

Research Article

Tactile Modulation of Emotional Speech Samples

**Katri Salminen,¹ Veikko Surakka,¹ Jani Lylykangas,¹ Jussi Rantala,¹
Teemu Ahmaniemi,² Roope Raisamo,¹ Dari Trendafilov,² and Johan Kildal²**

¹Tampere Unit for Computer Human Interaction, School of Information Sciences,
University of Tampere, Kanslerinrinne 1, 33014 Helsinki, Finland

²Nokia Research Center, Itämerenkatu 11-13, 00180 Helsinki, Finland

Correspondence should be addressed to Katri Salminen, katri.salminen@uta.fi

Received 19 December 2011; Revised 1 June 2012; Accepted 27 June 2012

Academic Editor: Antonio Krüger

Copyright © 2012 Katri Salminen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditionally only speech communicates emotions via mobile phone. However, in daily communication the sense of touch mediates emotional information during conversation. The present aim was to study if tactile stimulation affects emotional ratings of speech when measured with scales of pleasantness, arousal, approachability, and dominance. In the Experiment 1 participants rated speech-only and speech-tactile stimuli. The tactile signal mimicked the amplitude changes of the speech. In the Experiment 2 the aim was to study whether the way the tactile signal was produced affected the ratings. The tactile signal either mimicked the amplitude changes of the speech sample in question, or the amplitude changes of another speech sample. Also, concurrent static vibration was included. The results showed that the speech-tactile stimuli were rated as more arousing and dominant than the speech-only stimuli. The speech-only stimuli were rated as more approachable than the speech-tactile stimuli, but only in the Experiment 1. Variations in tactile stimulation also affected the ratings. When the tactile stimulation was static vibration the speech-tactile stimuli were rated as more arousing than when the concurrent tactile stimulation was mimicking speech samples. The results suggest that tactile stimulation offers new ways of modulating and enriching the interpretation of speech.

1. Introduction

In daily communication we acquire emotion-related information via several senses. Recently the investigation of the utilization of the sense of touch in technology contexts (e.g. cars, tablets, mobile phones, etc.) has been very active. Current mobile phones use routinely direct tactile manipulation for operation. Increasingly phones also utilize tactile feedback. Tactile feedback in mobile phones can be used and is aimed at mediating basically same type of information that is mediated in human-human communication. As mobile phones are still used also for conversations a natural question to come in mind is how speech and tactile information might function together.

Human touch system is designed so that via it we can get cognitive, social, and emotional information. From early studies by Harlow [1–3] we learned that tactile information is closely related to the human emotional system. People use the sense of touch when they aim, for instance, to communicate affection or to get someone's attention in a

socially acceptable manner [4]. It has also been shown that people are capable of sending and receiving emotion-related information (e.g., expressions of anger or love) through touch (e.g., [5]). In addition, studies have shown that the mediation of emotional information via tactile technology works well. Recently Smith and MacLean [6] and Bailenson et al. [7] studied how well participants could identify haptically presented emotions. In their studies one participant (i.e. the sender) used a force-feedback device to create haptic messages which would communicate certain emotions from a list (e.g., anger, sadness, and happiness). The other participant's (i.e., the receiver's) task was to try to identify intended emotion-related contents from tactile stimulations the sender had generated. The results suggested that haptic stimulation generated and interpreted this way can communicate emotion-related information at a better chance level.

Also speech can mediate emotion-related information both in human-human and human-technology contexts. For example, there is evidence that synthetically (i.e., speech

synthesizers) generated speech samples with positive and negative content are in general rated as pleasant and unpleasant, respectively [8, 9]. Thus, both speech and tactile stimulations have the potential to evoke emotional experiences in humans. Although currently available vibrotactile technologies in mobile phones enable utilizing tactile modality in conjunction with speech, the potential to enrich speech communication with simultaneous tactile stimulation is largely unmapped area in human-technology interaction (HTI).

In one study by Chang et al. [10] the participants used a device prototype capable of converting hand pressure into vibration to complement speech while, for example, talking with another participant. Their results showed that the participants used tactile stimulation for emphasizing spoken messages and to indicate turn-taking behavior. In general the authors argued that their results showed that the participants can use tactile channel to transfer meaningful information simultaneously with speech. However, the potential to modulate the emotion-related responses to speech with the sense of touch in mobile contexts has not been studied.

Several studies have shown that processing of auditory and tactile information is known to have crossmodal effects on each other. Information from tactile and auditory modalities is integrated in an early phase of information processing chain and evokes responses partly in the same areas of the brain [11–14]. Further, studies have shown that auditory stimuli can affect the perception of tactile stimuli and vice versa. For example, in a study by Bresciani et al. [15] tactile taps and auditory beeps were presented to the participants who were instructed to report how many tactile taps they perceived. The results showed that the number of auditory beeps systematically modulated the perception of tactile taps regardless of the actual number of the taps. In another study by Ro et al. [16] the task of the participant was to verbally report whether or not they felt a tactile stimulus which was either accompanied by a sound or not. Their results showed that the auditory stimuli increased the detection rate especially when the tactile and auditory stimuli had the same frequency. Thus, by providing simultaneous auditory and tactile signals one can improve the cognitive performance of humans.

In addition, previous studies suggest that the form of the tactile and auditory signals has an effect to the processing of the multimodal stimulus. Especially synchronizing the signals coming from our environment affects the perception of the stimuli. A view referred as assumption of unity suggests that the brain treats information coming from different modalities as coming from the same source or object only if they share similar amodal properties. Most important of these properties is temporal coincidence [17]. Further, previous studies suggest that synchronizing tactile and auditory stimuli can affect human cognitive processing. For example, in a study by Gillmeister and Eimer [18] the task was to indicate whether a single trial contained an auditory signal near the perceivable threshold. A tactile stimulus was always delivered in the trial, and the auditory signal was presented randomly in half of the trials. The results showed that synchronous tactile stimuli improved the detection of

near-threshold auditory stimuli. Similarly, in another study [19] synchronizing the auditory and tactile signal improved the detection rate of the stimuli. In another experiment by Gillmeister and Eimer [18] the participants were asked to judge the intensity (i.e., loudness) of auditory tones. The results clearly showed that intensity judgments were considerably higher in the presence of synchronous tactile stimulation than in asynchronous or no tactile stimulation trials.

Interestingly, there is also evidence suggesting that tactile information can improve the understanding of spoken syllables especially when synchronized with the speech [20]. The participants' task was to perform a forced-choice syllable decision task in which the syllables were either accompanied by additional noise or not. The syllables were presented in three modality conditions: auditory only, congruent mouthing, or incongruent mouthing manually felt from speaker's face. The results showed that congruent mouthing improved the amount of correct responses when compared with auditory only condition and incongruent mouthing but only when the syllables were accompanied by noise. In addition, there were more correct responses in auditory only condition than in incongruent mouthing condition. Based on these results the authors argue that manual tactile information relevant to speech gestures can improve auditory speech perception.

