Vibrotactile Rendering of Human Emotions on the Manifold of Facial Expressions

Shafiq ur Réhman and Li Liu
Email:{shafiq.urrehman, li.liu}@tfe.umu.se
Department of Applied Physics and Electronics (TFE),
Umeå University, 901-87 Umeå, Sweden.

***Abstract*— Facial expressions play an important role in every day social interaction. To enhance the daily life experience for the visually impaired, we present the Facial Expression Appearance vibroTactile System (FEATS), which uses a vibrotactile chair as the social interface for the visually impaired. An array of vibrating motors are mounted spatially on the back of an office chair. The Locally Linear Embedding (LLE) algorithm is extended to compute the manifold of facial expressions, which is used to control vibration of motors to render emotions. Thus, the chair could provide the visually impaired with on-line dynamic emotion information about the person he/she is communicating with. Usability evaluation of the system is carried out. The results are encouraging and demonstrate usability for the visually impaired. The user studies show that perfect recognition accuracy of emotion type is achieved by the FEATS.**

***Index Terms*— locally linear embedding, vibrotactile interface, visually impaired, manifold of facial expressions, personal mimic gallery**

I. INTRODUCTION

Human face can serve an important role in both human to human and human to machine communication [1]. Facial features can not only be used in face identification but also in emotion recognition [2], [3]. People use facial expression information to switch between conversation topics to avoid conflicts, to track their attempts to change emotional states or reactions of others, and to determine attitudes of individuals. Missing information from facial expressions makes the visually impaired extremely difficult to interact with others in social events. The visually impaired always has to rely on hearing to get other's emotional information since the voice is their main information source.

According to [4] more than 65% information carried out through nonverbal during face-to-face communication. It is difficult to understand complex emotions from voices alone, particularly, when the speaker is silent there is no way to get emotion information at all. Furthermore, human speech is not a reliable indicator of emotions either. For instance, it is rather difficult to recognize the emotion of fear from voices alone (i.e. only 42.93% correct recognition rate [5]). To enhance the visually impaired's social interactive ability, it is important to provide them with reliable, "on-line" emotion information. This paper addresses the challenging topic of *how to display dynamic facial expressions through a vibrotactile chair.*

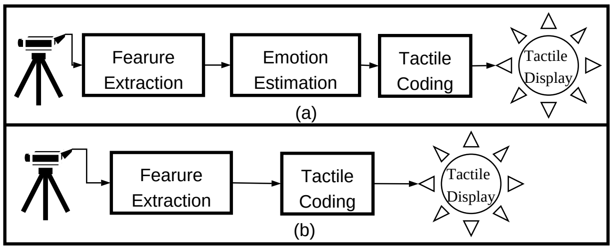


Figure 1. *Schematic diagram of our vibrotactile emotion rendering system: (a) TEARS system (FACS-based), (b) FEATS system (LLE-based).*

To avoid information overloading resulted from a direct translation from visual to tactile information, the visual signals have to be simplified and reduced in the order of $10000 : 1$. This implies that one has to employ a semantic representation of facial expressions. Previously, in our Tactile Facial Expression Appearance System (TEARS) [6], we used such a strategy to render human emotions: *estimate emotion by classifying facial expressions and then render directly the compact emotion information*. Specifically, we extract human emotion information from lips based on the assumption that the shape of the lips reflects the emotion associated with it [7]. The lips are extracted through an accurate and robust lip tracking technique [8] [9]. To characterize human emotions from the lips, the Ekman's Facial Action Coding System (FACS) [10] was used. In FACS, the action units are the smallest visibly discriminating changes in facial display. Combinations of action units can be used to describe complex emotion expressions [7]. In [6], action units were computed from dynamics of lip contours and were used to classify the underlying emotion into one of three types of emotion *happiness*, *surprise* and *sadness*, which are closely related to quite dissimilar shapes of the lips. The complete system diagram is shown in Fig. 1(a), consisting of four blocks: *feature extraction block*, *emotion estimation block*, *tactile coding block* and *tactile display block*.

This paper is an extension of the work presented in the paper "Manifold of Facial Expression for Tactile Perception" by Shafiq ur Réhman et al., which appeared in the Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP2007), 2007, Chania, Crete, Greece. © 2007 IEEE.

This work is supported by Swedish Research Council (Vetenskapsrådet (www.vr.se)).

