Skin Pattern Sonification
Using NMF-based Visual Feature Extraction and Learning-based PMSon

Yoon Chung Han

Media Arts and Technology,
University of California, Santa Barbara,
3309 Phelps Hall, Santa Barbara, CA USA
yoon@mat.ucsb.edu

Byeong-jun Han

School of Electrical Engineering,
Korea University
Seongbuk-gu, Anam-ro 145, Seoul, Korea
hbj1147@korea.ac.kr

ABSTRACT

This paper describes the use of sonification to represent the scanned image data of skin pattern of the human body. Skin Patterns have different characteristics and visual features depending on the positions and conditions of the skin on the human body. The visual features are extracted and analyzed for sonification in order to broaden the dimensions of data representation and to explore the diversity of sound in each human body. Non-negative matrix factorization (NMF) is employed to parameterize skin pattern images, and the represented visual parameters are connected to sound parameters through support vector regression (SVR). We compare the sound results with the data from the skin pattern analysis to examine how much each individual skin patterns are effectively mapped to create accurate sonification results. Thus, the use of sonification in this research suggests a novel approach to parameter mapping sonification by designing personal sonic instruments that use the entire human body as data.

1. INTRODUCTION

Every living organism has their own unique body patterns and skin textures. The position, size and condition of the body, are all factors contributing to skin textures diversity. Due to the distinct patterns, some of skin patterns or textures such as fingerprints, iris, earlobe and face are commonly used for security and identification, which are called biometric data. It is useful to identify people by comparing unique visual features and patterns of the biometric data. This data provides key information that is a valuable resource in the creation of personalized musical instruments, data sonification, and interactive sound art pieces.

In our previous research, Digiti Sonus, we explored various visual features of fingerprints and transformed the data of the features into sound. [1] The static data of fingerprints was turned into sound, and each individual could have their own sonic signature since the visual features of each fingerprint was unique. Thus, Digiti Sonus suggested a novel technique that uses the sonification of biometric data to create diverse audio results and enable personalized sound expression in both the art and scientific fields.

In this paper, we came up with an idea of expanding the boundary of body data, not just limiting the data to certain areas of the body, but opening it up to the larger area of the entire human body. Instead of limiting biometric data to such things as the fingerprints or iris, we expanded the spectrum of data to include the skin of the entire human body. The diversity of skin patterns and textures from each different body parts is a good resource to prove that it creates subjectivity of sound from one body. Therefore, each individual body could become a single and unified instrument creating different sounds by exploring the position and condition of the body.

However, skin patterns are not as clear as other biometric data such as fingerprints, iris, or earlobe. It varies depending on the conditions and locations of the body parts. Therefore, analyzing the subtle difference of skin patterns is particularly challenging because many techniques for image processing are needed concurrently. In order to solve this problem, we employed non-negative matrix factorization (NMF) [2] to learn the acquired skin pattern image and discover new image features. In addition, parameter sonification mapping (PMSon) allowed us to create a connection between the learned features and the parameters by sound. In order to optimize and discover diverse aspects of representation learning and parameter mapping, we carried out experiments under the acquired skin pattern image dataset using many configurations. We will explain more details about this below.

The aim of this research has two main goals: 1) to propose a multimodal interface for users to explore their sonic difference using skin patterns, and create various sound compositions using NMF-based visual feature extraction and learning-based PMSon 2) to suggest a possible way to allow visually impaired people to be able to explore their bodies through sound.

This paper is structured as follows: Section 2 presents related works, and a brief background on the skin patterns and the techniques to analyze the skin patterns will be discussed in Section 3. Section 4 explains how we link sound to skin pattern data using machine learning techniques. Section 5 describes the implementation of a skin pattern sonification framework with examples and its results, and finally Section 6 concludes our research and suggests future works.

2. RELATED WORKS

Across a broad range of scientific fields, there have been various experiments and studies on body data sonification. The resources of most studies are body data, which can be efficiently categorized by two main factors: static data and dynamic data. Static data are still images and patterns from the body, and dynamic data contains the movements and performances (e.g. rhythm, articulation, tempo, etc.) of the body. In this research, we chose the static data of the body instead of the dynamic data because of the unique static characteristics of skin/body patterns that can be easily mapped to data sonification. Training a machine to recognize the unique characteristics of those static data is an active field of study these days, and human-computer interfaces improves it toward further levels.

