ThermalSense: A Sensory Substitution System for Enabling Perception of Thermal Information

IDDO YEHOSHUA WALD, University of Bremen, Germany and Reichman University, Israel
YONI BRONSTEIN, Reichman University, Israel
AMBER MAIMON*, Ben Gurion University of the Negev, Israel and Reichman University, Israel
AMIR AMEDI*, Reichman University, Israel

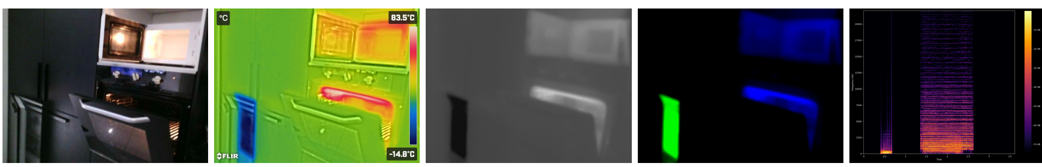

Fig. 1. The ThermalSense mapping process high level stages (from left to right). A standard image; A thermal image taken by the FLIR camera, overlaid on the original image; The thermal image converted to grayscale in the range of 0-100℃ and cleared of overlayed content; The image mapped to the chosen hot and cold color palette; The spectrogram of the soundscape generated from the mapped image by the EyeMusic algorithm.

We present the development process, design considerations and generalized system engineering for the ThermalSense sensory substitution system. The system utilizes visual-to-auditory sensory substitution to convey thermal properties through auditory cues. In an online proof of concept study evaluating the system, participants demonstrated a high level of accuracy in recognizing and localizing thermal information which is otherwise invisible in visual scenes. The ability of users to gain proficiency within a short training period was accompanied by a positive user experience evaluation. These results provide an indication of the system's potential to enable an extended sense, serving as an extension of visual perception to the range of thermal information beyond its natural capacities. The proposed system architecture, together with these findings, set the ground for the further development of this and similar systems, as well as general advances in the study of sensory perception.

CCS Concepts: • **Human-centered computing → Interaction devices**; **Sound-based input / output**; **Interaction techniques**; **Human computer interaction (HCI)**; **Mixed / augmented reality**; **Visualization systems and tools**.

Additional Key Words and Phrases: Multisensory integration, Sensory substitution, Thermal perception, Human computer interaction, NeuroHCI, Sensory augmentation, Sensory extension

*Both authors contributed equally to this research.

Authors' Contact Information: Iddo Yehoshua Wald, University of Bremen, Digital Media Lab, Bremen, Germany and Reichman University, Herzliya, Israel, wald@uni-bremen.de; Yoni Bronstein, Reichman University, Herzliya, Israel, yonbro22@gmail.com; Amber Maimon, Ben Gurion University of the Negev, Be'er Sheva, Israel and Reichman University, Herzliya, Israel, maimon.amber@gmail.com; Amir Amedi, Reichman University, Herzliya, Israel, amir.amedi@runi.ac.il.

ACM 2573-0142/2025/6-ARTEICS002
https://doi.org/10.1145/3733048

ACM Reference Format:
Iddo Yehoshua Wald, Yoni Bronstein, Amber Maimon, and Amir Amedi. 2025. ThermalSense: A Sensory Substitution System for Enabling Perception of Thermal Information. *Proc. ACM Hum.-Comput. Interact.* 9, 4, Article EICS002 (June 2025), 21 pages. https://doi.org/10.1145/3733048

1 Introduction

Our perception of the world is limited by the capacities of our senses. The limits vary from one person to another, and we are often fascinated by these interpersonal differences in experiencing the world, such as with color blindness. However, interpersonal differences are dwarfed in comparison to the sensory experiences of other species. Bats, for example, can perceive higher auditory frequencies than humans and localize using audition alongside vision, an ability sometimes described as ”seeing through sound”[47]. Visual capabilities in general are extremely diverse throughout the animal kingdom. While humans can process only three channels of color (red, green and blue), the mantis shrimp can process 12 channels of color, detect UV (ultraviolet), and polarized light which it uses for communicating [61].

In nature, there is a known interplay between thermal and visual properties [51] such that there is a direct physical connection between heat and infrared frequencies. This results in some forms of thermal information perception being, in essence, an extension of visual information perception. Yet humans are incapable of accessing these frequencies with their natural visual sensory organs. Like the mantis shrimp mentioned above, some animals have developed the ability to sense or see in the infrared frequencies. This is seen in some species of snakes, such as rattlesnakes and pythons [13], which identify their prey by its body heat, and in the black fire beetle, known for its ability to identify forest fires from a distance using infrared receptors [14]. Thermal vision, though prevalent throughout the animal kingdom, is considered one of the paradigm examples of speculative sensory expansion for humans.

Thermal vision can be seen among the superpowers of Superman, in design fictions at the interface of science and design like Superflux’s ”Song of the Machine” [60], and in many more examples from popular culture. Over the last two decades, the interest in expanding our spectral vision range has been growing due to the unique features the thermal spectrum offers. Among these features are: invariance to light conditions, enabling a glimpse into the past by allowing recognition of residue heat, differentiating material based on its unique reflective behavior, and more [3]. In this work, we present the ThermalSense system for extending the human visual range to the infrared thermal spectrum, via sensory substitution. We also offer a user evaluation, a generalized procedure for specifying similar systems and engineering considerations.

2 Related work

2.1 Sensory substitution

Sensory substitution is defined by the APA Dictionary of Psychology as ”[t]he perception of a stimulus normally analyzed by one sense through the activity of another sense”. In previous research, sensory substitution systems and devices such as the vOICe and the EyeMusic have been used for conveying visual information through auditory signals [35, 59]. Visual-to-auditory sensory substitution devices (SSDs) such as these acquire visual images through a camera, which are then processed using an algorithm that converts the images into soundscapes, thus enabling humans to ”see” through sound. By converting visual information into soundscapes, the technology allowed individuals to develop extraordinary sensory abilities through perceptual learning, leading to a high level of accuracy in object recognition, localization, shape and color recognition [1, 7, 22]. These abilities were achieved not only by sighted users but also by congenitally blind individuals [36, 57].

Neuroimaging studies have demonstrated the brain's ability to adapt and respond to such new or substituted senses within tens of hours of training [2, 7, 22, 57, 58]. After extensive training with visual-to-auditory SSDs, visual areas in the brain begin to respond to soundscapes (even in the congenitally blind), corresponding to the known specializations within the visual cortex. This strengthens the understanding that users of the system are indeed ”seeing through sound”. Phenomenologically, users have reported the development of a degree of automaticity for this form of perception following extended use of such systems [39].

The concept of seeing through sound by creating soundscapes can be considered to be based on data sonification, though their purpose and context of use are different in essence. Sonification is a representation of non-auditory data through auditory cues [23, 66]. While sonification has been utilized for some thermal and temperature-related purposes, mainly in the domain of weather and climate data [12, 17, 18, 20, 33], there is a fundamental difference between data sonification and sensory substitution. Namely, data sonification is used mainly to interpret and understand data transmitted through sound (the nature of the information perceived is auditory). However, sensory substitution is geared towards perceiving sensory information, such as vision, by means of audition. Ultimately, the information processed by the brain in data sonification is generally auditory, while the information processed by the brain in visual-to-auditory sensory substitution is visual (sonification poses as a mediator). There is a qualitative difference in the nature of the experience. The soundscapes used in sensory substitution are presented in a way that corresponds to the manner in which the brain is accustomed to perceiving information from the substituted sense, in this case, visual information. In this way, the information conveyed through sensory substitution is processed in the brain as visual information (not auditory), as seen in numerous behavioral and neuroimaging studies employing visual-to-auditory sensory substitution.

2.2 Sensory extension

Sensory substitution has most commonly been employed for rehabilitation of people with a sensory disability. Yet in recent years, sensory augmentation has been explored in attempts to create novel or enhanced senses [34]. In the past, people have been enhancing their sensory capabilities through the use of technology in various ways. Specifically in regards to vision, microscopes were developed to see into the micro-scale, telescopes to see into the distance, and as technology evolved, researchers also began applying tools such as displays, thermal cameras and more [4, 5, 31]. While these approaches commonly consider our sensory systems and brain function as a given, the sensory substitution based approach relies on neuroscientific studies such as the ones discussed above, and treats our senses as ”programmable” systems that can be developed and enhanced in themselves[7, 22].

