Colorophone 2.2 – A Spatial, Real-Time Color-to-Sound Sensory Substitution Device *Supplementary Material*

Dominik Osiński, *Member*, *IEEE*, Marta Łukowska, Weronika Kałwak, Michał Wierzchoń and Dag Roar Hjelme

THIS document includes the supplementary material for the paper Colorophone 2.2. It contains descriptions of the superpixel-auxel position calibration process, software implementation of 2.1 and 2.2 versions of the system, system controls, details regarding the experimental results and the list of selected usability issues backed up with the user quotes.

I. Superpixel-auxel position calibration

Fig. 1. illustrates how the azimuth angle is defined together with the resulting ITD difference between the left and right ear.

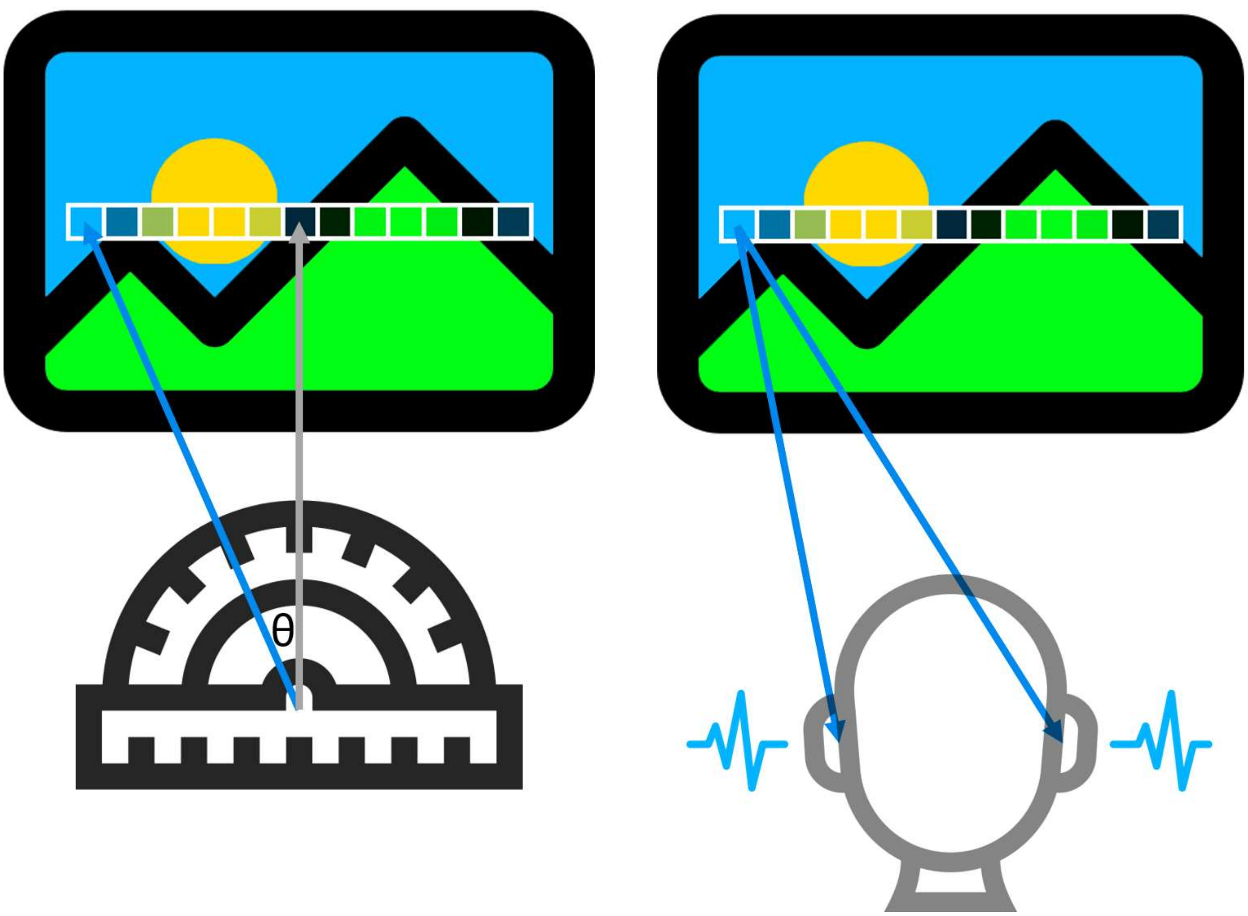


Fig. 1. The image on the left presents the way in which the azimuth angle θ is defined for the leftmost superpixel and consequently for the associated auxel. The image on the right presents the visualization of the ITD between the left and the right ear for the leftmost auxel. Mark that the sound reaching the right ear is delayed in relation to the sound reaching the left ear.

