ThermalWear: Exploring Wearable On-chest Thermal Displays to Augment Voice Messages with Affect

Abdallah El Ali [1], Xingyu Yang [2], Swamy Ananthanarayan [3], Thomas Röggla [1], Jack Jansen [1], Jess Hartcher-O'Brien [2], Kaspar Jansen [2], Pablo Cesar [1,2]
[1] Centrum Wiskunde & Informatica (The Netherlands), [2] Delft University of Technology (The Netherlands), [3] University of Oldenburg (Germany)
a.el.ali@cwi.nl, x.yang-6@student.tudelft.nl, s.ananthanarayan@uol.de, t.roggla@cwi.nl, jack@cwi.nl, j.hartcher-obrien@tudelft.nl, k.m.b.Jansen@tudelft.nl, p.s.cesar@cwi.nl

ABSTRACT
Voice is a rich modality for conveying emotions, however emotional prosody production can be situationally or medically impaired. Since thermal displays have been shown to evoke emotions, we explore how thermal stimulation can augment perception of neutrally-spoken voice messages with affect. We designed ThermalWear, a wearable on-chest thermal display, then tested in a controlled study (N=12) the effects of fabric, thermal intensity, and direction of change. Thereafter, we synthesized 12 neutrally-spoken voice messages, validated (N=7) them, then tested (N=12) if thermal stimuli can augment their perception with affect. We found warm and cool stimuli (a) can be perceived on the chest, and quickly without fabric (4.7-5s) (b) do not incur discomfort (c) generally increase arousal of voice messages and (d) increase / decrease message valence, respectively. We discuss how thermal displays can augment voice perception, which can enhance voice assistants and support individuals with emotional prosody impairments.

Author Keywords
thermal; affect; emotion; voice; prosody; wearable; display; chest

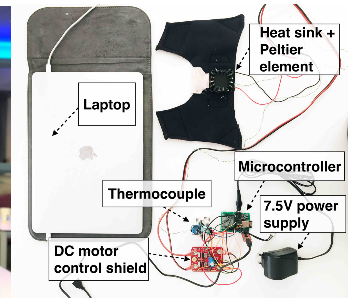
CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** *Haptic devices;* User studies;

INTRODUCTION
Human voice is a rich modality for conveying emotions [61, 63, 38], whether verbal or non-verbal [60]. This process however can be impaired due to situational factors (e.g., ironic speech [1]), remote communication settings where much non-verbal information is missing [30, 32], or medical conditions such as Autism Spectrum Disorder (ADS) [13, 56]). Within HCI, it has been shown that thermal displays can evoke emotions in isolation [20, 58, 73] or to augment media [67, 47, 58]. Emotions are often stated to be connected to sensed variations in temperature

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
DOI: https://doi.org/10.1145/3313831.3376682

(a) Male and female stretchy neoprene wetsuit tops.



(b) ThermalWear hardware components.

Figure 1: Our ThermalWear prototype and components. *Best seen in color.*

[58, 74], where warm temperatures are found to be comfortable and pleasant [29] and promote greater social proximity [24], while colder temperatures perceived as being uncomfortable [66]. This provides an opportunity to explore how and to what extent thermal displays can augment human voice with affect information, when it is lacking in emotional prosody.

While there has been much HCI research focused on augmenting media (from images to social media post) with thermal stimulation [47, 3, 42, 67], only recently has there been work on augmenting voice, and this was specifically geared towards thermal interaction with voice-based intelligent agents [31]. Within the haptics community, there has been work on augmenting remote speech communication [65], where they found that warm and cold thermal messages were useful in communicating positive and negative meanings. However in their work, the voice message per se was not augmented, rather thermal stimulation was being sent by the sender during remote speech. In this work, we designed and built ThermalWear, a wearable upper (central) chest (i.e., Sternum) thermal display that can augment voice messages that have missing emotional prosody. Our work fills what we believe is an important research gap in voice augmentation research (artificial or human impairment). Given the importance of body location in emotional processing, and the promise of wearables to provide feedback at any time or place, it is important to carefully consider where on the body thermal stimuli should occur, and what the interaction with fabrics are. To this end, we draw upon Nummenmaa et al.'s [49] bodily map of emotions, Zeagler's [77] wearability guidelines, and prior work that has shown that the central (upper) chest is an on-body site where the skin is thermally sensitive [50, 15].

In this paper, we ask: **(RQ1:)** Is it feasible to design a wearable upper-chest thermal display that ensures accurate temperature perception, thermal comfort, and fast thermal stimuli detection times suitable for augmenting brief voice messages? Specifically, in Experiment 1 we consider the effects of contact medium (fabric, no fabric), thermal intensity (low, high), and direction of change (warm, cool) on perceived temperature, perceived comfort, and detection times in wearable upper-chest thermal displays. **(RQ2:)** How does wearable thermal stimulation affect emotion perception from neutrally-spoken voice messages? Specifically, in Experiment 2 we focus on valence and arousal ratings of synthesized speech audio that lack emotional prosody (i.e., neutrally spoken). Our exploratory work offers two primary contributions: (1) we design and develop ThermalWear[1], a wearable prototype that thermally stimulates the upper chest, and evaluate it with respect to evoked temperature and comfort and (2) we present empirical findings that show neutral voice messages can be effectively augmented with affect through thermal stimulation. Below we start with a review of related work.

RELATED WORK

There are several strands of research that have strongly influenced the design and motivation of ThermalWear. We begin by examining the theory behind emotional prosody and the factors that can inhibit its comprehension. Next, we examine the link between temperature and thermoception and the use of wearables to convey emotion on the body. Lastly, we explore current HCI research in this area, to better identify where our work fits in this landscape.

Emotional Prosody Production and Processing

Emotional prosody, or the ability to interpret another person's feelings by listening to their tone of voice, is crucial for effective social communication. Research coming from neuroimaging studies provides evidence (cf., [7]) that, relative to neutral speech, "hearing" emotional prosody enhances hemodynamic responses (e.g., rapid delivery of blood to active tissues) within the mid superior temporal cortex (m-STC) [14]. This has been shown to aid decoding of emotional prosody, regardless of the type of emotion expressed [70]. This process however can be impaired due to situational factors (e.g., ironic speech [1]), remote communication settings where non-verbal information is often missing [30], or medical conditions. Two well known medical conditions that result in aberrant emotional prosody processing include Autism Spectrum Disorder (ADS) [13, 56] and right hemisphere lesions known as Aprosodias [17]. Autistic children for example have considerable difficulty in labeling emotional intonations [36]. Furthermore, it has also been shown that emotional prosody processing is more prone to errors with age [57] and stress [51]. Although there are a variety of therapies (e.g., learning to read body language cues) for improving emotional prosody comprehension, our work offers a complementary technological approach. From this perspective, our aim with ThermalWear is to lay the groundwork for designing wearable systems that use thermoception to help convey the tone of a person's voice.

[1]https://github.com/cwi-dis/ThermalWear

Body, Wearables, and Emotion

Research has shown that emotions are not only felt in the body, but that somatosensory feedback (which includes thermoception) can actively trigger conscious emotional experiences. In a sense, these models of embodied emotion state that we understand others' emotions by simulating them in our own bodies [48]. With respect to temperature, this occurs at low physiological levels and is dubbed as "temperature contagion" [10]. Building on the approach that emotions are bodily experiences, Nummenmaa et al. [49] propose a bodily map of emotions that is stated to be culturally universal [69] and applicable across child development [23]. In this model, discernible sensation patterns associated with different emotions have a strong correspondence with large changes in physiological functions [26]. Importantly, the most basic emotions were associated with sensations of elevated activity in the central, upper chest area, which likely correspond to changes in heart rate, breathing, and other vitals [41]. Moreover, prior work has also shown that the central (upper) chest is an on-body site where the skin is thermally sensitive [50, 15]. This leads us to consider the upper chest area as a promising site for investigation.

Given the importance of body location in emotional processing, and the promise of wearables to provide feedback at any time or place, it is important to carefully consider where on the body thermal stimuli should occur, and what the potential interaction with fabrics may be. With the rise of smartwatches, perhaps the most obvious area for thermal feedback is the wrist. Although researchers have explored the utility of thermal bracelets [52] for providing spatio temporal feedback, it has not been studied from the perspective of conveying emotional prosody. Given the physiological aspects of emotional perception, an alternative garment based solution might be more applicable. However, the use of clothing requires other considerations, namely thermal intensity. For example, Halvey et al. [22] showed that the presence of clothing requires higher intensity thermal changes for detection, but that these changes are more comfortable than direct stimulation on skin. Technical aspects aside, there is also the role of fashion and personal style in the design of wearable systems. As Devendorf el al. [11] point out, users are often less interested in wearing "screens" on their clothing and expect different aesthetics than smartphones or watch based displays. We discuss these issues in further detail and better motivate our design in the description of our prototype.

Thermal Displays and Emotion Perception

As alluded to earlier, thermal sensation is one intrinsic aspect of sensory and perceptual experience [40]. It is tied with several experience facets, including cognitive [45], emotional [12], and social [16, 40] phenomena. The capability of thermal stimuli to evoke emotions has been demonstrated in isolation [20, 58, 73], or to augment media [67, 47, 58]. The latter is part of a growing trend in HCI across a range of interaction scenarios, including conveying (emotional) information through temperature displays [67], improving material properties of virtual content [54], or communicative functions [25, 16, 39].