In this work, we applied sensory substitution techniques for extending visual capabilities, providing the users with an extended sense for perceiving information in the infrared range. While the concept of extended senses has been discussed in literature [15], very few such systems have been realized [15, 25, 55, 65]. One relevant example used sensory substitution for extending the visual spatial range, representing visual information from behind the user through auditory soundscapes. The research showed that the visual and auditory information can be seamlessly integrated into a combined understanding of the expanded visual field [55]. Eagleman and Perrotta [15] report that they have used a haptic feedback bracelet to achieve an expansion of the auditory range to the ultrasonic range, and of the visual range to ultraviolet, and Hsu et al. [25] used virtual reality game settings for simulating echolocation achieved by perception of ultrasound frequencies, as well as an extended visual range.

The definition of sensory extension in the SSD domain as extending the user's sensory perception, overlaps in many ways with the sensory extension definition used as part of the conceptual space for Sensorimotor Realities charted out by Vatavu[63]. The six-axis space allows for situating this work in relation to others, in the broader context of computer-mediated reality environments. Particularly on the scale of sensory mediation (correlating with Mann's mediated reality XYR continuum[40]) this work is situated at the scale's far end as sensory extension, which is defined as "enhancing abilities beyond the possibilities offered by the human anatomy and biology". We further discuss the system as part of this framework in the discussion.

We chose to use the EyeMusic algorithm for the purpose of carrying out sensory expansion, since it has been proven to be an effective algorithm for inducing visual perception (in behavioral studies) and the corresponding activation of visual areas in the brain (in neuroimaging studies). The algorithm's ability to represent differentiation based on colors, also allowed us freedom in exploring different implementations as described in the following section.

With the repurposing of the EyeMusic algorithm, we developed a training program for enabling the perception of thermal information in a scene via soundscapes, and a test for evaluating this ability. We were specifically interested in exploring the research question of whether and to what extent the system enables users to perceive otherwise invisible thermal information. An online evaluation with 24 participants demonstrated the effectiveness of this method, paving the way for further research on the experience of extending one's sensory perception, and development of related applications, as well as providing a better understanding of the underlying perceptual mechanisms.

It is important to differentiate the contribution made in this work from the EyeMusic system. The scientific novelty of the current work is in leveraging the known abilities of the EyeMusic as an SSD to present a case of extended senses - such that the sensory system could potentially perceive information that is out of the scope of its perception as sensory information. In this case, visual information (infrared light) that is not normally perceived by the human visual system could potentially be perceived as visual cues by augmenting it via audition. The corresponding user study we carried out demonstrating the possibility of extending sensory perception using this method is meaningful both for the HCI and neuroscience communities. Specifically conveying thermal information via sensory substitution has not to our knowledge been empirically tested before. As such, the motivations of the study are twofold - contributing to the development of thermal representation systems; and as an initial validation of a sensory substitution based system for enabling extended sensing.

3 Design and implementation

The rationale behind the method we developed was to apply the EyeMusic algorithm's conversion over a thermal image in order to convey thermal information in a way proven to induce visual perception via sound. The intention was to use the audio layer as an extension of natural visual perception. Hence, the soundscape is played while viewing a standard visual scene (as a supplement, without substituting the natural image with a thermal image, or a standalone audio representation).

The EyeMusic algorithm downsamples and translates captured images into 1,500 pixels (50x30), which are then correspondingly transformed pixel by pixel into intricate auditory scenes termed soundscapes [10]. The sweep line algorithm sweeps from left to right to represent visual object location, shape and color, with the x-axis represented by time, and the y-axis represented by pitch manipulations. Specifically higher pitches represent spatially higher elements in the visual scene, while lower pitches represent the lower elements. Colors are conveyed by the algorithm as different instruments, each has a different timbre - commonly referred to metaphorically as the "colors" of sound [1, 32]. The algorithm employs tones in the pentatonic scale, shown to be pleasant and harmonious even during the presentation of numerous simultaneous sounds [1].

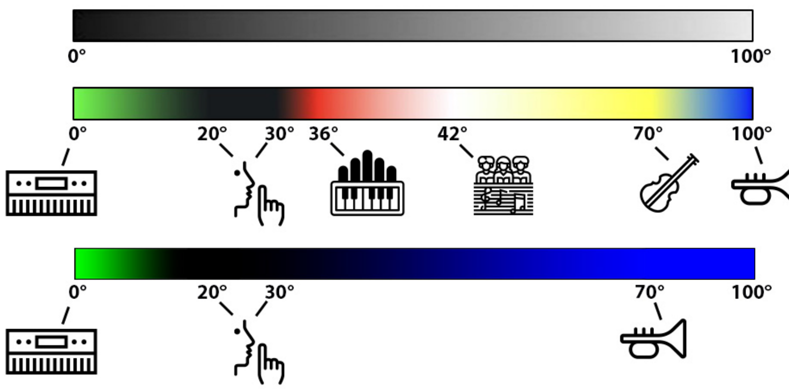

Fig. 2. Different mappings employed in the study. Mapping to grayscale (top); Mapping according to the high-resolution approach (middle); And the mapping used in the system - mapping according to an adapted vision-inspired approach (bottom).

The algorithm implemented in the current study applied this technique to map pixels of "cold" temperatures in the range of 0-20℃ (32-68°F) to the sounds of Rapman's reed, and "hot" temperatures in the range of 30-100℃ (86-158°F) to the sounds of brass instruments. We chose to represent the 21-29℃ range with the absence of sound in order to display "neutral" temperatures without adding unnecessary noise (see Figure 2).

3.1 Design considerations and process

There were several options and considerations taken into account when designing the final ThermalSense algorithm that was used, which were also evaluated through several short iterations. Most notable is the mapping chosen for representing thermal information.

To direct the mapping of instruments to sounds, audio samples of all the instruments used by the EyeMusic algorithm representing the different colors were produced. A professional musician (also educated in music technology) was then asked to rate the different instrument samples by their associated temperatures and to explain the rationale behind his proposed scale. The audio feature he found most associable was the sounds' "brightness". Brightness of sound is a subjective definition, described by him as "a sound being 'brighter' when the higher overtones composing its timbre are more prominent in their level in comparing the fundamental frequency". The association to brightness also resonated with the nature of thermal information as a light frequency which strengthens this choice as one that potentially builds on crossmodal correspondences. The different colors were subsequently sorted from "cold" to "hot" in the following order (from least to most bright): green, red, white, yellow, blue. It is important to note that the colors serve only as a method for transitioning from image to audio. As such, crossmodal correspondences related to the colors themselves are irrelevant and were not taken into consideration.

For defining the represented information, the temperature range was divided into categories based on research pertaining to temperature perception to better understand the information that should be represented. The categories ultimately defined were very cold =< 0℃ (32°F), cold <20℃ (68°F), neutral <30℃ (86°F), warm <42℃ (107.6°F), hot <70℃ (158°F), very hot <100℃ (212°F), and burning hot >100℃. Consequently, according to 0° being water freezing point, the range between 0-20° is considered as noxious cold [30, 52], the range around 30° degrees is considered to be neutral

[45], 43° and above is considered noxious hot [41, 50], above 70° is considered to be an immediate danger [43], while above 100° is past the boiling point of water.

Two different approaches to mapping were considered, each with its own advantages and disadvantages. One can be considered to be ”high-resolution” and the other ”vision-inspired”. For both approaches, temperatures in the ”neutral” range were filtered out, as they were determined to introduce unwanted noise, without adding meaningful thermal information.

The high-resolution approach involves representing each of the temperature ranges using a different instrument (see Figure 2). The advantages of this approach are that it allows for a more distinctive representation of each temperature's sound, with a clear separation between meaningful temperature ranges using the instruments. The disadvantage lies in the fact that it requires more extensive training to recognize and associate thermal properties with five different instruments, one per range. Another approach considered is the vision-inspired approach. The vision-inspired approach considers the response to infrared frequency as similar to how the red, green, and blue light receptors in the eye function. That is in the sense that a receptor is sensitive to the specific frequency, and the stimulus increases in relation to the radiation intensity. In this case, the sound of only one instrument would be required for representing heat. The advantage of this approach is that it simplifies the training process, but compromises on the resolution in the differentiation between sounds representing temperatures, in addition to losing the association of sounds to meaningful temperature categories.

