hear rather than an assessment they can interpret.

We propose communicating personal health information to users through ambient music. The idea is to develop a system that converts biobehavioral signals to musical models to convey the status of health seamlessly. In this study, we approach health using a broad lens: in most cases not referring to any particular aspect of health, but rather leaving that for the participant to interpret the specific dimension of health. In doing so, we hope to gather data about all aspects of health, allowing our method to convert all kinds of health data to music. The challenge is, however, that interpretation of music is highly subjective. A melody that sounds exhilarating to one may sound boring to another. Our goal is, therefore, to 1) identify individual factors that affect one's impression of music healthiness, and 2) identify what music characteristics relate to the perception of healthiness. We develop a two-step approach to achieve this goal. In the first step, we create a survey to acquire users' general impression of music healthiness. In the second step, we build customized melodies for each user based on the combination of music preferences acquired in step one. We deliver these melodies to 55 participants over 31 days and ask them to rate the perceived healthiness of each melody. We then analyze the patterns in the collected data to identify factors that affect users' characterization of healthiness in music.

We summarize the **primary contribution** of this paper as establishing the feasibility of using personalized music models to convey general health information. In doing so, we also provide useful insights into how different musical features impact the perceived healthiness of music, how perceptions of healthiness vary between users, what trends exist between users' impressions, and what influences (or does not influence) a user's perception of healthiness in a melody. We also develop and assess a method to present users with health data transcribed to music. Going forward, this method could be deployed into IoT health systems to help users better internalize their own health in an ambient manner. Using data collected by personal health trackers, this method could help raise awareness of users' wellness by finding opportune times to play the melodies through users' phones so that the message resonates with the user, without requiring prompting from the user. To the best of our knowledge, this is the first study that evaluates the feasibility of using personalized music models for communicating personal health information.

2 RELATED WORK

2.1 Ambient Displays

2.1.1 Phone-Based Ambient Displays. Many technologies have been developed to present users with relevant health information in an ambient manner. Early versions utilized phone screens to display data. The UbiFit Garden was one of the first of these displays [8] where different flowers and butterflies demonstrated the participant's activity level and progress on their weekly activity goals. Bewell [22] and Bewell+ [23] followed the success of the UbiFit Garden. In this new system, the scope was expanded beyond physical activity to also provide feedback on sleep and social interaction via an underwater ecosystem where various fish and their activities showed how healthy the user was in each of the three dimensions. BeWell+ provided several improvements focusing on more personalized score calculation and suggesting interventions to the dimension of wellness where users needed the most help. MONARCA system [16] went beyond general wellness and specifically targeted individuals with bipolar disorder. Part of this system was an ambient display, which showed the factors affecting patients' conditions by changing the size and color of speech bubbles on the user's phone screen. Overall, these phone-based ambient displays have proven to promote healthier behavior over a wide range of data types and through many metaphorical representations.

2.1.2 Physical Ambient Displays. In recent years, more pervasive ambient displays have been developed, each containing a physical component, allowing them to exist beyond a phone screen and become even more integrated into a user's life. For example, the health bar [28] intended to help desk workers to take a break from sitting by changing the light color to red to encourage standing. MoodLight [44], a system that mapped the user's affect to the color of light, used EDA to determine the user's level of arousal, displaying warmer lights when the user is aroused, and cooler when the user is not. Howel [17] used an armband containing several colored rings, each correlating to a different range of heart rates. These were used to help users understand what phase of exercise they are in during a cardio workout. Each of these ambient displays managed to successfully provide data in an accessible manner, while becoming more pervasive than a simple phone screen, indicating understandable, pervasive, and unobtrusive presentations of data are worth developing.

2.2 Non-Visual Communication

Despite their success, the ambient displays discussed in the previous section remain focused, at least in part, on visual representation. Research has been done into displaying information using the other senses. Yet, limitations exist for most senses which prevent them from presenting complex information well.

2.2.1 Haptics. Haptics are used to passively portray information from many devices such as a computer mouse click, force applied to a joystick [21], or notifications on a phone or smartwatch. Researchers have embedded haptics into different locations such as floors [49]. This use case could have unique applications, such as assisting the blind with self-navigation. Despite its prevalence, haptic displays are challenged by communicating highly complex data [15], often requiring support from other presentation modalities, such as audio or vision. Thus, haptics alone is not ideal for delivering data as complex as health assessments.

2.2.2 Olfactory. Studies have been conducted to deliver smells using wearables [1, 11]. In [1], a necklace capable of producing basic scents such as a tea tree, peppermint, or a rose was used. The findings showed the perceived intensity of smell varied greatly among participants. A following study [11] assigned different smells to indicate different notifications. Findings in [36] determined smells can trigger memories, stimulate, build expectations, identify and locate, and influence mood. However, olfaction has major limitations. First, it is very challenging to create specific smells [53]. It is far harder to create smells than colors or music which can be assembled through their sub-components. Additionally, smell is a highly personal experience [11], every person will interpret a smell differently, possibly triggering different moods and memories. As such, while olfactory displays may be useful in displaying complex information in the future, the technology requires significant advancements first.

2.2.3 Taste. Of all the senses, communicating data through taste appears to be the least researched. It is possible to mechanically create tastes using electrical stimulation of the tongue [38]. Through a 21-participant study, researchers proved smell, taste, and color can all be combined to alter the perceived taste of a drink [38]. Data Cuisine [45] created chocolate coffins with different fillings to represent causes of death. Despite the ability to invoke specific tastes, the sense is still rarely used to communicate data, possibly because triggering it requires eating food. This lack of research could be due to sustainability, preservation, or the limited ability of humans to consume [52].

2.2.4 Auditory. Musical ambient displays have been proposed in smart homes to inform party hosts of events in other rooms [48]. Researchers propose using sensors to gather information about the events in each room. The music will then play for the host to understand what is occurring elsewhere in the party. In one of the most relevant work, researchers investigated whether an ambient display can utilize different musical genres to represent various affective states [4]. Affect data was mapped to a valence-arousal score and played via music. The clips were altered through various filters and alterations to the melody (such as skipping or repeating notes). Results indicated commonalities between participants in how music was perceived. Given the inherently ambient nature of music and the success of [4] in identifying commonalities in how music represents emotions, we believe utilizing sonification can create an ambient display capable of providing users with easily understood and accessible health data. Our work builds off this study, advancing to more complex data, and discrete health levels, while also reducing the musical features to basic musical components and heavily accounting for individual differences.

2.3 Sonification

Research into the fields of sonification and auditory displays indicates several strengths and flaws of the field as a whole that must be considered in the context of our study. To start, these systems are useful because they do not require the user to be facing a certain direction or interacting with a specific device, rather the sounds can be played and as long as the user is near the source, they can be heard [12]. However, being more ubiquitous and pervasive naturally creates the issue of privacy, as anyone in the area of the device can hear it. This may be solved with personalization, which has been shown to assist visually impaired users with technology [3] and may challenge unintended users from correctly deciphering the audio. Another limitation is considering sonification and music to be the same [32]. While music is the art form that combines pleasing tones, sonification focuses on translating a data source into a sound. However, the inclusion of music in sonification can have positive impacts, as straight conversion to noise without considering music will likely make the data exhausting to hear. Thus, sonification often has a trade-off: musicality vs. information. As one increases, the other likely decreases. Thus, effective sonification methods should determine a balance between music and data.

Sonification has been applied in different domains. For example, it has been used to augment presentations about the dangers of alcohol [50, 51]. These presentations mapped data about alcohol into synthesized sounds, auditory icons, and earcons. In another interesting use case [31], researchers sonified crime rates against women in India using audio clips of women's screams and mapping the frequency, amplitude, and timbre of the audio. They ran a questionnaire study and found the sonified data could meaningfully portray the violence. In rehabilitation, musical variables have been manipulated to guide users in movements e.g., raising their arm until the music changes [32–34]. However, in all these settings the sonified data is simply being played for the participant to hear. While there have been attempts to compose music from physiological data over a period of time [37], no attempt has been made to present it in a method intended for internalization and reflection, which could spur larger behavior change.

Some studies specifically target behavior change with sonification. In Go-With-The-Flow [43], participants performed physical activity, either in silence or with sonified music playing. Participants reported feeling more motivated after the sonified condition than in silence, indicating the sounds help users internalize their own physical activity. Similarly, Ambient Walk [6], a tool that monitors users' walking pace and breaths to play sounds, made users feel more mindful of their movement while using the tool. Lastly, sonification was used to invoke behavior change in dancers [20] by guiding them about their movement through audio.

Outside of sonification, music's success in swaying behavior can easily be observed through the success of music therapy and audio marketing. Music therapy has been shown to invoke changes in behavior, emotion, and physiology [9]. Marketers are capable of influencing consumers to make purchases [27] by invoking specific emotions through music. Therefore, when utilized correctly, music can sway users into making behavioral changes.

3 STUDY DESIGN

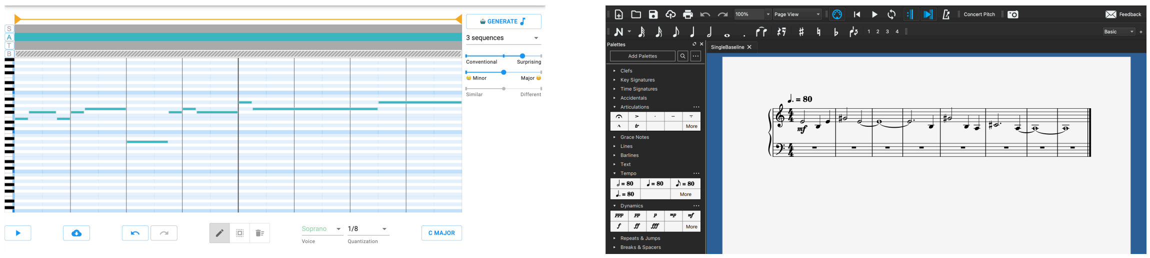
Our goal was to determine the feasibility of communicating personal health information through music using personalized models for each user. To achieve this, we designed a study where participants were asked to complete the model creation process. During this process, users completed two surveys which were deployed over Qualtrics and delivered to the participants through their university email. The first contained audio files of melodies manipulated using the musical features: tempo, pitch, key, dynamics, and smoothness. After listening to each file, participants were asked to provide the wellness level (provided through a five-point Likert scale containing a range of unhealthy and healthy options) that best represented the tune. We then used those responses to combine musical features and make healthy and unhealthy-sounding music for each individual participant. We deployed the generated melodies in personalized surveys in which participants provided their interpretation of the melodies.

Our approach involved evaluating the participant responses and models on an individual level and combined with all other participants. The personal level allowed us to determine how well the user could interpret music allowing us to create their personalized model, while the group analysis let us determine the uniqueness (or lack thereof) of each model. The remainder of this section describes the technical steps and surveys deployed during this study.

3.1 Health Levels

Throughout the study and analysis, we frequently use the terms health and wellness levels. We define five distinct wellness levels, each represented by a different point on the Likert scale. These five levels depict the varying statuses of healthiness ranging from very unhealthy (1) to very healthy (5). The five-point Likert scale was chosen based on the music ambient display created in [4], which assessed musical features in affect using a similar scale. We also confirmed a five-point Likert scale had been applied to many types of health as documented in [29].

These wellness levels are highly subjective, as they exist solely in the participant's interpretation of the music. Even when we attempt to incite specific health levels, the exact score is the user's decision. The decision to describe health in this manner was done intentionally, as we wanted our system to be applicable to all types of wellness, rather than influencing participant responses by asking them to consider a specific type of health in most questions. Using our method, once the user's models are complete, real health data could be applied to the



(a) A sample melody generated using Bach CoCoCo

(b) The melody from our first survey, created in MuseScore

Fig. 1. The software used to create, modify, and export the compositions

user's interpreted wellness levels. As an example, a common daily goal is to walk 10,000 steps. These musical models could be deployed so that with every additional 2,000 steps, the melody depicting the healthiness of the individuals' steps changes to a melody that the user associates with a higher wellness level. As the user gradually walks more throughout the day, they will hear increasingly healthier-sounding music.

It is worth noting, we use a loose definition of health and wellness in our study. Participants are not given a specific type of health, but rather are free to approach health and wellness as it applies to them. As such, the term health in our study is a general term, encompassing the overarching well-being of an individual and the well-being elements of their lives they prioritize, as those are the types of health that participants would default to thinking of. While our definition is a general term, we find that existing literature also deploys a general approach. In their review of health technology for marginalized communities, the study in [30] identifies a quarter of papers that do not focus their technology on a specific aspect of health, but rather discuss general health.