Many machine learning-based techniques have been used for body data sonification. There is a research on vocal EEG sonification using Kernel Regression Mapping Sonification (KRMS) for optimized mappings between data features and the parameter space of Parameter Mapping Sonification. [4] KRMS enhances the possibility of extended mappings by connecting localized points in input space with a specific output. SoniScan [5] allows users to sonify brain scan data for augmented diagnosis methods used in analyzing Alzheimer's dementia. This research examines the correlation between the brain image data and the sonification as a useful foundation for physicians to diagnose diseases and monitor patient treatments. For high-dimensional datasets, the Crystallization Sonification model [6] is designed to provide information about the intrinsic data dimensionality and the global data dimensionality. Sound allows users to display the clustering in high-dimensional datasets. EmotionFace [7] is a software interface for visually displaying the self-reported emotion expressed by music. It can be viewed as a facial expression whose auditory connection is the time synchronized, associated music.

One of our goals in this research is to suggest a new multimodal interface that can examine the skin data for both visually impaired people as well as for those who have normal vision by using body data sonification. Due to the power of sound, experiments on data sonification for the visually impaired have been active. Research about psychophysical scaling of sonification mappings [8] shows that there seem to be some data-to-display mappings where the majority of visually impaired participants disagree with the polarities preferred by sighted listeners. Thus, according to this research, the sonification for visually impaired people should be different. Gaze-contingent audio-visual substitution [9] is a good example that shows how the new auditory system helps in everyday life of the visually impaired. The prototype combines eye-tracking with depth measuring and sonification techniques, and the blind and visually impaired might learn to perceive gaze-dependent sound visually through the system. Furthermore, there is a system that transforms images to acoustic signals for visually impaired people [10]. Image information can be acquired through an audio signal in the interactive system, using machine learning algorithms which include object recognition and the classification of regions in the images.

A review of the related works shows that researchers have mostly focused on the sonification of human body gestures for scientific applications and medical uses. So far we have not yet found research that connects sound and skin pattern data as a new interface for musical expression. Although skin patterns are not as distinct as other biometric data, it has a lot of potentials for creating a personalized sound interface due to the personal difference of the body. Moreover, the previous research in skin pattern data sonification has not been actively experimented, which shows that this research has novelty. In order to resolve these issues, this paper suggests a new multimodal interface that creates a possible approach to sonify skin patterns for both scientific applications and creative multimedia applications.

3. SKIN PATTERN ANALYSIS

Various visual features of skin pattern and textures can possibly be mapped to the parameters of sonification. In this section, the characteristics of skin patterns and techniques to analyze the visual features of skin patterns are described in greater detail.

3.1. Challenges in Analyzing Skin Patterns

Skin patterns and textures have been deeply analyzed in the fields of dermatology and physiology. In medical use, many advanced equipment and techniques diagnose the conditions of the skin to examine for things like disease, cancer, or any ailments that might cause critical health problems. Unlike other distinct biometric patterns such as the fingerprint, iris, and earlobe, the visual features of skin patterns are vague and unclear. These features are also dependent on many factors such as age, gender, diet, hydration, amount of collagen and hormones and skin cares. [11] As skin ages, it becomes thinner and more easily damaged with the appearance of wrinkles. Thus, it is very subtly and continuously changing. In these days, advanced image processing techniques can analyze the texture of the skin and extract their distinct features. Current methodologies and techniques used to analyze and interpret skin patterns and textures have been developed enormously and are largely based on the human visual system.

Despite these innovations, there are many limitations to analyzing and reading the skin pattern data due to the visual analysis and complex patterns of the data. Also, it needs to be concurrently examined due to the unstable conditions and subjectivity of the skin pattern data over time. It is hard to analyze the data using only visual representation. Even though diagnostic accuracy is higher than ever, users might not be able to examine the results easily. Advanced computer processing techniques may allow physicians or doctors to discover disease in new ways; however the complex visual data might make it slow to discover the results.

One potential method to solve this problem is by means of translating the analyzed data into a new format that can be easily recognized, via the human auditory system. The human brain reliably matches visual and auditory signals from the same source even though the modalities are different from each other [9]. Alais and Burr [13] presented research that shows that mainly visual cues are used to localize an audio-visual stimulus if vision is unrestricted. In contrast, for visually impaired people, sound dominates vision. This shows that information coming from dominated sense affects the brain and the other senses. And secondary senses can support the dominated sense to better understand the data. Therefore, multisensory integration improves the readability of information. There have been already large numbers of multimodal interface design studies aiming to allow users to move toward the new level of perception and experience. Data sonification is an active topic in those studies and offers a new perspective for understanding spatial, complex and temporal data.