In these interaction scenarios, emotions are often stated to be connected to sensed variations in temperature [58, 74], where warm temperatures are stated to be comfortable and pleasant [29] and promote greater social proximity [24], while colder tempera-

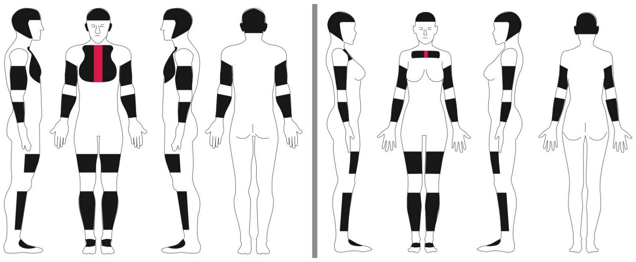
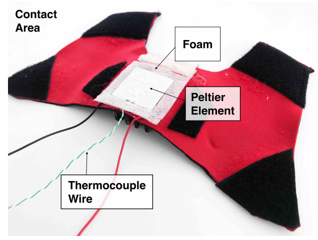
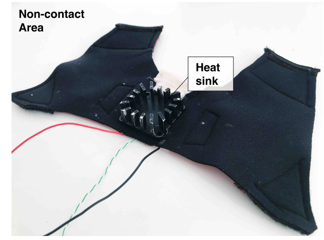


Figure 2: Overlay of three wearability body maps for male and females [77]: Motion Impedance, Garment Manufacturing, and Social Acceptability. Magenta fill highlights the stimulation site based on [49], where we ensured the exact same Peltier element position for males and females. *Best seen in color.*



(a) Vest inner layer.



(b) Vest outer layer.

Figure 3: Vest inner and outer components. *Best seen in color.*

tures are perceived as being uncomfortable [66]. This in line with previous research suggested that warmth tends to be associated with positive factors; cold, on the other hand, conveys negative factors [74]. More specifically, Salminen et al. [59] found that a 6°C change in temperature (especially when warm) was rated as unpleasant, arousing, and dominant. However a 4°C increase was still rated as arousing and dominant but pleasant.

In recent years, thermal feedback has been used to augment digital media including images [3, 47], social media posts [67], and videos [42]. There has even been research into thermal feedback on mobile phones beyond basic 'yes-no' detection of stimuli [72]. However, only recently has there been work on augmenting voice, and this was specifically geared towards thermal interaction with voice-based intelligent agents [31]. Within the haptics community, there has been work on augmenting remote speech communication [65], where they found that warm and cold thermal messages were useful in communicating positive and negative meanings. However here, the voice message per se was not augmented, rather thermal stimulation was being sent by the sender during remote speech. From a wearable design perspective, SWARM [71], an intelligent scarf designed to help a user reflect on their own emotional state, is perhaps similar to our work. However, heat here was used as a general, less granular indicator of mood and emotion. In our work, we specifically explore the use of a wearable thermal display that can augment voice messages that have missing emotional prosody.

THERMALWEAR PROTOTYPE

To build ThermalWear, several considerations went into the design, including wearability factors and choice of body location for thermal stimulation, textile fabrication, and choices related to hardware design and components. These are described below.

Wearability and Body Location

To design our ThermalWear prototype, we drew on Zeagler's [77] wearability guidelines. Our choices for on-body location came down to a balance between the desired use of the wearable device and the affordances different parts of the body offer. Among the 13 wearability body maps, three were relevant: Motion Impedance, Garment Manufacturing, and Social Acceptability. The maps can be found here: http://wcc.gatech.edu/content/wearable-technology-affordances-body-maps. To factor in these considerations, we overlay the three maps to find common acceptable areas, which leads to a new body map shown in Figure 2. From the new map, the potential areas are head, forearm, upper arm, upper chest, the front of thigh, the front of shin and instep. To narrow down the design space, we considered prior work on the bodily map of emotions [49], as alluded to in our related work, and chose the central, upper chest as our thermal stimulation site.

Textile Fabrication

Inner and Outer Vest Layers

Two stretchy neoprene wetsuit tops (a male version and a female version) are used (shown in Figure 3a). The upper chest part of a diving vest is cut out and used to place the Peltier element. The vest is thick enough to hold the Peltier element without wrinkles. The Peltier element and the heat sink are connected through the hole with thermal glue. The Peltier element is embedded on the inner side, while the heatsink is attached on the other side. A piece of silk is sewed on the vest and can cover the Peltier element if needed. For the shirt, a front opening is made, through which the Peltier element can touch the skin on the upper chest. An aluminum frame on the edge of the opening is sewed on so the structure is stable even if the shirt is stretched. Hooks and loops are sewed on the shoulder, the lower sides of chest and right next to the front opening. The inner and outer vest layers are shown in Figure 3b.

Fabric Construction and Material

For ThermalWear, we consider the feasibility of using fabric as a suitable contact medium. Generally, adding an extra fabric layer between actuators and skin always weakens the thermal stimulus [22], while increasing comfort. The fabric texture should not interfere with the thermal sensation, given our context of emotion perception through voice communication. A main requirement for a fabric is that it should have high thermal conductivity, and be able to provide a stable structure for embedded electronics. While the conductivity of polymers and yarns are all similar, the way of constructing the fibre into a piece of fabric dominates the thermal conductivity. Different construction and finishes will lead to different thickness and roughness. If the fabric is thick and rough enclosing lots of air, the conductivity is much lower. Majumdar [44] showed that thermal conductivity decreases with increase in fabric thickness. Moreover, research has shown that knitted fabrics have lower values of thermal conductivity in comparison with woven fabrics [2] because they are generally thicker than that of woven fabrics. Therefore, considering the thermal conductivity, woven fabrics are the better choice.

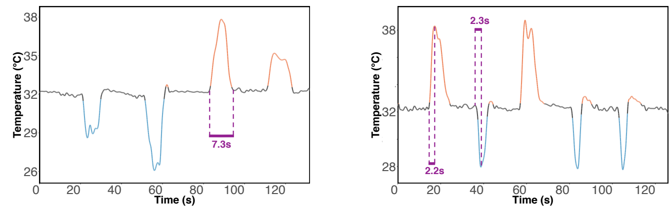
For the outer layer, we choose elastic fabric, which is a knitted and stretchy compression fabric, which enables the actuators to sit close to the human body. Additionally, such a fabric is suitable to make garments across multiple sizes, making it suitable for user studies. For the inner layer, given that the fibre composition does not strongly affect thermal conductivity, we ran early tests on material choice between cotton, wool, silk and polyester. For these tests, we applied 4V across fabrics, providing either warm or cold stimulation. Based on this early feedback, we found that silk provides the least irritation, is less sensationally stimulating than polyester, and allows heat to spread evenly (in contrast with cotton). Thermal conductivity of our silk fabric ranged between 0.08-0.1 (W/mK). Given this, silk with elastane is chosen, which is a lightweight silk that includes Lycra in the weft.

Hardware Design and Components

As in prior work [20, 52], we used Peltier elements (TEC1-12706[2] thermoelectric modules) as primary thermal actuators, given their capability of providing fast changes in temperature for creating cool or warm stimuli. For our first prototype to be used in Experiment 1, modules were controlled using a Proportional, Integral, Derivative (PID) controller to ensure accurate temperature control. Each module for this first experiment was driven through a motor driver (L298N Dual H Bridge DC stepper motor[3]) that was controlled by a custom built ESP8266 ESP-12 microcontroller that is Arduino compatible.

Our ThermalWear hardware components for Experiment 2 are shown in Figure 1b. It consists of our custom ESP8266 ESP-12 microcontroller, a 3.3v regulator (so it can be powered with a standard 5V-16V power supply), a MAX31850 thermocouple to digital converter, a K-type thermocouple, a DC motor driver (DC Motor Control Shield BTN8982TA), and a Peltier element attached to a heat sink, that connects to a laptop through a Serial (RS-232) to USB cable with 115200 baud rate. The board communicates to the Android device via a local HTTP server enabled through a REST API. A thermocouple records the real-time temperature (100ms intervals) of the touching area and sends it to the board. The motor driver enables the change of output value in terms of voltage and direction to offer different thermal stimuli. For synchronization, our control board acted as proxy to the Peltier element, where all timestamps were generated and recorded on the mobile device. For Experiment 1, the Peltier element is powered by a 9V power supply (which differs from the 7.5V power supply shown in the figure). The thermocouple reads the real-time temperature of the touching area. For the first prototype, we did preliminary testing with the PID parameters to ensure quick temperature changes and a frequent set point shift. To this end, we set K_d=0, and tuned the P and I parameters. We create four thermal change conditions (intensity x direction of change): warm-high: 32+6°C, warm-low: 32+3°C, cool-high: 32-6°C, cool-low: 32-3°C.

Given the skin's effect on thermal readings, the rate of change (ROC) is affected given the skin's absorption of heat. Furthermore, within the context of voice messages, we need to ensure a fast ROC given the short duration of voice messages. Thus, preliminary explorations led us to experiment with different PID parameters. To increase the ROC, we tune the PID parameters[4], along with placement of a medium-sized heatsink. With these parameters, we were able to reach an ROC of 1°C/s. Our temperature plots for all four temperature conditions (with fabric) in Experiment 1 is shown in Figure 4a.



(a) Experiment 1 Peltier element on fabric temperature plot across four trials.

(b) Experiment 2 Peltier element on fabric temperature plot across five trials.

Figure 4: Real-time (serial port) temperature plots from sample experiment trials across both experiments. *Best seen in color.*

EXPERIMENT 1: ON-CHEST THERMAL PERCEPTION AND COMFORT

Our aim in the first experiment is to explore thermal perception across contact medium, thermal intensity, and direction of change. Below we present our design choices, experiment setup and procedure, our first results and how they influenced subsequent design decisions.