3.2 Music Generation

Our method relies on manipulating musical characteristics to represent health data. Several musical elements were manipulated throughout the study. Table 1 contains the musical terms relevant to this paper and their definitions. All music utilized in this study was created using Bach CoCoCo, which generates melodies using DNNs while accepting inputs from users [25]. Bach CoCoCo (Figure 1a) was chosen to generate the music because it uses AI to create the music but still gives users control over the creation process. For our study, the most important feature allows for the creation of multiple melodic lines that consider the other lines to form a pleasing melody but can be created one at a time. This feature was essential to creating our three-lined melodies because it meant that a line that fails to meet all the criteria (discussed in the next paragraph) could be recreated without needing to scrap the other existing melodic lines.

Once a tune was generated, it was evaluated simply to ensure the tune contained a melody. To contain a melody, we required it to meet three criteria. First, the melody had to have at least as many new notes as musical measures. This rule existed to prohibit tunes that only played a small selection of notes, or held a single note for a majority of the tune. Second, no notes could obviously clash, ensuring that the melody would not be inherently unpleasant to listen to. Lastly, the melody had to have multiple occurrences of notes that would change when the melody is transposed from the key of C-major to C-minor, ensuring that this musical feature included in the personalized models would be relevant for the music. If a melody did not meet these criteria, Bach CoCoCo generated a new melody.

If a melody met the criteria, it was recreated in MuseScore [10], as shown in figure 1b. MuseScore is a sheet music editor with a free-to-use version that provides access to a wide variety of music traits. This allowed for the easy conversion of melodies from Bach CoCoCo to a format granting greater access to modifications such as

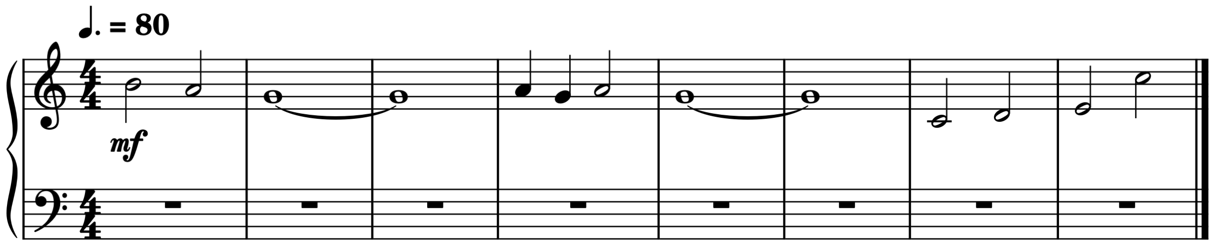


Fig. 2. Unaltered melody from the preferences survey

dynamics, tempo, and instrumental changes. It also provides an option to export sheet music to an MP3 so the music could be embedded into Qualtrics surveys without requiring any effort from the participant other than opening a survey.

Table 1. Relevant musical terms and their definitions.

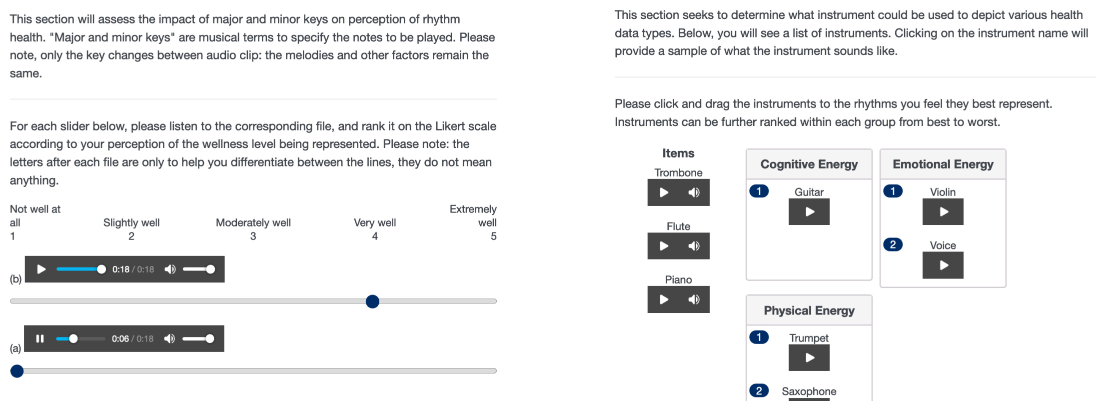
Music Term	Definition
Notes	The building blocks of music. Notes represent tones and their duration.
Melody	A sequence of notes in a pleasing manner.
Tempo	The speed of music.
Dynamics	The volume of music.
Pitch	How high or low a note is.
Octave	A collection of sequential notes after which the name of the note begins to repeat.
Key	Specifies specific tones that should replace an adjacent tone throughout the melody.
Staccato	Indicates that notes are shortened, leaving a longer break between each pair.
Legato	Indicates that notes are lengthened, leaving little to no space between each pair.

3.3 Acquiring General Healthiness Impression of Music Features

We created a survey consisting of several audio clips, each changing a different factor (tempo/speed, dynamics/volume, major vs. minor keys, etc.). After each audio clip, participants were asked to rate the clip on a Likert scale depicting their impression of the healthiness the music clip would represent. Participants were also presented with clips of different musical instruments and were asked to rank them according to the type of health energy they felt those instruments represent.

The goal was to identify the musical variables that created the impression of healthiness or unhealthiness in melodies. Split into several sections, the survey contained MP3 files of the same tune each altered by changing a single musical trait each time. For example, one audio file contained the melody in a major key and the following contained the same melody in a minor key. An unaltered version of the melody can be seen in figure 2. After each audio file users were asked to consider that the music represented health data and to fill out a five-point Likert scale demonstrating the wellness level they believed the music represented, as shown in figure 3a.

To determine our musical features, we used a literature review of sonification in physical therapy which identified the six most common auditory alteration categories to be: event-driven, loudness-related, timbral, pitch-related, spatial, and temporal [35]. In our study, not all of these categories are possible. Event-driven changes trigger when a condition is met, this is not practical for us to deploy in our study because participants are receiving the data through a survey completed on their own time. Similarly, we could not alter the spatial



(a) Assigning health scores to different levels of smoothness

(b) Assigning instruments to energy levels

Fig. 3. Sample of both question types in the preferences survey

features because the sound will originate from the user's device. Lastly, we did not include timbral changes because they may alter the quality of the audio, and we did not want to lower the quality as other features could become difficult to understand. Therefore, in our study five musical features: tempo, pitch, key, dynamics, and smoothness were manipulated in the survey. We will now briefly explain each feature (see table 2).

- Tempo - Six different tempos were included in the survey. The musical notation for tempo states how many of a certain type of note will be played in a minute. Table 2 shows all six tempos, both in musical notation and file length in seconds. This feature was included due to the temporal changes found in the literature review.
- Pitch - The files were manipulated so that the majority of a melody falls within a specific octave (range of pitches). In this survey, the notes ranged from C2 (low) to C5 (high). This range was selected as each range sounds unique but is not painful to listen to. Table 2 shows the octaves used in our study, and their respective frequency ranges [46]. This feature was included as it lies within the pitch-related category found in the literature.
- Key - The study used melodies in the keys C-major, and C-minor. Three notes differ between these two keys by a half step (a small change). Traditionally, minor keys are thought to make music sound sadder, however, this is not always the case. A famous example of this is Survivor's 'Eye of the Tiger' is in the key of C-minor [47], the same key as our study. Table 2 shows the frequency of the impacted notes in octave C4. We included the key as another interpretation of altering the pitch-related category. The literature review defined this category as an 'increase or decrease in perceived audio frequency' [35]. While initial thoughts will lead this variable to mean pitch, we realized it could also indicate the key of the music, which changes the frequency of three notes rather than all seven.
- Dynamics - Three different dynamics, namely pianissimo (quiet), mezzo forte (medium), and fortissimo (loud) were used. These dynamics were selected because they were extreme enough values to sound different, but not so extreme that the music became challenging or painful to hear. Sample decibels recordings for each dynamic are shown in table 2. These measurements were made from the same computer, with the speakers set to the same percentage. Exact decibels will vary between recordings based on these settings. We included dynamics in our study as an example of the loudness-related variables detected in the literature.

Table 2. Musical notation and real world equivalency for the musical features in this study.

Musical Notation	Real World Equivalency	Feature Name
Tempo		
Eighth Note = 80	49 seconds	Very Slow
Dotted Eighth Note = 80	33 seconds	Slow
Quarter Note = 80	25 seconds	Moderately Slow
Dotted Quarter Note = 80	18 seconds	Moderately Fast
Half Note = 80	14 seconds	Fast
Dotted Half Note = 80	10 seconds	Very Fast
Dynamics		
PP	63.4 dB	Quiet
MF	71.5 dB	Middle
FF	85.5 dB	Loud
Key		
Major	329.63, 440.00, & 493.88 Hz	Major
Minor	311.13, 415.30, & 466.16 Hz	Minor
Pitch		
C2	65.41 - 123.47 Hz	Very Low
C3	130.81 - 146.94 Hz	Low
C4	261.63 - 493.88 Hz	High
C5	523.25 - 987.77 Hz	Very High
Smoothness		
Staccato	1.11 seconds	Staccato
Not Staccato	2.22 seconds	Legato

- Smoothness - Users listened to the melody where every note was staccato, and one where the notes were not. The opposite of staccato is legato (smooth), not normal notes that we use in this study. However, during development, we realized the exported MP3 files of legato melodies had no apparent noticeable difference compared to melodies where the notes were both not staccato and not legato. Therefore, our personalized models use staccato and not staccato rather than staccato and legato, although the term 'legato' will be used for simplicity. The length of the same sustained note, both staccato and not are shown in table 2. Smoothness was included in our study as another example of temporal changes. While we determined the speed of the song overall is a strong example of a timing change, we realized the amount of space between notes might also have a significant impact on a user's perception of health.

Additionally our study also seeks to determine if evaluating multiple lines at once impacts the user's ability to depict health levels. The motivation for doing so was that music can, theoretically, express multiple health levels simultaneously as melody with multiple lines. To differentiate the different lines, we designed each to be played by a different musical instrument. In this initial survey, users were asked to listen to the same audio clip but played by different instruments. The possible instruments are flute, guitar, piano, saxophone, trombone, trumpet, violin, and voice. Participants were asked to classify each instrument as best representing a specific type of health energy: cognitive, physical, or emotional as seen in Figure 3b. In addition to sorting between the classified energy groups, users could sort within each group, specifying which instruments better classify each

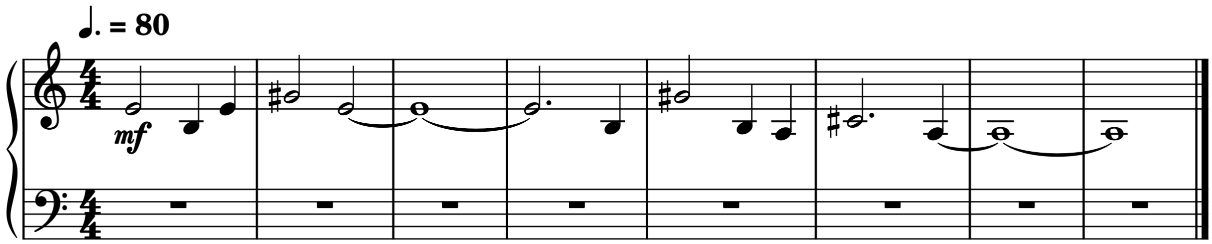


Fig. 4. Unaltered melody from the second survey

type of energy than other instruments within the same group. While the instrument question was always asked last, the sections focusing on musical features were randomized for each participant, and the audio files within each section were randomized as well. This was done to minimize the influence between audio files.