Therefore, the main hypotheses of this paper contain two parts: 1) auditory processing of skin pattern data will enhance understanding of the visual features of the data 2) Furthermore, exploring skin pattern data using sound can create a new musical interface that doesn't require extra traditional musical equipment or trained professional musical skills. Only using the body, users can create new combinations of sound to experience both artistic and experimental results.

3.2. Visual Feature Extractions of Skin Patterns

As we briefly described above, it is still hard to analyze personalized features from acquired skin pattern images because of their unclear patterns. Instead, each skin features can be mapped to parameters as subjective representations of data input, instead of distinct characteristics of each person.

One problem in the parameterization of skin features is that the visual features on the skin pattern images are not uniformly shaped. In other words, it is hard to define global skin features of a skin pattern image. For example, a palm image may contain clear wrinkles and strong palm lines. However, there are also so many small wrinkles distributed on the palm along with the clear palm lines, which is a challenging part to analyze deeply. Also, skin pattern images acquired from different body locations have the visual features in various numbers and distributed positions; for instance, data acquired from an arm can be different from those of legs because of different distributions of hairs, pores, and wrinkles on the skin. For the reason above, our method does not consider global skin features but only extracts locally distributed diverse skin features.

The most dominant features on the skin patterns are the wrinkles. In order to extract the wrinkles on skin pattern images, we applied a multispectral analysis proposed by Zhang et al.[14] Instead of using their original approach, however, we modified some parts of processing steps for extracting not only the major palm lines but also small wrinkles on the various body locations. The main improvements of the technique are: 1) Instead of using color mixture between red, green, blue, and NIR band, we only extracted saturation band in HSB color space, because the saturation information is clear enough to extract the wrinkles, and it is robust even if angles of camera or positions of external lights are changed; 2) We considered Gabor filter[14] with six parameters instead of the original Gabor function with four parameters in Zhang et al.'s method, and; 3) We increased the number of Gabor filters to sixteen for increasing accuracy to extract candidate wrinkles.

Figure 1 depicts the overall steps of multispectral analysis. In pre-processing stage, a skin pattern image is divided into multiple color bands. For example, in RGB color space, the image are divided into red, green, and blue bands. Since saturation band is able to discover most of wrinkles on the skin pattern image, our approach only considered the saturation band in HSB color space. Next, we apply Gabor filters onto the saturation image in order to discover directivity of the wrinkles on the image. The Gabor function we applied is defined as following:

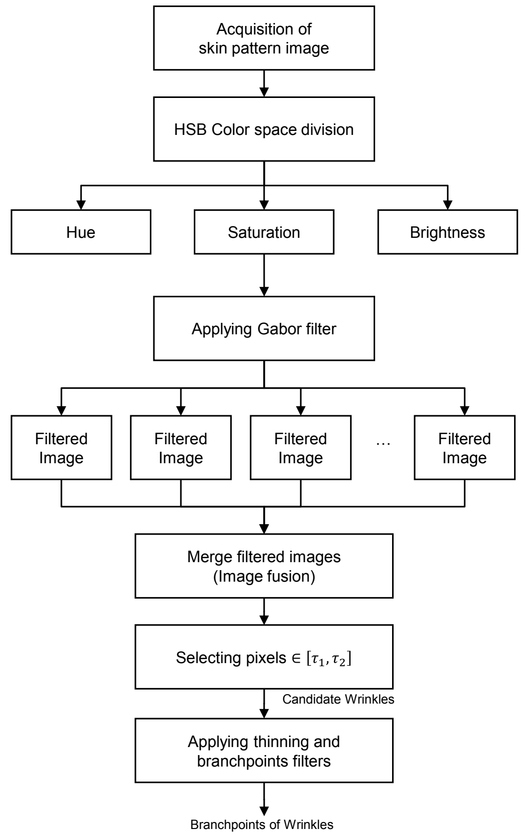

Figure 1. Overall process for Skin Pattern Analysis

$$g(x,y,\sigma,\theta,\lambda,\psi,\gamma) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} e^{i\left(2\pi\frac{x'}{\lambda}+\psi\right)} \quad (1)$$