To determine which of the approaches to use in the present study, implementations of both were created, and initial trials were conducted for each, consisting of a short training, test, and feedback. This experimentation was done with the goal of creating a system that could be effectively used for testing the basic underlying principles of conveying thermal information through auditory properties. Our first trials used the high-resolution approach, yet it was quickly apparent that users showed very little improvement in a single short training session, and the written feedback collected indicated users found it difficult to differentiate between the sounds, particularly at higher pitches. Following this, we developed several iterations based on the vision-inspired approach. The main challenge was in defining the range so that temperatures in the lower end of the spectrum would still be noticeable. We finalized our iterative process with an adapted vision-inspired approach - in order to also maintain a representation of the cold temperature range, we opted for using one instrument to represent heat and another to represent cold. We also narrowed the spectrum range to 70℃, sacrificing the ability to differentiate temperatures in the ”very hot” range, to allow a better distinction between temperatures in the represented range, yet still enable recognition of the ”very hot” range of temperatures at which heat can cause immediate physical damage [43]

3.2 Technical implementation

We propose a generalized six stage procedure for characterizing thermal auditory representation systems (see Figure 3). The stages are as follows: Image acquisition - generating a thermal image of a scene using a thermal camera or sensors; Removal of non thermal data - removing any added information from the acquisition source, such as scales, technical data, watermarks, etc.; Conversion to unified format - color mapping of the thermal image into grayscale gradient from min to max temperatures; Remapping - Remapping the grayscale gradient to a new gradient based on a chosen implementation. Different mappings can be applied, using more or less colors and gradient distributions; Sonification - Associating each color with a sound. Based on each pixel's color and location, a soundscape of the scene is generated; Representation - Playing back the generated soundscape to the user. We further discuss possibilities and considerations of designing such systems in the system engineering subsection under the discussion.

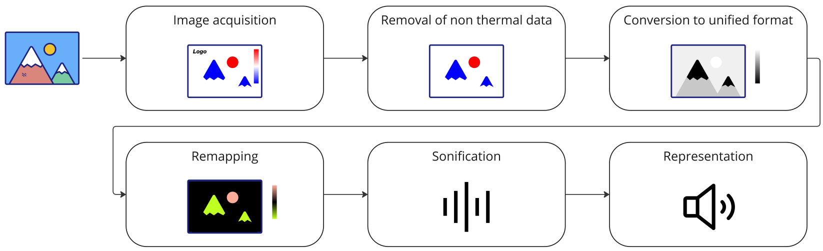

Fig. 3. High-level system architecture diagram for characterizing thermal auditory representation systems.

Following from this, in the implemented ThermalSense system presented (see Figure 1), Image acquisition was performed using a FLIR C5 thermal camera [16]. The FLIR camera and software offer a range of representations to enable different modes and methods of post processing, and settings that aid in the differentiation and recognition of temperatures to suit specific needs. For our purposes, we generated alongside each standard image the thermal image in a grayscale mapping between 0-100℃, covering the entire range that we initially considered representing (see Figure 2). The chosen grayscale representation is of a Temperature Linear scale - meaning the temperature is mapped linearly, rather than presenting the raw signal received by the camera, which is not proportional to temperature and needs to be mapped according to the environmental conditions.

Removal of non-thermal data was performed for both thermal and standard images, stripping off all overlaying data such as logos and scales. This was done using Adobe Photoshop in a semi-automated process - a recorded action was created, applying content aware filling to the locations where logos, scales or numerical data appeared. The action was then applied to all images, which were later reviewed for validation. Being that the FLIR was able to produce grayscale images, the unified formatting stage was already integrated in the image acquisition stage. Remapping of the grayscale thermal images was done using a dedicated script coded in Python to a new color palette as described above, for enabling sonification using the EyeMusic algorithm (see Figure 2). The two colors used for representing 0℃ and 70℃ were green and blue, to match the "coldest" and "hottest" sounds as determined in the design considerations subsection. Following this, sonification was performed using the EyeMusic algorithm, which generated the representing soundscape as an audio file. The algorithm translated the green pixels to the sounds of a rapman's reed, and blue to the sounds of brass instruments [1]. This process was used to produce a bank of standard images and corresponding audio files that represent the thermal information in the visual scene (see Figure 1), later represented in the experiment using the users' own computers and headphones.

Our validation procedure included several stages. While different measures were taken through the development of each of the stages, a few methods were used for validating the system as a whole. First, a number of simple graphic figures were generated and run through the conversion stage and the stages following. These simple images included single lines or basic geometric shapes in a single or very few colors, which were run through the various stages, validating the performance of the color conversions as well as the sonification algorithm, verifying that each delivers the expected result. We then tested the flow using images with clear temperature associations (e.g tray of ice, or boiling water), while sampling temperatures in the frame using FLIR's toolkit and verifying the expected output throughout the different stages as well. Through this validation procedure for example we recognized the need for applying a Temperature Linear scale mapping. We repeated this procedure using images representing real-world elaborate scenes and edge cases - including recognition of human figures in an environment, recognition of smaller temperature differences, very warm or cold environment temperatures, wide fluctuations in temperature ranges within a single image and more. This allowed us to test the system's functionality both technically and performance wise for different mappings that were tested. Finally, each of the images chosen to be used in the experiment was individually checked by examining the standard images next to the products of each stage from thermal image to sound side by side.

Fig. 4. Two examples of images used in the test. On the left are the standard images, and on the right are thermal images mapped according to the ThermalSense algorithm (blue is hot and green is cold).

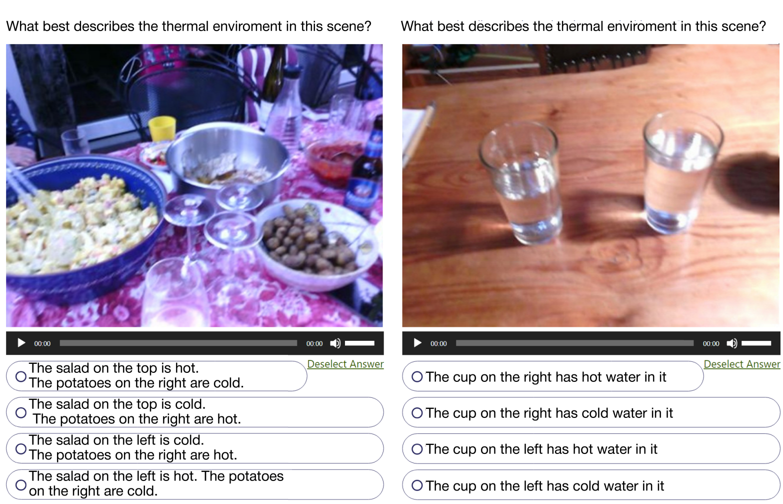

Fig. 5. Screen captures from the final test stage of the experiment.

3.3 User evaluation

3.3.1 Procedure. The study included 24 (18 male, 6 female) participants (mean age 29.1 ± 7.2). The participants were recruited from the Prolific online platform [49] and received monetary compensation for their participation in the study. The research protocol was approved by the university psychology department's IRB committee and was given ethical clearance number P_2023153. All participants signed an informed consent form before starting the experiment. All participants reported normal vision, hearing, and neurological function, as well as fluency in English.

The interface used in the present study took inspiration from the online version of the EyeMusic self training program presented in Buchs et al. [10] who developed an online program for learning to use the EyeMusic sensory substitution system. Participants were required to use an audio enabled desktop computer, and standard headphones.

The experiment included five sections: baseline test, familiarization, training, a final test and a personal reflection. The baseline and final tests consisted of 10 standard images (the same in both stages) whereas in the baseline the participants were presented with the image alone, and in the final test each image was accompanied by the corresponding soundscape. The images were chosen to be ones devoid of clear visual thermal clues i.e. ”temperature transparent” (see Figure 4). The questions were 4AFC, asking participants ”What best describes the thermal environment in this scene?”, with answer options composed of: one correct option (1 point); two partially correct options, representing either correct location or correct temperature determination (0.5 points); and one wrong option, in both the location and temperature domains (0 points) (see Figure 5).

Upon completion of the baseline test, participants continued to the familiarization stage in which they were given an overview of the experiment flow, basic information about the algorithm and its features, alongside several images with matching soundscapes that demonstrate the algorithm’s use: a ”very hot” and a ”very cold” diagonal line (displaying the full audio spectrum), a ”very hot” and ”very cold” rectangular shapes, and five images and soundscapes displaying items easily associable with their respective temperatures - two for a single ”very cold” or ”very hot” item (an ice tray and a boiling kettle), two for multiple ”very cold” or ”hot” items (beer bottles and tea cups), and one with both ”cold” and ”hot” items (see Figure 1).

Next was the training stage. In this stage participants were presented with images and their matching soundscapes. They were instructed to choose between two answers, the one that ”best describes the thermal scene” of the standard image accompanied by the corresponding soundscape. The first five questions related to the location of thermal information in the scene, while the next five related to the temperatures represented. The questions were 2AFC, where one answer was correct and the other was incorrect (see Figure 6) This section included a total of 10 questions.