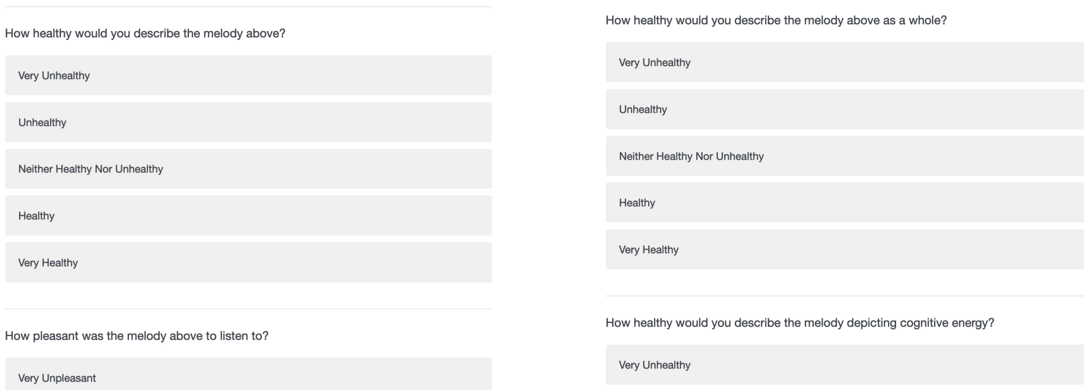
3.3.1 Creating Personalized Models. We use participants' ratings of healthiness impression to build customized models and further evaluate how well participants were able to distinguish the healthy music combinations from unhealthy ones. We combine multiple features based on the user's responses to the first survey. In this work, we define personal models to be the collection of what each participant chose as the healthiest and unhealthiest sounding feature for each musical characteristic. Building these models on a personal level was inspired by inScent [11] which let participants decide what meaning each scent should have. Much like smell, music is a highly personal experience with every person enjoying different types of music. As such, we decided a musical-based system must be personalized so that it can be accurately and easily understood.

For each musical feature in survey one, we identified the option participants chose as the healthiest and the unhealthiest. For example, if 'very fast' was ranked as the healthiest of all tempos, and 'very slow' as the least healthy of all of the tempos, we recorded those tempos as the settings in the healthy and unhealthy models. In the case a user provided the same level of healthiness for multiple choices, the system would decide the respective feature from among these. To prevent overlap between the healthy and unhealthy models, the features were not chosen randomly. Instead, it began by looking at one end of the spectrum, and worked to the other end, taking the first value it reached that was included in the tie. For each feature, determining healthy and unhealthy began on opposite sides of the spectrum, so overlaps are not possible.

To demonstrate, imagine a user ranked the pitches such that 'very low' and 'very high' were tied for the unhealthiest option, while 'low' and 'high' were tied for the healthiest. The method began by finding the feature for the healthy model, starting at 'very high.' Because 'very high' was not a viable option, it will move to 'high' which is a possible choice, and save that as the healthy pitch. It would then repeat this process for the unhealthy model, starting with 'very low.' Since the 'very low' pitch was rated as one of the unhealthiest features, it would be selected. This would result in 'high' being the pitch for the user's healthy model, and 'very low' being the pitch for the user's unhealthy model.

3.4 Healthiness Interpretation of Personalized Music Models

Once the personalized models were created, the interpretation survey was generated using them. The goal of this survey was to determine how the user will interpret music where the features are combined. The audio files in this survey use elements from the user's healthy and unhealthy models on the same melody (shown in figure 4). Every audio clip contains either the healthy or unhealthy setting for each musical feature described in the preferences survey. Participants analyzed these audio files using a five-point Likert scale, the same method as the preferences survey. Additionally, for this study, we added an additional question to this survey. After each audio



(a) Sample questions for the single-line melodies

(b) Sample questions for the triple line melody

Fig. 5. Question types in the interpretation survey

Table 3. Features of all seven compositions.

Music Feature	H	U	A	B	C	D	E
Tempo	Healthy	Unhealthy	Unhealthy	Healthy	Healthy	Unhealthy	Unhealthy
Pitch	Healthy	Unhealthy	Unhealthy	Healthy	Unhealthy	Unhealthy	Healthy
Key	Major: Healthy Minor: Unhealthy	Major: Unhealthy Minor: Healthy	Major: Healthy Minor: Unhealthy	Major: Healthy Minor: Unhealthy	Major: Unhealthy Minor: Healthy	Major: Unhealthy Minor: Healthy	Major: Healthy Minor: Unhealthy
Dynamics	Healthy	Unhealthy	Healthy	Unhealthy	Healthy	Unhealthy	Healthy
Smoothness	Healthy	Unhealthy	Unhealthy	Unhealthy	Healthy	Healthy	Unhealthy

file, participants were asked to complete a 5-point Likert scale for how pleasant the melody was. These questions were included because we suspected pleasantness would serve as a strong indicator of perceived healthiness. This portion of the survey can be seen in figure 5a.

The interpretation survey was comprised of two melodies: one with a single melodic line, and another with three. The single-line melody was always played on the piano. It assessed how users perceive wellness when all musical factors from the first survey were combined. In some instances, all factors indicated the same level of wellness, while other clips assessed how participants will perceive conflicting information. In all, there were seven audio clips containing a single melody line. One contained all 'healthy' features, one contained all 'unhealthy' features, and five contained conflicting information. For ease of reference, these will henceforth be called composition H, U, A, B, C, D, and E respectively. The settings that made up each composition can be seen in table 3. When discussing healthy and unhealthy features these are the respective options for each musical element that were saved in the user's model to indicate healthy or unhealthy. Using the example from the previous section, 'very slow' would be the model's unhealthy tempo feature, while 'fast' would be the model's healthy tempo feature.

The second melody, shown in figure 6 was a three-line melody, where three different sets of notes were played simultaneously. During the study, users were asked to listen to this melody four different times, twice where each line was using all the features from the healthy model, and twice again with unhealthy features. For each level of healthiness, it was played once solely on the piano, and again by a different instrument that the user specified in the previous survey. When each line was represented by a different instrument, users were asked to further break down the melody, and state the wellness level of each type of energy. The three-lined melodies were played on both the piano and instruments so any inconsistencies compared to the single-line melodies can be

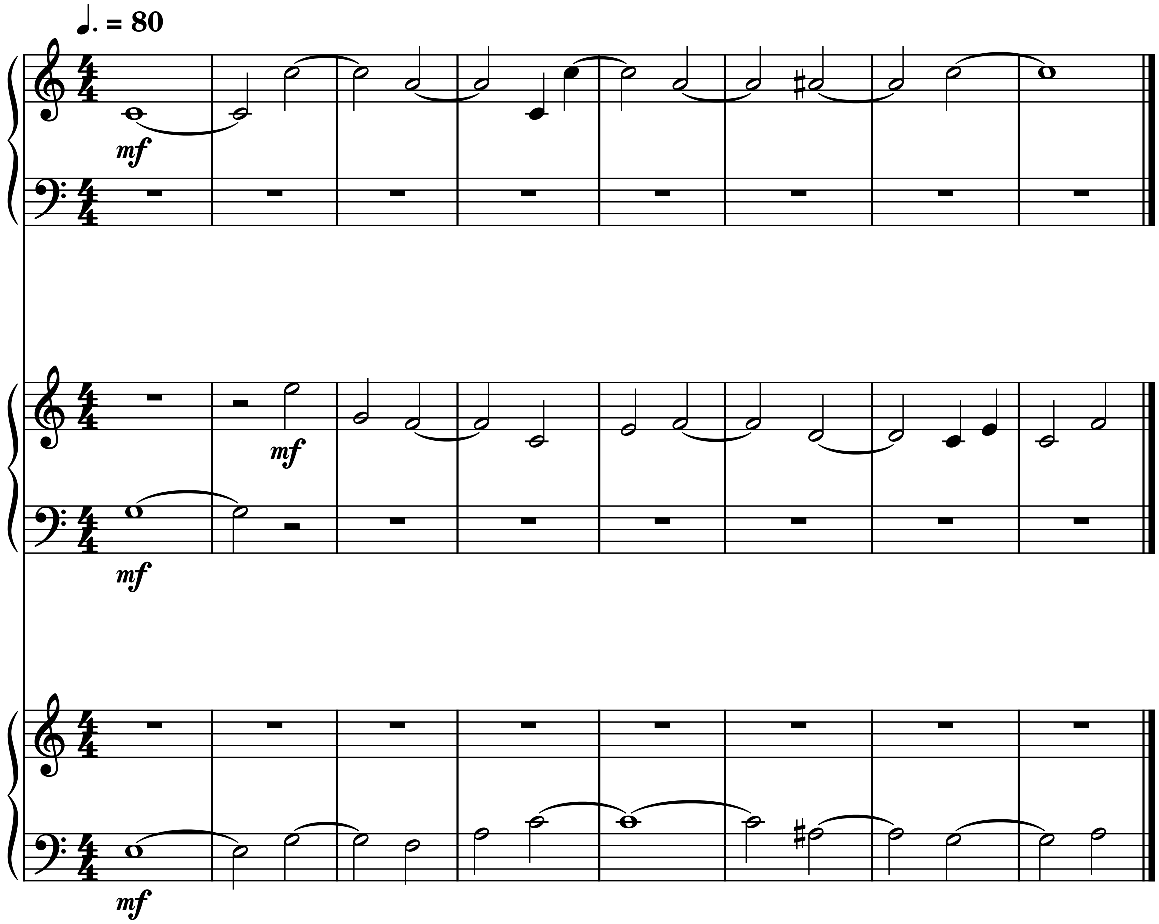


Fig. 6. Unaltered three-line melody in the second survey

determined to be a result of the instrument changes, or the change to multiple lines. As with the other question types, a picture of the survey asking questions from this section is shown in figure 5b. While the single-line melodies were always played before the three-line melodies, all compositions within each group were provided in a random order for each participant.

4 STUDY PROCEDURE

4.1 Recruitment and Participants

We recruited 55 participants at a large university in the mid-Atlantic United States, and 52 (13 male, 39 female) were included in the data analysis. In the first round, participants were sent the general healthiness impression survey to fill out. We then used responses to the first survey to generate models and melodies based on each participant's preferences. These melodies were incorporated in an individualized interpretation survey and sent to participants for rating. The surveys were deployed and left open for participants to complete at their own pace over 31 days. During this time, 55 participants completed the general healthiness survey (henceforth called survey 1) and 50 completed the individualized interpretation survey (henceforth called survey 2). Two responses from each survey had to be removed because participants provided the same answer for every question, indicating they may have been focused on completing the study for compensation rather than providing useful data. The participants with insufficient answers in the first survey had to have their second survey removed as well because the first survey's responses impacted the creation of the second. However, the survey 1 responses for the problematic survey 2 participants were left in for data analysis because even if the participant did not take the second survey seriously, they appeared to do so with the first. Additionally, one response to each survey had to be removed, because a participant took the study twice. Therefore, during data analysis, 52 responses to the first survey and 45 for the second were included. Upon completing survey 2, the participants were automatically flagged to receive $10 in compensation.

To understand the effect of demographics on music preference and interpretation, we also collected information about race and ethnicity. Thirty-two participants self-identified as white, eighteen as Asian, five as Black or African American, and one as Hispanic; thirty-four participants identified as not Hispanic or Latino, twelve as non-black, non-white Hispanic or Latino, eight as white Hispanic or Latino, one each middle eastern and Asian, and lastly two participants opted not to provide their race. Participants were further asked about their music backgroundbecause theories identify music background as a potential influencing factor of music taste [40]. In particular, users provided musical information regarding music genre preferences, music education, the frequency with which they listen to music, and activities they do while listening to music. Figure 7 shows the musical demographics of participants. The majority of participants listened to music every day, and more than half indicated they can read music and/or play an instrument. The most popular genres of music were pop, rap, and rock.

5 ANALYSIS AND RESULTS

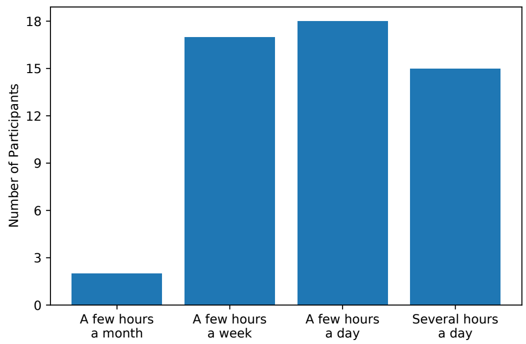
To analyze the effectiveness of and future considerations for sonification-based health applications, we investigate the following questions:

R1: What music features (single or combined) make the melodies to be perceived as healthy or unhealthy?
R2: How well can the combination of multiple aspects of health (e.g., physical, emotional, and cognitive) be expressed in one melody? Is the healthiness of a single health aspect more easily communicated with the users than multiple?
R3: How common are the healthiness impressions of music features among participants?
R4: What external factors affect the perceived healthiness of music? To what degree do demographics, music background, and frequency of listening to music affect participants' impression of music healthiness?