In equation (1), x and y are two-dimensional coordinates specified by $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$, σ is a standard deviation for Gaussian function, θ is Gabor function's angle, λ represents wavelength of sinusoidal factor, ψ is phase offset, and γ is ellipticity of support of the Gabor function. Among these parameters, we fixed σ, λ, ψ, and γ, because only angle θ of Gabor filter was an important factor for detecting wrinkles on the skin pattern image. With $(\sigma,\lambda,\psi,\gamma) = (4,8,0,1)$, hence, the reduced form of Gabor function is defined as following:

$$g(x,y,\theta) = e^{-\frac{x'^2+y'^2}{32}} e^{i\frac{1}{4}\pi x'} \quad (2)$$

We divided the angle θ into sixteen steps in (2), in other words, θ in Gabor function was set to $j\pi/8$, where $j = 0,1,\ldots,15$. Figure 2(b) and (c) show examples of filtered images created on the steps in (2). On the other hand, since

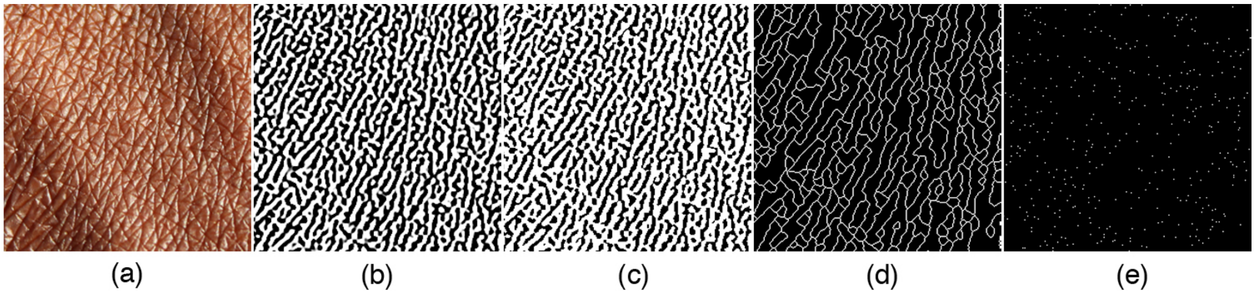

Figure 2. An example of skin pattern analysis: (a) original image, (b, c) filtered images, (d) selected wrinkle candidates (e) selected branch points (from left to right)

each filtered image shows only partial wrinkles, it is impossible to select entire wrinkles without merging all the filtered images. One simple solution to merge the filtered images is to average all the filtered images. This guarantees that all the energy of each pixel in the merged image is bound within the bound of Gabor filter response. After merging those filtered images, the wrinkle candidate pixels are selected within a predefined statistical range. Figure 2(d) shows an example of the selected wrinkle candidate pixels.

We bound a standard deviation of each pixel within the range of $[0, \infty)$, because the statistical distribution of saturation values of pixels in the original image cannot be predicted. In our approach, we assumed that the saturation values follow Gaussian distribution and we selected the bounding parameter as the standard deviation. After selecting the candidate pixels, our method applied thinning filters to extracted skeleton structures of candidate pixels only, and selected ridges and valleys of the extracted skeletonized structure. Figure 2(e) depicts an example of skeletonization and selected branch points.

3.3. NMF-based Visual Parameterization

The skin pattern pixels acquired from branch points of the local skin patterns, described in Section 3.2 need to be represented by limited numbers of parameters. However, in general, it is hard for sonification designers to parameterize such local skin patterns with lack of knowledge on dermatology. Instead of using such preliminary knowledge (dermatological information or related experience), we applied non-negative matrix factorization (NMF) [17] onto a group of local skin patterns and extracted common patterns from them. The reason why we adopted NMF instead of other well-known factorization methods, such as principal component analysis (PCA), is that it is not required to learn statistical distribution of such local skin patterns.

Local skin pattern images should be extracted before applying NMF in order to acquire basis patterns from a group of local skin pattern images. For a skin pattern image $\mathrm{I}\ (x, y) \in \mathbb{R}^{N_x \times N_y}$, branch point $P_k = (x_k, y_k)$, where $k = 1, 2, \dots, N$, is extracted using the visual feature extraction approach described in Section 3.2. In our next step, local skin pattern $I_k(x, y)$, where $x \in [x_k - N, x_k + N]$ and $y \in [y_k - N, y_k + N]$, is selected. In our approach, we set N as 5, therefore the size of each local skin pattern image becomes $31 \times 31 = 961$ dimensions. Also, we set a number of latent variables in NMF to 100. Figure 3 shows examples of selected local skin pattern images and basis vectors trained by NMF.