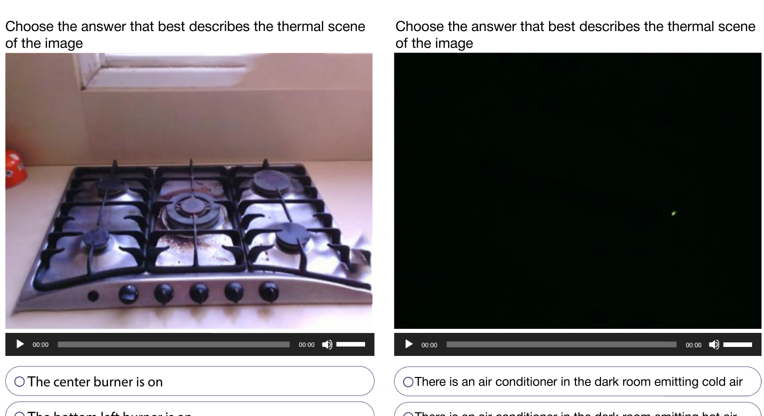

Fig. 6. Screen captures from the training stage of the experiment. On the left - a question related to the location of thermal information. On the right - a question related to the temperatures represented. Images were slightly modified for readability.

Participants who scored 70% or under on the test were required to repeat the training stage until they passed.

Upon passing the training, participants could continue to the final test, which was identical to the baseline test, with the addition of soundscapes for each image. Finally, the participants answered a number of subjective response questions during the personal reflection section composed of two open questions ("what did you experience?" and "what did you perceive") and the User Experience Questionnaire (UEQ) [54]. The full procedure and the files used are available at https://osf.io/msxfh/.

3.3.2 Results. The results of the study revealed that in the baseline test, participants were able to detect the thermal properties within a visual scene without the system at an average accuracy of 58%, while in the final test they could detect the thermal properties within a visual scene with the system at an average accuracy of 87% (see Figure 7). To assess the normality of the study results, a Shapiro-Wilk test was conducted on the differences between the baseline and test groups. The results did not indicate a significant deviation from normality. Next, a paired t-test was performed to assess the mean differences between the groups. Results of the paired-t test indicated that there is a significant large difference between baseline ($M = 5.8$,$SD = 1$) and test ($M = 8.7$,$SD = 1$), $t = 10.1$, $p < .001$. Furthermore, a post-hoc power analysis was conducted, revealing an effect size (Cohen's d) of 2.89, and a power of 1. This indicates that the study had the maximal level likelihood of detecting an effect, given the effect size observed and sample size.

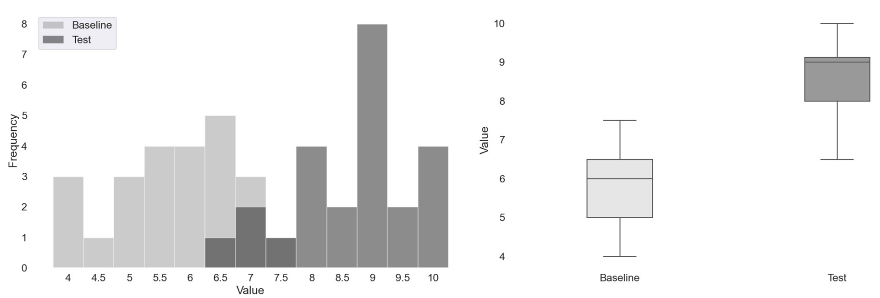

Fig. 7. The distribution of scores frequency in the baseline and test stages (left). A box plot representing the baseline and test scores (right).

The great majority (17) of the participants passed the training stage on their first attempt, and only a few (7) passed on the second, with an average of 1.29 attempts. No participant required a third round of training. The median time for completing all stages was 25:41 minutes. It is worth mentioning that this is representative of the whole time participants were taking part in the experiment, as logged by the Prolific system, and may include self initiated breaks and other events beyond the procedure itself.

Participants were equally capable of detecting thermal information, and localizing it within the scene. To assess this, we summed and compared between the number of temperature mistakes (hot vs. cold) and the number of location mistakes (eg, right vs. left). Participants on the final test made 23 temperature only mistakes, and 23 location only mistakes out of the whole 240 answered questions (10 questions x 24 participants). Only 8 mistakes in both domains simultaneously were recorded (temperature and location together). The accuracy of the participants who needed to repeat the training more than once was seen to be lower with respect to location than their accuracy with respect to temperature (49% accuracy vs 74% respectively). This is in contrast to the participants who passed the first time and showed no significant difference with respect to accuracy-related to location vs. temperature (89% vs 93%).

While overall in the training there were no specific questions that stood out as particularly easy or difficult, three questions stood out in particular on the final test with a higher percentage of mistakes (see Figure 8). On image *a*, 27% of the participants made mistakes (9 in location, 4 in temperature, 2 of which were mistaken in both). On image *b*, 25% of the participants made mistakes (3 in location, 9 in temperature, and 2 in both). On image *c*, 23% of the participants made mistakes (11 partial mistakes, 6 in location 5 in temperature, and 1 in both). On the rest of the questions, all mistake ratings were below 15%.

Participants ranked the system as above average on pragmatic quality and excellent on hedonic quality as based on the UEQ. The results of the UEQ are scored on a Likert scale of 1 (lowest) to 7 (highest). Rankings are then scaled to negative (-3) and positive (+3). Scaled values above 0.8 indicate a positive evaluation, 0.8 a neutral evaluation, and below 0.8 a negative evaluation [53]. Results are then compared to a benchmark data set that contains data from 21,175 people from 468 studies [54]. All factors except complicated-easy received a positive evaluation. All of the scaled ratings are presented in Table 1. Scaled means were compared to existing values from the benchmark data set to arrive at an Above average ranking for Pragmatic quality, an Excellent ranking for Hedonic quality, and an Excellent ranking overall (see Table 2). The positive assessment is corroborated by the open answers of some participants regarding their experience, for example, participants stated that "This was super interesting!", "This experiment was fun to do. It challenged my memory in a fun manner", and "This was really interesting and I've never done anything like it. Once I got my head around the positioning, the different instruments and the pitch I found working out the temperatures on the photos easier than I thought it would be".

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale
1	1.7	0.8	0.9	24	obstructive	supportive	Pragmatic Quality
2	0.5	2.3	1.5	24	complicated	easy	Pragmatic Quality
3	1.7	1.3	1.1	24	inefficient	efficient	Pragmatic Quality
4	0.9	3.2	1.8	24	confusing	clear	Pragmatic Quality
5	2.0	1.1	1.0	24	boring	exciting	Hedonic Quality
6	2.5	0.7	0.8	24	not interesting	interesting	Hedonic Quality
7	2.3	1.1	1.0	24	conventional	inventive	Hedonic Quality
8	1.4	1.0	1.0	24	usual	leading edge	Hedonic Quality

Table 1. The scaled results of the participants' evaluations on the UEQ

Scale	Short UEQ Scale	Comparison to benchmark	Interpretation
Pragmatic Quality	1.188	Above average	25% of results better, 50% of results worse
Hedonic Quality	2.052	Excellent	In the range of the 10% best results
Overall	1.620	Excellent	In the range of the 10% best results

Table 2. Comparison of the participants' UEQ results to the benchmark dataset

3.3.3 Evaluation discussion. The fact that most participants required only one brief training session indicates that the system is relatively intuitive. This sets the current system apart from many traditional sensory substitution devices and systems that require extensive training and perceptual learning in order to be efficiently used.

The tendency of participants who failed the training on the first trial to make more mistakes in recognizing location in comparison to thermal properties may imply that localization was more challenging to learn using the algorithm. Yet, it is worthwhile mentioning that the order of the presented questions was consistent for all participants, such that the location recognition questions preceded the temperature recognition questions. This could account for some of the difficulty experienced by those who repeated the training. The fact that in the final test, there were the same number of mistakes made in locations as in temperatures, also supports this hypothesis.