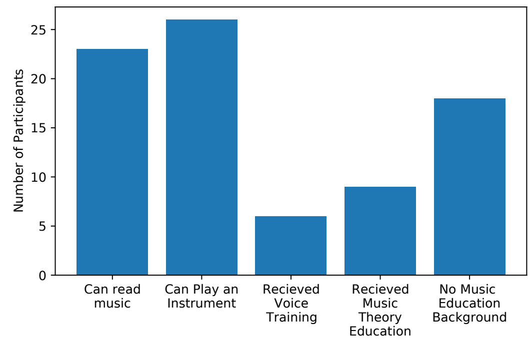
5.1 What Music Features Make the Melodies Sound Healthy or Unhealthy?

To create personalized melodies aligned with users' music healthiness impression, we must first identify how the healthiness of each musical feature was perceived. We began by analyzing participant responses to the general healthiness impression survey.

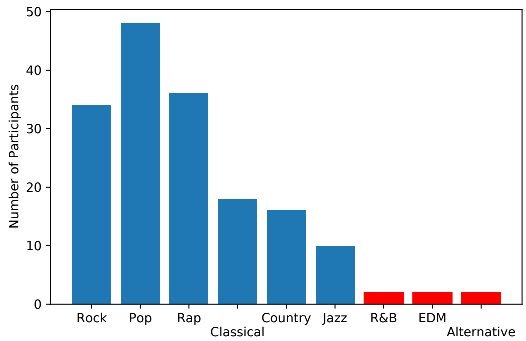
5.1.1 The Impact of Single Musical Features On Healthiness Impression. In the first survey, participants provided their impressions of how each musical feature impacted the healthiness sound of the melody. To provide these impressions, participants ranked the features on a five-point Likert scale representing very unhealthy (1) to very healthy (5). In the following, we analyzed participants' responses to each feature and compared them to the other features within the same category. In doing so, we focused on finding significant differences between responses to the various variables and defining the trends within each musical feature. The mean scores and distribution of responses for each feature are shown in table 4 and figure 8 respectively.



(a) Frequency participants listen to music



(b) Formal music education background participants have



(c) Music genres participants listen to
(Written in genres that multiple participants listed are in red)

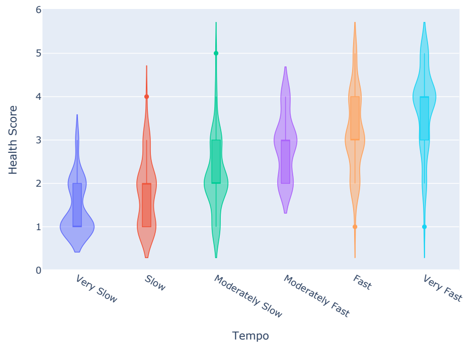
Fig. 7. Musical Demographics of Participants

Table 4. Summary of health scores for each individual musical feature

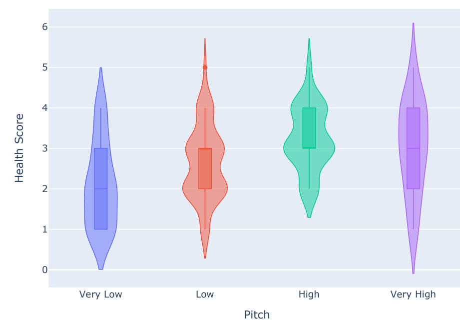
	Tempo						Pitch				Key		Dynamic			Smoothness	
Feature	Very Slow	Slow	Moderately Slow	Moderately Fast	Fast	Very Fast	Very Low	Low	High	Very High	Major	Minor	Quiet	Middle	Loud	Staccato	Legato
Mean	1.442	1.788	2.212	2.865	3.173	3.712	2.096	2.635	3.231	3.039	3.327	2.019	2.442	3.096	3.039	2.769	3.25
Standard Deviation	0.608	0.776	0.847	0.715	0.923	0.936	1.034	0.908	0.783	1.137	0.785	0.852	1.110	0.934	1.084	0.983	0.86

Tempo. The tempo ratings followed the trend that faster is healthier (figure 8a). A Kruskal-Wallis test comparing responses for each feature within tempo provided significance results ($H = 156.230$, $p = 6.293 * 10^{-32}$), indicating there were differences in how participants interpreted the various melody speeds. The very slow tempo provided the lowest average score of any feature in the study. Health score gradually increased as the tempo gets faster, with the fastest tempo being the one participants perceived to be the healthiest. Interestingly, as the mean score increased, the standard deviation tended to as well. This may indicate that while most participants considered faster tempos to be healthier, they disagreed on the degree of healthiness.

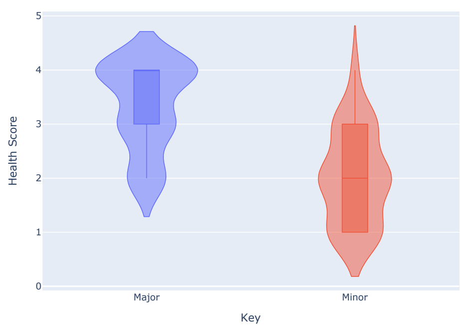
Pitch. As with tempo, a Kruskal-Wallis test showed there are differences in the interpretation of healthiness based on pitch ($H = 34.419$, $p = 1.616 * 10^{-7}$). However, the pitch did not follow as clear of a trend as the tempo (figure 8a). Generally, participants perceived higher pitches as healthier. However, when the pitch becomes too



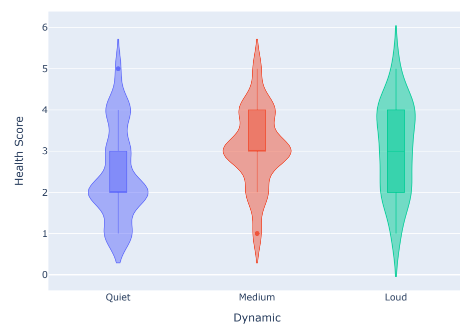
(a) Tempos



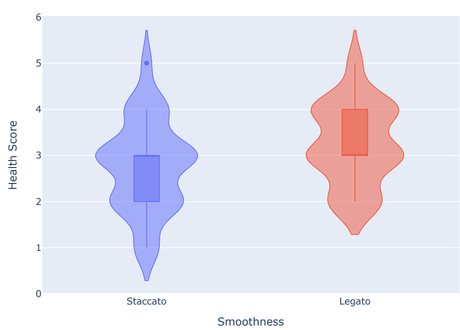
(b) Pitches



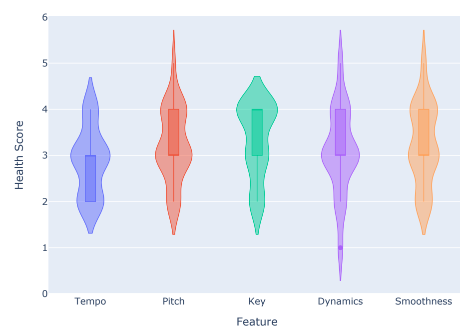
(c) Keys



(d) Dynamics



(e) Smoothness



(f) Baseline melody present across all features

Fig. 8. Perceived Healthiness of Different Music Features

high, the impression of healthiness drops. As a result, the 'high' pitch variable was perceived to be the healthiest value. This is notable because the value we call 'high' is significant in music as well. This octave, beginning with a note called 'Middle C' is considered to be the center of the music scale, and it commonly appears in music. Therefore, while participants perceived higher pitches as healthier, they may also have interpreted common pitches as healthier.

Key. Key also had a significant impact on participants' interpretation of healthiness. A Kruskal-Wallis test between the major and minor keys indicated a level of significance (H = 40.424, p = $2.044 * 10^{-10}$). Participants strongly associated the major key with a higher level of healthiness (figure 8c). On average, the switch from major to minor decreased the perceived healthiness by 1.3 points, a change larger than all other features except for tempo.

Dynamics. Similar to the previously discussed musical features, changes in dynamics also resulted in changed impressions of health (H = 12.256, p = 0.002). As shown in figure 8d, participants considered quiet melodies to be less healthy than those at a loud or middle volumes (H = 12.188, p = $4.809 * 10^{-4}$). This is particularly notable because a Kruskal-Wallis test showed no significant difference between the middle and loud dynamics (H = 0.031, p = 0.860). Thus, while participants perceived quiet melodies as unhealthy, they did not perceive significant differences between the healthiness of middle and loud dynamics.

Smoothness. Lastly, smoothness also impacted users' perceptions of music's healthiness (H = 6.304, p = 0.012). As shown in table 4 and figure 8e, participants generally found legato models healthier than staccato models. That is to say, smoother melodies appeared to be perceived as healthier.

Baselines. We concluded our analysis of the first survey with a comparison across each feature category. Within each set of features, there was one musical composition that was present in all five categories. This composition was included to ensure participants responded similarly to a melody if they encounter it multiple times. Between the five feature categories, our baseline compositions were: moderately fast, high, major, middle, and legato. These compositions can be seen in direct comparison in figure 8f. When comparing these musical features with a Kruskal-Wallis test, there was a significant difference (H = 78.445, p = $3.719 * 10^{-16}$), indicating user responses may not be consistent over time. However, when 'moderately fast' was removed from the set, the significance disappears (H = 2.313, p = 0.510). Because there was no significance between these four baselines, we can assume participants perceived them the same way, indicating there is little to no difference between them. Due to the randomized order in which the baselines were presented, we must assume the tempo baseline was only impacted by the other tempo melodies being perceived as healthier. Therefore, we concluded that participant perceptions frequently remain stable, but may also vary given the context in which the melody is delivered.

5.1.2 The Impact of Combined Music Features on Healthiness Perception. The analysis of the initial survey provided fascinating insights into how musical features impacted the perceived healthiness of music. We were further interested in understanding whether these effects remain when musical features are combined. To determine this, we analyzed the responses to our second survey through the context of each of the 7 musical compositions described in table 3. These compositions each combined features from users' healthy and unhealthy models described previously. We analyzed these compositions using a generalized linear mixed model. This allowed us to determine the average impact each musical feature had on the overall healthiness and whether this difference was significant while reducing the impact of individual differences between the compositions. As such, our model used the participant's rating as the dependent variable, a binary representation for each of tempo, dynamic, pitch, smoothness, and key as either 'healthy' or 'unhealthy' as independent variables, and the composition itself as the random effect to minimize the impact of the different melodies had on one another. All

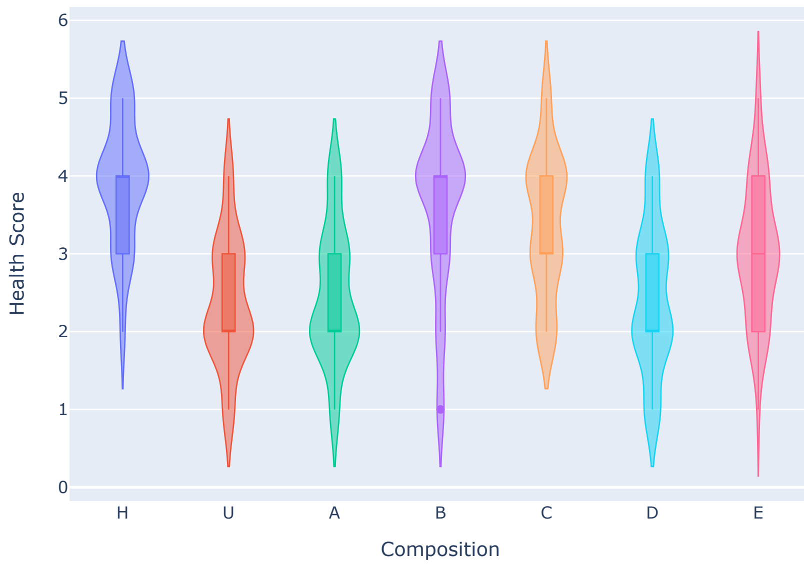


Fig. 9. Distribution of participants' perceptions of healthiness for each composition with combined musical features

Table 5. Summary of participant perceptions for the musical compositions in survey 2

	Composition H	Composition U	Composition A	Composition B	Composition C	Composition D	Composition E
Mean	3.911	2.378	2.467	3.622	3.378	2.422	3.022
Standard Deviation	0.793	0.806	0.842	1.093	0.912	0.917	0.866

of our mixed models were created using the statsmodels package in python [42]. While we will not discuss the exact scores of these compositions as a whole in analysis, they can be seen in figure 9 and table 5.