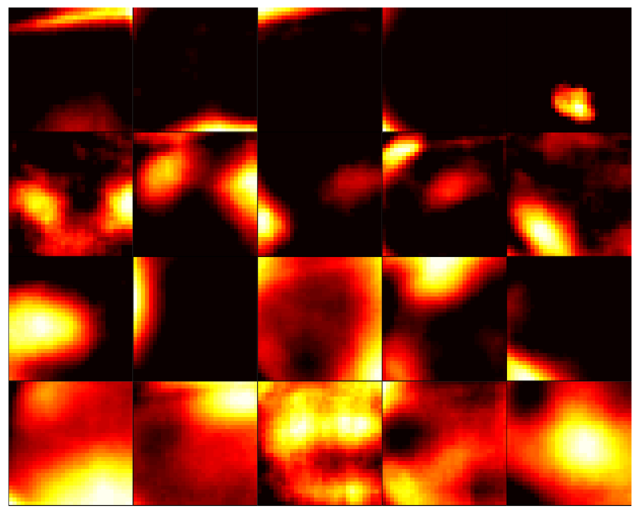

Figure 3. Selected basis images from local skin pattern images by NMF.

4. SKIN PATTERN SONIFICATION

4.1. Parameter Mapping Sonification (PMSon)

The visual features of skin patterns extracted from the process described above contain the dataset that is feasible to be mapped to parameters in data sonification. In order to sonify our data, we used Parameter Mapping Sonification (PMSon), which is the most common and easiest way to sonify data using and controlling multiple parameters. PMSon involves the association of information with auditory parameters for the purpose of data display. [16] Multivariate data is suitable to be displayed in PMSon because sound is inherently multidimensional. PMSon has been used in a wide range of applications, products and fields in these days. In order to select methods of mapping the data, sonification designers should examine possible mapping strategies to scale the data to the parameters. The mapping strategies should consider how the data is fitted for auditory display, which sound parameters should be used, and how to scale the data to the chosen parameters. It is also based on the cognitive science due to the auditory stimuli of the human body. In order to find the best

PMSon method, F. Grond and J. Berger suggest a general design process of effective PMSon [16]; data is selected and prepared from data domain, and through the mapping function, it is tuned to the parameters of sound synthesis. Data processing moves towards the sections of parameter domain, signal domain and human perception, which move backward to the previous steps to tune the process for achieving the best results. As the process suggests, the design of PMSon has interplay of connection between the data and the signal domains. Iterating the process on both worlds is the key point in creating effective sonification.

In this research, we mapped the visual features of skin pattern data into sound parameters by machine learning techniques. As we described above, we employed NMF to parameterize skin pattern images, and the represented visual parameters are connected to sound parameters through support vector regression (SVR). This will be described in Section 4.3.

4.2. FM Synthesis

In our approach on sound synthesis, we employed FM synthesis [17] to create diverse sound by controlling a small number of input parameters. FM Synthesis is easy to create diverse timbre of sound because of simplicity of implementation. In the basic FM synthesis, a modulator oscillator modulates the frequency of a carrier oscillator. By controlling a small number of input parameters, it creates a large range of output sounds. This was a main reason why we adopted FM synthesis. Our Max/MSP implementation has six input parameters: Carrier frequency, harmonicity ratio, duration, length, modulation index, and amplitude. Carrier frequency determines the fundamental carrier frequency of sound. Harmonicity ratio is defined as modulator frequency / carrier frequency, which controls what frequencies are mixed in either harmonic or inharmonic relationship. Duration is the playing length of sound and length determines the number of division on the envelopes. Modulation index and amplitude use envelopes, which is implemented with function object in Max/MSP. Modulation index is defined as modulator amplitude / modulator frequency, which affects the relative strength of partials to control the brightness of sound.

One limitation in our approach is that envelopes of modulation index and amplitude are represented by limited number of samples, instead of using polynomial representations. In our approach, thus we tried to set the sampling resolution of envelopes as high as possible for preserving original shapes of envelopes in the polynomials representation. Detailed information is presented in Section 5.

4.3. Learning Sonification Parameters

In order to automatically map the visual features to sonification parameters, we employed support vector regression (SVR) [18] as a mapping method. For representing multidimensional sonification parameters in FM synthesis, as presented in section 4.2, it is essential to generate multiple regression models to be mapped in the parameters. In the training phase, we assumed that pairs (I_k, S_k) of training data are provided, where I_k is an extracted visual feature, S_k is a sound parameter, and $k = 1,2, \ldots, K$ is an index of the pairs of training data. If those two values (I_k, S_k) are acquired, users can simply enter it to the SVR method. In other words, sonification designers can train his/her own skin pattern image-to-sonification and generate skin pattern sonification models using this methodology as a simple and easy way. Detailed explanation on the parameters and configurations for training the method will be provided in Section 5.