Woodworth's formula [1], describes the relationship between the horizontal position (azimuth, θ) of the auxel and the interaural time difference (ITD) that the spherical head model generates.

$$ITD = \frac{a}{c}(sin\theta + \theta),$$

where a is the sphere's radius, c is the speed of sound (343 m/s), and θ is the azimuth. For the average head radius value, we used 87 mm from an estimation of a spherical head model based on anthropometry [1]. Woodworth's formula is a relatively universal and straightforward way to generate horizontally spatialized sound. The practical implementation of Woodworth's formula in digital systems involves phase shifting of the used digital signals. The relationship between the azimuth of the virtual sound source and the corresponding sample delay for the digital waveform sampled with 44100Hz is presented in Fig. 2.

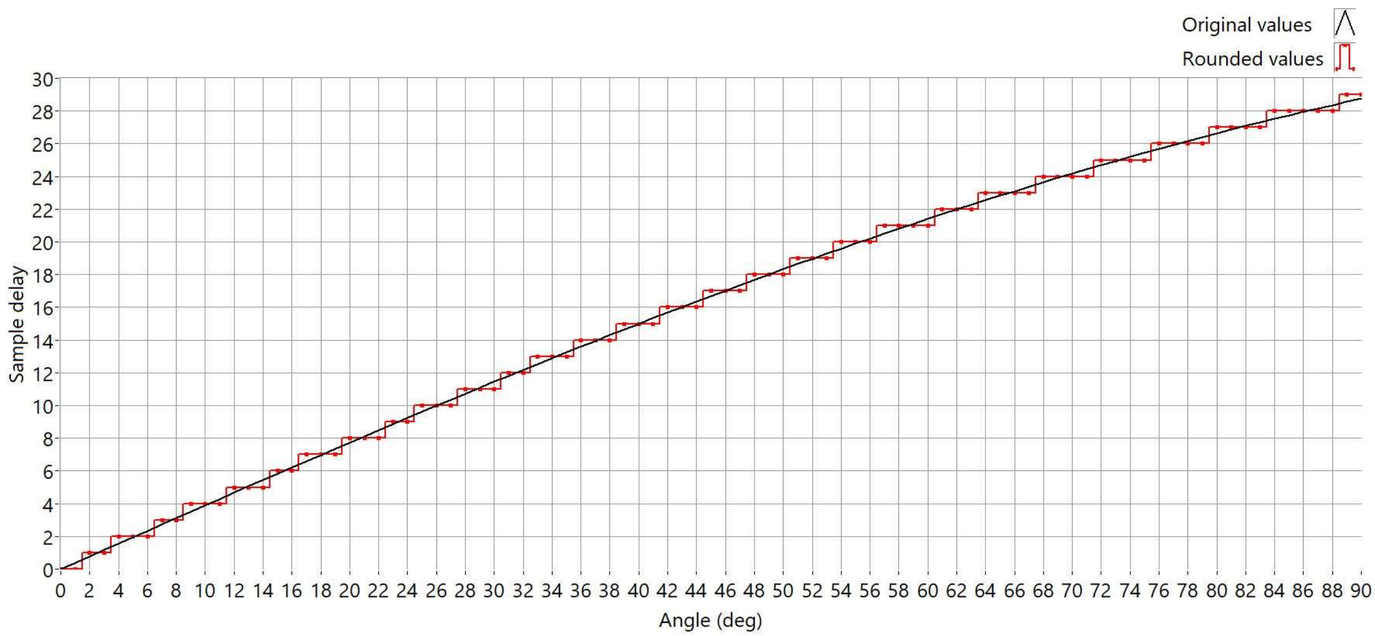


Fig. 2. Graphical presentation of the sample delay generated based on Woodworth's formula for 44100Hz sampling frequency of the sound signal.

Based on the sound spatialization model described above, the coordinates and sizes of the superpixels have been determined in such a way that the center of every superpixel aligns with the angle corresponding to the given sample delay. The superpixels' height is equal to the width of the central superpixel. The defined superpixels are adjacent to each other to ensure no 'blind spots'. In both studies, the systems have been configured so that the position of the center of every superpixel corresponds precisely to the multiples of the 2 sample delays between the left and right sound channels.

The process of superpixel-auxel calibration is different for non-distortion and distortion lenses. Both cases are described in the following sections.

A. Colorophone 2.1

The camera used in the pilot study has 80° FoV and a non-distortion lens. The distortion-free lens enables the calculation of the pixel's angular position based on the following mathematical formula:

$$\beta = \tan^{-1}\left(\frac{b}{a}\right)\tan\alpha,$$

where β is the angular position of a given pixel, α is half the FoV of the camera, a is half of the horizontal resolution of the camera in pixels, and b is the position of the pixel counted from the optical axis of the camera (see Fig. 3. for the simplified graphical representation of the geometrical relationships).