Taken together, it seems that there is only a little information on how speech and tactile stimulation are related to each other. Given that synchronized tactile and auditory or speech signals can affect human cognitive processes we selected a starting point where we first produced tactile signals that accurately mimicked the amplitude changes of the speech signal to study the effects of concurrent tactile stimulation on the emotion-related ratings of speech in a general level. Next, we proceeded in studying whether synchronizing the amplitude changes of speech and tactile signal is the optimal way to produce the concurrent tactile stimulation to evoke emotion-related experiences in the user. For this purpose the amplitude changes of the tactile signal were either mimicking the amplitude changes of the speech sample in question (i.e., congruent stimulation) or amplitude changes of another speech sample used in the study (i.e., incongruent stimulation).

To measure emotional responses the dimensional theory of emotions was used as the frame of reference for rating the effects of stimuli. According to Bradley and Lang [21–23] and Schlosberg [24] there are three basic bipolar dimensions that cover well the dimensional emotion space while rating different types of stimuli. The dimensions are valence, arousal, and dominance. Using these dimensions affective ratings can be collected with a special set of rating scales where the valence dimension varies from unpleasant to pleasant, the arousal dimension varies from relaxed to arousing, and the dominance dimension varies from feeling of being controlled by the stimulus to the feeling of being in control of the stimulus [21–23]. Although the dimensional theory of emotions suggests that the dimensions are related to a motivational tendency to approach or withdraw from emotion evoking stimulus, this tendency has not been

measured frequently. Recently in some studies this fourth dimension has been taken into consideration by asking also ratings of the approach-withdrawal tendency with a bipolar rating scale varying from avoidable to approachable [25–27]. In earlier studies these four dimensions were used to evaluate how participant reacts to haptic only stimulation [28, 29]. The results have shown that varied haptic stimulus parameters (e.g., amplitude and continuity) evoke different ratings in previously mentioned four scales so that, for example, stimuli with high amplitudes are rated as more arousing and dominant than stimuli with low amplitudes.

To summarize, the present aim was to study if the emotion-related ratings of speech are affected by tactile stimulation. A handheld prototype device with four vibrating actuators produced the tactile stimulation. The participants' task was to rate the stimuli with the scales of pleasantness, arousal, approachability, and dominance. In Experiment 1 the participant's task was to rate speech only and speech tactile stimuli. The amplitude of the speech signal was mimicked as accurately as possible with tactile actuators. In Experiment 2 the speech samples were presented to the participants with both congruent and incongruent tactile stimulation. Congruent stimuli consisted of speech samples that were paired with tactile stimulation derived from that particular speech sample. Incongruent stimuli consisted of speech samples that were paired with tactile stimulations created on the basis of another speech samples.

Experiment 1. The aim of Experiment 1 was to study how simultaneously presented tactile stimulation affects to the subjective ratings of short speech samples. For this purpose, speech samples were presented to the participant with and without tactile stimulation. The participants rated all the stimuli, speech only and speech tactile, one at a time using the rating scales for pleasantness, arousal, dominance, and approachability.

2. Methods

2.1. Participants. Twelve voluntary male participants took part in the study (mean age 27, range 21–41 years). All the participants were students at the University of Tampere and recruited via e-mail. They did not receive any compensation for their participation. Nine of the participants were right handed, and three were left handed by their own report. Seven of the participants told that they hold a mobile phone in their right hand, and five told holding it in their left hand while typing a text message. All had normal or corrected to normal vision, and normal hearing and sense of touch by their own report. All the participants were fully informed for the purpose of the study prior to the experiment. They were also informed that they were able to abort the experiment at any point without a specific reason. A consent form was signed by all the participants.

2.2. Apparatus. A haptic device prototype was used in the experiment (Figure 1). The prototype was chosen for two reasons. First, its actuator solutions were more capable of producing similar variations in vibrotactile amplitude and

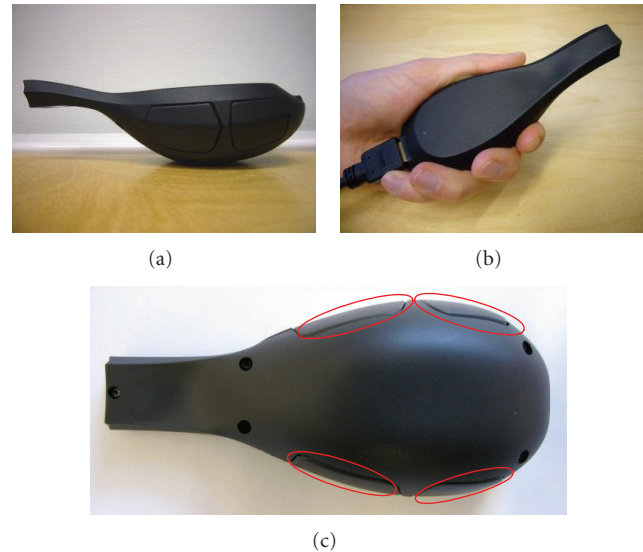


FIGURE 1: (a) Side view of the prototype device, (b) a person holding it, and (c) the bottom-up view with circles showing the placement of the actuators.

frequency than the rotating mass vibrators commonly used in standard mobile phones. Second, the shape of the prototype was designed so that the actuators would stimulate the whole palm area when the device was vibrating. The device was equipped with four Minebea Linear Vibration Motor actuators (LVM8, Matsushita Electric Industrial Co., Japan). Two actuators were located on the right side and two on the left side of the device. Actuation of these motors was based on a small electromagnetic weight which moves down when driving signal is applied and backs up using a spring when no signal is present. The resulting rapid movement in opposite directions creates the vibration. The LVM8 actuators were mounted inside separate buttons in the device in order to isolate the vibration from the body of the device. Thus, the actuation was localized to four specific areas on the device. Usable driving frequencies of 120–180 Hz and a resonant frequency of 155 Hz were measured from the LVM8 actuators after mounting. The LVM8 actuators can be driven using audiosignal which makes it easy to modify input using audiosynthesis software. A controller box was connecting the prototype device to a laptop computer with HDMI and USB connections. The controller box was designed to reduce the amount of cables between the prototype device and the computer. For more technical details, see [30].

The stimuli were controlled with an external Gigaport HD USB sound card. Pure Data Audio synthesizer software (PD, version 0.41.4) was used to create the tactile stimuli and to control the stimulus presentation. An Acer Netbook laptop computer recorded the ratings which were given with a standard computer mouse (Figure 2).

2.3. Stimuli. Two stimulus modalities were used in the experiment: speech only and speech tactile (see Table 1). The speech-only stimuli were selected from speech synthesizer named Loquendo text-to-speech. Speech acts refer to a

TABLE 1: The speech samples used in the experiment.