Before reviewing this analysis, we must first acknowledge the inconsistencies with the 'key' variable. Due to an error in model creation, all participants were classified as considering the major key to being healthier than the minor key. In reality, 42 participants believed this, 5 saw no difference, and 5 believed the minor key sounded healthier. Due to the binary nature of this variable we were still able to analyze the models by using the anonymous identifiers within the surveys to track which users received the wrong key and adjusting the analysis as necessary.

Tempo. Similar to what results alluded to in the previous section, our analysis indicated tempo had the largest impact on the perceived healthiness of a combined melody. On average, healthy tempos were classified 0.778 points higher than unhealthy tempos, providing a significant difference ($z = 5.825$, $p = 2.9 * 10^{-9}$).

Pitch. Pitch also provided a significant impact between healthy and unhealthy compositions ($z = 3.491$, $p = 2.407 * 10^{-4}$). On average, a composition using a healthy pitch was classified 0.452 points higher than one using an unhealthy pitch.

Table 6. Summary of the impact each musical element had on a melody when they are combined (section 5.1.2). Significance indicates changing the musical element from healthy to unhealthy results in a lower perceived healthiness.

Musical Feature	P-Value	Significant
Tempo	$2.9 * 10^{-9}$	Yes
Pitch	$2.407 * 10^{-4}$	Yes
Key	0.035	Yes
Smoothness	0.247	No
Dynamics	0.310	No

Key. The last variable that significantly impacted the composition's perceived healthiness was key (z = 2.111, p = 0.035). Using the healthy key resulted in an average improvement of 0.194 points over the unhealthy key.

Smoothness. While the effect produced by smoothness was not significant (z = 1.157, p = 0.247), healthy smoothness was, on average, classified 0.146 points higher than unhealthy. Due to the lack of significance participants likely did not associate changes in smoothness to hold much impact on healthiness.

Dynamics. Dynamics also did not provide a significant impact (z = 1.015, p = 0.310), but on average healthy dynamics were classified 0.109 points higher than unhealthy dynamics. While this could mean dynamics do not have an impact on perceptions of health, we believe another explanation is more likely. In our study, the participants completed surveys on their own computers. It is possible that between the beginning of the first survey and the end of the second, participants changed the volume on their machines. This could have been done during a study or during the period between them. Regardless, it may have impacted how users perceived the various dynamics.

While definitive trends did exist within the data, exact interpretations varied greatly between participants, as will be shown in section 5.3. Thus, while we can create assumptions to predict how users will respond to different musical features, building personalized models remain necessary to portray the necessary information. A summary of results from this section can be seen in table 6.

5.2 How Well Can Multiple Aspects of Health Be Expressed In One Melody?

We experimented with the idea of presenting health information related to multiple health factors such as physical, emotional, and cognitive in one single melody. We implemented this concept by developing three-line melodies, where each line is played by a different instrument. Each instrument represented a unique aspect of health (e.g., physical health), as specified by the participant in the first survey. These health aspects, represented as energy types in our survey, function the same way as the single-line melodies; they were manipulated through the user's personalized models to demonstrate various health levels.

To test the feasibility of combining multiple health factors in one melody, we first analyzed user perceptions of instruments and the types of health energy they represented. We will then discuss how well the overall health (all three aspects) could be received through the three-line melodies.

5.2.1 Association Between Instruments and Types of Health Energy. In the first survey, participants were asked to listen to a melody played by various instruments. These instruments were categorized to represent one of three types of energy: cognitive, emotional, and physical. In addition to categorizing the instruments, users could sort them, indicating the instruments they felt 'best' represented the energy compared to all other instruments. Figure 11b shows the distribution of how many times each instrument was assigned as the best instrument in each category. Cognitive and physical energy both showed a strong consensus with a clear preference for piano

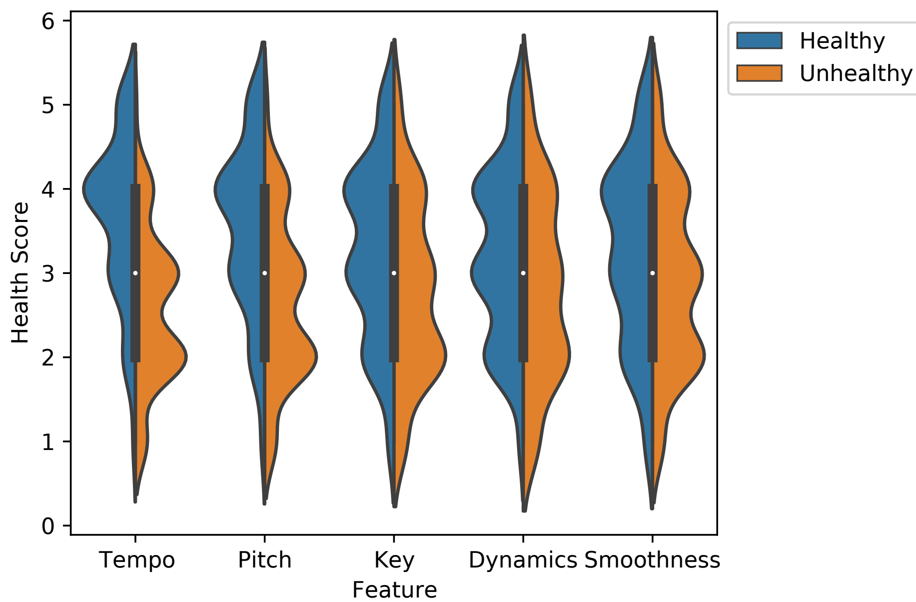
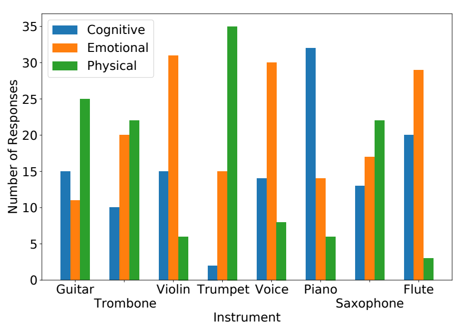
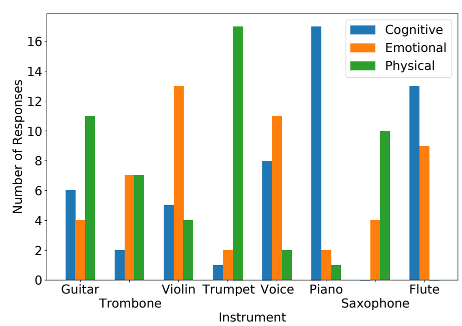


Fig. 10. Distribution of compositions using the healthy and unhealthy options for each feature category



(a) Number of times an instrument was classified as each type of energy



(b) Number of times participants classified an instrument as the best instrument

Fig. 11. Number of times participants assigned an instrument to a type of energy

and trumpet respectively. While not as common as piano and trumpet, guitar and flute were frequently sorted into these categories as well. Emotional energy also showed instruments that are popular, although the difference is not as profound as trumpet and piano. In this case, participants often associated the violin and voice with emotional energy.

Although the association of instruments to emotional energy looked similar when ignoring rank (figure 11a), there is one very large difference. Considerably more participants assigned the flute to emotional energy than

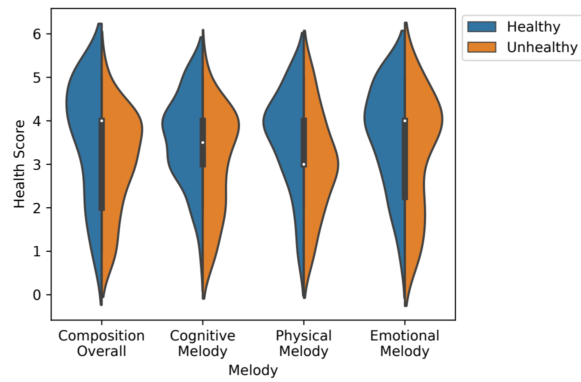


Fig. 12. Perceived healthiness of the three-line instrumental melody, and each individual line

cognitive energy, despite the fact that more participants ranked it as the best instrument for cognitive energy than emotional. This difference is notable because all instruments other than flute remained most commonly associated with the same energy type. However, flute is the third most popular choice for emotional energy in both representations. This substantial drop, may not be due to the flute itself, but due to the apparent uncertainty that exists within emotional energy, especially when compared to cognitive and physical energy which have clearly defined peaks.

Overall, these results showed trends in associating instruments with types of energy. While most energy types were associated with multiple instruments, cognitive and physical both had well-defined instruments that were most commonly associated with them. Yet, emotional energy was far less definitive.

5.2.2 Healthiness Perception of Single vs. Three Line Melodies. After establishing their instrument preferences, participants listened to and interpreted the three-line melodies. The study included four of these melodies, two using instruments and two using only the piano. Participants were only asked to interpret individual levels when using the instrumental compositions. However, the piano compositions were included to act as a baseline, so any differences between the instrumental composition and single line could be narrowed down to being a result of the instruments, or the transition from one line to three.

The healthiness level was intended to be identical across the overall melody and each individual line. We created generalized linear mixed models to test for a difference between all three lines in both melodies. For these models, we again used the participant's rating as the dependent variable. The energy being depicted (cognitive, physical, emotional, and overall) was used as the independent variable. The participants themselves were used as the random effect because, in this specific analysis, we were concerned with consistency within each user rather than isolating the impact of musical elements from each other. We found no significant difference between any pair of energy within the healthy or unhealthy composition. The distributions for all types of energy and healthiness can be seen in figure 12.

The success of the three-line instrumental melodies lessens when comparing across different melodies. Using generalized linear mixed models, we compared the instrument and piano three-lined melodies to each other, and to the single-line melodies H and U which used the same musical features. Same as in the previous models, the user rating was used as the dependent variable. Meanwhile, the composition type (instrumental, three-line piano, and single-line) was used as the independent variable. The participants were again used as the random variable for these models. There was no significant difference between the two types of three-lined melodies ($z = 1.743$, p

Table 7. Summary of the all significance tests conducted on the three lined melodies (section 5.2.2). Significance indicates the two elements (different melodic lines in the top section, and different instrumental compositions in the bottom three rows) were perceived differently by participants.

Comparison	P-Value	Significant
Overall Energy & Cognitive Energy	0.842	No
Overall Energy & Emotional Energy	0.464	No
Overall Energy & Physical Energy	0.842	No
Cognitive Energy & Emotional Energy	0.351	No
Cognitive Energy & Physical Energy	1.000	No
Emotional Energy & Physical Energy	0.351	No
Three-Lined Instrument & Three-Lined Piano	0.081	No
Three-Lined Instrument & Single-Lined	0.284	No
Three-Lined Piano & Single-Lined	0.005	Yes

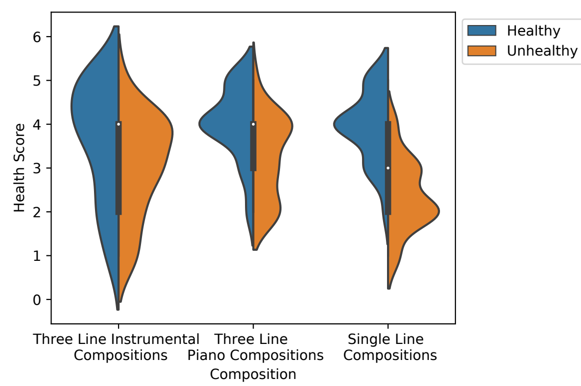


Fig. 13. The perceived healthiness of all three melodies using the same features

= 0.081), or the instrument and single-lined compositions (z = 1.072, p = 0.284). However, there was a difference between piano and single-lined compositions (z = 2.815, p = 0.005). The distributions of responses to each of these compositions can be seen in figure 13. These results indicate that the addition of instruments did not appear to impact the perceived healthiness, but the conversion from one line to three might have.

Thus, while instrument meanings may be generalizable, the three-line melodies may lead to different health perceptions than intended. Even though participants can accurately interpret healthiness from sonified data in single and three-lined melodies, further research is needed to determine whether users can accurately depict the same sonified data from single and three-line melodies that both use the same personalized models. Different unique personalized models may be required for the three-line melodies as well as the single-lined melodies. A summary of results discussed in this section can be found in table 7.