Contact Author's email: shafiq.urrehman@tfe.umu.se

Although the system was demonstrated successfully but we find some weak points in our previous system:

- *only three (can be extended to six) prototypic expressions, i.e. only emotion types without any emotion intensity information, can be displayed. In contrast, human beings have much rich emotion spectrum;*
- *the user has to learn to associate 'designed vibrotactile patterns' with three types of emotions.*

To overcome these problems, a natural question is: *how to render 'rich' facial expressions in an intuitive way?* To be able to render rich facial expressions, we take away the "emotion estimation" block from our previous prototype system and build a new system, Facial Expression Appearance for vibroTactile System (FEATS) as shown in Fig. 1(b). In FEATS, facial features are directly rendered on the back of users. This provides the users with rather large degree of freedom to explore emotions with their skins. However, the simplified system does not lower its technical difficulty and it is even more challenging technically to the contrary.

To let the users understand the rendered emotion information, one has to build an intuitive tactile display. The major requirement for such intuitive display of emotions is to characterize how facial expressions evolve during emotion events. Ekman's FACS is state of the art system to characterize facial expressions (which is claimed to be able to display more than 7000 facial actions [10]). We find that to enable saltatory display of emotions, FACS is no longer suitable for the purpose due to

- *FACS consists of 44 action units, which are too many to display tactually;*
- *The discrete nature of AUs is not suitable for displaying continuous emotions;*
- *Automatic recognition of action units is still a hard image analysis problem.*

Some global, analytical, and semantic representation for all possible facial expressions must be considered. Our technical approach is to use manifold of facial expressions for tactile rendering. In this work, we show how to compute manifold of facial expressions and render the manifold on a chair with vibrators. To enable rendering of a live video, an extended Locally Linear Embedding (LLE) coding algorithm is developed. The algorithm has been implemented to generate manifold visualization pattern for our Facial Expression Appearance for vibroTactile System (FEATS).

This paper is organized in seven sections. After an overview of related research activities in section II, a brief introduction of manifold for facial expressions is presented in section III. Section IV describes the our extended LLE algorithm in details. In section V, we provide a system architecture for vibrotactile rendering of emotions for the visually impaired. Section VI presents user studies of our current prototype system (FEATS). Finally, the paper is concluded in section VII.

II. RELATED WORK

In the late 1960's researchers showed that subjects could identify simple objects by getting low-resolution "image" projected on their backs [11], [12]. The subjects were seated in a dental chair with a 20×20 matrix of 1 mm thick circular vibrotactile stimulaters spaced 12 mm apart. The matrix was connected to a television camera that could be freely moved around by the subjects. Images from the camera were converted to tactual information, causing the tactile stimulaters to vibrate if they were in an illuminated area of the camera's image. After some training, the persons were able to identify numerous objects. More recently, vibrotactile displays for the fingertips were presented, such as reported by [13], where a 12×4 tactile element array, suitable for three fingertips, moved about a 220×220 mm area, thus giving a virtual resolution of 4000 elements. The fingertips can perceive much finer details than the back but have smaller areas, so the environment had to be explored bit by bit.

Some researchers developed vibrotactile interfaces for the specific part of the body while others tried the whole body suits fitted with multiple tractors [14]–[16]. Tan et al. [17] have introduced the idea of using an office chair embedded with vibrotactile units on the back to get touch signals from users. Lindeman and Cutler Robert [18] experimented with DC-motors mounted on the chair as a letter recognition system. They concluded that the accuracy in letter recognition could be achieved by raising the performance of identifying the position of tactors, and the apparent motion mode could be helpful to display the letter in less time. Jones et al. [19] used a vibrotactile interface for presenting directional information as a navigational aid systems and argued " *...confirmed that the tactile display can be used as navigation aid outdoors and that the vibrotactile patterns presented can be interpreted as directional or instructional cues with almost perfect accuracy"*. Vibrotactile interfaces also achieved success as communication systems for pilots and astronauts to access the information about body tilt to individuals with balance disorders (i.e. vestibular dysfunction [20]) [21], [22].

It can be summarized from previous research results

- *Only the static visual stimulus or forms, like text, contour of object, 3D shaped object etc, are used for testing. High-level dynamic visual forms, like facial expressions, which we have targeted for now, are not yet tested.*
- *There is no straightforward way to convert visual information into tactile representation. The tactile sense is an inferior information carrier to the visual sense. To avoid information overload that would result from a direct translation from a visual to the tactile information, the visual signals have to be simplified and reduced (i.e. dimensionality reduction).*
- *To develop an effective tactile display of visual information both psychological and physiological knowledge of human touch sense is needed.*

It is challenging to have an intuitive rendering of rich

Figure 2. Setup for collection of Personal Mimics Gallery.