Stimulus in English and Finnish	Emotional content	Duration (sec)	Graphical presentation of amplitude variations
“how beautiful” “oi miten kaunis”	Positive	1.78	
“nice to hear” “hauska kuulla”	Positive	1.47	
“how ugly” “miten ruma”	Negative	1.48	
“sad to hear” “ikävä kuulla”	Negative	1.5	



FIGURE 2: Picture of the experimental setup.

collection of prerecorded emotional speech samples spoken by an actor. We chose four short speech samples. The samples were spoken by a native Finnish male voice. Two of the samples had a positive content, and two had a negative content. Also the tones of the speech samples were different so that the speech samples with positive content had a clear positive tone and the samples with negative content had a negative tone. The stimuli with positive content, and the stimuli with negative content were selected so that they represented bipolarity of the valence dimension. For this purpose we selected the positive and negative samples in a way that they could be seen as opposites of their semantic meaning (see Table 1).

Half of the stimuli were presented with a concurrent tactile vibration. The amplitude of the tactile stimulation followed the original amplitude of the speech. So, it was extracted from the speech sample and presented simultaneously with the speech. A 160 Hz sine wave with varying amplitude was used in each of the tactile stimulations. A fixed frequency value was used due to the LVM8 actuator’s narrow range of perceivable frequencies. The frequency of 160 Hz was chosen based on piloting which showed that the

resonant frequency of 155 Hz resulted in a distracting audible noise. With 160 Hz there was noticeably less leakage of noise although the vibration felt equally strong. The amplitude level of the tactile feedback was set by using an envelope follower that took the speech signal as an input and returned its root mean square (RMS) value as an output.

The amplitude level for the tactile stimulation was set 22 times in a second based on 1000 sequential audio samples. We found the rate of 22 Hz to be sufficient for getting a perceivable tactile estimate of intensity changes in the speech stimuli. Using this feedback synthesis the amplitude of the tactile stimulation followed the original amplitude of the speech stimulus in real time. Finally, the tactile feedback signal was driven to the four LVM8 actuators. A total of 8 different stimuli (4 speech only and 4 speech tactile) was used in the study. The stimuli were presented in random order.

2.4. Procedure. When the participant arrived in the laboratory, the equipment and the environment were introduced to him. He was told that the purpose of the experiment was to study subjective experiences evoked by short speech samples with four rating scales. The participant was also told that some of the speech samples had only speech, but in some samples the prototype vibrated during the speech.

A soft foam cushion was under the arm of the hand holding the prototype to prevent muscle fatigue. Participants were instructed to hold the prototype on the same hand they reported holding a mobile phone while typing a text message (see Figure 1). A computer mouse was used to give ratings, and it was operated with the other hand. The participant was told that when the prototype device would be finalized, they would be able to hear the speech from the prototype. However, currently this was not yet the case, and the participants wore a Peltor HTB 79A hearing protector headset from which they could hear the speech component of the stimuli. It also blocked the noise from the prototype while producing the tactile stimulation. The participant was

instructed to keep the gaze on the laptop display during the experiment. In the center of the display the participant could see instructions related to the stimulus presentation and for giving ratings.

The experiment was divided in four experimental blocks. In each block the participants' task was to rate the 8 stimuli using one of the emotion-related rating scales (i.e., pleasantness, arousal, approachability, or dominance). The ratings were given by selecting a number on the display with the mouse from nine checkboxes labeled from -4 to $+4$. On each of the scales 0 represented neutral experience (e.g., neither unpleasant nor pleasant). The order of the blocks was Latin square counterbalanced.

Before each experimental block there was a practice session to familiarize the participant with the stimuli and the rating scales to be used in the following experimental block. In a practice session, four different stimuli (two speech only and two speech tactile) were evaluated with one rating scale. The stimuli in the practice sessions were different than the stimuli in the experimental sessions. The practice session proceeded as follows. In the beginning of each trial the participant clicked an on-screen stimulus initiation button, and after a 2000 ms interval a stimulus was presented. During the stimulus presentation the participants were instructed to listen to the stimulus carefully and not to clench the prototype. The stimulus offset was followed by a 2000 ms interval after which a rating scale appeared on the screen and the participant was able to respond. After giving the rating, a new trial was initiated 2000 ms after the participant had clicked the stimulus initiation button again. This procedure was repeated until the participant had rated all the four practice stimuli with one rating scale. After the practice session the participant continued to rate the eight experimental stimuli with the same scale. Thus in each block there were 4 practice trials and 8 experimental trials. After the participant had rated all the stimuli with one scale, he proceeded to rate the stimuli with the next scale in the similar manner. This was repeated until all the stimuli were rated with all the four scales. Thus, the total number of the experimental trials was 32. Conducting the experiment took approximately 40 minutes.

2.5. Data Analysis. A Wilcoxon signed-ranks test was used for pairwise comparisons. First pairwise comparisons were conducted to study the effects of the emotional content of the speech (i.e., positive versus negative) to the ratings. Then pairwise comparisons were used to study the effects of the concurrent tactile stimulation (i.e., speech only versus speech tactile).

3. Results

3.1. Pleasantness. Means and standard error of the means (SEMs) for the ratings of the stimulus pleasantness are presented in Figure 3. For the effects of the emotional content of the speech the results showed that the speech only stimuli with positive content were rated as significantly more pleasant than the stimuli with negative content $T = 3.01$, $P < 0.01$, Cohen's $d = 2.70$. Also, the speech tactile stimuli

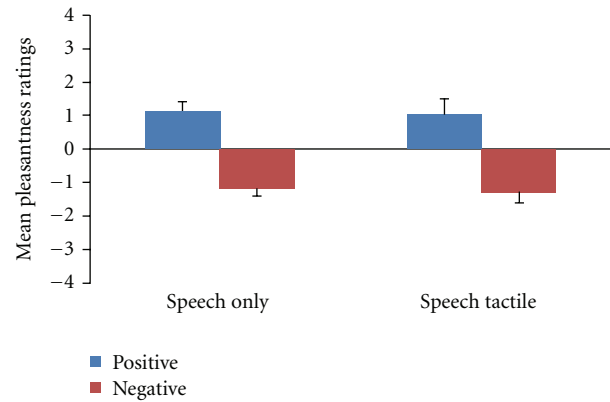


FIGURE 3: Means and SEMs for the ratings of the pleasantness of the stimuli by the emotional content of the speech sample and the stimulus presentation modality.

with positive content were rated as significantly more pleasant than the speech tactile stimuli with negative content $T = 2.99$, $P < 0.01$, Cohen's $d = 1.80$. For the effects of the concurrent tactile stimulation the results did not show any statistically significant differences.

3.2. Arousal. Means and SEMs for the ratings of the stimulus arousal are presented in Figure 4. For the effects of the emotional content of the speech the results showed that the speech only stimuli with negative content were rated as significantly more arousing than the speech only stimuli with positive content $T = 2.27$, $P < 0.05$, Cohen's $d = 1.01$. The speech tactile stimuli were not rated differently from each other in respect to the emotional content of the speech.

For the effects of the concurrent tactile stimulation the results showed that the speech tactile stimuli with positive content were rated as significantly more arousing than the speech only stimuli with positive content $T = 3.07$, $P < 0.01$, Cohen's $d = 1.88$. Also, the speech tactile stimuli with negative content were rated as significantly more arousing than the speech only stimuli with negative content $T = 2.82$, $P < 0.01$, Cohen's $d = 1.39$.