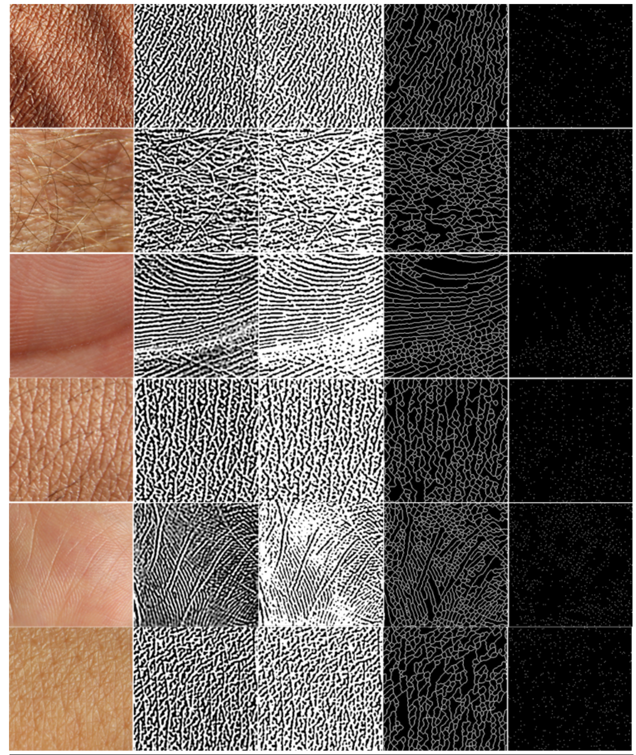

Figure 4. Selected examples of skin pattern images acquired from six different participants along with the process of skeletonization to extract visual features.

5. EXAMPLES AND RESULTS

5.1. Skin Pattern Samples and Analysis

We collected skin pattern data samples from twenty participants for our initial implementation; they were chosen based on ages, gender, races, and body conditions. The participants consisted of six Asians, three African Americans, four Indians, six Caucasians, and one Iranian. Ages varied; three early twenties, ten late twenties, four early thirties, and three late thirties. Lastly there were eleven males and nine females. The overall body conditions of the participants were good. The captured body parts were mostly from hands and arms. Some of data were from legs, back, neck, and face.

Twenty numbers of participants is obviously not enough to normalize and generalize the final data ranges and numbers of parameters; however we first needed to capture a small number of data to pre-determine the range of parameters, which attempts to prove our hypothesis in the initial stage with simple implementations. Based on the results and pre-determined results, we will conduct a formal user study with larger number of users in the near future.