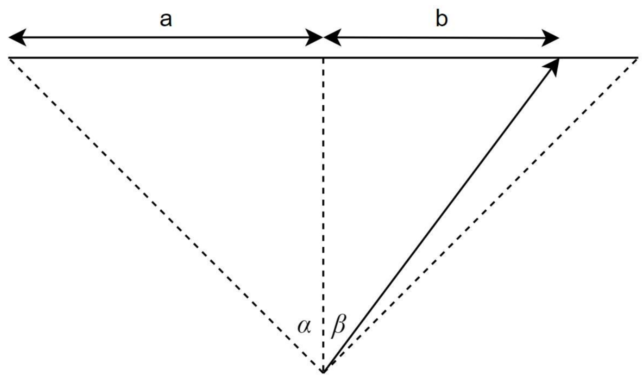


Fig. 3. Simplified graphical representation of the geometrical relationships of the angular pixel position for non-distortion lenses.

The described geometrical relationships allow for theoretical superpixel parameter calculation. We present superpixel and auxel parameters in Table I. Since all of the presented parameters are symmetrical along the vertical axis of the image, we will only show the parameters from the right side of the image. A similar presentation convention will be followed for every calibration table. The correctness of the superpixel-auxel calibration has been experimentally verified by using a dedicated camera calibration program that inserts vertical black lines in the center of every superpixel (see Fig. 4).

TABLE I
CALIBRATION PARAMETERS FOR COLOROPHONE 2.1 SYSTEM WITH NON-DISTORTION LENS FOR 15 SUPERPIXELS

Superpixel (spx) parameters for auxels (ax)	**Ax 0**	**Ax 1**	**Ax 2**	**Ax 3**	**Ax 4**	**Ax 5**	**Ax 6**	**Ax 7**
Theoretical spx center (°)	0	5.1	10.3	15.5	20.7	26.1	31.5	37.1
Spx center (°)	0	5.1	10.4	15.5	20.7	26	31.5	37.3
Spx width (px)	16	18	18	18	20	22	26	30
Spx height (px)	16	16	16	16	16	16	16	16
Sample delay to the central ax	0	2	4	6	8	10	12	14

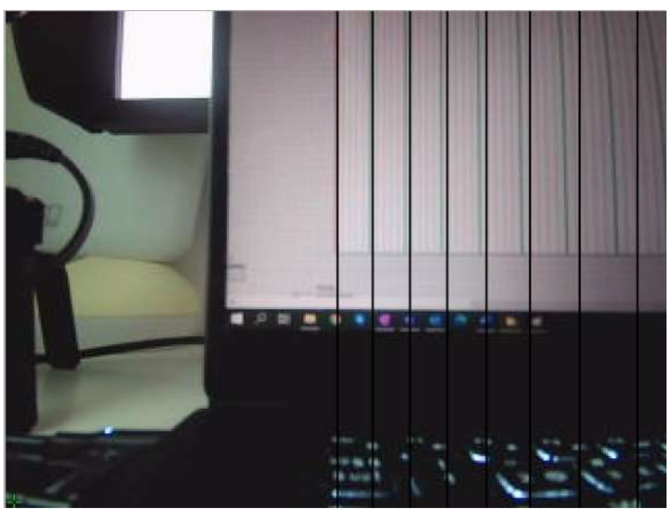


Fig.4. Colorophone 2.1 – a quick experimental verification of the camera calibration results.

B. Colorophone 2.2

The camera used in the main study has a 140° field of view and a distortion lens. The distortion requires individual experimental adjustment for every superpixel. The calibration results for 23 superpixels are presented in Table II.

TABLE II
CALIBRATION PARAMETERS FOR COLOROPHONE 2.2 SYSTEM WITH DISTORTION LENS FOR 23 SUPERPIXELS

Superpixel (spx) parameters for auxels (ax)	**Ax 0**	**Ax 1**	**Ax 2**	**Ax 3**	**Ax 4**	**Ax 5**	**Ax 6**	**Ax 7**	**Ax 8**	**Ax 9**	**Ax 10**	**Ax 11**

Theoretical spx center (°)	0	5.1	10.3	15.5	20.7	26.1	31.5	37.1	42.9	49	55.3	64
Spx center (°)	0	5	10.3	15.4	20.9	26.1	31.6	37	42.9	49	55.3	63.5
Spx width (px)	24	25	25	25	25	26	26	26	27	30	30	43
Spx height (px)	24	24	24	24	24	24	24	24	24	24	24	24
Sample delay to the central ax	0	2	4	6	8	10	12	14	16	18	20	23