3.3. Approachability. Means and SEMs for the ratings of the stimulus approachability are presented in Figure 5. For the effects of the emotional content of the speech the results showed that the speech only stimuli with positive content were rated as significantly more approachable than the speech only stimuli with negative content $T = 3.07$, $P < 0.01$, Cohen's $d = 1.64$. Also, the speech tactile stimuli with positive content were rated as significantly more approachable than the speech tactile stimuli with negative content $T = 2.51$, $P < 0.05$, Cohen's $d = 1.16$.

For the effects of the concurrent tactile stimulation the results showed that the speech only stimuli with positive content were rated as significantly more approachable than the speech tactile stimuli with positive content $T = 2.26$, $P < 0.05$, Cohen's $d = 0.61$. The difference between the speech only stimuli with negative content and the speech tactile stimuli with negative content was not statistically significant.

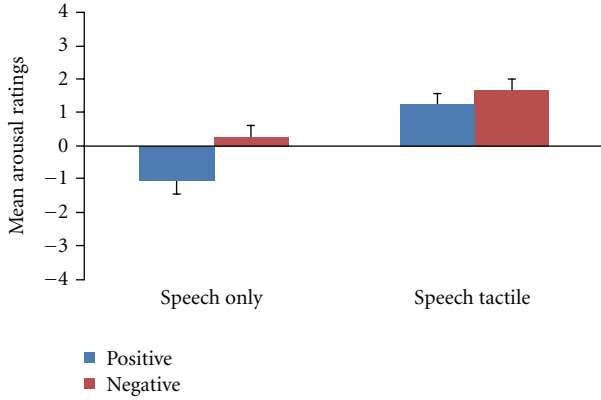


FIGURE 4: Means and SEMs for the ratings of the stimulus arousal.

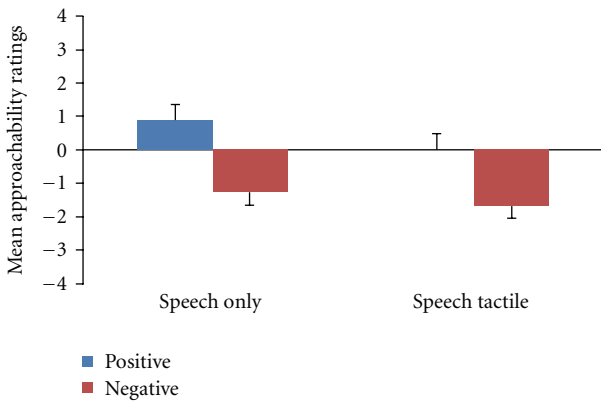


FIGURE 5: Means and SEMs for the ratings of the stimulus approachability.

3.4. *Dominance.* Means and SEMs for the ratings of the stimulus dominance are presented in Figure 6. For the effects of the emotional content of the speech the results showed no statistically significant differences in the ratings. For the effects of the concurrent tactile stimulation the results showed that the speech tactile stimuli with negative content were rated as significantly more dominant than the speech only stimuli with negative content $T = 2.21, P < 0.05$, Cohen’s $d = 0.86$. The difference between the speech tactile stimuli with positive content and the speech only stimuli with positive content was not statistically significant.

4. Summary

For the effects of the emotional content of the speech the results showed that both speech only and speech tactile stimuli with positive content were rated as more pleasant and approachable than the stimuli with negative content. In addition, the speech only stimuli with negative content were rated as more arousing than the speech only stimuli with positive content. The emotional content of the speech had no statistically significant effects for the ratings of the dominance. For the effects of the concurrent tactile stimulation the results showed that the speech tactile stimuli were in general rated as significantly more arousing and more

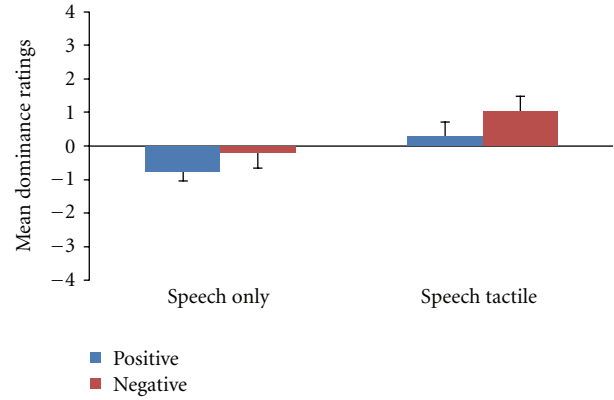


FIGURE 6: Means and SEMs for the ratings of the stimulus dominance.

TABLE 2: Emotional content of the speech sample, and the congruency of the concurrent tactile stimulation.

Speech	Tactile congruent	Tactile incongruent	
“nice to hear”	“nice to hear”	“sad to hear”	“I see”
“I see”	“I see”	“nice to hear”	“sad to hear”
“sad to hear”	“sad to hear”	“nice to hear”	“I see”

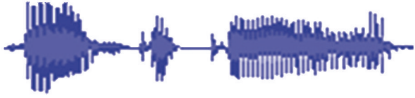

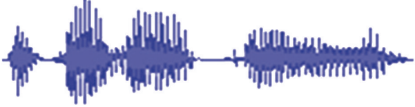
dominant than the speech only stimuli. The speech tactile stimuli were also rated as less approachable than the speech only stimuli. For the ratings of the pleasantness there were no significant differences between the speech tactile and speech only stimuli.

Experiment 2. The purpose of Experiment 2 was to study whether extracting the amplitude changes of the tactile signal from the concurrent speech sample is the optimal way to modulate the emotion-related responses related to speech. For this purpose, we varied the congruency of the tactile stimulation. By this we refer to whether the speech sample was presented simultaneously with the tactile stimulation extracted from that specific speech sample in question, or with a tactile stimulation extracted from one of the other speech samples (Table 2). In addition, we decided to use slightly longer speech samples and include neutral speech samples in Experiment 2.

5. Methods

5.1. *Participants.* Sixteen voluntary participants (eight female) took part in the study (mean age 23, range 19–30 years). All the participants were students at the University of Tampere. They were recruited from computer science courses and received a course credit from their participation. Fifteen of the participants were right handed, and one was left handed by their own report. Eight of the participants told that they hold a mobile phone in their right hand and seven told holding it in their left hand while typing a text message. All had normal or corrected to normal vision, and normal hearing and sense of touch by their own report. All the participants were fully informed for the purpose of

TABLE 3: The stimulus parameters varied in the experiment.

Stimulus in English and Finnish	Emotional content	Duration (sec)	Graphical presentation of amplitude variations
“nice to hear” “hauska kuulla”	Positive	1.5	
“I see” “vai niin”	Neutral	1.5	
“sad to hear” “ikävä kuulla”	Negative	1.5	

the study prior to the experiment. They were also informed that they were able to abort the experiment at any point without a specific reason. A consent form was signed by all the participants.

5.2. Apparatus and Procedure. The device prototype, the experimental setup, and the procedure were similar to Experiment 1.