With respect to the questions that stood out as difficult in the final test (see Figure 8), it is interesting to note that the mistakes correspond to elements that are particularly difficult about each image. Image *a* was difficult with respect to location, indicating that participants were able to identify the temperature of the water poured through the faucet into the sink, but had more difficulty determining whether both the sink and faucet were hot or only one of the two. Image *b* was particularly difficult with respect to temperature, indicating that participants found it a challenge to discern between hot and cold presented in relatively close proximity. Image *c* was nearly equally difficult with respect to location and temperature, this image, like image *b*, had representations of both hot and cold thermal properties in the same image, yet here they were not localized in a distinct location of the scene, and hot and cold elements were represented simultaneously. This

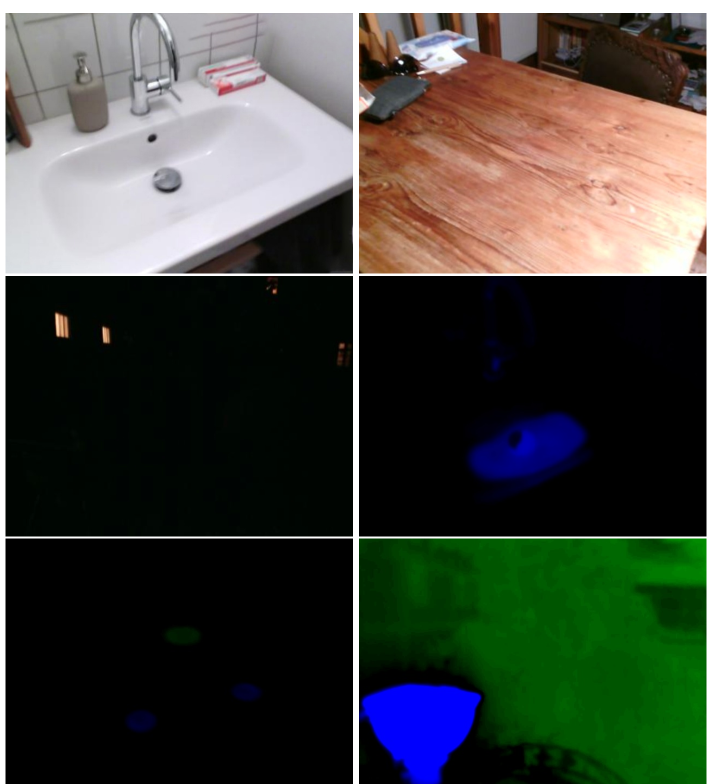

Fig. 8. From left to right: a. The image depicts a sink and faucet, both of which are hot after hot water was pouring through and on them. b. This image depicts a table, upon which had been three cups, two hot and one cold. The question referred only to the two hot ones. c. This image depicts an entirely pitch black scene, of a hot barbecue, on a cold night.

was also one of the only two images presented that had an environmental temperature not within the ”neutral” range, but rather in the ”cold” range. We believe that with more extensive training, these edge-case difficulties and confusions experienced by some people can be overcome, based on the improvement shown by those who failed the first round in the training stage, subsequently passing on the second attempt. This initial analysis provides a glimpse into the cognitive processes and mechanisms utilized by users of the system. A more extensive study of the different challenges users face would contribute to gaining a deeper understanding, characterization, and definition of the challenges posed by learning and using the system, thereby guiding future development.

An important aspect to note considering the effectiveness of the training program is the user experience. As described above, achieving effective sensory substitution involves perceptual learning, and SSDs usually require extensive training in order to become usable [35, 39]. As such, compliance is a key element in designing effective training programs in which users would be willing or interested to participate. This is a drawback shared by many traditional sensory substitution devices and systems, making them less practical for rehabilitation and even less feasible for use in sensory augmentation. For this reason, we find the result of the User Experience Questionnaire and user feedback to be especially encouraging: The UEQ results showed that the system was perceived as both pragmatic and pleasing. The learning process was described by a few participants as follows: ”at first it was pretty hard for me to understand all of these sounds, but I slowly started to understand them, so I think I've done pretty well.” and ”I've perceived that these sounds at first sound very similar to each other, but then I realized how they're different from one another”. We believe this shows the system provides a balance, requiring relatively minimal effort for reaching practical results in using it, and making the actual use and learning process enjoyable. This balance potentially enables an effective device that also supports the extensive use and training required for perceptual learning.

4 Discussion

In this work, we presented an initial proof of concept for providing an extended sensory experience by employing sensory substitution and augmentation techniques, namely conveying thermal information through sound. Our methodology consisted of mapping thermal images into soundscapes using the EyeMusic visual-to-auditory sensory substitution algorithm. This method allowed for representing information ”invisible” to the human visual perception range through an alternate sensory route. The findings show that participants' ability to identify thermal properties conveyed through the SSD reached 87% (from a baseline of 58%) after very brief training. In addition, the algorithm was effective at conveying the thermal properties and the locations of thermal information through auditory properties, thereby further indicating the possibility of expanding visual perception. Finally, the participants' subjective ratings of the system for user experience indicated that they found it both pragmatic and hedonically pleasing, receiving a score of Excellent in both parameters [54].

In an effort to frame this work in a broader context, and allow for the characterization and comparison with other systems, we positioned our system on the Sensorimotor realities conceptual scale. The scale was suggested as part of the computer-mediated reality environments framework [63]. This conceptual space for computer-mediated perception and motor action, is defined along six dimensions. The first dimension, sensory mediation, specifies the nature and amount of mediation of user perception. It ranges from sensory contraction (refusing the user's sensory abilities) through diminution and amplification to sensory extension (enhancing abilities beyond the possibilities offered by the human anatomy and biology). The second dimension, motor mediation, represents the mediation of the user's motor actions, also ranging from contraction to expansion of motor abilities. The third dimension, virtuality, denotes the amount of virtual content presented to the user, based on Milgram and Kishino's Reality-Virtuality continuum [42], ranging from the physical world to the virtual world. The fourth dimension, imaginarity, expresses the degree to which mental imagery is needed for effective operation. Higher imaginarity requires users to rely more on imagination and memory. The fifth and sixth dimensions, body augmentation and environment augmentation, specify the use of wearables and the integration of technology into the physical environment, respectively, to mediate sensorimotor abilities.

As described in the sensory extension subsection, our definition of extended senses resonates with the one used in the scale's Sensory Mediation axis, and it is unmediated in relation to Motor mediation. The system can be described as mixed in it's Virtuality. Imaginarity is particularly interesting for our use case - We suggest that in the case of a sensory substitution system, while it may originally require mental imagery, following training and the process of perceptual learning, the system could require little to no imagery. Body augmentation is mixed, as headphones are required for providing auditory feedback, and the system is non augmented in relation to Environmental augmentation.

4.1 System Engineering

Based on the proposed six stage procedure for characterizing thermal auditory representation systems (as described in the technical implementation subsection and Figure 3), we propose an implementation roadmap for future systems, including engineering considerations for each of the elements. Moving forward, the system should naturally be developed as a device. We offer our insights also for such real-time implementation based on our experience in the development so far.

Image acquisition - While this stage can rely on other sources such as existing databases or some configurations of sensors, a thermal camera is most likely to be the tool for acquisition. The variety of thermal cameras available is vast, ranging from small and cheap modules that can connect to a microcontroller, through solutions that can connect to a mobile phone, to high end professional equipment that offers advanced features, and the considerations as for choosing a solution may vary depending on the application. With respect to resolution, it is important to note that the acuity of vision through sound is generally lower than through the visual system, and as such, there is little gain to be had from the use of a high-resolution camera. To give an indication of suitable resolutions, the algorithm we employed downsamples the image to a 50x30 pixel image based on previous work, while similar systems such as The vOICe SSD reach a resolution as high as 176×64 [9]. Similarly, regarding frame rate, for a continuous feedback system frame rate could be rather low as most sonification algorithms represent a single frame at a time. Our system performs a 2s sweep, while The vOICe defaults to 1s, such that both of these algorithms require a much lower frame rate than 9fps, which poses a common limitation in thermal imaging due to US security regulations. Some thermal cameras also include more advanced features such as different color schemes and temperature linear scaling which may simplify some of the following stages as indicated below. If possible, it is useful to capture also the standard image taken for validation purposes.

Removal of non-thermal data - Thermal images may include information beyond the thermal content. Such information could be a visual legend of the color gradients applied, temperature readings, technical information on the image shot, watermarks and more, and could be a consideration in choosing an acquisition method. This information introduces noise that is not of a thermal origin into the auditory cues at later stages, and as such should be removed. Since often such information is consistent in location and characteristics, its removal is feasible in several ways. For non-real time applications such as the one used in the present study, graphic tools can be used, and new artificial intelligence based tools enable very high quality results with minimal effort. Real-time and wearable systems could apply simpler techniques such as averaging the surrounding pixels of an area, or mapping a predefined block to a temperature representation which is not sonified later on in the process. These methods will result in loss of data, yet it is likely to be minimal, and based on our experience would be preferred to the representation of added noise.

Conversion to unified format - The conversion to a unified format such as grayscale is done in order to produce an image that includes all the information one may wish to represent in a consistent manner, irrespective of the sonification algorithm to be applied. While this stage may not always be necessary for producing a functional system, it is good practice, especially when designing a new system and considering different mappings and implementations, as it simplifies testing and comparing different mappings. We suggest using a grayscale gradient that covers the entire temperature range that could be relevant for the application, keeping in mind that taking subsets of that range would be possible at a later stage. It is important to verify that the gradient actually represents the desired data - the infrared readings from the sensor are often not directly mapped to temperature readings and should be translated to a linear scale. As mentioned above, the camera we used already performed the linear scaling and grayscaling, though this might not be the case when using simpler cameras or basic modules, and as such, conversion may be required.