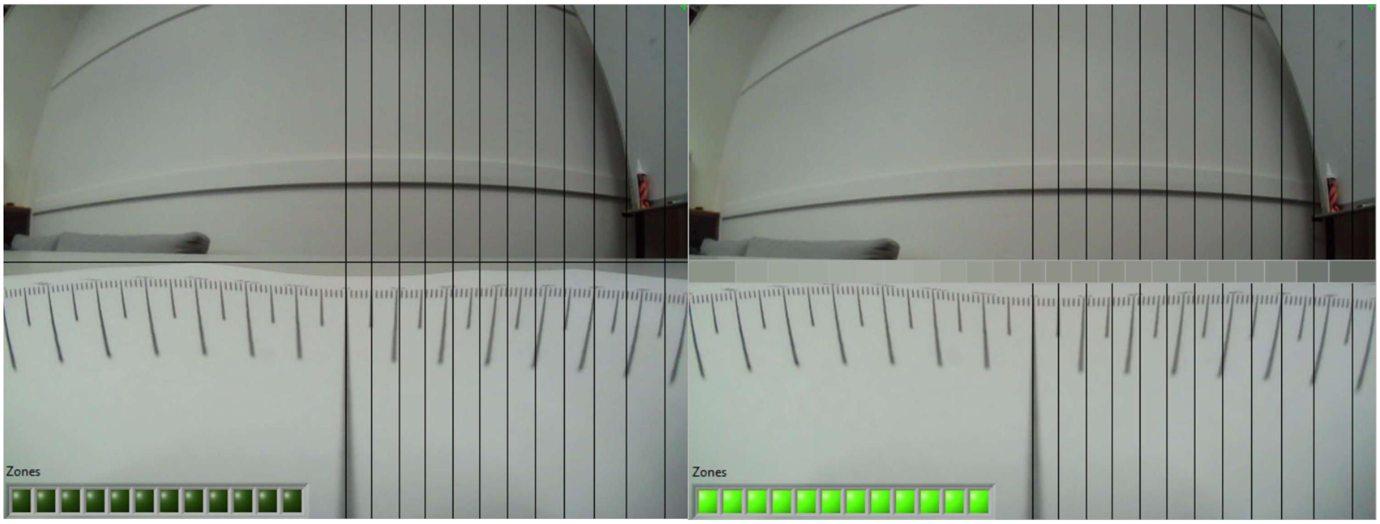


Fig. 5. Calibrated system showing calibration lines and associated superpixels for 23 auxels. Mark that presented images do not come from the actual calibration procedure.

We have also performed the calibration for the cases where the position of the center of every superpixel corresponds precisely to the delay value of 1 sample. The result of this calibration is presented in Table III.

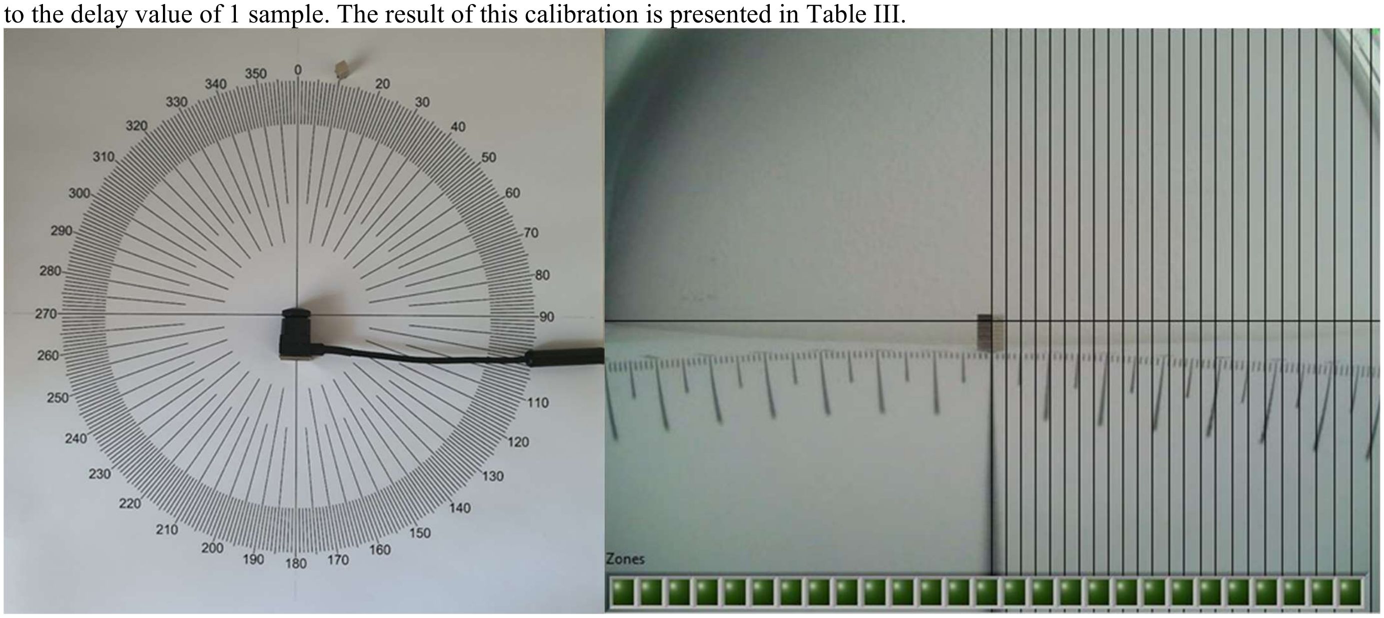


Fig. 6. Camera calibration setup for the camera equipped with a distortion lens for 49 auxels.