Temperature choice

Human skin sits at a neutral temperature of 30-36°C. We chose 32°C for the touching area between skin and the Peltier element as this is within the so-called neutral zone of thermal sensation [64, 75]. Lab tests with three participants also gave an average of 31.8°C touch area after 2 min. For stimuli intensity, we draw on Halvey et al. [20], who presented thermal stimuli in conjunction with images or music, and chose ±6°C as warm or cool changes. Here, we also investigate this choice by considering ±6°C as high intensity. We furthermore test ±3°C as low intensity, given this was found to be perceivable within brief reasonable detection times, and also to ensure participant comfort [75, 67].

To evaluate the thermal perception and comfort of our ThermalWear prototype, we conducted a 2x2x2 within-subjects controlled laboratory experiment, with three independent variables: 2 (Contact Medium: Fabric vs. NoFabric) x 2 (Thermal Intensity: Low (3°C) vs. High (6°C)) x 2 (Direction of Change: Cool vs. Warm). For this experiment, we measured the following: (a) Perceived thermal intensity on a 7-point Likert scale with 4 being neutral intensity (following prior work [67]) directly in the app interface (b) Perceived comfort on a 7-point Likert scale with 4 being neither comfortable nor uncomfortable in the app interface (c) Reaction times of when thermal stimulus was felt (d) Participant (think aloud) self-reports (e) Room and skin temperature at the beginning of the experiment session.

Experiment Setup and Procedure

The experiment was conducted in an air-conditioned room, where the average room temperature across all sessions was 23°C (SD=0.26). The Arduino control circuit, the Android application, and laptop were connected via WiFi. The real-time temperature of contact side of the Peltier element is monitored through the

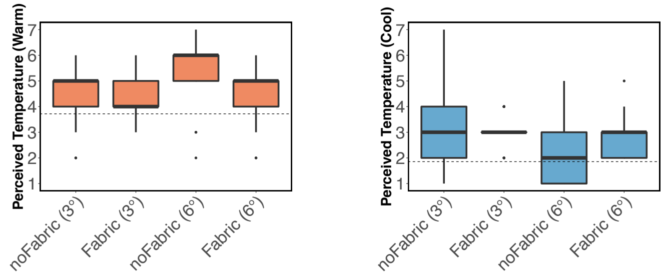
[2] http://www.thermonamic.com/TEC1-12706-English.pdf

[3] https://www.oddwires.com/l298n-dual-h-bridge-dc-stepper-motor-controller-module-for-arduino/

[4] Our final set of PID parameters for Experiment 1 are as follows: K_P=300 for 35°C, K_P=400 for 38°C, K_P=700 for 29°C, and K_P=700 for 26°C without fabric and 650 with fabric.

(a) Participant in Experiment 2. (b) Apparatus.

Figure 5: Basic experimental setup. *Best seen in color.*



(a) Temperature ratings for warm stimuli. (b) Temperature ratings for cool stimuli.

Figure 6: Boxplots for perceived temperature ratings (1-very cold to 7-very hot).

laptop. A thermometer is used to measure room temperature. An Empatica E4 wristband was used to collect the skin temperature of participants before the experiment. Paper forms on the table included: experiment instructions, informed consent form, participant information form, and Android application instructions. Study was approved by our institute's ethics committee. Figure 5 shows a study participant and our basic setup.

The participant was sat at a desk upon which there was an Android smartphone. For each experiment session, participants were welcomed, asked to read and sign the informed consent form, and then fill in the information form. During this time, their skin temperature was collected by the Empatica E4 wristband. Afterwards, they were asked to wear the ThermalWear prototype (where the Peltier element was located on their upper chest), and thereafter given a demonstration. Here, participants could get familiar with the thermal stimuli and Android application by experiencing the coolest and warmest stimuli.

After the demonstration, participants were presented with 16 trials that presented the unique combinations of experiment conditions twice (8x2). Eight trials were experienced directly on the skin (NoFabric condition), and the other eight through a silk patch (Fabric condition). The order of contact medium, thermal stimulus, and thermal intensity was rotated using a Latin square. At the start of each condition the stimulators were set to a neutral starting temperature of 32°C for 10 sec so as to adapt the skin to this temperature. Previous studies used delays between 10 sec [73] to two minutes [22] for re-adaptation, to avoid thermal adaptation effects of the skin [28]. After the adaptation period, all 16 stimuli were presented in random order. A stimulus presentation had a duration of approximately 7s, followed by a return to the neutral temperature.

After each trial, participants had to indicate when they felt the stimulus, by pushing the "Feel it!" button. This button triggers data to be stored locally on the mobile device. Two Likert-scale questions appeared on the screen (Figure 5a) asking participants to rate the stimulus in terms of intensity (1-"very cold" to 7-"very hot") and comfort (1-"very uncomfortable" to 7-"very comfortable"). These questions are similar to what was used in prior work [22]. Participants were given a two-minute break between each 8-trial set. This was to ensure sufficient time to switch between Contact Medium conditions, and to lower desensitization. At the end of the experiment, a short semi-structured interview was conducted. Participants were asked to recall the stimuli they experienced, and discuss the different conditions. The whole process was recorded by an audio recorder. The entire session took approximately 45 min. Participants were rewarded with a monetary voucher for participation.

Participants

Twelve participants (6 m, 6 f) aged 24-32 (M=26.1,SD=2.9) were recruited based on apriori power and effect size[5]. Most (10/12) were master's students. Average height was 172 cm (SD=7.8), average weight was 62.2 kg (SD=11.5), and average skin temperature was 31.4°C (SD=1.9). While there may be gender-specific differences in emotion interpretation [9, 4], some research has found no gender effect for touch-based emotion communication [27], and neither for thermal stimulation [74]. Therefore, we do not test this.

Results

We consider the effects of the three factors (Thermal Intensity, Direction of Change, Contact Medium) on each 7-point Likert-scale measure: perceived temperature ratings and perceived comfort ratings. We additionally analyze stimulus detection times. Shapiro-Wilk normality tests (all at $p<0.001$) showed that participants' warm and cool ratings, fabric and no fabric comfort ratings, and warm and cool stimuli detection times are not normally distributed, so therefore we conduct non-parametric statistical tests. Since such tests do not support computing interaction effects, we considered the aligned rank transform (ART) test [76]. This however is not applicable to our data due to high proportion of ties, and risks of inflation of Type I errors and severe loss in statistical power [53, 43].

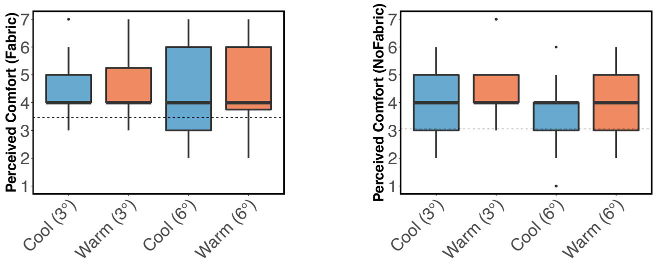
Perceived Temperature Ratings

Box plots for perceived temperature ratings for warm and cold stimuli across contact medium and thermal intensity factors are shown in Figure 6a and Figure 6b, respectively. The dashed line indicates the mean line across conditions.

Warm ratings As we compare four matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal intensity and contact medium conditions on warm temperature ratings ($\chi^2(3)=22.2$), $p<0.001$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between noFabric-6°C and Fabric-6°C ($Z = -3.4$, $p<0.001$, $r = 0.5$) and between noFabric-3°C and noFabric-6°C ($Z = -3.2$, $p<0.05$, $r = 0.46$).

Cool ratings As we compare four matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal intensity and contact medium

[5] For effect size f=0.3 under $\alpha = 0.05$ and power $(1-\beta) = 0.95$, with 16 repeated measurements within factors, we need 12 participants.



(a) Comfort ratings for fabric condition.

(b) Comfort ratings for no fabric condition.

Figure 7: Boxplots for perceived comfort ratings (1-very uncomfortable to 7-very comfortable).

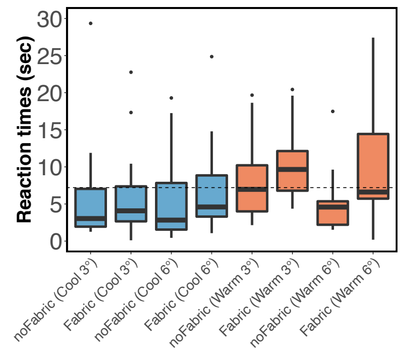


Figure 8: Detection times across contact medium, thermal intensity, and direction of change conditions.

conditions on cool temperature ratings ($\chi^2(3)$=12.5), p<0.05). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between noFabric-6°C and Fabric-6°C (Z = 2.1, p<0.05, r = 0.31), between Fabric-3°C and NoFabric-6°C (Z = 2.4, p<0.05, r = 0.35), and between NoFabric-3°C and NoFabric-6°C (Z = 2.9, p<0.05, r = 0.41).

Perceived Comfort Ratings

Box plots for perceived comfort ratings for Fabric and NoFabric across direction of change and thermal intensity factors are shown in Figure 7a and Figure 7b, respectively.

Fabric ratings As we compare four matched groups within subjects, we directly performed a Friedman rank sum test. Here however, we did not find a significant effect of thermal intensity and direction of change on Fabric comfort ratings ($\chi^2(3)$=2.8), p=0.42). This indicates that participants found the thermal stimulation neither comfortable nor uncomfortable irrespective of fabric as a contact medium, given the uniform median ratings of 4 (IQR=2).