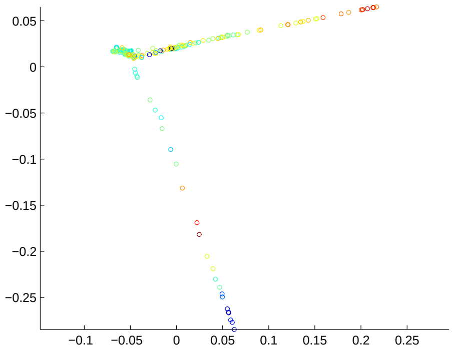


Figure 3. Personal mimics face images mapped into the embedding space described by the first two coordinate of LLE. Each branch shows the cluster of same facial expression with various intensity.

facial expressions. To render high dimensional visual information on a 2D tactile display one has to employ effective and efficient dimensionality reduction techniques. The nonlinear dimensionality reduction technique, Locally Linear Embedding (LLE) algorithm [23] is applied in this work.

III. MANIFOLD OF FACIAL EXPRESSIONS

The Locally Linear Embedding (LLE) algorithm is based on a simple geometric intuition: *The algorithm computes a low dimensional embedding with the property that nearby points in the high dimensional space remain nearby and similarly co-located with respect to one another in the low dimensional space* [23].

In the simplest formulation of LLE, we view all N images $\{\mathbf{X}_i\}$ sampled from a smooth underlying manifold. We identify K nearest neighbors per image point, as measured by Euclidean distance. Reconstruction errors are then measured by the cost function:

$$E(\mathbf{W}) = \sum_i |\mathbf{X}_i - \sum_j W_{ij}\mathbf{X}_j|^2 \tag{1}$$

The weight W_{ij} summarizes the contribution of the jth image to the ith reconstruction. LLE constructs a neighborhood preserving mapping based on the computed weight W_{ij}. Each high dimensional input $\mathbf{X}_i$ is mapped to a low dimensional output a_i representing global internal coordinates on the manifold. This is done by choosing the d-dimensional coordinates of each output a_i to minimize the embedding cost function:

$$\Phi(\mathbf{a}) = \sum_i |a_i - \sum_j W_{ij}a_j|^2 \tag{2}$$

a_i are obtained by solving a sparse $N \times N$ eigenvalue problem [23].

Locally linear embedded (LLE) algorithm has been applied to characterize and visualize facial expressions [24]. Unlike the early work [24], where 58 facial feature points have to be extracted and used for the input to LLE, we directly apply LLE to sub-sampled facial images. To ensure there is sufficient data (such that the manifold is well-sampled), a large number of video frames are collected. To create a compact and efficient manifold of facial expressions, a wearable camera (positioned to frontal face) is used to record videos. The users are instructed to perform a series of six typical emotions. Such a setup ensures almost no global motion (see Fig. 2) but with efficient facial mimic. The sub-sampled video frames are used to compute the manifold. Fig. 3 shows the manifold of facial expressions in the 2D LLE space. One can see that neutral facial expressions act as the central reference points and similar expressions are points in the local neighborhood on the manifold. The basic emotional expressions with increasing intensity become curves on the manifold extended from the center.

Although LLE is good for visualization, however, it yields a map defined only on the training data and is not suitable for analysis and/or visualization of new data. In next section we show how to extend the standard LLE for visualization of new video on-line.

IV. EXTENDED LLE ALGORITHM

To analyze new videos, the extended LLE coding algorithm is proposed in this section. The extended LLE coding algorithm uses face recognition technique to retrieve stored LLE codes for individual video frames. The developed scheme works as following:

(1) A personal facial mimic video database called Personal Mimic Gallery *(PMG) is created and maintained first. The videos are recorded so that video frames in the PMG cover all natural facial expressions for the person. Unlike traditional face databases where data can be in any order, in the PMG the video frames are ordered according to their frame sequence number.*

(2) The manifold of facial videos in the PMG is calculated using the standard LLE technique i.e. project individual frames onto the embedded space. The LLE codes (projection) are tagged to every frame and stored for later retrieval.

(3) When the new facial video is given, the face recognition module checks/recognizes (i.e. using the mouth lip information) which video frame stored in the PMG best

Figure 4. Face identification module results of two "persons" with similar lips but with different appearances of eyes. Top row shows consecutive frames from the new video (test sequence) and bottom frames are "identified" (matched) from the PMG.

matches the input frame. The index information from the face recognition module is used to retrieve the stored LLE codes.

(4) The manifold of the input video is rendered using developed vibrotactile patterns.

Our extended LLE algorithm consists of two steps: *LLE coding* and *frame indexing*.