TABLE III

CALIBRATION PARAMETERS FOR COLOROPHONE 2.2 SYSTEM WITH DISTORTION LENS FOR 47 SUPERPIXELS

Superpixel (spx) parameters for auxels (ax)	**Ax 0**	**Ax 1**	**Ax 2**	**Ax 3**	**Ax 4**	**Ax 5**	**Ax 6**	**Ax 7**	**Ax 8**	**Ax 9**	**Ax 10**	**Ax 11**	**Ax 12**
Theoretical spx center (°)	0	2.6	5.1	7.7	10.3	12.9	15.5	18.1	20.7	23.4	26.1	28.8	31.5
Spx center (°)	0	2.6	5.1	7.7	10.3	12.9	15.5	18.1	20.8	23.4	26.1	28.8	31.6

Spx width (px)	12	12	12	12	12	12	12	12	12	12	12	13	13
Spx height (px)	12	12	12	12	12	12	12	12	12	12	12	12	12
Sample delay to the central ax	0	2	2	3	4	5	6	7	8	9	10	11	12
Superpixel (spx) parameters for auxels (ax)	**Ax 13**	**Ax 14**	**Ax 15**	**Ax 16**	**Ax 17**	**Ax 18**	**Ax 19**	**Ax 20**	**Ax 21**	**Ax 22**	**Ax 23**	**Ax 24**	
Theoretical spx center (°)	34.3	37.1	40	42.9	45.9	49	52.1	55.3	58.6	62.1	65.6	69.3	
Spx center (°)	34.3	37	40	43	45.9	48.9	52.1	55.2	58.6	62.1	65.6	69.3	
Spx width (px)	13	13	13	13	14	14	14	14	14	15	15	15	
Spx height (px)	12	12	12	12	12	12	12	12	12	12	12	12	
Sample delay to the central ax	13	14	15	16	17	18	19	20	21	22	23	24	

II. Software

This section contains a description of the developed software modules for the pilot study and the main study. The dedicated software packages include programs that were used during the training of spatial color sonification and the main Colorophone application in various versions adjusted for the needs of the researchers, users, and assistants. All software has been developed in LabVIEW 2020. Most of the developed programs show real-time indicators of the processing time of the current iteration of the loops engaged in the information processing. It allows for easy troubleshooting in case of program malfunction.

A. Colorophone 2.1

Here, we describe software modules used in the pilot study and their functions.

1) VAS generator 2.1

The developed software generates a virtual acoustic space (VAS). As presented in Fig. 7, the program enables individual control of the auditory color representation for every auxel. The program contains two tabs: image and sound tab (See Fig. 8-9 presenting programs' graphical user interfaces - GUIs). The superpixels and processing times of the loops are located outside tab structures to be visible independently of the chosen active tab.

The image tab contains configuration data obtained from the calibration of the system with a non-distortion camera. The width of the superpixels can be set up here, and the size of every superpixel numbered from -7 to 7 will reflect the configuration data. The size of the central superpixel should be defined as half of its actual size due to the vertical symmetry of the acoustic space and its implementation. Based on the information about the center of every superpixel, horizontal resolution, and the camera's FoV, the program calculates corresponding sample delays.

The sound tab contains paths to the .wav files used to represent auditory color components, an independent control to the amplitude of the white auditory color component, and the value of the power coefficient used for perceptual compensation of the generated sound amplitude.

The program is implemented as two independent loops. Data between the loops is transferred by using local variables. Color boxes native to the LabVIEW programming environment have been chosen as input controls for color data.

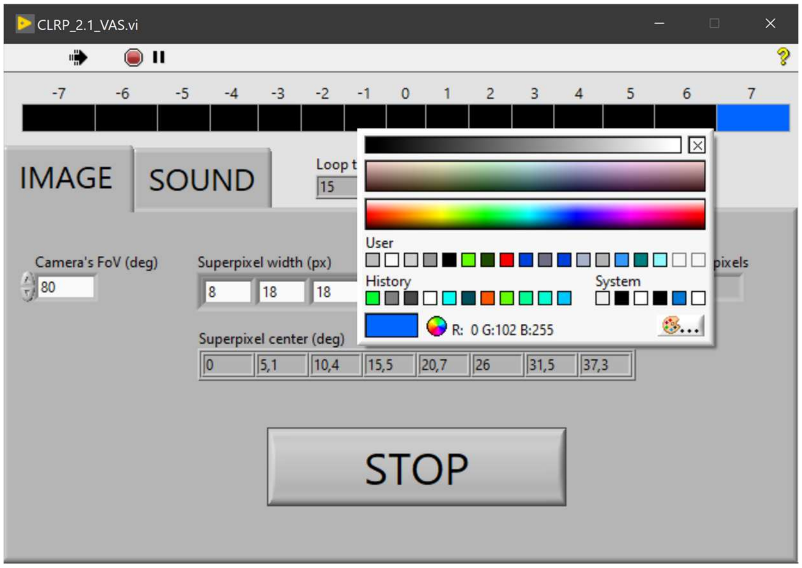


Fig. 7. VAS generator 2.1 GUI – choosing a color for the rightmost superpixel.