5.3. Stimuli. Two stimulus modalities were used in the experiment: speech only and speech tactile (see Table 3). First parts (i.e., beginnings) of each sentence were varied in their emotion-related content while the final part was always the same. The beginning had either positive, negative, or neutral content. All the beginnings of the speech samples were 1500 ms long. The positive and negative beginnings were selected from the samples used in Experiment 1 that is, “nice to hear” and “sad to hear.” There was no clear neutral counterpart for these speech samples, and therefore we created one with the Loquendo synthesizer. By this we also aimed to create neutral speech stimulation without emotional prosody evident in the prerecorded samples spoken by an actor. So, the neutral speech sample “I see” in Finnish was chosen because it had no emotion-related content. Even though the neutral beginning was created with the synthesizer, the identity of the speaker was clearly the same as in the prerecorded samples. Thus, the neutral speech sample did sound quite natural, not robotic.

As the motivation was to use longer sentences instead of short phrases, a controlled confirmation and ending was needed. The ending of all speech samples was always “everything seems to go as usual.” The ending was 3000 ms long. This type of ending was chosen because it was presumed to be rather neutral in its emotional content, and it semantically continued smoothly the three possible beginnings of the sentences.

The tactile stimulation was presented only during the beginning part of the sentence (i.e., the first 1500 ms). The tactile stimulation for positive and negative speech samples (i.e., “nice to hear” and “sad to hear”) was produced similarly as in Experiment 1. The tactile stimulation for the neutral speech sample was a vibration with a static frequency and

amplitude. The amplitude of the static tactile vibration created for the neutral speech sample was the mean of the amplitudes of the tactile stimuli extracted from both positive and negative speech samples. The frequency for all the tactile stimuli was 160 Hz.

Because in the current study we wanted to know whether the congruency of the tactile stimulation had an effect on the ratings, a tactile stimulation presented simultaneously with a speech sample was either the tactile stimulation extracted from the current speech sample, or tactile stimulation extracted from one of the two other speech samples. So, for example, the positive speech sample was repeated with tactile stimulation derived from positive speech sample (congruent), with tactile stimulation derived from negative speech sample (incongruent), and with static vibration (incongruent). Thus, a total of 12 stimuli (3 speech only, 3 speech tactile with congruent tactile stimulations, and 6 speech tactile with incongruent tactile stimulations) were used in the study.

5.4. Data Analysis. Three Friedman tests were conducted in order to test whether varying the congruency of the tactile stimulation affected the ratings of a speech sample. Then, four Friedman tests were conducted in order to test whether varying the emotional content of the speech sample affected the ratings differently between the four concurrent tactile stimulation categories used (i.e., speech only, tactile derived from positive speech sample, static vibration, and tactile derived from negative speech sample). If the Friedman test revealed statistically significant differences between the ratings of the stimuli, a Wilcoxon signed-ranks test was used for pairwise comparisons.

6. Results

6.1. Pleasantness. Means and standard error of the means (SEMs) for the ratings of the stimulus pleasantness are presented in Figure 7. Varying the congruency of the tactile stimulation affected the ratings when the emotional content of the speech sample was positive $\chi = 11.2$, $P < 0.05$. In this case the speech only stimuli were rated as more pleasant than the congruent speech tactile stimuli $T = 2.62$, $P < 0.01$, Cohen’s $d = 0.78$. Other pairwise comparisons were

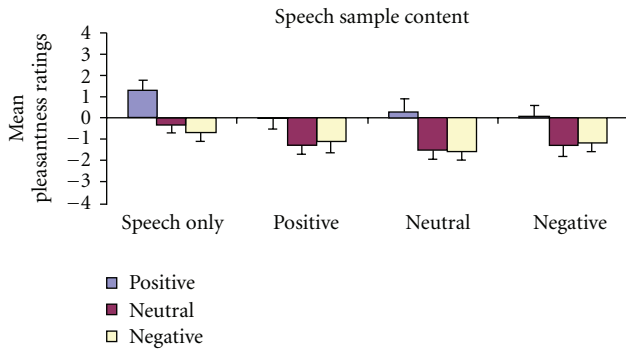


FIGURE 7: Means and SEMs for the ratings of the pleasantness by the speech sample content and the tactile stimulation.

not statistically significant. Varying the congruency of the tactile stimulation had no statistically significant effects on the ratings of neutral or negative speech samples.

Varying the emotional content of the speech affected the ratings of speech only stimuli $\chi = 15.0$, $P < 0.001$. It also affected the ratings of the speech tactile stimuli when the concurrent tactile stimulation was a static vibration $\chi = 6.45$, $P < 0.05$. However, when the concurrent tactile stimulation was derived from positive or negative speech samples varying the emotional content of the speech sample had no statistically significant effect on the ratings of the speech tactile stimuli. The results of the pairwise comparisons can be seen in Table 4.

6.2. Arousal. Means and SEMs for the ratings of the stimulus arousal are presented in Figure 8. Varying the congruency of the tactile stimulation affected the ratings when the emotional content of the speech sample was positive $\chi = 18.8$, $P < 0.001$, when the emotional content of the speech sample was neutral $\chi = 26.3$, $P < 0.001$, and when the emotional content of the speech sample was negative $\chi = 26.1$, $P < 0.001$. The results of the pairwise comparisons can be seen in Table 5. Varying the emotional content of the speech had no statistically significant effects to the ratings of the stimuli.

6.3. Approachability. Means and SEMs for the ratings of the stimulus approachability are presented in Figure 9. Varying the congruency of the tactile stimulation had no statistically significant effect on the ratings of the approachability. However, varying the emotional content of the speech affected the ratings of the speech only stimuli $\chi = 21.7$, $P < 0.001$. It also affected the ratings of the speech tactile stimuli when the concurrent tactile stimulation was derived from positive speech sample $\chi = 15.2$, $P < 0.001$, when the concurrent tactile stimulation was static vibration $\chi = 13.6$, $P < 0.001$, and when the concurrent tactile stimulation was derived from negative speech sample $\chi = 11.7$, $P < 0.05$. The results of the pairwise comparisons can be seen in Table 6.

6.4. Dominance. Means and SEMs for the ratings of the stimulus dominance are presented in Figure 10. Varying

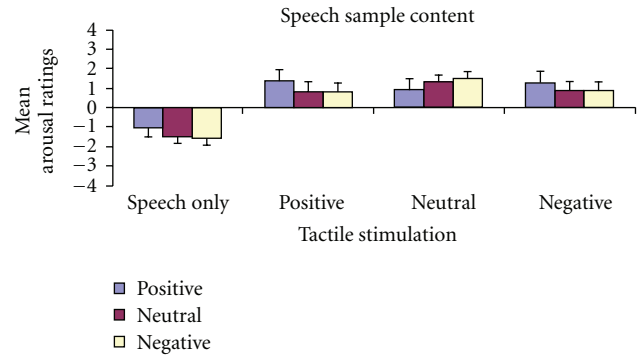


FIGURE 8: Means and SEMs for the ratings of the stimulus arousal.