Remapping - In this stage, the grayscale gradient is mapped to the new color gradient, to be sonified using the chosen algorithm. As discussed in depth in the design considerations subsection, the manipulation of this gradient may allow for a lot of flexibility and adjustment of the system to fit a desired application. By introducing more or less colors and changing the map ranges, one can choose to represent different temperature ranges, enhance or decrease the resolution, and control how much "noise" the system allows for (e.g. filtering out room temperature). It is also important to consider at this point the balance between the intuitiveness and learnability of the system vs its range and resolution.

The Design considerations subsection offers several insights to guide the design of an effective mapping. In addition, we propose future investigations considering principles of sensory substitution and perceptual learning could assist in informing this decision as well as similar ones for other SSDs: Our current decision to utilize a representation of instruments that distinguishes only between hot and cold thermal properties has advantages and disadvantages. It would be worthwhile exploring both the more physical approach of representing only heat (all infrared signals) on the one hand, and revisiting the high-resolution approach on the other, and comparing the different methods after more extensive training. Such an evaluation should also include an assessment of the ability to differentiate between temperatures in the same range (a factor that was not assessed in the present study). This evaluation could perhaps rely on common practices from psychophysics for evaluating just noticeable differences (JND).

Sonification - While the reasons for choosing the EyeMusic were described in detail in previous sections, other algorithms could be used for this purpose. In the case of a single color mapping, some algorithms that are agnostic to color could be used (e.g. The vOICe). For using a diverse color scheme, other algorithms that can be used include: ColEnViSon[8], Colorphone [48], and SoundSight (who also presented a similar implementation)[21]. Each of these color-incorporating algorithms uses different methods and approaches that can be explored. It is important to consider the representation as well at this stage. For example aside from the ColEnViSon and some variations of SoundSight, all of these require stereo sound. Some of the algorithms may also require more complex computation than the EyeMusic implementation, though generally speaking, they all rely on relatively simple procedures.

It is also worthwhile considering the design of variations of the algorithm specific for representing thermal information. Such implementations could take into consideration crossmodal correspondences known to exist between temperatures and auditory properties such as pitch [56], beyond the ones so far considered. One such relevant crossmodal correspondence that could be leveraged involves the sounds of pouring water, as research indicates that people throughout their lifespan develop the ability to recognize the thermal properties of pouring water (hot or cold) through the sound made by it being poured [6, 67]. Other abilities our brains inherently excel in could also be leveraged, such as language, similar to the TopoSpeech visual-to-auditory system that conveys spatial information through a combination of language and auditory properties, thus shortening the learning curve by tapping into the intuitiveness of language [37, 38, 46].

Representation - The audio playback of the soundscape to the user can be achieved using headphones, speakers or transducers. One consideration to keep in mind is the practical element of supplementary audio that can disturb hearing in real-life scenarios. For this, means of playback that do not block the ear, or use only one ear might be more fitting. Examples include directional speakers such as the ones used in VR, or bone conduction headphones (though these may affect the perception of some frequencies). With that said, the fact that the system enables an expansion of the visual range rather than a substitution of vision, can ameliorate this drawback significantly. As an addition to visual information and not a substitute, thermal auditory systems can potentially be activated only when needed rather than as a constant stream of information, thereby mitigating its disturbance to other senses. Such an implementation could be operated by the user, or defined by specific rules such as predefined temperature thresholds, or other criteria. This is in addition to the fact that it is likely that the system would become more transparent following training [39].

4.2 Use cases

Beyond the novelty of this proof of concept as a sensory extension system, future development of similar thermal information augmentation systems could prove highly valuable in numerous use cases. Some literature grounded use cases where thermal sensory information could be warranted and would be specifically fitting for extended sensing include: Low lighting, where thermal information can replace vision (e.g indicate the location of people or animals); Scenarios where thermal information is hidden yet meaningful, such as in cases of gas leakage detection [27], or occupational scenarios such as firefighting where recognizing dangerously hot objects is important while obstructing one's vision with a thermal camera is not practical; Prostheses users and robotic surgery scenarios, in which a device is used as an extension of the user's own body that suffers from the lack of thermal feedback; Generally in medicine, where temperatures may provide meaningful information yet physical touch could be undesired; Virtual reality, in which thermal information is infamously difficult to convey [56], and even agriculture where thermal imaging is used for example in recognizing water stress in plants, and an augmented sense could enhance the perception and intuition of agricultural workers in their everyday work [62]. Below we expand on several of these use cases, suggesting guidelines for how the system engineering could be applied for such applications.

In the case of virtual environments, thermal feedback is thought to potentially contribute to immersive user experience [11, 29], yet the implementation of thermal feedback systems can be challenging. As such the use of thermal related cross modal correspondences was previously suggested as a realistic solution [19, 56]. A simplified implementation could be carried out by assigning a temperature variable to different objects in the environment. A unified format view could then be generated by translating temperature parameters to a grayscale color, and then applying a sonification algorithm to the acquired image. Such a functionality could be developed as a Unity package for example, enabling the addition of thermal information perception to existing projects, allowing developers to make a room cozier with a fire-place, surround a castle with a river of freezing cold water, or make a game weapon overheat. Considering the visual properties of thermal information, a more advanced implementation could be applied, leveraging the platform's existing light and color tools for creating a complex and realistic thermal representation.

Further exploring the use case of prosthetics, the lack of physical thermal sensation is detrimental for natural interaction with objects in the world, and has an impact on the embodiment of the prosthetic limb [44]. There have been several attempts to deliver this information via direct thermal [26, 44] or tactile feedback [28] on the user's body. The basic implementation for this use case could be a simple one, where for example a thermal camera module (e.g MLX90640, or AMG8833) and a microcontroller using BLE audio (e.g ESP32) would be mounted on glasses frames with embedded speakers (e.g Bose Frames). The camera and microcontroller could also be mounted in a less intrusive location such as on the neck or chest, and a more advanced implementation could explore in some cases integration into the prosthetic itself. These solutions are unique as they can be retrofitted, don't require contact with the body, are rather cheap to implement, and add the benefit of the extended sense beyond the tactile sensation.

The latter implementation, using glasses frames and a microcontroller, or a similar one, could naturally adapt to many of the use cases suggested above. To exemplify some of the variations and the options suggested in the previous section, a technician could have such a setup that reacts only to extreme temperatures and warns them of dangerous elements in their environment. Or a chef could have their system tuned for differentiating between small temperature differences, allowing (after training) for accurate control over the temperatures of elements in a dish without damaging it with a thermometer or physically touching the food.

4.3 Basic science application

Tying into the more basic science end of this work, there are also several promising avenues to explore with respect to the neural correlates and underlying mechanisms in the brain of such an experience. The first relates generally to the possibility of developing extended senses and the adaptations in the brain following perceptual learning [24]. Second is the goal of exploring thermal perception in the brain. Temperature perception remains relatively poorly understood and the theoretical localization in the brain of hot versus cold sensation, in particular, is not well-defined [64]. Being able to achieve thermal perception through auditory stimulation would enable experimentation exploring this subject. The findings of such research can have significant implications for our understanding of numerous aspects of thermal perception and perception in general, among them the evolution of our perceptual processes, the development of these processes throughout the lifetime, and more. Insights which in their turn will help in the development of new interfaces and interactions.

5 Conclusion

In this work, we detail a novel methodology for applying a sensory substitution approach to represent thermal information. We present the details of the ThermalSense system development, including the choices made in acquiring thermal imaging data, the procedure for its processing, classification to meaningful temperature ranges, considerations for information to represent, two proposed mappings for enabling an effective representation, in addition to detailing the rationale and process underpinning these choices, and a training program. These novel methodologies and techniques aim to enable the development of real-life implementations, as well as offer insights for developing similar systems for sensory extension, and for thermal perception in particular.

The results of the study demonstrate that the ThermalSense system effectively conveys thermal information through auditory properties, using a visual-to-auditory sensory substitution algorithm. The system equips its users with the ability to perceive temperature information in a visual scene, thus highlighting its potential for extending the natural human visual range. The system was shown to require minimal training and be both pragmatic and hedonically pleasing. As such, this study serves as a proof of concept and springboard for a variety of applications and further implementations. It also provides a foundation for future research into questions concerning basic scientific principles, human experience and interaction.

Acknowledgments

This work was supported by a European Research Council Consolidator grant (773121 NovelExperiSense), a European Research Council POC grant (101123434 TouchingSpace360), and a European Union Horizon 2020 research and innovation program grant (101017884 GuestXR) - all to AA. The work was also partially supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 Project-ID 329551904 "EASE - Everyday Activity Science and Engineering," University of Bremen - to IYW.