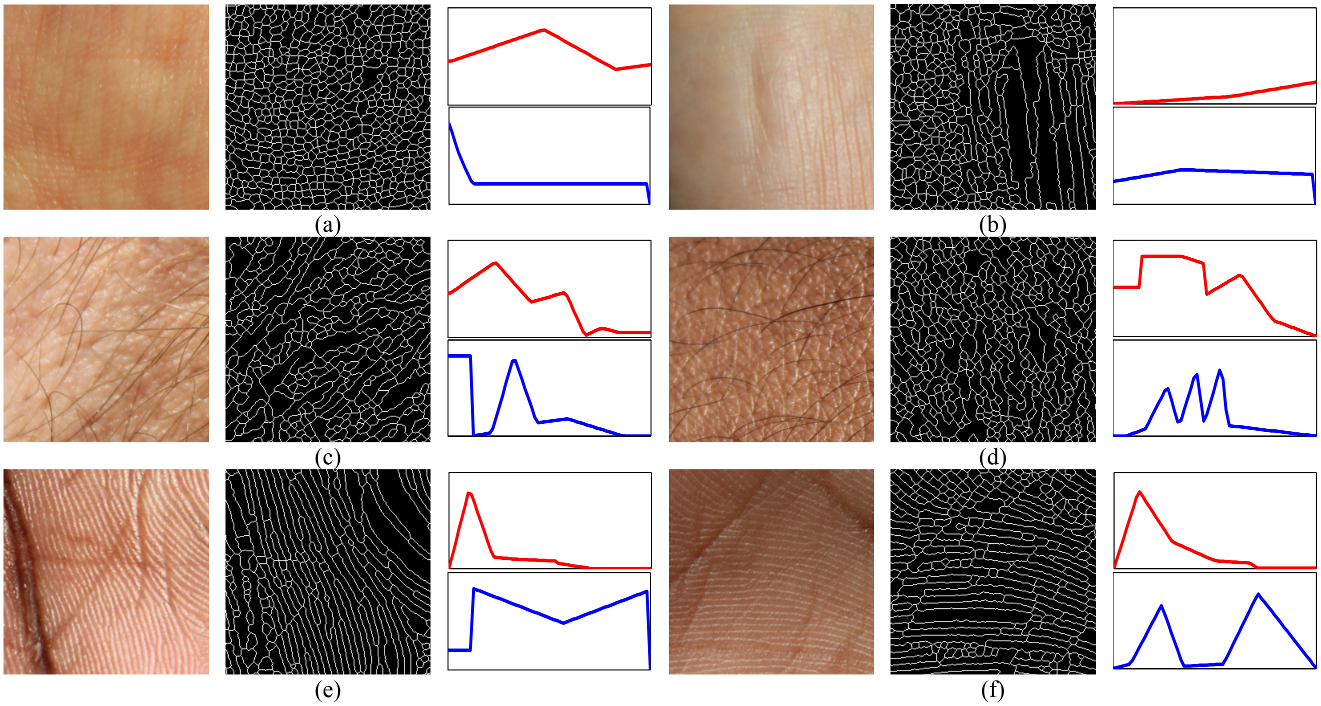

Figure 5. Each skin pattern image groups consist of (left) a raw image, (center) a skeletonized image to find branch points, (top-right, red-lined) modulation envelope, and, (bottom-right, blue-lined) amplitude envelope. (a) and (b) are comparably plain images and generated simple envelops of both modulation and amplitude. (c) and (d) are the most complex images with some noises (hairs) and formed envelopes with dramatic changes. Finally, (e) and (f) are not simple; however show constant patterns in their raw images.

In sonification parameter learning step, as described in Section 4.3, we employed a library for support vector machines (LIBSVM)[19], which is a well-known SVM and SVR tool and widely used in diverse research areas related to machine learning and pattern recognition. We used common C-SVR model and selected radial basis functions (RBFs) as kernel functions to attempt to obtain statistically generalized results. LIBSVM itself enables automatic data scaling for input data into normal distributions; however, it is generally not feasible on the logarithmically represented data such as frequency or amplitudes. Therefore, most of our sonification presets can be converted into log-scale, and we transformed frequency, modulation indices, and amplitude indices into log-scaled before training the SVR models.

5.2. Results and Discussion

Figure 5 shows a selected result of skin pattern sonification, of which original images were acquired from diverse body parts, such as palms, arms, and legs. Mostly, the body parts that consist of spread distribution of wrinkles, showed lower number of branch points, and therefore, such skin textures were sonified in much simpler way. On the other hand, skin pattern images from body parts that consist of complex skin patterns such as many wrinkles, pores, hairs, etc., generated very unusual sound using our approach. This implies that the number of branch points from visual feature extraction in section 3 is directly related to the complexity of sonification and timbre of sound.

In the case of local skin pattern images, used as the visual features in regression model, which predict sonification parameters, it seems that the complexity of local images was related to the shapes of modulation index and amplitude envelopes. If the local skin pattern images themselves are very complex and are represented with complicated combination of learned basis vectors, the envelopes are also formulated into very complex shapes. Consequently, modulations and amplitude dramatically have been changed in FM synthesis steps. In contrast, it was hard to recognize major changes from the sound results created by simple local images in modulations and amplitudes.

However, few number of skin pattern data affected the insufficient results to show the enough mapping ranges for learning-based sonification. Our examples and demonstrations briefly showed how the learning-based PMSon could be created using the skin pattern data; however, accuracy and diversity of results should be backed up by increasing the number of input data. Achieving larger number of the data will allow us to increase the accuracy and performance of parameter mapping based on learning process. More data we can acquire from many participants, the more various results of sonification they can receive. Thus, it is the most crucial part in this kind of learning-based analysis and implementation.

6. CONCLUSION AND FUTURE WORKS

Skin patterns contain very subtle, irregular, and complex data and it is challenging to examine and understand the data with their unique visual features. For normal people who want to create their own sonic signature or explore new musical experiment, skin patterns are promising resources because of their unique personal information and distinct characteristics of the patterns. However, it is hard to examine the multivariate and complex data only by eyes. Sonification seems to be a promising approach for reading and analyzing data since it can support the visuals with multisensory experience. Thus, the development of data sonification for analyzing skin pattern data is important for advancement of auditory display, multimodal interface, and even medical applications.

The most challenging part of data sonification is to design appropriate strategies to connect between the dataset and parameters of sound for effectively delivering the data by sound. Multidimensional and complex data acquired from complicated images requires the better ways to read it. As we experimented above, machine learning is able to enhance the ways to understand how to retrieve the right formats of the data, and to find appropriate connections between sorted data and audio parameters.