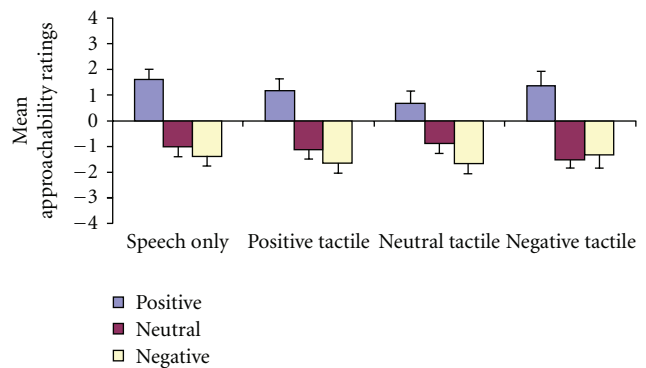


FIGURE 9: Means and SEMs for the ratings of the stimulus approachability.

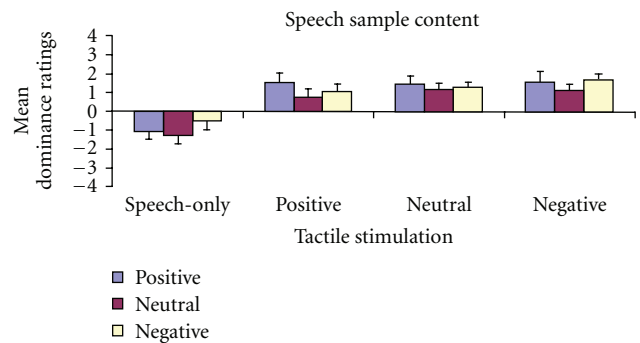


FIGURE 10: Means and SEMs for the ratings of the stimulus dominance.

the congruency of the tactile stimulation had a statistically significant effects to the ratings of the stimuli when the emotional content of the speech sample was positive $\chi = 19.8$, $P < 0.001$, when the emotional content of the speech sample was neutral $\chi = 14.8$, $P < 0.01$, and when the emotional content of the speech sample was negative $\chi = 20.4$, $P < 0.001$. Varying the emotional content of the speech had no statistically significant effects to the ratings. The results of the pairwise comparisons can be seen in Table 7.

TABLE 4: Pairwise comparisons between the speech samples within the four concurrent tactile stimulation categories. In each cell there is Wilcoxon’s T value, stars to indicate the P value, the emotional content of the speech sample rated as more pleasant, and Cohen’s D value.

Concurrent tactile stimulation	Emotional content of the speech sample		
	Positive versus neutral	Positive versus negative	Neutral versus negative
Speech only	$T = 3.0^{**}$ Positive $d = 1.13$	$T = 2.9^{**}$ Positive $d = 1.21$	ns
Derived from positive speech sample	ns	ns	ns
Static vibration	$T = 2.5^*$ Positive $d = 0.97$	$T = 2.4^*$ Positive $d = 0.98$	ns
Derived from negative speech sample	ns	ns	ns

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and ns: non-significant.

TABLE 5: Pairwise comparisons between the concurrent tactile stimulation categories within each of the speech samples used. In the table the tactile stimulation derived from positive speech sample is referred as pos and the tactile stimulation derived from negative speech sample as neg.

Emotional content of the speech sample	Concurrent tactile stimulation					
	speech-only versus pos	speech-only versus static vibration	speech-only versus neg	pos versus static vibration	pos versus neg	static vibration versus neg
Positive	$T = 2.60^{**}$ pos $d = 1.22$	$T = 2.59^*$ static vibration $d = 1.03$	$T = 2.68^{**}$ neg $d = 1.18$	ns	ns	ns
Neutral	$T = 3.10^{**}$ pos $d = 1.44$	$T = 3.55^{***}$ static vibration $d = 2.10$	$T = 3.39^{***}$ neg $d = 1.56$	ns	ns	ns
Negative	$T = 3.14^{**}$ pos $d = 1.56$	$T = 3.32^{***}$ static vibration $d = 2.28$	$T = 3.29^{***}$ neg $d = 1.65$	$T = 1.98^*$ static vibration $d = 0.43$	ns	$T = 1.98^*$ static vibration $d = 0.41$

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and ns: non-significant.

TABLE 6: Pairwise comparisons between the speech samples within the four concurrent tactile stimulation categories.

Concurrent tactile stimulation	Emotional content of the speech sample		
	positive versus neutral	positive versus negative	neutral versus negative
Speech-only	$T = 3.43^{***}$ positive $d = 1.71$	$T = 3.46^{***}$ positive $d = 2.02$	ns
Derived from positive speech sample	$T = 3.11^{**}$ positive $d = 1.38$	$T = 3.28^{***}$ positive $d = 1.56$	ns
Static vibration	$T = 2.26^*$ positive $d = 0.93$	$T = 3.04^*$ positive $d = 1.44$	$T = 2.21^*$ neutral $d = 0.56$
Derived from negative speech sample	$T = 3.23^{***}$ positive $d = 1.64$	$T = 2.77^{**}$ positive $d = 1.28$	ns

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and ns: non-significant.

TABLE 7: Pairwise comparisons between the concurrent tactile stimulation categories within each of the speech samples used.

Emotional content of the speech sample	Concurrent tactile stimulation					
	speech-only versus pos	speech-only versus static vibration	speech-only versus neg	pos versus static vibration	pos versus neg	static vibration versus neg
Positive	$T = 3.11^{**}$	$T = 2.74^*$	$T = 3.0^{**}$			
	pos $d = 1.61$	static vibration $d = 1.55$	neg $d = 1.42$	ns	ns	ns
Neutral	$T = 2.84^{**}$	$T = 3.0^{**}$	$T = 2.95^{**}$			
	pos $d = 1.12$	static vibration $d = 1.42$	neg $d = 1.44$	ns	ns	ns
Negative	$T = 2.03^*$	$T = 3.0^{**}$	$T = 2.93^{**}$			
	pos $d = 0.90$	static vibration $d = 1.29$	neg $d = 1.48$	ns	ns	ns

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and ns: non-significant.

7. Summary

Varying the tactile stimulation had the following effects to the results. In general, all the speech tactile stimuli were rated as significantly more dominant and arousing than the speech only stimuli. In addition, the speech tactile stimuli where the tactile stimulation was static vibration were in some cases rated as more arousing than the speech tactile stimuli where the tactile stimulation was derived from positive or negative speech samples. This effect, however, was independent of the congruency of the tactile signal and speech. The effect of the stimulus congruency became relevant for the ratings of the pleasantness. Congruent speech tactile stimuli were rated as less pleasant than the speech only stimuli when the emotional content of the speech was positive. Varying the tactile stimulation had no statistically significant effects for the ratings of approachability.