NoFabric ratings As we compare four matched groups within subjects, we directly performed a Friedman rank sum test. Here as well, we did not find a significant effect of thermal intensity and direction of change on noFabric comfort ratings ($\chi^2(3)$=5.5), p=0.14). This indicates that participants found the thermal stimulation neither comfortable nor uncomfortable irrespective of using skin as the direct contact medium, given the uniform median ratings of 4 (IQR=2).

Thermal Stimuli Detection

Box plots for detection times across contact medium, thermal intensity, and direction of change conditions are shown in Figure 8.

Warm stimuli As we compare four matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal intensity and contact medium conditions on warm temperature ratings ($\chi^2(3)$=32.2), p<0.001). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between noFabric-6°C and noFabric-3°C (Z = -3.8, p<0.001, r = 0.6), between noFabric-3°C and Fabric-3°C (Z = -2.4, p<0.05, r = 0.4), between noFabric-6°C and Fabric-3°C (Z = -3.6, p<0.001, r = 0.52), and between noFabric-6°C and Fabric-6°C (Z = -3.3, p<0.001, r = 0.5).

Cool stimuli As we compare four matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal intensity and contact medium conditions on cool stimuli detection times ($\chi^2(3)$=13.6), p<0.05). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between noFabric-6°C and Fabric-6°C (Z = -2.2, p<0.05, r = 0.32), and between noFabric-3°C and Fabric-6°C (Z = -2.8, p<0.05, r = 0.41).

Thermal Stimuli Design Decisions

Our temperature ratings indicate that adding fabric lowers cool stimuli perception for higher thermal intensity (-6°C), especially when added to lower thermal intensity (-3°C). Furthermore, when considering contact on the naked skin, not surprisingly, the higher thermal intensity is perceived to be colder. Despite that thermal perception differs across individuals, our findings showed that all participants were able to feel the different thermal intensities for both cold and warm stimuli when felt on the skin. As in prior work [52, 28], our results overall also support the consistent finding that perception of cold thermal stimuli (whether participant sensitivity or detection times) is greater than warm stimuli. However, while higher thermal intensity made a few participants remark on the intense feeling, the comfort ratings indicate overall that this comfort is not different across contact medium. Moreover, adding a piece of silk weakens the subjective intensity of the temperature ratings, prolong detection times, and as mentioned, does not significantly increase perceived comfort. Also, since the thermal stimuli should be detectable for most people within an acceptable range, the low intensities tested cannot be used. This led us to make the following design decisions for our second experiment: for warm stimuli, warm stimuli should have a thermal intensity of 6°C (38°C). However, for the cool stimuli, we chose to decrease the high thermal intensity from -6°C (26°C) to -4°C (28°C), to ensure participant comfort.

For warm stimuli with fabric, average detection times range from 10.3s (low intensity) to 10.1s (high). For cool stimuli, average detection times range from 6.1s (low intensity) to 7.9s (high). For warm stimuli Without fabric, average detection times range from 8s (low intensity) to 4.7s (high). For cool stimuli, average detection times range from 5.5s (low intensity) to 5s (high). For Experiment 2, participants will experience the thermal stimuli in the context of voice messages. Since these messages are sudden and brief in duration, the stimuli should also be perceived quickly. When appraising stimuli for valence and arousal, participants are likely to relate the two cues together. Considering that fabric in general has "dulling" effects on thermal perception [22], and for us silk prolongs detection times without significantly increasing comfort, we chose to test thermal stimuli directly on the skin.

EXPERIMENT 2: EFFECTS OF ON-CHEST THERMAL STIMULI ON VOICE MESSAGE AFFECT

The goal of the second experiment is to explore how thermal stimuli on the upper chest influences emotion perception of neutrally spoken voice messages. Here, we aim to answer RQ2: can thermal stimuli on the upper chest influence perceived valence and arousal ratings of neutrally spoken voice messages? To test this, we first needed to ensure we have a corpus of neutrally spoken voice messages. Thereafter, taking the findings from Experiment 1, we ran a controlled laboratory study that investigates two factors: message valence (positive vs. negative) x thermal stimuli (warm vs. cool vs. baseline (no thermal stimulation)).

Generating Neutral Voice Stimuli

EU Emotion Voice Dataset

To gather suitable voice messages for testing, we explored whether there are existing datasets that contain valence and arousal ratings. The most suitable dataset is the recently published **EU-Emotion Voice Database** [37]. This dataset contains audio-recordings of 54 actors where stimuli are labeled with emotion expression, emotion intensity, valence, and arousal. The last two labels are relevant to our study.

From the list of EU-Emotion Voice stimuli, we first narrowed down our choice to audio recordings spoken in UK English. Thereafter, we looked at the mean valence and arousal ratings for positive and negative valence messages (i.e., whether the message script was positive or negative). We narrowed down this set to 12 voice recordings: six positive and six negative messages. This was based on three criteria: (a) mean valence ratings for positive messages must be greater than 3, and mean valence ratings for negative messages must be less than 3, and (b) mean arousal ratings for both positive and negative ratings should be between 3-4 (c) the message script had to be realistic if considered as a voice message heard on a mobile device, and (d) speaker genders had to be balanced as much as possible. Also, voice stimuli were carefully matched on intensity (all rated high according to EU-Emotion Voice) and duration (∼2s). The selected 12 messages with associated Emotions, Valence, and Arousal from the EU-Emotion Voice database are shown in Table 1.

Research coming from neuroimaging studies provides much evidence (cf., [7]) that, relative to neutral speech, "hearing" emotional prosody enhances hemodynamic responses within the mid superior temporal cortex (m-STC) [14], which aids in decoding of emotional prosody, regardless of the type of emotion expressed [70]. Since the goal of our work is to investigate whether thermal stimuli can influence valence and arousal perception of voice messages, we cannot test these voice messages directly, as they contain emotional prosody information which may bias participant perceptions. Given this, we draw on recent advances in speech generation, using WaveNet [68], a deep learning based generative model for raw audio. For each voice message in Table 1, we used Google Cloud Text-to-Speech[6], which is based on WaveNet, to generate a neutrally spoken counterpart. To this end, we also leave the default WaveNet pitch values of 0, to minimize injection of emotionality, which plays a role in non-tonal languages such as English [55]. The WaveNet codes (which were matched by gender to the original stimuli) are shown in Table 1. Generated voice message durations ranged between 1-3 sec.

[6]https://cloud.google.com/text-to-speech/

Annotation Study

While WaveNet generative audio does contain prosody information, from our initial observations there is a lack of emotional prosody. Moreover, while the synthesized voices for WaveNet had a mean opinion score (MOS) of 4.21 $\pm$0.081, as the authors state, sometimes it had unnatural prosody by stressing wrong words in a sentence [68]. To validate our observations, we ran an intermediary annotation study with independent raters (N=7). We created a survey where we allowed raters to hear each of the 12 neutrally spoken voice messages, and provide valence and arousal ratings. The order of voice messages was randomized. Rater responses were collected on a 9-point Likert-scale, and later transformed to a 5-point Likert-scale to allow comparison with the original valence and arousal ratings. Inter-rater Reliability (IRR) was assessed using a two-way mixed, consistency, average-measures Intra-Class Correlation (ICC) [19, 46] to verify the degree to which raters were consistent in their valence and arousal ratings. Both valence (ICC = 0.91) and arousal ratings (ICC = 0.81) had high degree of agreement amongst raters, which suggests a minimal amount of measurement error and high statistical power [8].

The mean (SD) and median (IQR) of valence and arousal ratings of neutrally spoken voice messages are shown in Table 1. These results show that our neutral voice messages had an overall neutral arousal level, and while valence was lowered for positive messages, this was not so for negative messages. This is not surprising given the valence of the message script itself, however it was important to ensure a neutral arousal level. By looking at the individual mean ratings, we decided to exclude messages 2,6,9 as the arousal ratings were either too high or too low. To ensure a balanced design, we furthermore excluded message 12, as the mean valence was decreasing, and going further away from neutral. The final list of 8 chosen voice messages are shown in **bold** in Table 1, and included as **supplementary material**.

ThermalWear Prototype Updates

For experiment 1, it took ∼6s to reach the highest thermal intensity. However, our voice messages are on average 2s in duration, which may create a mismatch in presentation time and thus requires a higher ROC. For this study, we forego the PID controller when the Peltier element reaches a setpoint, and instead apply a constant maximum output value to drive the element until it reaches the target temperature. Since this higher output value means a higher current is needed, we also replaced the motor driver with a more efficient one (DC Motor Control Shield BTN8982TA[7]). To avoid the increased heat generated due to resistance[8] brought by the higher current, we decrease the voltage applied from 9V to 7.5V.

We additionally added a buffer range to the lower output value, to avoid pushing the temperature over the target set point if we

[7]https://www.infineon.com/cms/en/product/evaluation-boards/dc-motorcontr_btn8982/

[8]In equation: $QR_v = I^2 * R_v/2$, where R_v is resistance and I is the current applied.)