References

[1] Sami Abboud, Shlomi Hanassy, Shelly Levy-Tzedek, Shachar Maidenbaum, and Amir Amedi. 2014. EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restorative Neurology and Neuroscience* 32, 2 (2014), 247–257. https://doi.org/10.3233/RNN-130338

[2] Sami Abboud, Shachar Maidenbaum, Stanislas Dehaene, and Amir Amedi. 2015. A number-form area in the blind. *Nature communications* 6, 1 (2015), 6026.

[3] Yomna Abdelrahman and Albrecht Schmidt. 2018. Beyond the visible: Sensing with thermal imaging. *Interactions* 26, 1 (2018), 76–78.

[4] Yomna Abdelrahman, Albrecht Schmidt, and Pascal Knierim. 2017. Snake view: exploring thermal imaging as a vision extender in mountains. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 1067–1071. https://doi.org/10.1145/3123024.3124450

[5] Yomna Abdelrahman, Pawel Wozniak, Pascal Knierim, Niels Henze, and Albrecht Schmidt. 2018. Exploration of Alternative Vision Modes Using Depth and Thermal Cameras. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM '18)*. Association for Computing Machinery, New York, NY, USA, 245–252. https://doi.org/10.1145/3282894.3282920 event-place: Cairo, Egypt.

[6] Tanushree Agrawal and Adena Schachner. 2023. Hearing water temperature: Characterizing the development of nuanced perception of sound sources. *Developmental Science* 26, 3 (2023), e13321.

[7] Amir Amedi, Shir Hofstetter, Shachar Maidenbaum, and Benedetta Heimler. 2017. Task Selectivity as a Comprehensive Principle for Brain Organization. *Trends in Cognitive Sciences* 21, 5 (May 2017), 307–310. https://doi.org/10.1016/j.tics.2017.03.007

[8] Codruta Orniana Ancuti, Cosmin Ancuti, and Philippe Bekaert. 2009. ColEnViSon: Color Enhanced Visual Sonifier-A Polyphonic Audio Texture and Salient Scene Analysis. In *International Conference on Computer Vision Theory and Applications*, Vol. 1. SCITEPRESS, 566–572.

[9] Andreja Bubic, Ella Striem-Amit, and Amir Amedi. 2010. Large-scale brain plasticity following blindness and the use of sensory substitution devices. *Multisensory object perception in the primate brain* (2010), 351–380.

[10] Galit Buchs, Benedetta Haimler, Menachem Kerem, Shachar Maidenbaum, Liraz Braun, and Amir Amedi. 2021. A Self-Training Program For Sensory Substitution Devices. *Plos One* 16, 4 (April 2021), 1–20. https://doi.org/10.1371AndJournal.Pone.0250281 Publisher: Public Library Of Science.

[11] Shaoyu Cai, Pingchuan Ke, Takuji Narumi, and Kening Zhu. 2020. Thermairglove: A pneumatic glove for thermal perception and material identification in virtual reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 248–257.

[12] Edward Childs and Ville Pulkki. 2003. Using multi-channel spatialization in sonification: a case study with meteorological data. Georgia Institute of Technology.

[13] TJARD DE COCK BUNING. 2015. Thermal Sensitivity as a Specialization for Prey Capture and Feeding in Snakes1. *American Zoologist* 23, 2 (08 2015), 363–375. https://doi.org/10.1093/icb/23.2.363 arXiv:https://academic.oup.com/icb/article-pdf/23/2/363/857173/23-2-363.pdf

[14] Zachary C. DeVries, Russell Mick, and Coby Schal. 2016. Feel the heat: activation, orientation and feeding responses of bed bugs to targets at different temperatures. *Journal of Experimental Biology* 219, 23 (12 2016), 3773–3780. https://doi.org/10.1242/jeb.143487 arXiv:https://journals.biologists.com/jeb/article-pdf/219/23/3773/1918551/jeb143487.pdf

[15] David M Eagleman and Michael V Perrotta. 2023. The future of sensory substitution, addition, and expansion via haptic devices. *Frontiers in Human Neuroscience* 16 (2023), 1055546.

[16] FLIR. 2023. FLIR C5. https://www.flir.com/products/c5/.

[17] John H. Flowers and Douglas C. Grafel. 2002. Perception of Sonified Daily Weather Records. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 46, 17, 1579–1583. https://doi.org/10.1177/154193120204601711 arXiv:https://doi.org/10.1177/154193120204601711

[18] John H Flowers, Laura E Whitwer, Douglas C Grafel, and Cheryl A Kotan. 2001. Sonification of daily weather records: Issues of perception, attention and memory in design choices. *Faculty Publications, Department of Psychology* (2001), 432.

[19] Alberto Gallace, Mary K Ngo, John Sulaitis, and Charles Spence. 2012. Multisensory presence in virtual reality: possibilities & limitations. In *Multiple sensorial media advances and applications: New developments in MulSeMedia*. IGI Global, 1–38.

[20] Scott St George, Daniel Crawford, Todd Reubold, and Elizabeth Giorgi. 2017. Making climate data sing: Using music-like sonifications to convey a key climate record. *Bulletin of the American Meteorological Society* 98, 1 (2017), 23–27.

[21] Giles Hamilton-Fletcher, James Alvarez, Marianna Obrist, and Jamie Ward. 2022. SoundSight: a mobile sensory substitution device that sonifies colour, distance, and temperature. *Journal on Multimodal User Interfaces* 16, 1 (March 2022), 107–123. https://doi.org/10.1007/s12193-021-00376-w

[22] Benedetta Heimler, Ella Striem-Amit, and Amir Amedi. 2015. Origins of task-specific sensory-independent organization in the visual and auditory brain: neuroscience evidence, open questions and clinical implications. *Current Opinion in Neurobiology* 35 (Dec. 2015), 169–177. https://doi.org/10.1016/j.conb.2015.09.001

[23] Thomas Hermann. 2008. Taxonomy and definitions for sonification and auditory display. (2008).

[24] Shir Hofstetter, Wietske Zuiderbaan, Benedetta Heimler, Serge O. Dumoulin, and Amir Amedi. 2021. Topographic maps and neural tuning for sensory substitution dimensions learned in adulthood in a congenital blind subject. *NeuroImage* 235 (July 2021), 118029. https://doi.org/10.1016/j.neuroimage.2021.118029

[25] Yu Lun Hsu, Chien-Ting Lu, Li-Chun Lu, Chih-Heng Tam, Yu-Chieh Sun, and Ting-Kang Wang. 2024. AnimalSense: Understanding Beyond-human Sensory Capabilities of Animals via VR Games. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.

[26] Francesco Iberite, Jonathan Muheim, Outman Akouissi, Simon Gallo, Giulio Rognini, Federico Morosato, André Clerc, Magnus Kalff, Emanuele Gruppioni, Silvestro Micera, et al. 2023. Restoration of natural thermal sensation in upper-limb amputees. *Science* 380, 6646 (2023), 731–735.

[27] Mohd Shawal Jadin and Kamarul Hawari Ghazali. 2014. Gas leakage detection using thermal imaging technique. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*. IEEE, 302–306.

[28] Meghan C Jimenez and Jeremy A Fishel. 2014. Evaluation of force, vibration and thermal tactile feedback in prosthetic limbs. In *2014 Ieee Haptics Symposium (Haptics)*. IEEE, 437–441.

[29] Seung-Won Kim, Sung Hee Kim, Choong Sun Kim, Kyoungsoo Yi, Jun-Sik Kim, Byung Jin Cho, and Youngsu Cha. 2020. Thermal display glove for interacting with virtual reality. *Scientific reports* 10, 1 (2020), 11403.

[30] Beat Knechtle, Zbigniew Waśkiewicz, Caio Victor Sousa, Lee Hill, and Pantelis T. Nikolaidis. 2020. Cold Water Swimming—Benefits and Risks: A Narrative Review. *International Journal of Environmental Research and Public Health* 17, 23 (Jan. 2020), 8984. https://doi.org/10.3390/ijerph17238984 Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.

[31] Pascal Knierim, Thomas Kosch, Gabrielle LaBorwit, and Albrecht Schmidt. 2020. Altering the Speed of Reality?: Exploring Visual Slow-Motion to Amplify Human Perception using Augmented Reality. In *Proceedings of the Augmented Humans International Conference*. ACM, Kaiserslautern Germany, 1–5. https://doi.org/10.1145/3384657.3384659

[32] Shelly Levy-Tzedek, Dar Riemer, and Amir Amedi. 2014. Color improves "visual" acuity via sound. *Frontiers in Neuroscience* 8 (2014). https://www.frontiersin.org/articles/10.3389/fnins.2014.00358

[33] PerMagnus Lindborg. 2018. Interactive sonification of weather data for the locust wrath, a multimedia dance performance. *Leonardo* 51, 5 (2018), 466–474.