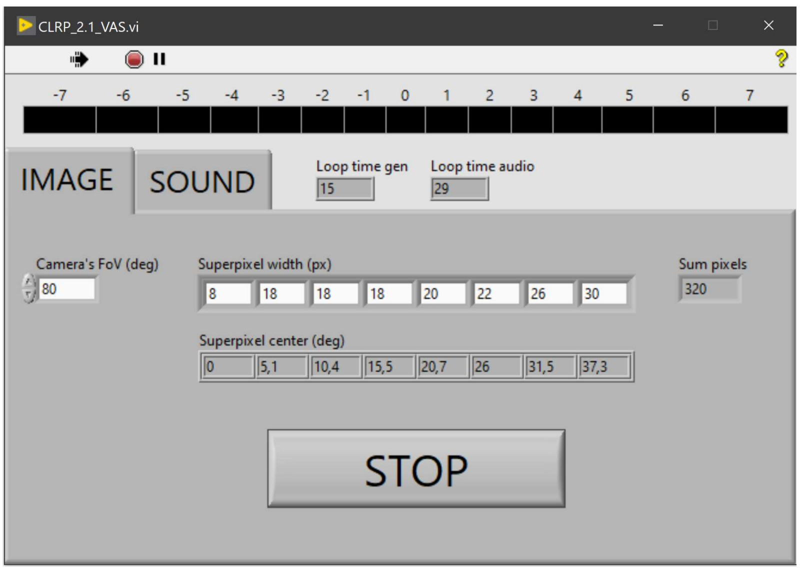


Fig. 8. VAS generator 2.1 GUI – image tab.

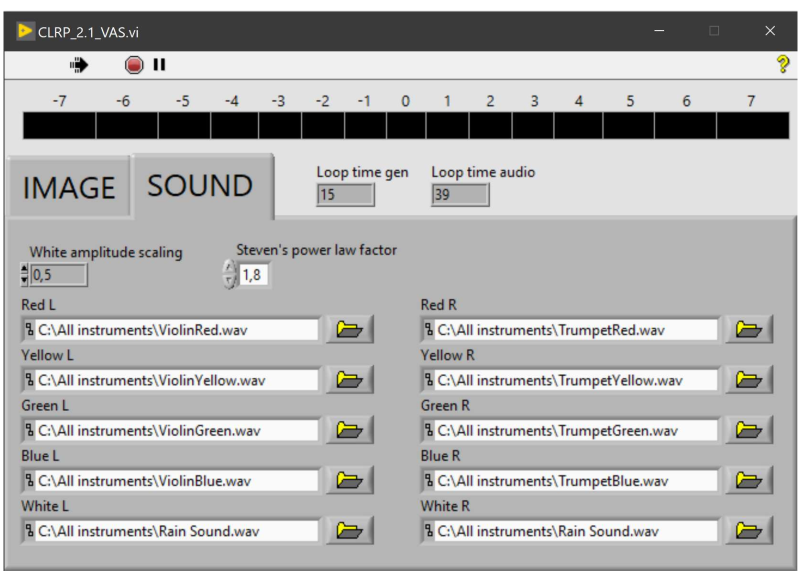


Fig. 9. VAS generator 2.1 GUI – sound tab.

2) Colorophone 2.1 researcher

The Colorophone 2.1 researcher program was used for the initial training during the experiments that did not involve participants' mobility. The program allows for switching on and off individual auxels, which enables the gradual introduction of the spatial color sonification mode. Auxels are symmetrically turned on or off on both sides of the vertical axis of the processed image by clicking on the green Boolean controls nested in the image (See Fig. 10 for an example with 9 active auxels). Similarly to the VAS 2.1 program, the Colorophone 2.1 has two tabs, where the image tab presents the processed image, and the settings tab enables the configuration of the system by choosing camera modes and sounds associated with color components. The rest of the settings have an analogous function to the VAS 2.1 software. The Colorophone 2.1 is implemented as three parallel loops: first responsible for image data acquisition, second realizing data processing, and third realizing sound generation.