A. LLE Coding

A personal mimic gallery (PMG) stores all video frames $\{\mathbf{X}_i\}$ of personal facial mimic. The manifold of a PMG is calculated with the standard LLE algorithm. The projection of a video frame $\mathbf{X}_i$ onto the manifold by the LLE operation, $\mathbf{LLE}()$, is called the *LLE-code*, a_i, of the frame $\mathbf{X}_i$.

$$a_i = \mathbf{LLE}(\mathbf{X}_i) \tag{3}$$

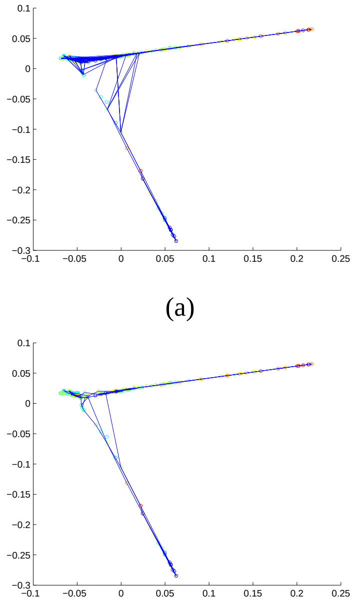
B. Frame Indexing

We assume that all frames in a PMG represent *"individual persons"*. Each *"person"* in the PMG is labelled with an index i. For the target frame $\mathbf{Y}_j$ from a test video, we see it as a "person", who has already been included in the PMG. The *person* can be identified by matching his/her mouth lips, $\mathbf{L}_{\mathbf{Y}_j}$ with the lips, $\mathbf{L}_{\mathbf{X}_i}$ of the face $\mathbf{X}_i$:

$$k \equiv \arg \max_i \mathbf{S}_1(\mathbf{L}_{\mathbf{Y}_j}, \mathbf{L}_{\mathbf{X}_i}) \tag{4}$$

where $\mathbf{S}_1$ is a similarity measure. In [9], we have explained how to match two lips in an optimal way. The *"person"* k is identified because his/her lips are the most similar to the lips of the face appearing in the frame $\mathbf{Y}_j$. Thus the input frame $\mathbf{Y}_j$ can be reconstructed from the frame $\mathbf{X}_k$

$$\hat{\mathbf{Y}}_j = \mathbf{X}_k \tag{5}$$



(a)

(b)

Figure 5. 2D representation of the first two coordinate of LLE embedded space for coded video sequence (a) without and (b) with frame reordering. Smooth frame rendering with less jerky effect (i.e. jumps from one cluster-branch to another) can be seen in (b) as compared to (a).

The index k is used to retrieve the LLE-code of frame $\mathbf{X}_k$ for frame $\mathbf{Y}_j$

$$b_j = \mathbf{LLE}(\mathbf{Y}_j) = a_k \tag{6}$$

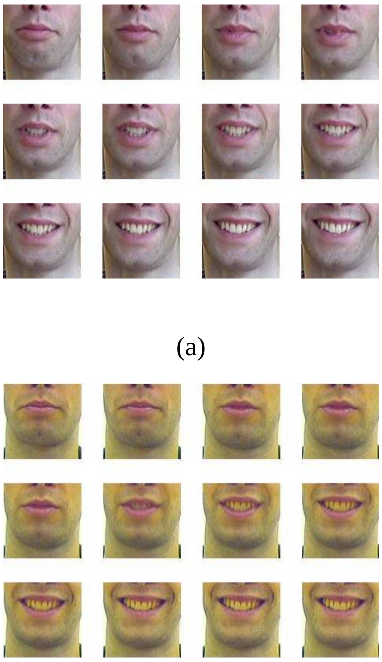
The new video $\{\mathbf{Y}_j\}$ is LLE-coded into $\{b_j\}$, which is used for rendering. In this way we can achieve real-time rendering of the manifold of on-line video.

The new video is reconstructed by selecting frames from the gallery according to the index from the person identification module using lipless method [9], i.e. selecting "persons" from the PMG based on the similarity between lips. Our experimental results show that this algorithm results in a non-smooth video sequence (i.e. jerky effect). The jerky reconstruction of new video can be explained, since the frames are evaluated individually, there is no sequential order among the retrieved frames. It can be seen from Fig. 4, the "person" identification process results in faces with similar lips but having different appearances of eyes. The reconstructed video sequence, based on the indexes from the recognition module in the LLE space, is shown in Fig. 5(a). One can see that the frames are jumping across two emotion branches of LLE coded representation.

C. Constrained Frame Indexing

To generate a natural-looking video, a *smoothness term* has to be included in the object function used in the optimization equation 4:

$$k \equiv \arg \max_i [\mathbf{S}_1(\mathbf{L}_{\mathbf{Y}_j}, \mathbf{L}_{\mathbf{X}_i}) + \alpha \mathbf{S}_2(\hat{\mathbf{Y}}_{j-1}, \mathbf{X}_i)] \tag{7}$$



(a)

(b)

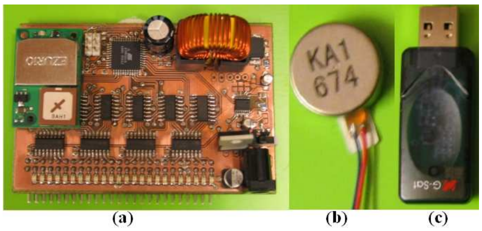
Figure 6. Frames of a smiling expressive lips in (a) the consecutive frames to be encoded (b) the reconstructed frames based on extended LLE coding with frame reordering.