[34] Fiona Macpherson. 2018. Sensory Substitution and Augmentation: An Introduction. In *Sensory Substitution and Augmentation*, Fiona Macpherson (Ed.). British Academy, 1–42. https://doi.org/10.5871/bacad/9780197266441.003.0001

[35] Shachar Maidenbaum, Sami Abboud, and Amir Amedi. 2014. Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews* 41 (April 2014), 3–15. https://doi.org/10.1016/j.neubiorev.2013.11.007

[36] Shachar Maidenbaum, Roni Arbel, Galit Buchs, Shani Shapira, and Amir Amedi. 2014. Vision through other senses: Practical use of Sensory Substitution devices as assistive technology for visual rehabilitation. In *22nd Mediterranean Conference on Control and Automation*. IEEE, Palermo, Italy, 182–187. https://doi.org/10.1109/MED.2014.6961368

[37] Amber Maimon, Iddo Yehoshua Wald, Meshi Ben Oz, Sophie Codron, Ophir Netzer, Benedetta Heimler, and Amir Amedi. 2023. The Topo-Speech sensory substitution system as a method of conveying spatial information to the blind and vision impaired. *Frontiers in Human Neuroscience* 16 (2023). https://doi.org/10.3389/fnhum.2022.1058093

[38] Amber Maimon, Iddo Yehoshua Wald, Adi Snir, Meshi Ben Oz, and Amir Amedi. 2024. Perceiving depth beyond sight: Evaluating intrinsic and learned cues via a proof of concept sensory substitution method in the visually impaired and

sighted. *Plos one* 19, 9 (2024), e0310033.
[39] Amber Maimon, Or Yizhar, Galit Buchs, Benedetta Heimler, and Amir Amedi. 2022. A case study in phenomenology of visual experience with retinal prosthesis versus visual-to-auditory sensory substitution. *Neuropsychologia* 173 (2022), 108305. https://doi.org/10.1016/j.neuropsychologia.2022.108305
[40] Steve Mann, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang. 2018. All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv preprint arXiv:1804.08386* (2018).
[41] N. A. Martin and S. Falder. 2017. A review of the evidence for threshold of burn injury. *Burns* 43, 8 (Dec. 2017), 1624–1639. https://doi.org/10.1016/j.burns.2017.04.003
[42] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.
[43] N J Mogale. [n. d.]. MASTER OF MEDICINE (PAEDIATRICS). ([n. d.]).
[44] Jonathan Muheim, Francesco Iberite, Outman Akouissi, Rachel Monney, Federico Morosato, Emanuele Gruppioni, Silvestro Micera, and Solaiman Shokur. 2024. A sensory-motor hand prosthesis with integrated thermal feedback. *Med* 5, 2 (2024), 118–125.
[45] Kimberly J. Nahon, Mariëtte R. Boon, Fleur Doornink, Ingrid M. Jazet, Patrick C. N. Rensen, and Gustavo Abreu-Vieira. 2017. Lower critical temperature and cold-induced thermogenesis of lean and overweight humans are inversely related to body mass and basal metabolic rate. *Journal of Thermal Biology* 69 (Oct. 2017), 238–248. https://doi.org/10.1016/j.jtherbio.2017.08.006
[46] Ophir Netzer, Benedetta Heimler, Amir Shur, Tomer Behor, and Amir Amedi. 2021. Backward spatial perception can be augmented through a novel visual-to-auditory sensory substitution algorithm. *Scientific Reports* 11, 1 (2021), 11944. Publisher: Nature Publishing Group UK London.
[47] G. Neuweiler. 1990. Auditory adaptations for prey capture in echolocating bats. *Physiological Reviews* 70, 3 (July 1990), 615–641. https://doi.org/10.1152/physrev.1990.70.3.615 Publisher: American Physiological Society.
[48] Dominik Osiński, Marta Łukowska, Dag Roar Hjelme, and Michał Wierzchoń. 2021. Colorophone 2.0: A Wearable Color Sonification Device Generating Live Stereo-Soundscapes—Design, Implementation, And Usability Audit. *Sensors* 21, 21 (2021). https://doi.org/10.3390AndS21217351
[49] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (March 2018), 22–27. https://doi.org/10.1016/j.jbef.2017.12.004
[50] Elena Peltz, Frank Seifert, Roberto DeCol, Arnd Dörfler, Stefan Schwab, and Christian Maihöfner. 2011. Functional connectivity of the human insular cortex during noxious and innocuous thermal stimulation. *NeuroImage* 54, 2 (Jan. 2011), 1324–1335. https://doi.org/10.1016/j.neuroimage.2010.09.012
[51] Toky Rakotoarivelo and Bruno Malet-Damour. 2023. Exploring the Interplay between Thermal and Visual Perception: A Critical Review of Studies from 1926 to 2022. *Buildings* 13, 4 (2023), 879.
[52] Jane Saycell, Mitch Lomax, Heather Massey, and Mike Tipton. 2019. How cold is too cold? Establishing the minimum water temperature limits for marathon swim racing. *British Journal of Sports Medicine* 53, 17 (Sept. 2019), 1078–1084. https://doi.org/10.1136/bjsports-2018-099978
[53] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40. https://doi.org/10.9781/ijimai.2017.445
[54] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103. https://doi.org/10.9781/ijimai.2017.09.001
[55] Shira Shvadron, Adi Snir, Amber Maimon, Or Yizhar, Sapir Harel, Keinan Poradosu, and Amir Amedi. 2023. Shape detection beyond the visual field using a visual-to-auditory sensory augmentation device. *Frontiers in Human Neuroscience* 17 (2023). https://doi.org/10.3389/fnhum.2023.1058617
[56] Charles Spence. 2020. Temperature-Based Crossmodal Correspondences: Causes and Consequences. *Multisensory Research* 33, 6 (Oct. 2020), 645–682. https://doi.org/10.1163/22134808-20191494 Publisher: Brill.
[57] Ella Striem-Amit and Amir Amedi. 2014. Visual Cortex Extrastriate Body-Selective Area Activation in Congenitally Blind People "Seeing" by Using Sounds. *Current Biology* 24, 6 (March 2014), 687–692. https://doi.org/10.1016/j.cub.2014.02.010
[58] Ella Striem-Amit, Laurent Cohen, Stanislas Dehaene, and Amir Amedi. 2012. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 3 (2012), 640–652. Publisher: Elsevier.
[59] Ella Striem-Amit, Miriam Guendelman, and Amir Amedi. 2012. 'Visual' Acuity of the Congenitally Blind Using Visual-to-Auditory Sensory Substitution. *PLOS ONE* 7, 3 (March 2012), 1–6. https://doi.org/10.1371/journal.pone.0033136 Publisher: Public Library of Science.
[60] Superflux. 2011. Song of the Machine. https://superflux.in/index.php/work/song-of-the-machine/

[61] Hanne H. Thoen, Martin J. How, Tsyr-Huei Chiou, and Justin Marshall. 2014. A Different Form of Color Vision in Mantis Shrimp. *Science* 343, 6169 (2014), 411–413. https://doi.org/10.1126/science.1245824

[62] R Vadivambal and Digvir S Jayas. 2011. Applications of thermal imaging in agriculture and food industry—a review. *Food and bioprocess technology* 4 (2011), 186–199.

[63] Radu-Daniel Vatavu. 2022. Sensorimotor realities: Formalizing ability-mediating design for computer-mediated reality environments. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 685–694.

[64] M. Vestergaard, M. Carta, G. Güney, and J. F. A. Poulet. 2023. The cellular coding of temperature in the mammalian cortex. *Nature* 614, 7949 (Feb. 2023), 725–731. https://doi.org/10.1038/s41586-023-05705-5 Number: 7949 Publisher: Nature Publishing Group.

[65] Iddo Yehoshua Wald, Donald Degraen⊠, Amber Maimon⊠, Jonas Keppel, Stefan Schneegass, and Rainer Malaka. 2025. Spatial Haptics: A Sensory Substitution Method for Distal Object Detection Using Tactile Cues. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–12.

[66] Bruce N Walker and Michael A Nees. 2011. Theory of sonification. *The sonification handbook* 1 (2011), 9–39.

[67] Mohr Wenger, Amber Maimon, Or Yizhar, Adi Snir, Yonatan Sasson, and Amir Amedi. 2024. Hearing Temperatures: Employing Machine Learning for Elucidating the Cross-Modal Perception of Thermal Properties Through Audition. *Frontiers in Psychology* 15 (2024), 1353490.