#	Message Valence	Voice Message Script	Gender (WaveNet Code: en-GB)	$Emotion_{EU}$	$Valence_{NEUT}$ M (SD)	$Valence_{EU}$ M	$Arousal_{NEUT}$ M (SD)	$Arousal_{EU}$ M
1	Positive	**It's going to be great.**	f (C)	Excited	3.5 (0.6)	4.2	2.3 (0.9)	3.5
2		It's wonderful to see you.	m (B)	Happy	3.9 (0.6)	3.9	3.6 (0.7)	3.6
3		**Hmm, I love chocolate.**	m (B)	Happy	3.2 (0.3)	4.3	2.6 (1.1)	3.5
4		**You've done really well.**	f (C)	Proud	3.7 (0.9)	4.1	3.1 (1.2)	3.4
5		**I knew I could do it.**	f (C)	Proud	2.9 (0.8)	4.6	3 (0.8)	3.5
6		I'm going to bake a pirate cake for the children.	f (A)	Excited	3.9 (0.7)	4.3	3.7 (0.7)	3.6
	$Mean_{ALL}$ (SD)				3.5 (0.4)	4.2 (0.2)	3.1 (0.6)	3.5 (0.1)
	$Median_{ALL}$ (IQR)				3.6 (0.6)	4.3 (0.2)	3 (0.8)	3.5 (0.1)
	$Mean_{FIN}$ (SD)				3.4 (0.3)	4.3 (0.21)	2.8 (0.4)	3.5 (0.04)
	$Median_{FIN}$ (IQR)				3.4 (0.4)	4.3 (0.2)	2.8 (0.5)	3.5 (0.05)
7	Negative	**What do you think you are doing?**	m (B)	Angry	1.9 (0.8)	1.6	3.5 (0.8)	3.7
8		**Ugh...cover your mouth when you sneeze.**	f (C)	Disgusted	1.9 (0.4)	1.7	3.6 (0.8)	4.1
9		I've lost everything.	m (D)	Sad	2 (1)	2.7	1.6 (0.5)	3.9
10		**I tried so hard.**	f (C)	Disappointed	2.2 (0.5)	2.5	2.6 (0.8)	3.6
11		**That wasn't very nice of me.**	m (D)	Ashamed	2.4 (0.7)	2	2.1 (0.8)	3.2
12		Look what you've done?	f (C)	Angry	1.9 (0.4)	2.1	3.1 (0.9)	3.4
	$Mean_{ALL}$ (SD)				2 (0.2)	2.1 (0.4)	2.8 (0.8)	3.7 (0.3)
	$Median_{ALL}$ (IQR)				2 (0.3)	2.1 (0.6)	2.9 (1.2)	3.6 (0.4)
	$Mean_{FIN}$ (SD)				2.1 (0.3)	2 (0.4)	3 (0.7)	3.6 (0.3)
	$Median_{FIN}$ (IQR)				2 (0.4)	1.9 (0.5)	3.1 (1.1)	3.6 (0.3)

Table 1: The selected 12 messages with associated Emotions, WaveNet Code, Valence, and Arousal from the EU-Emotion Voice (EU) database, as well as results (NEUT) from our annotation study (N=7).

have a non-stop current. For heating, we define two buffer ranges. Between 35°C to 37.8°C, the output value is 300. When the temperature is between 37.8°C to 38°C, the output value is zero. For cooling, only one buffer range is set and the output value is zero when the temperature is between 28.3°to 28°C. With pilot testing, we find that the 7.5V is acceptable, if there are two breaks during the experiment to cool down the Peltier element. With this, our thermal stimuli take approximately 2s to reach the target temperature, and remain constant at that temperature for another approximately 2s. The temperature plots for five trials for one participant are shown in Figure 4b.

Experiment Setup and Procedure

Our second experiment was also conducted in an air-conditioned room, where the average room temperature across all sessions was 22.7°C (SD=0.42). Setup and procedure was identical to Experiment 1, except participants were asked to wear noise-cancelling headphones to hear the messages. For this experiment, given the 3 (Thermal Stimuli) x 2 (Positive vs. Negative Message Valence) design and that we test eight voice messages, each session consisted of 24 trials. Given our hardware setup, we provide a 2-minute break after 8 trials, resulting in two breaks per session. Study was also approved by our institute's ethics board.

Thermal stimuli were activated simultaneously with audio messages, for a total duration of 4s (2s onset and remained 2s at target temperature). As in the first study, we again set an adaptation period of 10s. The order of thermal stimuli and message valence was rotated using a Latin square. After each trial, participants had to fill out a 9-point SAM emotion rating [5] Likert-scale on the Android app for valence (1-"very negative" to 9-"very positive") and arousal (1-"very calm" to 9-"very excited"), shown as Manikin icons. A 9-point SAM scale was used here due to findings from prior work that 5-points did not leave enough expressivity for participant ratings [67]. At the end of the experiment, participants were additionally given the Comfort Rating Scale (CRS) [34, 33] to assess overall perceived comfort of our ThermalWear prototype. Finally, as before, participants were given a short semi-structured interview at the end of the session and rewarded. The whole session took approximately 45 min.

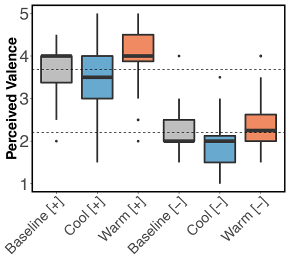
Participants

Twelve participants (6 m, 6 f) aged 24-32 (M=27.1,SD=3.7) were recruited based on apriori power and effect size[9]. Most (11/12) were students, however many were recruited from a different host institute than for the first study. Average height was 173.3 cm (SD=8.6), average weight was 64.4 kg (SD=14.6), and average skin temperature was 30.8°C (SD=1.7). When asked about how frequently they use voice messages in messaging applications (e.g., WhatsApp, FB Messenger, WeChat, etc.), three mentioned daily usage, three mentioned frequent usage, three stated sometimes, and the last three stated they rarely do. All participants were fluent in English, to ensure no misinterpretations when listening to voice messages.

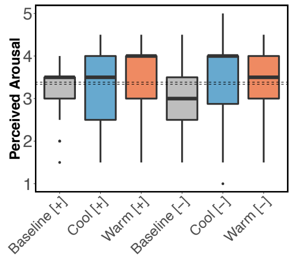
Results

We consider the effects of the two factors (Thermal Stimuli, Message Valence) on each (transformed) 5-point Likert-scale measure: valence and arousal. We then present CRS results to assess wearability of ThermalWear. Box plots for valence and arousal ratings across thermal stimuli and message valence factors are shown in Figure 9a and Figure 9b, respectively. Upper and lower dashed lines show the mean for positive and negative messages, respectively. Shapiro-Wilk normality tests showed that participants' valence ratings for positive ($p<0.001$) and negative ($p<0.05$) messages and arousal ratings for positive ($p<0.05$) and negative ($p<0.05$) messages are not normally distributed, and thereafter non-parametric statistical tests were conducted.

[9] For effect size f=0.3 under α = 0.05 and power (1-β) = 0.95, with 24 repeated measurements within factors, we need 10 participants.



(a) Perceived valence across positive and negative voice messages.



(b) Perceived arousal across positive and negative voice messages.

Figure 9: Boxplots for valence and arousal ratings of voice messages.

Valence

Positive Voice Messages As we compare three matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal stimuli on valence ratings ($\chi^2(2)$=19.0), p<0.001). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences between baseline and warm (Z = 2.4, p<0.05, r = 0.4), baseline and cool (Z = -2.8, p<0.05, r = 0.3), and warm and cool (Z = 3.8, p<0.001, r = 0.4).

Negative Voice Messages As we compare three matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal stimuli on valence ratings ($\chi^2(2)$=11.7), p<0.05). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between baseline and cool (Z = 3.5, p<0.05, r = 0.4), and warm and cool (Z = 2.9, p<0.05, r = 0.3).

Arousal

Positive Voice Messages As we compare three matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal stimuli on valence ratings ($\chi^2(2)$=11.3), p<0.05). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between baseline and warm (Z = 3.1, p<0.05, r = 0.5), and warm and cool (Z = 2.2, p<0.05, r = 0.3).

Negative Voice Messages As we compare three matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of thermal stimuli on valence ratings ($\chi^2(2)$=13.4), p<0.05). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed significant differences only between baseline and warm (Z = 3.2, p<0.001, r = 0.5), and baseline and cool (Z = 2.9, p<0.05, r = 0.4).

Subjective Feedback

One participant did not provide verbal feedback, and therefore omitted from subsequent analysis. Most participants (9/11) stated that for positive messages, warm stimuli make the message even more positive. P3 stated that: "If it's warm, I can sense the positive emotion is sincere." For cool stimuli, almost all participants (9/11) found that cool stimuli lowered the valence of the positive message (P1: "I feel it's not that positive anymore.") and made it less sincere (P8: "He is not speaking sincerely."). For negative messages, almost half of participants (5/11) did not believe warm stimuli made the message worse. With cool stimuli, around half (6/11) found it exacerbated the negativity (P2: "It makes it more angry or more severe."). Few participants (3/11) mentioned that the rate of change was more important than the temperature in interpreting speaker arousal. Other participants (2/11) stated that when the message content itself lends no clues about valence, they turn to the thermal stimuli for interpretation.

Comfort Assessment

With respect to comfort assessment of our ThermalWear prototype, we used the Comfort Rating Scales [34, 33]. On each scale, the lower the score, the more it contributes to overall subjective comfort. The results for each factor are as follows: Emotion (Md=13, IQR=12.3), Attachment (Md=10.5, IQR=6.6), Harm, (Md=3, IQR=10), Perceived Change (Md=9.5, IQR=9.5), Movement (Md=11, IQR=8.8), Anxiety (Md=8.5, IQR=5.8). Given that ThermalWear is an early prototype with many wires and a bulky Peltier element module, it was not unexpected to see divided scores. While participants may have some concerns about feeling secure ('Anxiety') wearing such a device, we believe it is important that our current prototype scores low on Harm, which is a legitimate concern due to thermal pain thresholds of ∼15°C and ∼45°C for cold and hot stimulation [28].