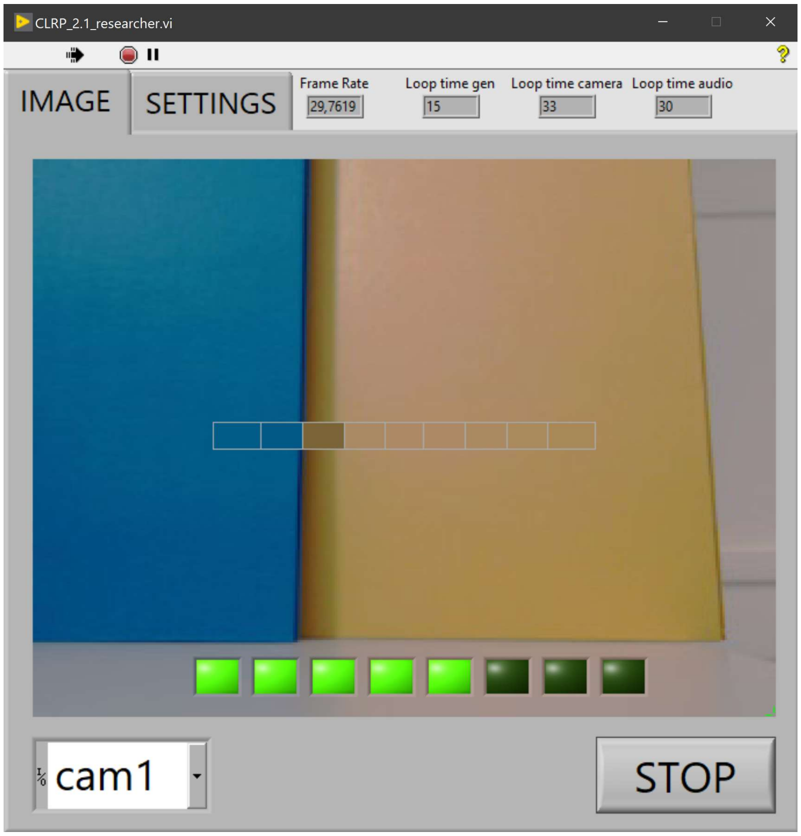


Fig. 10. Colorophone 2.1 researcher GUI.

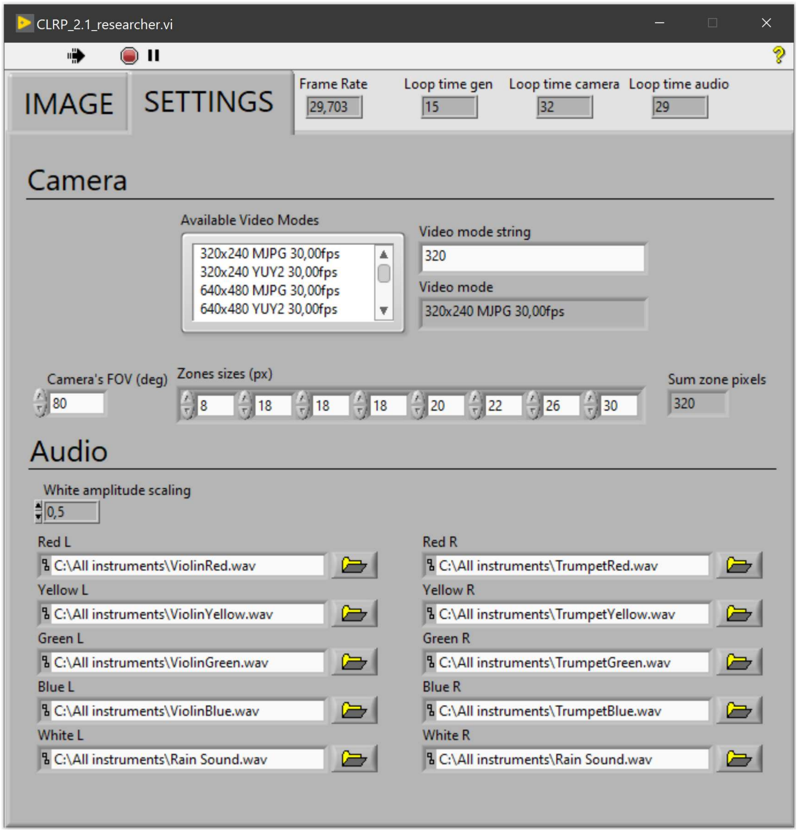


Fig. 11. Colorophone 2.1 researcher GUI.

3) *Colorophone 2.1 user*

The Colorophone 2.1 user program has been used for experiments that involve participants' mobility. This realization of the system includes a simplified GUI, the possibility to switch between spatial and non-spatial modes (by using volume buttons), and inversion mode (toggling between white and black color components).