5.3 How Common Are the Healthiness Impressions of Music Features Among Participants?

To determine the commonalities among participants in their healthiness impression of music features, we analyzed the models that were generated from the initial survey responses. We focused on more frequent models that

Table 8. The number of occurrences and settings for all personalized models that were generated by multiple participants

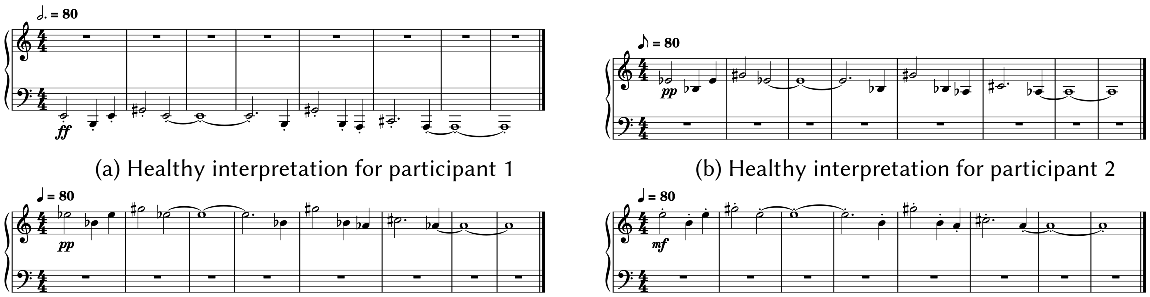
Number of Participants	Tempo	Pitch	Key	Dynamics	Smoothness
Healthy					
3	Moderately Fast	High	Major	Middle	Legato
3	Fast	Very High	Major	Loud	Legato
2	Very Fast	Very Low	Major	Loud	Legato
2	Very Fast	Very High	Major	Middle	Legato
2	Very Fast	Very High	Major	Loud	Staccato
2	Very Fast	Very High	Major	Loud	Legato
2	Very Fast	Low	Major	Quiet	Legato
2	Very Fast	High	Major	Middle	Staccato
2	Very Fast	High	Major	Loud	Legato
2	Moderately Slow	High	Major	Middle	Legato
2	Fast	Very Low	Major	Middle	Legato
Unhealthy					
4	Moderately Slow	Very Low	Minor	Quiet	Staccato
3	Very Slow	Very Low	Minor	Quiet	Staccato
3	Slow	Very Low	Minor	Loud	Staccato
3	Slow	Very High	Minor	Loud	Staccato
2	Very Slow	Very High	Minor	Quiet	Staccato
2	Very Slow	Low	Minor	Loud	Staccato
2	Slow	Very Low	Minor	Quiet	Legato
2	Slow	Low	Minor	Quiet	Staccato

were common among participants as listed in table 8, but all models generated throughout the study are included in Appendix A.

For both healthy and unhealthy models, the most popular choices agreed along several factors. Aligned with figure 8, almost all popular healthy compositions believe the melody should be fast and played in a major key generally in a higher pitch, while the unhealthy composition should be the opposite while also (for the most part) being staccato. However, participants were unable to agree on which dynamic indicated what level of health. While they mostly agreed healthy models should be in a loud or middle dynamic, they could not agree if unhealthy models should use quiet or loud.

While common trends are observed in individualized music models, our 52 participants generated a large sample of models with 39 unique healthy and unhealthy models. Models differed between participants, indicating participants may have perceived the healthiness of music very differently from others. As an example, figure 14 shows the healthy and unhealthy models for two participants. The two disagreed on every feature in the healthy model and only agree on the tempo and pitch of the unhealthy model. The vast number of models indicates few users may abide by every trend but are individualized in their exact preferences.

However, while there was a wide range of preferences for different music features, there were also shared models between participants. In all, almost half of all healthy models were identical to at least another model for a different participant. Only 28 (53.9% of) participants had unique healthy models. Identical unhealthy models were slightly less common, with 31 (59.6% of) participants having unique unhealthy models. Given the number of variables the participants had to choose from, there were a total of 288 unique possible models that could have



(a) Healthy interpretation for participant 1
(b) Healthy interpretation for participant 2
(c) Unhealthy interpretation for participant 1
(d) Unhealthy interpretation for participant 2

Fig. 14. The single line melody from survey 2 based on two participant's personalized models

been designed. It is highly unlikely that our participant pool of 52 users would have so much overlap if each user had completely unique preferences. Thus, we must conclude that, while there is still great variation in the personalized models, they are not truly unique to the individual.

5.4 What External Factors Affect the Perceived Healthiness of Music?

Due to the presence of trends, yet strong individuality in our data, we further analyzed the impact of demographics, musical background, and frequency of listening to music on the perceived healthiness of music. We also compared responses to the pleasantness of the music to the ratings of health in the second survey to determine if the two are related.

5.4.1 Demographics. We analyzed all 45 Likert scale questions in our study between different demographic options using Kruskal-Wallis tests. Of the 45 tests to detect differences in responses between the genders, one test came back significant ($H = 5.832$, $p = 0.016$) . This test indicates female participants perceived the healthiness of physical energy within the all-healthy three-line instrumental melody to be healthier than male participants. Similarly, two tests for race showed significance. Ironically, one of the two tests is for the perceived healthiness of physical energy within the all-unhealthy three-line instrumental melody ($H = 12.124$, $p = 0.007$) , which indicates participants that self-identify as Hispanic or Black or African American considered this melody to be healthier than other participants. Another significant test within the race demographic was caused by white participants perceiving the healthiness of the quiet dynamic as lower than participants from other races ($H = 13.572$, $p = 0.004$). Due to the small number of significant tests (2.22%), we conclude that the discrepancies within the general demographics are due to small sample size or external features, such as personal preferences, that we are not accounting for in our study.

As was done with the general demographics, each music demographic was compared across all 45 Likert scale questions to detect any significant differences in responses. We detected three significant tests across all music demographics. Participants with a voice training background perceived composition U to be less healthy than participants with other music backgrounds ($H = 10.097$, $p = 0.039$). Two tests regarding the frequency with which participants listened to music were also significant. First, participants that listened to music several hours a day and a few hours a month perceived composition E to be less healthy than participants that listened to music a few hours a week or a few hours a day ($H = 9.657$, $p = 0.022$). This result is mildly notable, as participants that listen to very little music and participants that listen to a lot of music provided very similar responses, while participants that listen to moderate music provided a different perception. Lastly, participants that listened to music only a few hours a month perceived the very fast tempo to be less healthy than the other groups (H

= 8.49, p = 0.037). However, once groups with 2 or fewer participants were removed, this test was no longer significant (H = 5.018, p = 0.081). Thus, we can reach a similar conclusion to the general demographics that the small number of significant tests (1.11%) indicates they are also likely a result of a small sample size rather than actual demographic differences.

Due to the inconsistent and small number of positive tests, we must conclude these demographics did not appear to influence participant interpretations, and that each participant's responses are likely indicative of personal identity traits other than the ones we accounted for.

5.4.2 The Effect of Pleasantness on Music Healthiness. As discussed earlier, we theorized participants may associate healthier music with music they find more pleasing, while unhealthy music could be more associated with unnerving or unpleasant melodies. Therefore, for each melody in the second survey, we asked participants an extra 5-point Likert scale question, assessing how pleasant they found the melody. We used the Pearson correlation coefficient to compare the pleasantness scores to the health scores for each melody. In doing so, we hoped to identify the correlation between the participant's impression of healthiness and pleasantness.

Our results show perceived healthiness and perceived pleasantness of a melody are highly correlated. When analyzing all seven single-line melodies, all showed a significant correlation between health score and pleasantness. Both instrumental and piano triple-line melodies also showed a significant correlation between pleasantness and healthiness.

In all, while neither demographics information nor music background show any significant effect on the perceived healthiness of melodies, we see a strong correlation between how pleasant participants found a melody and how healthy they perceived it. This could provide an avenue for better predicting user responses, as an understanding of what they find pleasant could extend to explain what they find healthy.

6 DISCUSSION AND IMPLICATIONS

6.1 Clear Trends Within Participant Healthiness Impression of Music Features

Our results contain the clear existence of trends explaining how participants perceive the healthiness of music. All five categories of musical features showed a trend, with commonly perceived healthy and unhealthy values. These trends ranged from simple, almost linear trends where healthiness improved as the tempo increased, to more complex, musically ingrained trends, such as more musically common pitches seeming healthier. We also determine the strength each feature appears to have on the perceived health level of a melody when they are all combined. Some features such as tempo are strong indicators of how users will view the music, while features such as smoothness result in no significant difference.

At the most basic level, the presence of these trends indicates users can perceive healthiness through music, showing that our method is viable. This means future health monitoring systems can develop similar methods and deliver information about one's healthiness using music. We believe this will be especially useful to the field of health awareness. Being able to use sonification to present health data creates brand new avenues for users to receive their data. The delivery of this data can become more ubiquitous than ever before, simply requiring the user to be within hearing range of the speaker generating the noise. This can be presented regardless of the user's activities, possibly being developed into calculating the times and activities where the user will be most receptive to the data, rather than waiting for them to interact with a necessary device.

Displaying behavioral health data through music may allow users to better internalize their own health status. It may also aid users in sustained behavior change. Music is well-known and widely accepted for its ability to invoke emotions. Presenting users with their health status through music may trigger emotional connection, spurring them to make healthier decisions than if they were simply told their health information.

However, the presence of trends may indicate a potential flaw with using sonification: privacy. A great strength of sound-based models is also a large potential weakness: they can be heard from anywhere by anyone within the intended area. Initially, we believed the personalized models would prevent unintended actors from gaining information about the user's health. However, our data shows this may not be the case. The trends we identified mean overlap between user perceptions are more common than we thought, so an unintended listener may be able to partially, or fully, interpret the user's healthiness. While it is entirely possible someone hearing the melody may not understand the information it contains, or that it contains information at all, some users may feel uncomfortable taking that risk. Rather, these models may need to be played at times only the user or people they trust would be able to hear them. This means they could be played at times such as putting on headphones, or perhaps even as an alarm in the morning.

The pleasantness of music appeared to be a strong indicator of the perceived healthiness. This provides a background for future music-based information displays and may be used to create simpler music systems. Rather than needing to define a user's exact preferences, it may be possible to simply reduce the user's music tastes to what they find pleasant, or enjoy and infer their health level interpretations from those.

6.2 Multiple Aspects of Health Can Be Expressed Simultaneously Through Musical Melodies.

We identified and showed participants' ability to perceive consistent health levels within three-line melodies. We also showed different instruments are perceived to represent different types of health energy. This means it may be possible to provide multiple types of health information simultaneously.

However, our results indicate participant perceptions may differ between the three-line and single-line melodies. Specifically, when comparing each of the three compositions, participants perceived the piano-based three-lined melody as being healthier than those with a single line. This difference indicates user perception of features' healthiness in a three-line melody may function differently than in a single melodic line. To be practical, three-line melodies may also require their own models to be generated the same way as single-line melodies to better communicate health data.

Without the capability to provide multiple kinds of health data simultaneously, our music-based method would be forced to play the melodies in a sequential manner. Given that these melodies sometimes last as long as 49 seconds, this could become a very monotonous and drawn out process. Thus, future implementations will need to further delve into the problem of presenting multiple types of health data in tandem.

While our method may require personalized models to work, there may be other ways to represent multiple kinds of data simultaneously. As we saw in the feature trends in figure 8, users interpret tempo and pitch through relatively clear scales. It could be possible to use these scales to communicate the data. Rather than combining all five features, the features could be reorganized in attempts to invoke similar, multi-level trends. Then, each musical element could represent a different piece of health information. As an example, suppose someone abides by the commonalities we discussed and finds tempos and pitches to be healthier the faster and higher they get. Rather than using multiple lines, the tempo could simply demonstrate one kind of data, while the pitch represents another. The melody could be played very fast but very low, indicating the health data represented by tempo should be perceived as healthy, while the health data represented by pitch should be viewed as unhealthy.

Additionally, the consistent responses across the three-line melodies show results are unlikely to change as our definition of health becomes more specific. In the instrumental three-lined melodies, we narrowed each melodic line to a more specific type of health (cognitive, emotional, or physical). Our tests showed no significant differences between any lines, the overall healthiness of these melodies, and the healthiness of the identical single-line melodies. Thus, we can assume participants' perceptions did not change as health got narrowed to a more specific type.