Thermal Stimuli Effects on Valence and Arousal

Our findings show that with respect to valence ratings, thermal stimulation does indeed affect the valence of a neutrally spoken voice message **(RQ2)**. This however varies across cool and warm stimuli. First, while warm stimuli increase the valence of positive messages, this is not so for negative ones. Cool stimuli on the other hand, lower the valence of both positive and negative messages. The findings that warm stimuli and cool stimuli are associated with positive or negative emotions, respectively, is consistent with previous work that tested thermal stimuli in isolation [73, 59, 58]. Our findings however differ from Tewell et al.'s [67] work on thermally stimulating Facebook posts, where they do not find an effect of thermal stimulation on valence ratings, only for arousal. Furthermore, we show that thermal stimulation generally increases arousal, though not uniformly across warm and cool stimuli. Warm stimuli increase arousal for positive messages over baseline, however not cool stimuli. For negative messages, both warm and cool stimuli significantly increase arousal. This is in line with earlier findings by Salminen et al. [58] who show that warm and cool stimuli can increase arousal over neutral (baseline), but that warm stimuli are more arousing than cool stimuli. We however only observe this effect for positive voice messages, where both warm and cool stimuli are similarly arousing for negative messages.

GENERAL DISCUSSION

Study Limitations

There are naturally limitations to our work. First, we do not test individuals with impaired emotion prosody processing, and instead focus on processing of synthetic neutral voice messages. Second, given that ThermalWear is still an early prototype, it was expected that some participants did not rate it highly on the Comfort Rating Scales. Indeed, depending on participant body mass and which wetsuit top they wore, they sometimes had to hold the Peltier element close to their upper chest. Nevertheless, the focus of our study was to investigate temperature and affect ratings, and not on ensuring wearability. Third given the controlled nature of our studies, we did not look at in-the-wild factors such as ambient temperature and humidity [21]. Moreover, we acknowledge that hearing voice messages in the wild can come with auditory background noise. However, for this early prototype, we set aside requirements for usage across daily interactions (e.g., within remote speech dialogs), and leave realistic mobile scenarios for future work. Relatedly, we do not test real voice messages in an unconstrained setting. Previous work has shown that vocal information such as intentions, emotions, and attitudes are more competently transmitted when familiar voices are in play [62, 35]. However, this would be difficult to manipulate in the context of familiar voices, as artificially removing emotional prosody would likely create a stranger from a familiar voice. Another consideration is that we only look at valence and arousal ratings, but not more nuanced emotion interpretations (cf., [73]) – however, this was sufficient as we draw on the widely used dimensional model of emotions [74].

Towards Wearable Thermal Voice Augmentation Displays

Our work shows that it is feasible to design wearable thermal voice augmentation displays (**RQ1**), which crosses multiple application scenarios such as augmenting personal experience [73], emotion model mapping [74], using thermal feedback as a notification mechanism [72, 52], or augmenting voice message perception of intelligent voice assistants [31]. Our work adds to these findings, by showing that neutral spoken messages can be thermally augmented with the end goal of tuning the perceived valence and arousal of such voice messages, when emotional prosody may be situationally impaired (**RQ2**). These findings can be applied to voice assistants where emotional prosody is lacking (either due to insufficient technology advances or by design).

Furthermore, our work provides opportunities to support individuals with Autism Spectrum Disorder (ASD). Past research shows impairment in emotional prosody processing in individuals with ASD [56], and that this has a neural basis which results in increased reliance in cognitive control, attentional management, and reading of intentions [13]. This has implications for emotional prosody production and processing. For production, it has for example been shown that children and adolescents with high-functioning ASD have trouble expressing emotions during storytelling tasks [18]. While in our second experiment we do not compare our artificially generated neutral voices to how individuals with ASD speak, it does provide an opportunity to further investigate this. On the other hand, even though children with ASD show no deficit in perceiving prosody, they do exhibit atypical attention to emotional tone of voice [6]. Given the foregoing, our findings enable opportunities to further explore thermal augmentation of prosody-less speech, whether from human or machine.

Design Considerations for Thermal Voice Augmentation

There are several design considerations from our work. First, with respect to perceived temperature and comfort on the upper chest, we found that both warm and cool stimuli are easily perceived and detected within a few seconds time period. What we found is that including a silk fabric does provide comfort and safety (especially for higher (38°C) and lower (26°C) temperatures, however at the cost of detection time (especially for warm stimuli), which makes it unsuitable (at least within our current ThermalWear prototype) for use in augmenting brief voice messages.

Second, from our second experiment, we see that our findings support the general findings from prior work that thermal stimulation can increase perceived arousal [58, 67]. However in contrast to Tewell et al. [67] who looked at text messages and Salminen et al. [58, 59] who looked at thermal stimuli only, our work shows that neutral voice messages can also increase/decrease in valence when thermal stimulation is applied. Why might this be so? Lee and Lim [40] already found that the role of context is quite important for interpreting thermal stimuli augmenting media, and Wilson et al. [73] highlighted that the role of emotions can vary depending on which scenario thermal stimulation is used (e.g., happy memories, social closeness, etc.). Within Tewell et al.'s [67] work, they indeed did not control for the context, and participants may have found the task ambiguous. By contrast, our voice messages can be largely seen as a standalone unit of conversation, and it steers participants to rate the valence of the spoken message, and the arousal of the speaker. This attribution of intention to the message sender likely made the experimental setup less ambiguous than what was done by Tewell et al. [67], which is why we see clear effects on valence. However, despite such clear effects, from a comparative analysis between our neutral voices and the EU-Emotion Voice recordings, it appears that we cannot perfectly recreate the missing emotional prosody from neutral messages in a consistent manner (cf., FIN valence and arousal mean ratings for $NEUT$ and EU in Table 1) . While we did not expect that thermal stimulation can perfectly supplant emotional prosody, our results indicate that thermal stimulation is effective to influence (increase/decrease) valence and arousal in the right direction.

Lastly, we can speculate about the role of body location. We investigated the upper chest as a suitable on-body location for a wearable prototype given it is purported to relate across emotions (cf., [49]). However, we do not know how this differs across other candidate body location (e.g., forearms). Our choice came down to a balance between several factors: (a) choice and inclusion of fabric and overall wearability comfort (b) felt temperature perception, detection time and temperature comfort, and (c) the potential for thermally stimulating valence and arousal by choosing the most suitable body location. This is important, as for example when considering body location, we had to additionally factor in how much calibration is need, to what extent the forearms are overused as a site for stimulation, and to what extent do users find a site socially acceptable. For inclusion of fabrics, we can further explore how to ensure a sufficient ROC for differing fabrics, and importantly, how to design wearable thermal displays that participants can use on their own garments in the future.

CONCLUSION

We presented ThermalWear, a wearable prototype that thermally stimulates the central (upper) chest, with the end goal of thermally augmenting perception of voice messages with affect. We ran two studies, first investigating thermal perception and comfort (using fabric) on the upper chest, and then looking at the effects of thermal stimuli (warm, cool) and voice message content (positive, negative) on valence and arousal ratings of neutral AI-generated voice messages. We found that thermal chest stimulation can alter both arousal and valence ratings, supporting previous work that showed warm stimuli can increase valence, while cool stimuli lower valence. Our findings enable further research on how thermal augmentation can help design intelligent voice assistants, and aid communication to those with impairments (e.g., ASD) in emotional prosody processing.

REFERENCES

[1] Gregory A. Bryant. 2010. Prosodic Contrasts in Ironic Speech. *Discourse Processes - DISCOURSE PROCESS* 47 (10 2010), 545–566. DOI: http://dx.doi.org/10.1080/01638530903531972

[2] M. Abitha, Roshanara, and V. Subramaniam. 2015. Thermal Conductivity of Fabrics. In *Advances in Mechanical Engineering (Applied Mechanics and Materials)*, Vol. 813. Trans Tech Publications Ltd, 768–772. DOI:http://dx.doi.org/10.4028/www.scientific.net/AMM.813-814.768

[3] Moses Akazue, Martin Halvey, Lynne Baillie, and Stephen Brewster. 2016. The Effect of Thermal Stimuli on the Emotional Perception of Images. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4401–4412. DOI:http://dx.doi.org/10.1145/2858036.2858307

[4] Pascal Belin, Sarah Fillion-Bilodeau, and Frédéric Gosselin. 2008. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods* 40, 2 (01 May 2008), 531–539. DOI: http://dx.doi.org/10.3758/BRM.40.2.531

[5] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49 – 59. DOI: http://dx.doi.org/https://doi.org/10.1016/0005-7916(94)90063-9

[6] Patricia J. Brooks and Bertram O. Ploog. 2013. Attention to emotional tone of voice in speech perception in children with autism. *Research in Autism Spectrum Disorders* 7, 7 (2013), 845 – 857. DOI:http://dx.doi.org/https://doi.org/10.1016/j.rasd.2013.03.003

[7] Carolin Brük, Benjamin Kreifelts, and Dirk Wildgruber. 2011. "Emotional voices in context: A neurobiological model of multimodal affective information processing". *Physics of Life Reviews"* 8, 4 (2011), 383 – 403. DOI:http://dx.doi.org/https://doi.org/10.1016/j.plrev.2011.10.002

[8] Domenic V Cicchetti. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment* 6, 4 (1994), 284.