where α is a weight parameter and $\mathbf{S}_2$ is a similarity measure. The L-norm based distance measures cannot be used for $\mathbf{S}_2$ since they are all sensitive to minor spatial displacement and intensity changes. To measure the similarity, Euclidean distance measure in the LLE space is used i.e.

$$\mathbf{S}_2(\hat{\mathbf{Y}}_{j-1}, \mathbf{X}_i) \cong \mathbf{D}(\mathbf{LLE}(\hat{\mathbf{Y}}_{i-1}), \mathbf{LLE}(\mathbf{X}_i)) \quad (8)$$

where $\mathbf{D}()$ is Euclidean distance measure in the LLE space. To ensure smoothness in the *resulted video*, the dynamic programming technique is applied to optimize the sequence of frame indices. With forcing smoothing transition, the jerk effect is greatly reduced (which is clearly shown in Fig. 5(b)).

To test the feasibility of the extended LLE coding algorithm, we first recorded a personal mimic video gallery. The manifold of facial expressions extracted from personal mimic gallery (PMG) is used to analyze new videos. Fig. 6 shows two sequences of a mouth smiling in the video *'to be encoded'* and *'coded'* versions. The order of frames is from left to right then top to bottom. The selected frames (Fig. 6(b)), do not match exactly the input frames. However, the whole sequence looks (by visual inspection) more natural with less jerks. This demonstrates that our extended LLE-algorithm works fine to handle new videos.



(a) (b) (c)

Figure 7. Vibrotactile components of FEATS (a) Printed Circuit Board, (b) Tactor (vibration motor), (c) Bluetooth Module

V. VIBROTACTILE RENDERING

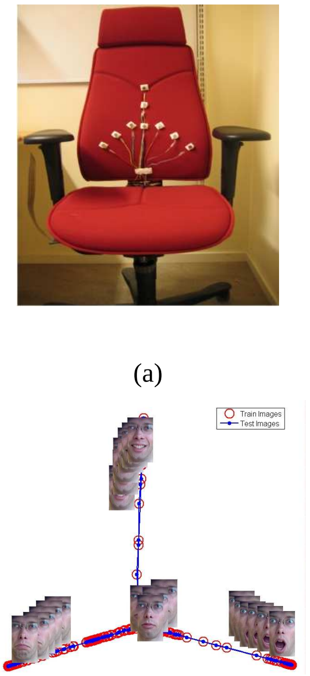
Vibrotactile sensation is generally produced by using tactically active sites stimulation to stimulate the skin surface at certain frequency using vibrating contactors. Two commonly used display techniques to generate vibration are, 1) *using a moving coil*, usually driven by a sine wave; 2) *a DC motor*, which is used in most of mobile phones. Other less common actuators are based on *piezoelectric benders*, *air puffs*, or *electrodes*. The vibrotactile display used here is based on a DC motor. The vibration is generated by rotating a wheel with a mounted eccentric weight. The wheel rotates at a constant speed. Switching it on and off yields the vibration.

Primarily, there are four types of parameters through which the vibrotactile sensation can be rendered, namely *frequency*, *amplitude (magnitude)*, *timing*, and *location*. To generate effective and efficient rendering one should follow the guidelines for designing vibrotactile displays [25]–[27]:

- *No more than 4 intensity (magnitude) levels should be used.*
- *No more than 9 frequency levels should be used for coding information and difference between adjacent levels should be at least* 20%.
- *The vibration on the human skin should be carefully dealt with and long durations might make users irritated.*
- *Vibration intensity can be used to control threshold detection and sensory irritation problem, similar to volume control in audio stimuli.*

Accordingly, a vibrotactile chair (called Facial Expression Appearance for vibroTactile System (FEATS)) has been built in our Lab (Fig. 7, 8). Two key modules of FEATS are *the vibrotactile control box (VCB)* and *the vibrotactile display box (VDB)*. The VCB consists of a single printed board (see Fig. 7). Vibrotactile Module communicates to PC using blue-tooth module. VDB consists of nine tactors mounted on the back of the chair (Fig. 8). The tatcors are vibration motor $C1234B028F$ manufactured by JinLong Machinery China [28]. The vibration motor has 10.0 mm diameter, 3.0 mm length and standard speed ~ 12000 rpm at 4.0 Volts.