6.3 There Are Commonalities and Differences in Healthiness Impressions of Music Features Among Participants.

Through an analysis of the customized music models for each user, we identified that our 52 participants created 39 different models for both health and unhealthy perceptions, indicating a wide variety of preferences. Yet, these models still appear to generally follow the trends established when analyzing participants' direct responses. Healthy models predominately utilize faster tempos, while unhealthy models use slower ones. Future research can attempt to identify why participants deviate from the common trends.

Upon identifying how and why participant responses fluctuate, these fluctuations could be targeted, thus vastly reducing the number of personal models. For example, we identified faster tempos are often healthier, but users' responses sometimes vary between the three fast tempos. Therefore, two fast tempos and two slow tempos could potentially be removed. Users would still be able to perceive healthiness in generally the same way, and there would not be as much variation in the models. However, if developed this way, the method may begin to sacrifice privacy on behalf of ease of use. Fewer models mean there will likely be more overlap, especially when users follow the trends discussed throughout this paper. More overlap means external viewers are more likely to understand the user's model, potentially gathering data the user does not want them to have.

6.4 Possible External Factors Affect the Perceived Healthiness of Music.

Analyzing the impact of demographics and music background on music healthiness perception portrayed few significant differences in participant responses. While there were the occasional signs of significance,these are likely a signal of small sample size due to their inconsistent appearances. Some demographic groups were very small, making them vulnerable to outlying data. It is possible other demographics, both musical and non-musical may be viable predictors of user responses, but not the ones we collected in our study. Should a predictor be found, this could be used to further generalize the data, decreasing the burden on the user to provide data to generate customized models.

The lack of significance in the impact of musical ability and demographics on music healthiness impression is notably interesting, as music psychologists note both can influence music preference [40]. Therefore, one would expect these would also influence the perceptions of how musical features are perceived as healthy or unhealthy. Yet, our results show no clear trend. This may indicate determining the healthiness of music relies on different existing knowledge than music taste. In some regards, this difference may make sense. Music taste tends to be a relatively intuitive decision. While this will sometimes require thought such as disliking a song due to disagreeing with the theme of the lyrics, it is common for one to decide quickly and easily determine what they like.

Determining health levels, however, may require more conscious effort. This study appears to be the first of its kind, requiring individuals to think of a musical melody as either healthy or unhealthy. While we assumed that participants would simply classify features they like as healthy and features they do not like as unhealthy, they may have approached the questions very differently, and used metrics other than musical preference to determine healthiness.

There is another possible factor that may have influenced how people perceive healthiness through the music in our study. Bach Cococo's music style is based on that of Johann Sebastian Bach [25], the famous 18th-century German composer. This means Bach CoCoCo's music, and our music by extension is based on the western music style, which are musical styles originating from Europe [41]. Western music follows the same set of rules, regardless of how different the melodies sound. As the name implies, non-western music originated outside of Europe and follows different rules depending on the geographical location the music originated from. After becoming accustomed to one style, other music styles may sound strange. While barriers between music styles have been broken over time, they were highly present at the time of Bach, meaning our music is based on the western music style rather than the more common modern-day combinations of styles. It is possible our method, which relies on western music and features may inadvertently exclude individuals from non-European cultures. Our participant pool suffered from a lack of diversity in this aspect, with only two participants citing Asian or middle eastern ethnicity. Thus, we recommend further research to ensure ethnicity does not influence user perceptions.

7 FURTHER LIMITATIONS AND FUTURE DIRECTIONS

The largest limitation not mentioned earlier in the discussion section is the fact that our study does not account for the impact of the melody itself. Although participants listened to three different melodies during the study, the melodies were never established in a way that there were no alterations other than the tune. Therefore, it is possible that changes that were attributed to musical traits may, in part, be due to changes in the tune itself. Future implementations should include multiple melodies, with the same number of lines, containing the same health models to account for this possibility.

Additionally, our second survey only analyzes 7 unique compositions. These were chosen to provide a wide enough range of possible interpretations while not being so taxing participants begin the survey, and stop without completing it. While some participants did leave surveys unfinished, our results likely would have been more decisive if we had included more of the 32 possible compositions in the study. In future studies, it could be possible to expand the selection of melodies.

This current work only considers user perceptions at two points in time, but user perceptions may change greatly over time. What sounds healthy while a person is at work may not sound healthy to a user relaxing at home, or even in the same context several weeks apart. A future study should account for this possibility and be designed to assess the impact and viability of portraying health data over time through a longitudinal study.

Future studies can also expand into more complex musical features. The features in this study were basic but can be applied to essentially any melody. However, there are other musical features that could be included in future work, such as slurs, chords, accents, crescendos, decrescendos, and vibratos. These features either did not make noticeable differences in MuseScore's MP3s, required more strategic placement in the melody to be effective (thus lowering the generalizability) or require a greater music theory background than the current research team is qualified to provide.

In this study, we did not include any biological data as the ultimate purpose was to investigate whether users can perceive healthiness through music. Our immediate next step is to develop a mobile application where participants can listen to personalized melodies generated from their real-time biological data. Using IoT devices, such as FitBit and personal phones, we plan to devise a study that collects users' biobehavioral data and creates music representing the healthiness of this data for users to listen to. We plan to focus on aspects of participants' lives they have direct control over, such as promoting more activity or sleep. We can then investigate how successful our method is in encouraging behavior change by analyzing subsequent data from those devices.

8 CONCLUSION

We propose an innovative use of sonification for delivering and communicating personal health information. We design and run a pilot study to capture the perceived healthiness of music features and evaluate how well the method can communicate health information through melodies. We find common trends among participants' healthiness impression of music features such as associating faster melodies with higher healthiness and slower melodies with lower healthiness. Analysis of personal and musical demographics showed little significance in participant responses, indicating they cannot be predictive of participants' music healthiness perception. However, we discovered a correlation between a melody's interpreted healthiness and perceived pleasantness, alluding that participants may have used pleasantness as a metric to determine healthiness. To the best of our knowledge, this is the first study that evaluates the feasibility of using personalized music models for communicating personal health information. While future research with a larger participant pool can replicate these results, we conclude that our method of using personalized music models to deliver health information is viable and offers multiple useful implications for designing personal health applications based on customized sonification of health data.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF-IIS-1816687).

REFERENCES

[1] Judith Amores and Pattie Maes. 2017. Essence: Olfactory Interfaces for Unconscious Influence of Mood and Cognitive Performance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 28–34. https://doi.org/10.1145/3025453.3026004

[2] Ishwarya Ananthabhotla and Joseph A. Paradiso. 2018. SoundSignaling: Realtime, Stylistic Modification of a Personal Music Corpus for Information Delivery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (Dec 2018), 1–23. https://doi.org/10.1145/3287032

[3] Md Mahfuz Ashraf, Najmul Hasan, Lundy Lewis, Md Rashadul Hasan, and Pradeep Ray. 2016. A Systematic Literature Review of the Application of Information Communication Technology for Visually Impaired People. *International Journal of Disability Management* 11 (2016), e6. https://doi.org/10.1017/idm.2016.6

[4] Luke Barrington, Michael J. Lyons, Dominique Diegmann, and Shinji Abe. 2006. Ambient Display using Musical Effects. In *Proceedings of the 11th international conference on Intelligent user interfaces - IUI '06*. ACM Press, 372. https://doi.org/10.1145/1111449.1111541

[5] Michał Bujacz and Paweł Strumiłło. 2016. Sonification: Review of Auditory Display Solutions in Electronic Travel Aids for the Blind. *Archives of Acoustics* 41, 3 (Sep 2016), 401–414. https://doi.org/10.1515/aoa-2016-0040

[6] Sixian Chen, John Bowers, and Abigail Durrant. 2015. "Ambient walk": a mobile application for mindful walking with sonification of biophysical data. In *Proceedings of the 2015 British HCI Conference*. ACM, Lincoln Lincolnshire United Kingdom, 315–315. https://doi.org/10.1145/2783446.2783630

[7] Sunny Consolvo, Predrag Klasnja, David W. McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A. Landay. 2008. Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*. ACM Press, 54. https://doi.org/10.1145/1409635.1409644

[8] Sunny Consolvo, Ryan Libby, Ian Smith, James A. Landay, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, and Louis LeGrand. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, 1797. https://doi.org/10.1145/1357054.1357335

[9] H. Covington. 2001. Therapeutic music for patients with psychiatric disorders. *Holistic Nursing Practice* 15, 2 (Jan 2001), 59–69. https://doi.org/10.1097/00004650-200101000-00009

[10] The MuseScore developer community. 2021. *MuseScore*. https://musescore.com/en

[11] David Dobbelstein, Steffen Herrdum, and Enrico Rukzio. 2017. inScent: a wearable olfactory display as an amplification for mobile notifications. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, Maui Hawaii, 130–137. https://doi.org/10.1145/3123021.3123035

[12] Alice Eldridge. 2006. Issues in Auditory Display. 12, 2 (2006), 16.

[13] Jaakko Erkkilä, Marko Punkanen, Jörg Fachner, Esa Ala-Ruona, Inga Pöntiö, Mari Tervaniemi, Mauno Vanhala, and Christian Gold. 2011. Individual music therapy for depression: randomised controlled trial. *British Journal of Psychiatry* 199, 2 (Aug 2011), 132–139. https://doi.org/10.1192/bjp.bp.110.085431

[14] Yuan Feng, Suihuai Yu, Dirk van de Mortel, Emilia Barakova, Jun Hu, and Matthias Rauterberg. 2019. LiveNature: Ambient Display and Social Robot-Facilitated Multi-Sensory Engagement for People with Dementia. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, 1321–1333. https://doi.org/10.1145/3322276.3322331

[15] J.P. Fritz and K.E. Barner. 1999. Design of a haptic data visualization system for people with visual impairments. *IEEE Transactions on Rehabilitation Engineering* 7, 3 (Sep 1999), 372–384. https://doi.org/10.1109/86.788473

[16] Mads Frost, Afsaneh Doryab, Maria Faurholt-Jepsen, Lars Vedel Kessing, and Jakob E. Bardram. 2013. Supporting disease insight through data analysis: refinements of the monarca self-assessment system. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 133–142. https://doi.org/10.1145/2493432.2493507

[17] Çağlar Genç, Yavuz Ali Ekmekçioğlu, Fuat Balci, Hakan Ürey, and Oguzhan Özcan. 2019. Howel: A Soft Wearable with Dynamic Textile Patterns as an Ambient Display for Cardio Training. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–6. https://doi.org/10.1145/3290607.3312857

[18] Michele Geronazzo, Federico Avanzini, and Federico Fontana. 2015. Use of personalized binaural audio and interactive distance cues in an auditory goal-reaching task. Georgia Institute of Technology.

[19] Michele Geronazzo, Federico Avanzini, and Federico Fontana. 2016. Auditory navigation with a tubular acoustic model for interactive distance cues and personalized head-related transfer functions: An auditory target-reaching task. *Journal on Multimodal User Interfaces* 10, 3 (Sep 2016), 273–284. https://doi.org/10.1007/s12193-016-0221-z

[20] T Grosshauser, C Spieth, and T Hermann. 2011. WEARABLE SENSOR BASED REAL-TIME SONIFICATION OF MOTION AND FOOT PRESSURE IN DANCE TEACHING AND TRAINING. *J. Audio Eng. Soc.* 1, 1 (2011), 11.

[21] Vincent Hayward, Oliver R. Astley, Manuel Cruz-Hernandez, Danny Grant, and Gabriel Robles-De-La-Torre. 2004. Haptic interfaces and devices. *Sensor Review* 24, 1 (Mar 2004), 16–29. https://doi.org/10.1108/02602280410515770

[22] Nicholas Lane, Mashfiqui Mohammod, Mu Lin, Xiaochao Yang, Hong Lu, Shahid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell. 2011. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. In *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*. IEEE. https://doi.org/10.4108/icst.pervasivehealth.2011.246161

[23] Mu Lin, Andrew T. Campbell, Tanzeem Choudhury, Nicholas D. Lane, Mashfiqui Mohammod, Xiaochao Yang, Hong Lu, Giuseppe Cardone, Shahid Ali, Afsaneh Doryab, and Ethan Berke. 2012. BeWell+: multi-dimensional wellbeing monitoring with community-guided user feedback and energy optimization. In *Proceedings of the conference on Wireless Health - WH '12*. ACM Press, 1–8. https://doi.org/10.1145/2448096.2448106

[24] Jack M. Loomis, Reginald G. Golledge, and Roberta L. Klatzky. 1998. Navigation System for the Blind: Auditory Display Modes and Guidance. *Presence: Teleoperators and Virtual Environments* 7, 2 (Apr 1998), 193–203. https://doi.org/10.1162/105474698565677

[25] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3313831.3376739

[26] Lin Lu, Jiayao Zhang, Yi Xie, Fei Gao, Song Xu, Xinghuo Wu, and Zhewei Ye. 2020. Wearable Health Devices in Health Care: Narrative Systematic Review. *JMIR mHealth and uHealth* 8, 11 (Nov 2020), e18907. https://doi.org/10.2196/18907

[27] Sylwia Makomaska. 2018. Audiomarketing - music as a tool for inderect persuasion. *Interdisciplinary Studies in Musicology* 10 (2018), 77–86. https://pressto.amu.edu.pl/index.php/ism/article/view/15020

[28] Victor Mateevitsi, Khairi Reda, Jason Leigh, and Andrew Johnson. 2014. The health bar: a persuasive ambient display to improve the office worker's well being. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 1–2. https://doi.org/10.1145/2582051.2582072

[29] Jason McGahan, Sean Lacey, Con Burns, Tim Gabbett, and Cian O'Neill. 2020. Variation in Training Load and Markers of Wellness. *Journal of Australian Strength & Conditioning* 27 (2020), 6–14.