[9] O. Collignon, S. Girard, F. Gosselin, D. Saint-Amour, F. Lepore, and M. Lassonde. 2010. Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia* 48, 1 (2010), 220 – 225. DOI: http://dx.doi.org/https://doi.org/10.1016/j.neuropsychologia.2009.09.007

[10] Ella A. Cooper, John Garlick, Eric Featherstone, Valerie Voon, Tania Singer, Hugo D. Critchley, and Neil A. Harrison. 2015. You Turn Me Cold: Evidence for Temperature Contagion. *PLOS ONE* 9, 12 (12 2015), 1–18. DOI:http://dx.doi.org/10.1371/journal.pone.0116126

[11] Laura Devendorf, Joanne Lo, Noura Howell, Jung Lin Lee, Nan-Wei Gong, M. Emre Karagozler, Shiho Fukuhara, Ivan Poupyrev, Eric Paulos, and Kimiko Ryokai. 2016. "I Don'T Want to Wear a Screen": Probing Perceptions of and Possibilities for Dynamic Displays on Clothing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 6028–6039. DOI: http://dx.doi.org/10.1145/2858036.2858192

[12] Lawrence E Williams and John A Bargh. 2008. Experiencing Physical Warmth Promotes Interpersonal Warmth. *Science (New York, N.Y.)* 322 (11 2008), 606–7. DOI:http://dx.doi.org/10.1126/science.1162548

[13] Inge-Marie Eigsti, Jillian Schuh, Einar Mencl, Robert T. Schultz, and Rhea Paul. 2012. The neural underpinnings of prosody in autism. *Child Neuropsychology* 18, 6 (2012), 600–617. DOI: http://dx.doi.org/10.1080/09297049.2011.639757 PMID: 22176162.

[14] T. Ethofer, D. Van De Ville, K. Scherer, and P. Vuilleumier. 2009. Decoding of Emotional Information in Voice-Sensitive Cortices. *Current Biology* 19, 12 (June 23, 2009), 1028–1033.

[15] Nicola Gerrett, Yacine Ouzzahra, and George Havenith. 2015. *Distribution of Skin Thermal Sensitivity*. 1–17. DOI: http://dx.doi.org/10.1007/978-3-319-26594-0_72-1

[16] Daniel Gooch and Leon Watts. Communicating Social Presence Through Thermal Hugs.

[17] Philip Gorelick and Elliott Ross. 1987. The aprosodias: Further functional-anatomical evidence for the organisation of affective language in the right hemisphere. *Journal of neurology, neurosurgery, and psychiatry* 50 (06 1987), 553–60. DOI:http://dx.doi.org/10.1136/jnnp.50.5.553

[18] Ruth B. Grossman, Lisa R. Edelson, and Helen Tager-Flusberg. 2013. Emotional Facial and Vocal Expressions During Story Retelling by Children and Adolescents With High-Functioning Autism. *Journal of Speech, Language, and Hearing Research* 56, 3 (2013), 1035–1044. DOI: http://dx.doi.org/10.1044/1092-4388(2012/12-0067)

[19] Kevin A. Hallgren. 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology* 8, 1 (2012), 23–34. DOI: http://dx.doi.org/10.20982/tqmp.08.1.p023

[20] Martin Halvey, Michael Henderson, Stephen A. Brewster, Graham Wilson, and Stephen A. Hughes. 2012a. Augmenting Media with Thermal Stimulation. In *Haptic and Audio Interaction Design*, Charlotte Magnusson, Delphine Szymczak, and Stephen Brewster (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 91–100.

[21] Martin Halvey, Graham Wilson, Stephen Brewster, and Stephen Hughes. 2012b. "Baby It's Cold Outside": The Influence of Ambient Temperature and Humidity on Thermal Feedback. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 715–724. DOI: http://dx.doi.org/10.1145/2207676.2207779

[22] Martin Halvey, Graham Wilson, Yolanda Vazquez-Alvarez, Stephen A. Brewster, and Stephen A. Hughes. 2011. The Effect of Clothing on Thermal Feedback Perception. In *Proceedings of the 13th International Conference on Multimodal Interfaces (ICMI '11)*. ACM, New York, NY, USA, 217–220. DOI: http://dx.doi.org/10.1145/2070481.2070519

[23] Jari Hietanen, Enrico Glerean, Riitta Hari, and Lauri Nummenmaa. 2016. Bodily maps of emotions across child development. *Developmental Science* 19 (02 2016), n/a–n/a. DOI:http://dx.doi.org/10.1111/desc.12389

[24] Hans IJzerman and Gün R. Semin. 2009. The Thermometer of Social Relations: Mapping Social Proximity on Temperature. *Psychological Science* 20, 10 (2009), 1214–1220. DOI: http://dx.doi.org/10.1111/j.1467-9280.2009.02434.x PMID: 19732385.

[25] Ken Iwasaki, Takashi Miyaki, and Jun Rekimoto. 2010. AffectPhone: A Handset Device to Present User's Emotional State with Warmth/Coolness. In *B-Interface*.

[26] C.E. Izard. 1991. *The Psychology of Emotions*. Springer US. https://books.google.nl/books?id=RPv-shA_sxMC

[27] Matthew J Hertenstein, Rachel Holmes, Margaret McCullough, and Dacher Keltner. 2009. The Communication of Emotion Via Touch. *Emotion (Washington, D.C.)* 9 (09 2009), 566–73. DOI: http://dx.doi.org/10.1037/a0016108

[28] L. A. Jones and H. Ho. 2008. Warm or Cool, Large or Small? The Challenge of Thermal Displays. *IEEE Transactions on Haptics* 1, 1 (Jan 2008), 53–70. DOI: http://dx.doi.org/10.1109/TOH.2008.2

[29] Kazuyuki Kanosue, Norihiro Sadato, Tomohisa Okada, Tamae Yoda, Sadamu Nakai, Kyoko Yoshida, Takayoshi Hosono, Kei Nagashima, Tomoko Yagishita, Osamu Inoue, Kaoru Kobayashi, and Yoshiharu Yonekura. 2002. Brain activation during whole body cooling in humans studied with functional magnetic resonance imaging. *Neuroscience Letters* 329, 2 (2002), 157 – 160. DOI: http://dx.doi.org/https://doi.org/10.1016/S0304-3940(02)00621-3

[30] M.R. Key. 1980. *The Relationship of Verbal and Nonverbal Communication*. Mouton. https://books.google.nl/books?id=Tbg98xnbIYwC

[31] Seyeong Kim, Yea-kyung Row, and Tek-Jin Nam. 2018. Thermal Interaction with a Voice-based Intelligent Agent. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW631, 6 pages. DOI: http://dx.doi.org/10.1145/3170427.3188656

[32] M.L. Knapp, J.A. Hall, and T.G. Horgan. 2013. *Nonverbal Communication in Human Interaction*. Cengage Learning. https://books.google.nl/books?id=rWoWAAAAQBAJ

[33] James F. Knight and Chris Baber. 2005. A Tool to Assess the Comfort of Wearable Computers. *Human Factors* 47, 1 (2005), 77–91. DOI: http://dx.doi.org/10.1518/0018720053653875 PMID: 15960088.

[34] J. F. Knight, C. Baber, A. Schwirtz, and H. W. Bristow. 2002. The comfort assessment of wearable computers. In *Proceedings. Sixth International Symposium on Wearable Computers,*. 65–72. DOI: http://dx.doi.org/10.1109/ISWC.2002.1167220

[35] Jody Kreiman and Diana Sidtis. 2011. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Wiley-Blackwell. DOI: http://dx.doi.org/10.1002/9781444395068

[36] Diana Van Lancker, Cathleen Cornelius, and Jody Kreiman. 1989. Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology* 5, 2-3 (1989), 207–226. DOI:http://dx.doi.org/10.1080/87565648909540433

[37] Amandine Lassalle, Delia Pigat, Helen O'Reilly, Steve Berggen, Shimrit Fridenson-Hayo, Shahar Tal, Sigrid Elfström, Anna Råde, Ofer Golan, Sven Bölte, Simon Baron-Cohen, and Daniel Lundqvist. 2019. The EU-Emotion Voice Database. *Behavior Research Methods* 51, 2 (01 Apr 2019), 493–506. DOI: http://dx.doi.org/10.3758/s13428-018-1048-1

[38] Petri Laukka. 2017. *Vocal Communication of Emotion*. Springer International Publishing, Cham, 1–6. DOI: http://dx.doi.org/10.1007/978-3-319-28099-8_562-1

[39] Wonjun Lee and Youn-kyung Lim. 2010. Thermo-message: Exploring the Potential of Heat As a Modality of Peripheral Expression. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 4231–4236. DOI: http://dx.doi.org/10.1145/1753846.1754131

[40] Wonjun Lee and Youn-kyung Lim. 2012. Explorative research on the heat as an expression medium: focused on interpersonal communication. *Personal and Ubiquitous Computing* 16, 8 (01 Dec 2012), 1039–1049. DOI: http://dx.doi.org/10.1007/s00779-011-0424-y

[41] ROBERT W. LEVENSON. 2003. Blood, Sweat, and Fears. *Annals of the New York Academy of Sciences* 1000, 1 (2003), 348–366. DOI: http://dx.doi.org/10.1196/annals.1280.016

[42] Markus Löchtefeld, Nadine Lautemann, Sven Gehring, and Antonio Krüger. 2014. ambiPad: Enriching Mobile Digital Media with Ambient Feedback. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI '14)*. ACM, New York, NY, USA, 295–298. DOI: http://dx.doi.org/10.1145/2628363.2628395

[43] Haiko Luepsen. 2017. The aligned rank transform and discrete variables: A warning. *Communications in Statistics - Simulation and Computation* 46, 9 (2017), 6923–6936. DOI:http://dx.doi.org/10.1080/03610918.2016.1217014

[44] Abhijit Majumdar. 2011. Modelling of thermal conductivity of knitted fabrics made of cotton and bamboo yarns using artificial neural network. *The Journal of The Textile Institute* 102, 9 (2011), 752–762. DOI: http://dx.doi.org/10.1080/00405000.2010.516929

[45] Catherine Mcconnell, Djordje Popovic, Chris Berka, Dan Levendowski, and Gene Davis. 2014. Somneo: Device with Thermal Stimulation for Modulating Sleep Architecture and Enhancing Neuro-Cognitive Function. In *8th International Conference on Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making Through Adaptive Systems - Volume 8534*. Springer-Verlag, Berlin, Heidelberg, 296–304. DOI: http://dx.doi.org/10.1007/978-3-319-07527-3_28

[46] Kenneth O McGraw and Seok P Wong. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological methods* 1, 1 (1996), 30.