In the current design, a Pulse Width Modulation (PWM) is used to send the varying amount of time voltage to the tactors. A micro controller, ATmega16L-16PI micro



(a)

(b)

Figure 8. (a)Vibrotactile chair and tactors arrangement (b) LLE coding patterns of 3 emotions happy, sad, and surprise.

controller [29] is used. It was programmed to generate PWM to control tactors. In this way we were able to active any one of nine tactors with certain frequency ($\sim$ 1 Hz to 1 KHz) and magnitude. The magnitude is determined by relative duration ratio of on-off signals. The host computer communicates at speed from 4800 bps (which could be modified) to the VCB through wireless communications.

The extended LLE coding algorithm is used to calculate manifold of facial expressions. On the manifold not only the type of emotion but also the intensity can be visualized. Using a 2D array of tactors to render emotions on the back of a chair, tactors are mounted based on the shape of the manifold. These tactors are arranged in three directions. Each set (three tactors) is used to present a specific tracing mode of tactile patterns. For example, emotion *happy* is rendered by three tactors lined from the bottom to upward, and emotion *sad* and *surprise* by other two side lines. Fig. 9 shows the vibration stimuli direction and arrangement of the tactors for specific type of emotion. Different branches correspond to the types of emotions and climbing each branch means the intensity of the emotion is getting strong. Obviously, location of tactors encodes both the *type* and *intensity* of emotions.

To enhance the effect of rendering, additional vibration information is encoded; i.e, the *emotion type* is coded by the frequency while *emotion intensity* is coded by the magnitude. To overcome the limitations, such as *stimuli masking* and *adaptation*, and to enhance the perception of vibrating stimuli, 3 different inter-switched frequency levels with $\sim$ 20% difference are used.

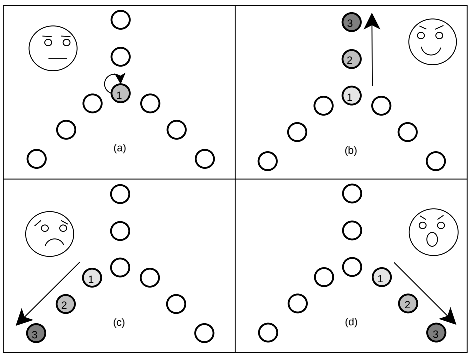


Figure 9. Visual depiction of the extended LLE algorithm based visualization vibrotactile pattern for a) normal b) happy c) sad d) surprise facial expressions. The directions of arrows show temporal order of activation of the tactors.

VI. USABILITY EVALUATION

To test efficiency and effectiveness of the vibrotactile rendering system of human emotions, we carried out a usability test.

1) Participants: Participants were recruited randomly from the campus of Umeå University, Sweden, including both students and staff members from different ethic groups. There were 28 people (18 are male and 5 female), aging from 18 to 64, being involved in the experiments. 70% of the participants had experience of vibrotactile stimuli (i.e. through the use of mobile phones and game consoles).

2) Procedure: First, we introduced the purpose of our experiments to the participants and instructed them how to use the system. Each participant was asked to seat and be relaxed in the vibrotactile chair with normal posture so that tactors could have indirect contact (i.e. through normal clothing) with the participant's back. The position of tactors was arranged so that it did not make any contact with the spine. In the training secession (20 sec. video clip with a framerate of 20 fps), a vibrotactile sequence was given to the participant along with visual picture of the pattern using the GUI. Each vibrating pattern was associated with specific type of videos. Two random video clips (i.e with 40 sec. duration each) were rendered using vibrotactile patterns in the testing phase. During the experiment, each participant was asked to write down the sequence-code of the presented vibrotactile pattern, so that it could be matched with the original pattern sequence. During the experiment, the participants hearing sense was also blocked to remove any auditory cues using head phones.

At the end, to measure levels of satisfaction, system usability and main effects later on, each participant was given a questionnaire to answer. Our subjective questionnaire used Likert style 7-point rating system, which scales from 1 to 7. Its values represent strong disagreement (negative) and strong agreement (positive) respectively. Fig.10 shows mean questionnaire responses to questionnaire questions. Each label on the x-axis in the figure represents a question where we used the following notations:

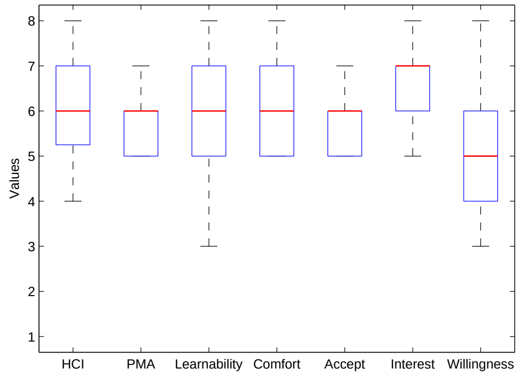


Figure 10. Mean Questionnaire Scores.