[30] Enid Montague and Jennifer Perchonok. 2012. Health and Wellness Technology Use by Historically Underserved Health Consumers: Systematic Review. *Journal of Medical Internet Research* 14, 3 (May 2012). https://doi.org/10.2196/jmir.2095

[31] Surabhi Nath. 2020. Hear Her Fear: Data Sonification for Sensitizing Society on Crime Against Women in India. In *IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*. ACM, Online India, 86–91. https://doi.org/10.1145/3429290.3429307

[32] Joseph Newbold, Nicolas E. Gold, and Nadia Bianchi-Berthouze. 2020. Movement sonification expectancy model: leveraging musical expectancy theory to create movement-altering sonifications. *Journal on Multimodal User Interfaces* 14, 2 (Jun 2020), 153–166. https://doi.org/10.1007/s12193-020-00322-2

[33] Joseph W. Newbold, Nadia Bianchi-Berthouze, and Nicolas E. Gold. 2017. Musical Expectancy in Squat Sonification for People Who Struggle with Physical Activity. In *Proceedings of the 23rd International Conference on Auditory Display - ICAD 2017*. The International Community for Auditory Display, University Park Campus, 65–72. https://doi.org/10.21785/icad2017.008

[34] Joseph W. Newbold, Nadia Bianchi-Berthouze, Nicolas E. Gold, Ana Tajadura-Jiménez, and Amanda CdC Williams. 2016. Musically Informed Sonification for Chronic Pain Rehabilitation: Facilitating Progress & Avoiding Over-Doing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 5698–5703. https://doi.org/10.1145/2858036.2858302

[35] Thomas H. Nown, Priti Upadhyay, Andrew Kerr, Ivan Andonovic, Christos Tachtatzis, and Madeleine A. Grealy. 2022. A Mapping Review of Real-Time Movement Sonification Systems for Movement Rehabilitation. *IEEE Reviews in Biomedical Engineering* (2022), 1–16. https://doi.org/10.1109/RBME.2022.3187840

[36] Marianna Obrist, Alexandre N. Tuch, and Kasper Hornbaek. 2014. Opportunities for odor: experiences with smell and implications for technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2843–2852. https://doi.org/10.1145/2556288.2557008

[37] Frederick Harry Pitts, Eleanor Jean, and Yas Clarke. 2021. *Rhythmanalysis, Concrete Abstraction and the Quantified Self: Sonification and Performance Research as Remediation of Data*. Research in Urban Sociology, Vol. 17. Emerald Publishing Limited, 209–226. https://doi.org/10.1108/S1047-004220210000017016

[38] Nimesha Ranasinghe, Thi Ngoc Tram Nguyen, Yan Liangkun, Lien-Ya Lin, David Tolley, and Ellen Yi-Luen Do. 2017. Vocktail: A Virtual Cocktail for Pairing Digital Taste, Smell, and Color Sensations. In *Proceedings of the 25th ACM international conference on Multimedia*.

ACM, Mountain View California USA, 1139–1147. https://doi.org/10.1145/3123266.3123440
[39] Kyle Rector, Keith Salmon, Dan Thornton, Neel Joshi, and Meredith Ringel Morris. 2017. Eyes-Free Art: Exploring Proxemic Audio Interfaces For Blind and Low Vision Art Engagement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 93 (sep 2017), 21 pages. https://doi.org/10.1145/3130958
[40] Thomas Schäfer and Peter Sedlmeier. 2009. What makes us like music? (2009), 4.
[41] Catherine Schmidt-Jones. 2017. *4.8: Classifying Music*. https://human.libretexts.org/Bookshelves/Music/Book%3A_Understanding_Basic_Music_Theory_(Schmidt-Jones)/04%3A_Defintions/4.08%3A_Classifying_Music
[42] Skipper Seabold and Josef Perktold. 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.
[43] Aneesha Singh, Stefano Piana, Davide Pollarolo, Gualtiero Volpe, Giovanna Varni, Ana Tajadura-Jiménez, Amanda CdeC Williams, Antonio Camurri, and Nadia Bianchi-Berthouze. 2016. Go-with-the-Flow: Tracking, Analysis and Sonification of Movement and Breathing to Build Confidence in Activity Despite Chronic Pain. *Human–Computer Interaction* 31, 3–4 (Jul 2016), 335–383. https://doi.org/10.1080/07370024.2015.1085310
[44] Jaime Snyder, Mark Matthews, Jacqueline Chien, Pamara F. Chang, Emily Sun, Saeed Abdullah, and Geri Gay. 2015. MoodLight: Exploring Personal and Social Implications of Ambient Display of Biosensor Data. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, 143–153. https://doi.org/10.1145/2675133.2675191
[45] Moritz Stefaner and Sussane Jaschko. n.d.. Data Cuisine. https://data-cuisine.net/
[46] B.H. Suits. 1998. Tuning. https://pages.mtu.edu/~suits/notefreqs.html
[47] Survivor. 1982. Eye of the Tiger.
[48] Quan T Tran and Elizabeth D Mynatt. n.d.. Music Monitor: Ambient Musical Data for the Home. (n.d.), 4.
[49] Yon Visell, Alvin Law, and Jeremy R. Cooperstock. 2009. Touch Is Everywhere: Floor Surfaces as Ambient Haptic Interfaces. *IEEE Transactions on Haptics* 2, 3 (Jul 2009), 148–159. https://doi.org/10.1109/TOH.2009.31
[50] Bartlomiej P Walus, Sandra Pauletto, and Amanda Mason-Jones. 2015. Evaluating the use of sonification and music to support the communication of alcohol health risk to young people: initial results. Georgia Institute of Technology.
[51] Bartlomiej P. Walus, Sandra Pauletto, and Amanda Mason-Jones. 2016. Sonification and music as support to the communication of alcohol-related health risks to young people: Study design and results. *Journal on Multimodal User Interfaces* 10, 3 (Sep 2016), 235–246. https://doi.org/10.1007/s12193-016-0220-0
[52] Yun Wang, Xiaojuan Ma, Qiong Luo, and Huamin Qu. 2016. Data Edibilization: Representing Data with Food. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, San Jose California USA, 409–422. https://doi.org/10.1145/2851581.2892570
[53] D.A. Washburn and L.M. Jones. 2004. Could olfactory displays improve data visualization? *Computing in Science Engineering* 6, 6 (Nov 2004), 80–83. https://doi.org/10.1109/MCSE.2004.66

APPENDIX A: ALL PERSONALIZED MODELS CREATED DURING THIS STUDY

Healthy						Unhealthy					
Number of Participants	Tempo	Pitch	Key	Dynamics	Smoothness	Number of Participants	Tempo	Pitch	Key	Dynamics	Smoothness
3	Moderately Fast	High	Major	Middle	Legato	4	Moderately Slow	Very Low	Minor	Quiet	Staccato
3	Fast	Very High	Major	Loud	Legato	3	Very Slow	Very Low	Minor	Quiet	Staccato
2	Very Fast	Very Low	Major	Loud	Legato	3	Slow	Very Low	Minor	Loud	Staccato
2	Very Fast	Very High	Major	Middle	Legato	3	Slow	Very High	Minor	Loud	Staccato
2	Very Fast	Very High	Major	Loud	Staccato	2	Very Slow	Very High	Minor	Quiet	Staccato
2	Very Fast	Very High	Major	Loud	Legato	2	Very Slow	Low	Minor	Loud	Staccato
2	Very Fast	Low	Major	Quiet	Legato	2	Slow	Very Low	Minor	Quiet	Legato
2	Very Fast	High	Major	Middle	Staccato	2	Slow	Low	Minor	Quiet	Staccato
2	Very Fast	High	Major	Loud	Legato	1	Very Slow	Very Low	Minor	Quiet	Legato
2	Moderately Slow	High	Major	Middle	Legato	1	Very Slow	Very Low	Minor	Loud	Staccato
2	Fast	Very Low	Major	Middle	Legato	1	Very Slow	Very Low	Major	Quiet	Staccato
1	Very Slow	High	Minor	Quiet	Legato	1	Very Slow	Very High	Minor	Quiet	Legato
1	Very Fast	Very Low	Minor	Middle	Legato	1	Very Slow	Very High	Major	Loud	Staccato
1	Very Fast	Very Low	Minor	Loud	Legato	1	Very Slow	Low	Minor	Middle	Staccato
1	Very Fast	Very Low	Major	Middle	Staccato	1	Very Slow	Low	Minor	Loud	Legato
1	Very Fast	Very Low	Major	Loud	Staccato	1	Very Slow	High	Major	Quiet	Staccato
1	Very Fast	Very High	Major	Quiet	Staccato	1	Very Fast	Very Low	Minor	Quiet	Staccato
1	Very Fast	Very High	Major	Quiet	Legato	1	Very Fast	Low	Minor	Quiet	Staccato
1	Very Fast	Low	Minor	Middle	Staccato	1	Very Fast	Low	Major	Quiet	Staccato
1	Very Fast	Low	Major	Loud	Legato	1	Slow	Very Low	Minor	Quiet	Staccato
1	Very Fast	High	Minor	Middle	Legato	1	Slow	Very Low	Minor	Middle	Legato
1	Very Fast	High	Minor	Loud	Staccato	1	Slow	Very High	Minor	Quiet	Staccato
1	Very Fast	High	Major	Quiet	Legato	1	Slow	Very High	Minor	Quiet	Legato
1	Slow	Low	Major	Quiet	Legato	1	Slow	Very High	Major	Quiet	Legato
1	Moderately Slow	High	Minor	Middle	Legato	1	Slow	Low	Minor	Quiet	Legato
1	Moderately Fast	Very High	Major	Middle	Legato	1	Slow	Low	Major	Loud	Legato
1	Moderately Fast	Low	Minor	Loud	Legato	1	Slow	High	Minor	Loud	Staccato
1	Moderately Fast	Low	Major	Quiet	Legato	1	Moderately Slow	Very Low	Minor	Loud	Staccato
1	Moderately Fast	High	Minor	Middle	Staccato	1	Moderately Slow	Very Low	Minor	Loud	Legato
1	Moderately Fast	High	Minor	Middle	Legato	1	Moderately Slow	Very High	Minor	Quiet	Staccato
1	Moderately Fast	High	Major	Middle	Staccato	1	Moderately Slow	Very High	Minor	Quiet	Legato
1	Fast	Very Low	Major	Quiet	Legato	1	Moderately Slow	Very High	Major	Quiet	Legato
1	Fast	Very Low	Major	Loud	Staccato	1	Moderately Slow	Very High	Major	Middle	Staccato
1	Fast	Very High	Major	Quiet	Legato	1	Moderately Slow	Low	Major	Loud	Staccato
1	Fast	Very High	Major	Middle	Legato	1	Moderately Slow	High	Minor	Quiet	Legato
1	Fast	Low	Major	Quiet	Legato	1	Moderately Fast	Very Low	Minor	Loud	Staccato
1	Fast	Low	Major	Loud	Staccato	1	Moderately Fast	Very High	Major	Quiet	Staccato
1	Fast	High	Major	Middle	Staccato	1	Fast	Very High	Minor	Middle	Staccato
1	Fast	High	Major	Loud	Legato	1	Fast	Very High	Minor	Loud	Staccato