[47] Mutsuhiro Nakashige, Minoru Kobayashi, Yuriko Suzuki, Hidekazu Tamaki, and Suguru Higashino. 2009. "Hiya-Atsu" Media: Augmenting Digital Media with Temperature. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. ACM, New York, NY, USA, 3181–3186. DOI: http://dx.doi.org/10.1145/1520340.1520453

[48] Paula M. Niedenthal. 2007. Embodying emotion. *Science* 316 5827 (2007), 1002–5.

[49] Lauri Nummenmaa, Enrico Glerean, Riitta Hari, and Jari K. Hietanen. 2014. Bodily maps of emotions. *Proceedings of the National Academy of Sciences* 111, 2 (2014), 646–651. DOI:http://dx.doi.org/10.1073/pnas.1321664111

[50] Yacine Ouzzahra, George Havenith, and Bernard Redortier. 2012. Regional distribution of thermal sensitivity to cold at rest and during mild exercise in males. *Journal of Thermal Biology* 37, 7 (2012), 517 – 523. DOI: http://dx.doi.org/https: //doi.org/10.1016/j.jtherbio.2012.06.003

[51] Silke Paulmann, Desire Furnes, Anne Ming BÃžkenes, and Philip J. Cozzolino. 2016. How Psychological Stress Affects Emotional Prosody. *PLOS ONE* 11, 11 (11 2016), 1–21. DOI:http://dx.doi.org/10.1371/journal.pone.0165022

[52] Roshan Lalitha Peiris, Yuan-Ling Feng, Liwei Chan, and Kouta Minamizawa. 2019. ThermalBracelet: Exploring Thermal Haptic Feedback Around the Wrist. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 170, 11 pages. DOI: http://dx.doi.org/10.1145/3290605.3300400

[53] Kathleen Peterson. 2002. Six modifications of the aligned rank transform test for interaction. *Journal of Modern Applied Statistical Methods* 1, 1 (2002), 100–109.

[54] Hendrik Richter, Doris Hausen, Sven Osterwald, and Andreas Butz. 2012. Reproducing Materials of Virtual Elements on Touchscreens Using Supplemental Thermal Feedback. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*. ACM, New York, NY, USA, 385–392. DOI: http://dx.doi.org/10.1145/2388676.2388765

[55] Emma Rodero. 2011. Intonation and Emotion: Influence of Pitch Levels and Contour Type on Creating Emotions. *Journal of Voice* 25, 1 (2011), e25 – e34. DOI: http://dx.doi.org/https: //doi.org/10.1016/j.jvoice.2010.02.002

[56] Gabriela Rosenblau, Dorit Kliemann, Isabel Dziobek, and Hauke R. Heekeren. 2016. Emotional prosody processing in autism spectrum disorder. *Social Cognitive and Affective Neuroscience* 12, 2 (08 2016), 224–239. DOI: http://dx.doi.org/10.1093/scan/nsw118

[57] Melissa Ryan, Janice Murray, and Ted Ruffman. 2009. Aging and the Perception of Emotion: Processing Vocal Expressions Alone and With Faces. *Experimental Aging Research* 36, 1 (2009), 1–22. DOI: http://dx.doi.org/10.1080/03610730903418372

[58] Katri Salminen, Veikko Surakka, Jukka Raisamo, Jani Lylykangas, Johannes Pystynen, Roope Raisamo, Kalle Mäkelä, and Teemu Ahmaniemi. 2011. Emotional Responses to Thermal Stimuli. In *Proceedings of the 13th International Conference on Multimodal Interfaces (ICMI '11)*. ACM, New York, NY, USA, 193–196. DOI: http://dx.doi.org/10.1145/2070481.2070513

[59] Katri Salminen, Veikko Surakka, Jukka Raisamo, Jani Lylykangas, Roope Raisamo, Kalle Mäkelä, and Teemu Ahmaniemi. 2013. Cold or Hot? How Thermal Stimuli Are Related to Human Emotional System?. In *Haptic and Audio Interaction Design*, Ian Oakley and Stephen Brewster (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 20–29.

[60] Disa A. Sauter, Frank Eisner, Andrew J. Calder, and Sophie K. Scott. 2010. Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Quarterly Journal of Experimental Psychology* 63, 11 (2010), 2251–2272. DOI: http://dx.doi.org/10.1080/17470211003721642 PMID: 20437296.

[61] Klaus R. Scherer. 2003. Vocal Communication of Emotion: A Review of Research Paradigms. *Speech Commun.* 40, 1-2 (April 2003), 227–256. DOI: http://dx.doi.org/10.1016/S0167-6393(02)00084-5

[62] Diana Sidtis and Jody Kreiman. 2012. In the Beginning Was the Familiar Voice: Personally Familiar Voices in the Evolutionary and Contemporary Biology of Communication. *Integrative Psychological and Behavioral Science* 46, 2 (01 Jun 2012), 146–159. DOI: http://dx.doi.org/10.1007/s12124-011-9177-4

[63] Emiliana Simon-Thomas, Dacher J Keltner, Disa Sauter, Lara Sinicropi-Yao, and Anna Abramson. 2009. The Voice Conveys Specific Emotions: Evidence From Vocal Burst Displays. *Emotion (Washington, D.C.)* 9 (12 2009), 838–46. DOI:http://dx.doi.org/10.1037/a0017810

[64] J. C. Stevens. 1991. *Thermal Sensibility*. Lawrence Erlbaum, Chapter 3, 61–90.

[65] Katja Suhonen, Sebastian Müller, Jussi Rantala, Kaisa Väänänen-Vainio-Mattila, Roope Raisamo, and Vuokko Lantz. 2012. Haptically Augmented Remote Speech Communication: A Study of User Practices and Experiences. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12)*. ACM, New York, NY, USA, 361–369. DOI: http://dx.doi.org/10.1145/2399016.2399073

[66] EUN-JUNG SUNG, SEUNG-SCHIK YOO, HYO WOON YOON, SUNG-SUK OH, YEJI HAN, and HYUN WOOK PARK. 2007. Brain activation related to affective dimension during thermal stimulation in humans: a functional magnetic resonance imaging study. *International Journal of Neuroscience* 117, 7 (2007), 1011–1027. DOI: http://dx.doi.org/10.1080/00207450600934432

[67] Jordan Tewell, Jon Bird, and George R. Buchanan. 2017. The Heat is On: A Temperature Display for Conveying Affective Feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1756–1767. DOI: http://dx.doi.org/10.1145/3025453.3025844

[68] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. In *Arxiv*. https://arxiv.org/abs/1609.03499

[69] Sofia Volynets, Enrico Glerean, Jari Hietanen, Riitta Hari, and Lauri Nummenmaa. 2019. Bodily maps of emotions are culturally universal. *Emotion* (07 2019). DOI: http://dx.doi.org/10.1037/emo0000624

[70] Sarah Wiethoff, Dirk Wildgruber, Benjamin Kreifelts, Hubertus Becker, Cornelia Herbert, Wolfgang Grodd, and Thomas Ethofer. 2008. Cerebral processing of emotional prosody – influence of acoustic parameters and arousal. *NeuroImage* 39, 2 (2008), 885 – 893. DOI: http://dx.doi.org/https://doi.org/10.1016/j.neuroimage.2007.09.028

[71] Michele A. Williams, Asta Roseway, Chris O'Dowd, Mary Czerwinski, and Meredith Ringel Morris. 2015. SWARM: An Actuated Wearable for Mediating Affect. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 293–300. DOI: http://dx.doi.org/10.1145/2677199.2680565

[72] Graham Wilson, Stephen Brewster, Martin Halvey, and Stephen Hughes. 2012. Thermal Icons: Evaluating Structured Thermal Feedback for Mobile Interaction. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 309–312. DOI: http://dx.doi.org/10.1145/2371574.2371621

[73] Graham Wilson, Gavin Davidson, and Stephen A. Brewster. 2015. In the Heat of the Moment: Subjective Interpretations of Thermal Feedback During Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2063–2072. DOI: http://dx.doi.org/10.1145/2702123.2702219

[74] Graham Wilson, Dobromir Dobrev, and Stephen A. Brewster. 2016. Hot Under the Collar: Mapping Thermal Feedback to Dimensional Models of Emotion. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4838–4849. DOI: http://dx.doi.org/10.1145/2858036.2858205

[75] Graham Wilson, Martin Halvey, Stephen A. Brewster, and Stephen A. Hughes. 2011. Some Like It Hot: Thermal Feedback for Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2555–2564. DOI:http://dx.doi.org/10.1145/1978942.1979316

[76] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proc. CHI '11*. ACM, New York, NY, USA, 143–146. DOI: http://dx.doi.org/10.1145/1978942.1978963

[77] Clint Zeagler. 2017. Where to Wear It: Functional, Technical, and Social Considerations in On-body Location for Wearable Technology 20 Years of Designing for Wearability. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (ISWC '17)*. ACM, New York, NY, USA, 150–157. DOI: http://dx.doi.org/10.1145/3123021.3123042