- HCI - *Is the human computer interface is easy to use?*
- Pattern Mapping Accuracy (PMA) - *Is the pattern mapping recognizable?*
- Learnability - *Is the training helpful?*
- Comfort - *Is this application comfortable to use?*
- Accept - *Is this application acceptable?*
- Interest - *Is this application interesting?*
- Willingness - *Are you willing to pay $50 to buy such a product?*

3) Experimental Results: The results in three aspects of usability are shown here.

Efficiency is calculated as effort required in order to accomplish the desired task [30]. An efficient system ought to require as little effort as possible. In our experiments, we used the reaction time of a user as an indicator of efficiency. The reaction time is measured by computing the delay between the time when a vibration signal is triggered and user response time. The delay time contains two parts: *cognitive time* and *response time*. Since the response time is approximately constant, we assume that the delay time is a reliable indicator of cognitive time.

Effectiveness is measured by recognition rate of both type and intensity of emotions. The recognition rate of ***emotion type*** from vibrotactile patterns was 100% for all the subjects. The recognition rate of ***emotion intensity*** was 95.1% for 97% of the participants and 93.75% for rest 3% subjects. An explanation of error in recognition could be due to cognitive overloading in mental mapping vibrotactile pattern and the intensity of emotions(mentioned by 10% of the subjects). This point has been confirmed in the second round of experiment. It is worth mentioning that subjects interest have significant effect on the recognition of emotion intensity. The recognition rate of emotion type was higher than that of emotion intensity.

User Satisfaction. Experimental results from the questionnaires indicated a high interest level for our application, mean score 6.3913. Participants gave an average score 5.7826 indicating the difficulty-easiness to recognize the vibration signals before training and an average score 6.5101 after training. They considered the training a helpful procedure. An average rating 5.7826 was given by participants to consider it a good application and showed an average rating 4.8696 for willingness to buy.

TABLE I.
USER'S WILLINGNESS

Source	F	P-value
Application interests	15.000	0.0003
Comfort	5.430	0.0245
Easy to use	6.130	0.0172

We performed an ANOVA analysis on the questionnaire. The results are listed in Tab.I. We found that user's previous vibrotactile stimuli had no effect on learnability, interest and willingness to buy. There was a significant effect of user's interests in application on willingness to buy the product. Similarly, there are effects of comfortable and easy to use on willingness.

VII. CONCLUSION AND FUTURE WORK

This paper addresses an important question: *how to render rich facial expressions in an intuitive way?* It is shown that manifold of facial expressions can be used as a compact and natural way to display human emotions for vibrotactile rendering. The topology of emotion branches is particularly suitable for intuitive rendering on the back of a person. In this work, the standard LLE algorithm has been extended to handle the problem of real-time coding of new videos. This makes it possible to use manifold of facial expressions for vibrotactile rendering of human emotions. The user tests from our work elaborate that it is extremely important to consider user's perceptual workloads when vibrotactile rendering experimental setups are designed. Current research work not only provides a concrete step towards building a vibrotactile interface but also introduces a new dimension to human-machine interaction.

REFERENCES

[1] M. Pantic and L. J. M. Rothkrantz, "Toward an Affect-Sensitive Multimodal Human-Computer Interaction," in *Proceedings of the IEEE*, vol. 91, no. 9, 2003, pp. 1370–90.

[2] Y. Wang, H. A.I., B. W.U., and C. Huang, "Real Time Facial Expression Recognition with Adaboost," in *Proceedings of IEEE 17th International Conference on Pattern Recognition*, 2004, pp. 3:926–929.

[3] Y.-L. Tian, T. Kanade, and J. Cohn, "Robust Lip Tracking by Combining Shape, Color and Motion," in *Proceedings of the 4th Asian Conference on Computer Vision*, 2000.

[4] M. Argyle, "Bodily Communication," *Methuen and Co, NewYork.*, 1988.

[5] T. Huang, L. Chen, and H. Taor, "Bimodal Emotion Recognition by Man and Machine," in *Proceedings of ATR Workshop on Virtual Communication Environments*, 1998.

[6] S. U. Réhman, L. Liu, , O. Lindahl, and H. Li, "Vibrotactile Chair: A Social Interface for Blind," 2006, pp. 117–120.

[7] Y. Tain, T. Kanade, and J. Chon, "Recognizing Action Units for Facial Expression Analysis," *IEEE Tran. Pattern Analysis and Mach. Intelligence*, vol. 23, pp. 97–115, 2001.

[8] S. U. Réhman and L. Liu, “Real-time Lip Tracking for Emotion Understanding,” in *Proceedings of Swedish Symposium on Image Analysis (SSBA2006)*, 2006, pp. 29–32.

[9] S. U. Réhman, L. Liu, and H. Li, “Lipless Tracking and Emotion Estimation,” in *Proceedings of IEEE 3rd int. Conf. on Signal-Image Technology & Internet-based Systems (SITIS’2007)*, 2007.