Varying the emotional content of the speech affected the ratings of pleasantness, but only when the stimulus was speech only or when the concurrent tactile stimulation was static vibration. In those cases the stimuli were rated adequately in respect to their emotional content. In addition, the results showed that all the stimuli with positive emotional content were rated as more approachable than the stimuli with negative or neutral emotional content. Interestingly, when the concurrent tactile stimulation was static vibration also the speech stimulus with neutral content was rated as more approachable than the stimulus with negative emotional content. The emotional content of the speech had no statistically significant effects for the ratings of the arousal or dominance.

8. Discussion

The results of both experiments showed that the speech tactile stimuli were rated as more arousing and dominant than speech only stimuli. This result, however, was not fully independent of the form of the tactile stimulation. The results of Experiment 2 showed that when the concurrent

tactile stimulation was static vibration, the speech tactile stimuli were experienced as more arousing than other speech tactile stimuli. In addition, in Experiment 1 the speech tactile stimuli were rated as less approachable than the speech only stimuli but in Experiment 2 the concurrent tactile stimulation had no effect on the ratings of approachability. However, the results of Experiment 2 suggested that congruent speech tactile stimuli were in some cases rated as less pleasant than speech only or incongruent speech tactile stimuli.

Then, the results of both experiments showed that the emotional content of the speech affected mostly the ratings of pleasantness and approachability. The stimuli with positive emotional content were in general rated as more pleasant and approachable than the stimuli with neutral or negative emotional contents. Further, the results of Experiment 2 showed that when static vibration was provided during speech the emotional content of the speech affected the pleasantness and approachability ratings more efficiently than when the stimulus was speech only or when the tactile stimulation was derived from positive or negative speech sample.

Previous studies [15, 18] suggest that by providing synchronous audiotactile or speech tactile signals the cognitive performance can be affected so that people, for example, can detect stimuli more accurately than when the signal is asynchronous. From this perspective it was reasonable to assume that temporal synchrony in amplitude variations of tactile and speech signals may have an effect on the emotion-related ratings of the stimuli as well. However, when looking at the ratings of congruent and incongruent speech tactile stimuli one can find only minor differences between the stimuli. Interestingly, they also show that at least with the current set of stimuli congruent speech tactile samples were in some cases rated as less pleasant than incongruent speech tactile samples. At this point, the reasons behind this result can only be speculated. As it can be seen from the results, both congruent and incongruent speech tactile stimuli were rated as rather neutral in respect to the pleasantness and approachability. This result was also

at least partly independent of the content of the speech despite the fact that the speech only samples were in general rated adequately in respect to their emotional content (e.g., negative speech samples were rated as unpleasant). Therefore, it seems possible that the amplitude changes used in the current study redirected the participant's attention away from the emotional message conveyed by the content of the speech sample, thus, making the experience neutral.

Then, even though the ratings of the congruent and incongruent speech tactile stimuli were rather similar, the form of the speech tactile signal did affect the ratings of the speech tactile stimuli. Speech-tactile stimuli with static vibrations were rated as more arousing than stimuli with congruent or incongruent tactile vibrations. They also had a clear effect on the ratings of the pleasantness and approachability of the speech unlike congruent and incongruent tactile vibrations. As both incongruent and congruent stimuli were rated rather similarly, it seems likely that the absence of amplitude (e.g., rhythm) changes in speech tactile stimuli with static vibrations caused the differences in the ratings. Intuitively it seems that the continuous static vibration elevated the level of arousal. From a theoretical framework [21–23] valence and arousal represent motivational parameters related to the general disposition to approach or avoid stimulation and the vigor of that tendency. Therefore, by elevating the level of arousal static vibration also activated the motivational system related to the general disposition to approach or avoid stimulation, and this way affected the experienced approachability of the speech sample. Thus, by providing static vibration simultaneously with speech we were able to create more arousing experience therefore intensifying the effects of the content of the speech sample to the experienced pleasantness and approachability of the stimulus.

There were also some similarities in the results when compared with previous studies. One previous study showed that in person-to-person communication participants used tactile stimulation to emphasize the content of the speech [10]. The current results show that tactile stimulation can modulate emotion-related responses to speech as the speech tactile stimuli were experienced as more arousing and as more dominant than speech only stimuli. This is in line with the results of the earlier studies. A novel finding in the current study was that with concurrent static tactile signal the pleasantness and approachability ratings of the content of the speech were clearly affected. This result as far as we know has not been obtained earlier. It suggests that by offering carefully selected tactile signal simultaneously with speech also other emotion-related dimensions than arousal can be affected.

One central difference emerged in the stimulus ratings between the two experiments. In Experiment 1 the participants rated the speech tactile stimuli as less approachable than the speech only stimuli. However, in Experiment 2 there were no differences between the approachability ratings of the speech tactile and speech only stimuli. This result seems to reflect observations from our previous studies [28, 29] in which the continuity of the stimuli has had an effect on the ratings of stimulus pleasantness and approachability. In general, continuous stimuli have been rated as less pleasant

and as less approachable than discontinuous stimuli. In Experiment 1 the prototype device was vibrating during the whole stimulus presentation while in Experiment 2 the prototype device vibrated only in the beginning of the stimulus. Hence, in Experiment 1 the tactile stimulation was continuous while in Experiment 2 the stimulation can be regarded as discontinuous. The results of the current study can, thus, be seen supporting the idea that the continuity of the vibrotactile stimulation is an important factor effecting the experienced pleasantness and approachability of the tactile stimulation.

From the interface design perspective the obtained results can be used to enrich emotion-related speech communication relatively easily. The static vibration was related to the elevated level of arousal and dominance as well as experiences of pleasantness and approachability. Therefore, at this point it seems that there is no need for a special algorithm detecting the emotional state of the user during conversation. Instead, the user can just send a static vibration whenever necessary simultaneously with emotional speech, for example, by squeezing the device or pushing a button. In general, the speech tactile stimuli shifted the experienced level of arousal and dominance. From a theoretical point of view [21–23] the enhancement in the subjective level of arousal can be easily seen as means to elevate the level of attentive behavior. Therefore, from application point of view, the results suggest that tactile cues can be used in mobile contexts to catch the attention during a conversation if wanted. Similarly, the static vibration cues can also be used to enhance the effect of the emotional content of the speech to the receiver. So, if one wants to make, for example, pleasant message more pleasant and approachable to the listener, static vibration works well for this purpose.

There were some restrictions in the current study. The experiment was conducted in a laboratory with a special prototype device. However, to maintain ecological validity the position of participant's hand mimicked accurately the position of a hand when a person is holding a mobile phone (i.e., the dominant hand's thumb was touching the actuators on the other side of the device, and the tips of the dominant hand's other fingers were touching the device on the other side). Therefore, all the participants received the tactile stimulation in the same parts of the hand. In addition, the vibrotactile actuators used to produce the tactile signal in the current study are similar as the vibrotactile actuators currently used in mobile phones. So, the amplitude changes varied in the study are reproducible with commercial products. Finally, it should be noted that the prototype was not capable of producing both tactile and auditory stimulation simultaneously. During the experiment the speech samples were presented to the participant via headphones. This may have some effects to the results. In real use cases, however, people often use headphones or a loudspeaker with mobile devices, for example, when walking and listening music or driving. Therefore, it seems that in future studies different user scenarios could be taken into account when studying the modulation of emotional speech with tactile stimulation.