The framework described in this paper, is a research exploring the better way of sonifying the body data images by using learning-based techniques. It is beneficial to not only easily retrieve the data but also suggest the new method to use sound as an interesting medium to explore the human body in an experimental way. Thus, this framework suggests a new way of creating sound using the body as both scientific method and artistic and experimental application.

We are still in the early stage in this research. We need to search the other visual features of the skin patterns, not only just wrinkles or pores, but also more dynamic and deep levels of the skin pattern data. More advanced image processing techniques might be required to retrieve more visual features of the skin patterns. Also, different machine learning techniques can be applied to sort more accurate results. Obviously more diverse sound synthesis such as additive, granular or wave terrain syntheses could be applied to create diverse timbre and combination of sound. Next step of this project is to conduct user study to various ranges of people to generalize and stabilize the range of the parameters and to enhance quality of this framework. Comparison between images and sound is a key part, and the results from the study may improve the existing analysis methods.

7. ACKNOWLEDGMENT

We thank participants who were willing to try our interface with permission using their body data.

8. REFERENCES

[1] Y. Han, B. Han, and M. Wright. Digiti Sonus: Advanced interactive fingerprint sonification using visual feature analysis. In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), Daejeon+Seoul, South Korea, 2013.

[2] P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory. In D. Rumelhart and J. McClelland (Eds.), Parallel distributed processing, vol.1, chapter 6, MIT Press, Cambridge, 1986, 194-281.

[3] G. Hinton, A Practical Guide to Training Restricted Boltzmann Machines. UTML TR 2010–003, University of Toronto. 2010.

[4] T. Hermann, *et al.*, Kernel regression mapping for vocal EEG sonification. In Proceedings of the 14th International Conference on Auditory Display (ICAD), Paris, France, 2008.

[5] A. Rogińska, K. Friedman, and H. Mohanraj, Exploring sonification for augmenting brain scan data, In Proceedings of the 19th ICAD, Poland, 2013.

[6] T. Hermann, and H. RITTER, Crystallization Sonification of High-Dimensional Datasets, ACM Transactions on Applied Perception, vol. 2, 2005, pp. 550–558.

[7] E. Schubert, EmotionFace: prototype facial expression display of emotion in music, In Proceedings of the 10th ICAD, Sydney, Australia, 2004.

[8] B. N. Walker and L. M. Mauney, Psychophysical scaling of sonification mappings: a comparison of visually impaired and sighted listeners, In Proceedings of the 7th ICAD, Paris, France, 2001.

[9] L. Twardon, *et al.*, Gaze-contigent audiovisual substitution for the blind and visually impaired, Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference, 2013, pp.129-136.

[10] M. Banf and V. Blanz, Sonification of images for the visually impaired using a multi-level approach, 4th Augmented Human International Conference (AH'13), 2013.

[11] A. Sparavigna and R. Marazzato, An image-processing analysis of skin textures, Skin research and technology, vol.16, 2010, pp.161-167.

[12] A. Kapur, *et al.*, A framework for sonification of vicon motion capture data, In Proceedings of the 8th International Conference on Digital Audio Effects (DaFX-05), Madrid, Spain, 2005.

[13] D. Alais and D. Blurr, The ventriloquist effect results from near-optimal bimodal integration, Current Biology, vol. 14, no. 3, pp. 257–262, 2004.

[14] D. Zhang, *et al.*, An online system of multispectral palmprint verification, IEEE Transactions on Instrumentation and Measurement, vol.59, no.2, Feb. 2010, pp.480-490.H. G. Feichtinger and T. Strohmer, Gabor Analysis and Algorithms, Birhaeuser, 1998.

[15] T. Hermann, A. Hunt, and J. G. Neuhoff, The sonification handbook, Logos Publishing House, Berlin, 2011.

[16] J. Chowning, The synthesis of complex audio spectra by means of frequency modulation, Journal of the Audio Engineering Society, 21(7), 1973.

[17] Daniel D. Lee and H. Sebastian Seung, Learning the parts of objects by non-negative matrix factorization, Nature, vol.401, no.6755, Oct. 1999, pp.788-791.

[18] H. Drucker, *et al.*, Support vector regression machines, Advances in Neural Information Processing Systems, vol.9, 1996, pp.155-161.

[19] C.-C.Chang and C.-J.Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology, vol.2, no.27, 2011, pp.1-27.