[10] P. Ekman and W. Friesen, “Facial Action Coding System: A technique for the measurement of Facial Movements,” *Consulting Psychologist Press, PaloAlto*, 1978.

[11] P. B. y Rita, C. Collins, F. Sauders, B. White, and L. Scadden, “Vision Substitution by Tactile Image Projection,” *Nature*, vol. 221, pp. 963–964, 1969.

[12] P. B. y Rita, “Brain Mechanisms in Sensory Substitution,” *Academic Press, NewYork*, 1972.

[13] T. Maucher, K. Meier, and J. Schemmel, “An Interactive Tactile Graphics Display,” in *Proceedings 6th Int. Symp. on Signal Processing and its App.*, 2001, pp. 190–193.

[14] P. B. Rita, K. Kaczmarek, M. Tyler, and J. Garcia-Lara, “Form Perception with a 49-point Electrotactile Etimulus Array on the Tongue,” *J. of Rehabilitation R. D.*, vol. 35, pp. 427–430, 1998.

[15] H. Yano, T. Ogi, and M. Hirose, “Development of Haptic Suit for Whole Human Body Using Vibrators,” *Trans. of the Virtual Reality Society of Japan*, vol. 3, pp. 141–148, 1998.

[16] R. Lindsman, R. Page, Y. Yanagida, and J. Sibert, “Towards Full-Body Haptic Feedback: The Design and Deployment of a Spatialized Vibrotactile Feedback System,” in *Proceedings of ACM Virtual Reality Software and Technology (VRST)*, 2004, pp. 146–149.

[17] H. Tan, I. Lu, and A. Pentland, “The Chair as a Novel Haptic User Interface,” in *Proceedings Workshop on Perceptual User Interfaces*, 1997, pp. 56–57.

[18] R. Lindeman and J. Cutler, “Controller Design for a Wearable, Near-Field Haptic Display,” 2003, pp. 397–403.

[19] L. Jones, B. Lockyer, and E. Piateski, “Tactile Display and Vibrotactile Pattern Recognition on the Torso,” *Advanced Robotics*, vol. 20, no. 12, pp. 1359–74, 2006.

[20] P. Kadkade, B. Benda, P. Schmidt, and C. Wall, “Vibrotactile Display Coding for a Balance Prosthesis,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. II, no. 4, pp. 392–399, 2003.

[21] S. Cardin, F. Vexo, and D. Thalmann, “Vibro-Tactile Interface for Enhancing Piloting Abilities During Long Term Flight,” *Journal of Robotics and Mechatronics*, vol. 18, no. 4, 2006.

[22] H. Veen and J. van Erp, “Tactile Information Presentation in the Cockpit,” *Lecture Notes in Computer Science*, vol. 2058, pp. 174–181, 2001.

[23] S. Roweis and L. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” *Science*, vol. 290, pp. 2323–2326, 2000.

[24] Y. Chang, C. Hu, and M. Turk, “Manifold of Facial Expression,” in *Proceedings IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, 2003, pp. 28–35.

[25] J. Craig, “Difference Threshold for Intensity of Tactile Stimuli,” *Perception and Psychophysics*, pp. 150–152, 1972.

[26] F. Geldard, “Cutaneous Stimuli, Vibratory and Saltatory,” *J. Investigative Dermatology*, pp. 83–87, 1977.

[27] D. Goff, “Differential Discrimination of Frequency of Cutaneous Mechanical Vibration,” *J. Experimental Psychology*, pp. 294–299, 1968.

[28] “Jinlong Machinery, CHINA,”
http://www.vibratormotor.com/.

[29] “Atmel Corporation,”
Atmel 16L ManualAtmel Corporation, San Jose, CA, 2006.

[30] X. Faulkner, “Usability Engineering,” *Palgrave Macmillan, 2002*.

Shafiq ur Réhman received his M.S. degree in 2004 from Department of Computer Science, Umeå University, Sweden. Currently he is pursuing his PhD degree in Applied Electronics at Department of Applied Physics and Electronics, Umeå University, Sweden. He is currently working on multimodal signal processing and user centered interface at Digital Media Lab. His research interests are computer vision, vibrotactile coding and multimedia technologies related applications.

Li Liu received the B.S. degree in Biochemistry from Nanjing University, China in 1986, and the degree of technical licentiate in Applied Physics from Linköping University, Sweden, 1994. In 1997 she received her PhD degree in Biophysics from Göteberg University, Sweden. She is now active researcher in the department of Applied Physics and Electronics, Umeå University, Sweden. Dr. Liu was awarded “Wallenbergs Excellent Woman Researcher”, by Knut och Alice Wallenbergs Foundation, 1999. Her research interests include tactile video for visually impaired persons, computer assessment of infant pain and e-health.