Next it would be possible to study in a laboratory how people use vibrotactile cues during longer conversations. Also, studies outside laboratory could provide an insight on how the users would use the tactile modulation of speech during their daily activities. In addition, it would be interesting to study how other than vibrotactile haptic cues (e.g., thermal stimulation or electrotactile stimulation) could modulate emotional responses related to speech.

In summary, our current results suggest that any concurrent tactile stimulation has an effect on the ratings of arousal and dominance of speech and that this effect is independent of the emotional content of the speech. Further, the results suggest that both the experienced arousal and approachability of a spoken message can be affected by concurrent static vibration. In addition, the continuity of the tactile stimulation had an effect on the ratings of approachability. The results suggest that discontinuous or brief static tactile stimulation should be used especially in the case one wants to create speech tactile stimuli experienced as approachable and pleasant.

Acknowledgments

This paper was a part of Mobile Immersion project funded by the Finnish Funding Agency for Technology and Innovation, TEKES and steered by Nokia Research Center. The author would also like to thank the volunteers who took part in the experiment.

References

- [1] H. F. Harlow, "The nature of love," *American Psychologist*, vol. 13, no. 2, pp. 673–685, 1958.
- [2] H. F. Harlow, *Learning to Love*, Albion, San Francisco, Calif, USA, 1971.
- [3] H. F. Harlow, R. O. Dodsworth, and M. K. Harlow, "Total social isolation in monkeys," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 54, no. 1, pp. 90–97, 1965.
- [4] S. E. Jones and A. E. Yarbrough, "A naturalistic study of the meanings of touch," *Communication Monographs*, vol. 52, pp. 19–56, 1985.
- [5] M. J. Hertenstein, D. Keltner, B. App, B. A. Buleit, and A. R. Jaskolka, "Touch communicates distinct emotions," *Emotion*, vol. 6, no. 3, pp. 528–533, 2006.
- [6] J. Smith and K. MacLean, "Communicating emotion through a haptic link: design space and methodology," *International Journal of Human Computer Studies*, vol. 65, no. 4, pp. 376–387, 2007.
- [7] J. N. Bailenson, N. Yee, S. Brave, D. Merget, and D. Koslow, "Virtual interpersonal touch: expressing and recognizing emotions through haptic devices," *Human-Computer Interaction*, vol. 22, no. 3, pp. 325–353, 2007.
- [8] M. Ilves and V. Surakka, "Emotions, anthropomorphism of speech synthesis, and psychophysiology," in *Emotions in the Human Voice. Volume III: Culture and Perception*, K. Izdebski, Ed., pp. 137–152, Plural Publishing, San Diego, Calif, USA, 2009.
- [9] M. Ilves, V. Surakka, and T. Vanhala, "The effects of emotionally worded synthesized speech on the ratings of emotions and voice quality," in *Proceedings of the Affective Computing and Intelligent Interfaces (ACII '11)*, pp. 588–598, October 2011.
- [10] A. Chang, S. O'Modhrain, R. Jacob, E. Gunther, and H. Ishii, "ComTouch: design of a vibrotactile communication device," in *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '02)*, pp. 312–320, June 2002.
- [11] M. S. Beauchamp, "See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex," *Current Opinion in Neurobiology*, vol. 15, no. 2, pp. 145–153, 2005.
- [12] J. Driver and T. Noesselt, "Multisensory interplay reveals crossmodal influences on "Sensory-Specific" brain regions, neural responses, and judgments," *Neuron*, vol. 57, no. 1, pp. 11–23, 2008.
- [13] J. J. Foxe, I. A. Morocz, M. M. Murray, B. A. Higgins, D. C. Javitt, and C. E. Schroeder, "Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping," *Cognitive Brain Research*, vol. 10, no. 1-2, pp. 77–83, 2000.
- [14] J. J. Foxe, G. R. Wylie, A. Martinez et al., "Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study," *Journal of Neurophysiology*, vol. 88, no. 1, pp. 540–543, 2002.
- [15] J. P. Bresciani, M. O. Ernst, K. Drewing, G. Bouyer, V. Maury, and A. Kheddar, "Feeling what you hear: auditory signals can modulate tactile tap perception," *Experimental Brain Research*, vol. 162, no. 2, pp. 172–180, 2005.
- [16] T. Ro, J. Hsu, N. E. Yasar, L. Caitlin Elmore, and M. S. Beauchamp, "Sound enhances touch perception," *Experimental Brain Research*, vol. 195, no. 1, pp. 135–143, 2009.
- [17] J. Vroomen and M. Keetels, "Perception of intersensory synchrony: a tutorial review," *Attention, Perception, and Psychophysics*, vol. 72, no. 4, pp. 871–884, 2010.
- [18] H. Gillmeister and M. Eimer, "Tactile enhancement of auditory detection and perceived loudness," *Brain Research*, vol. 1160, no. 30, pp. 58–68, 2007.
- [19] E. C. Wilson, C. M. Reed, and L. D. Braida, "Integration of auditory and vibrotactile stimuli: effects of phase and stimulus-onset asynchrony," *Journal of the Acoustical Society of America*, vol. 126, no. 4, pp. 1960–1974, 2009.
- [20] M. Sato, C. Cavé, L. Ménard, and A. Brasseur, "Auditory-tactile speech perception in congenitally blind and sighted adults," *Neuropsychologia*, vol. 48, no. 12, pp. 3683–3686, 2010.
- [21] M. M. Bradley, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [22] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.
- [23] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.
- [24] H. Schlosberg, "Three dimensions of emotion," *Psychological Review*, vol. 61, no. 2, pp. 81–88, 1954.
- [25] J. Anttonen and V. Surakka, "Emotions and heart rate while sitting on a chair," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, pp. 491–499, April 2005.
- [26] R. J. Davidson, "Cerebral asymmetry, emotion and affective style," in *Brain Asymmetry*, R. J. Davidson and K. Hugdahl, Eds., pp. 361–387, The MIT Press, Cambridge, Mass, USA, 1995.

- [27] V. Surakka, *Contagion and modulation of human emotions [Ph.D. thesis]*, Department of Psychology, Acta Universitatis Tamperensis, 1998.
- [28] K. Salminen, V. Surakka, J. Lylykangas et al., “Emotional and behavioral responses to haptic stimulation,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, pp. 1555–1562, April 2008.
- [29] K. Salminen, V. Surakka, J. Rantala, J. Lylykangas, P. Laitinen, and R. Raisamo, “Emotional responses to haptic stimuli in laboratory versus travelling by bus contexts,” in *Proceedings of the Affective Computing and Intelligent Interaction (ACII '09)*, pp. 1–7, September 2009.
- [30] J. Rantala, R. Raisamo, J. Lylykangas et al., “The role of gesture types and spatial feedback in haptic communication,” *IEEE Transactions on Haptics*, vol. 4, no. 4, pp. 295–306, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