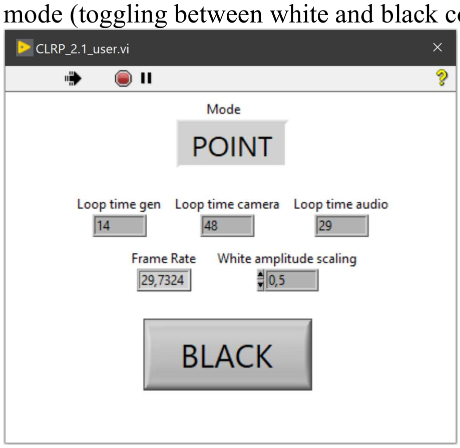


Fig. 12. Colorophone 2.1 user GUI.

B. *Colorophone 2.2*

Here, we describe software modules used in the main study and their functions.

1) *VAS generator 2.2*

The second version of the developed VAS generator had been modified to reflect the needs of the researchers working with the program. The layout is similar to the 2.1 version of the program. However, this version enables choosing only simple colors from a drop-down list (See Fig. 13.). Another modification is the possibility of presenting the black auditory color component as well as individual control of the amplitude scaling for achromatic auditory color components and individual perceptual compensation values for every auditory color component (See Fig. 14.). An alternative version of the VAS 2.2 program that enables presenting of complex color stimuli and dynamic auxel generation based on the 1, 2 or 3 sample delay has also been developed; however, it has not been used during the performed experiments. The program is implemented as three parallel loops, where the first one generates color data, the second loop generates sound signals, and the third one is responsible for GUI update.

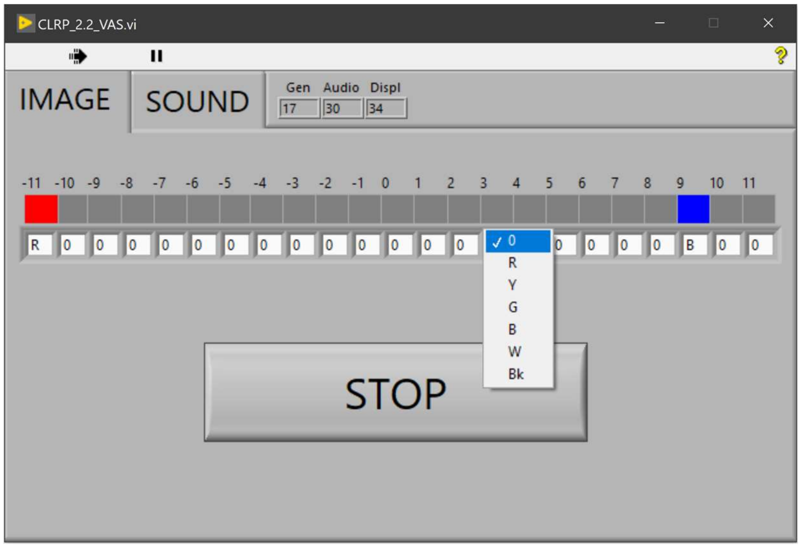


Fig. 13. VAS generator 2.2 GUI – image tab.

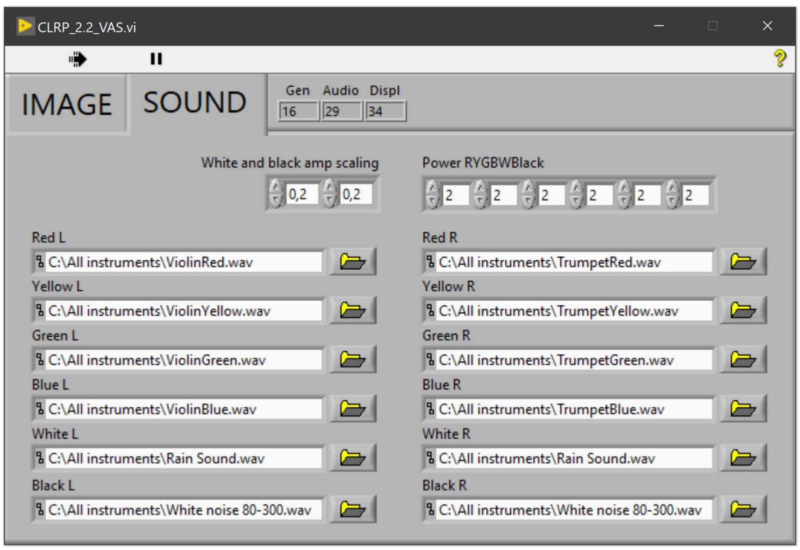


Fig. 14. VAS generator 2.2 GUI – sound tab.

2) Colorophone 2.2 researcher/user

The 2.2 version of the main Colorophone application is realized as on program that can operate in two modes (researcher and user mode). It provides the possibility to work in three different resolution modes (15, 23 or 47 auxels). The program enables flipping the image to provide flexibility of camera attachment. The external Bluetooth controller is integrated with the system to enable wireless control. The program features data logging of used modes and active color components. It is realized by 6 parallel loops that facilitate:

- User interaction via Bluetooth
- Image acquisition
- Color data generation
- Sound generation
- GUI update
- Data logging

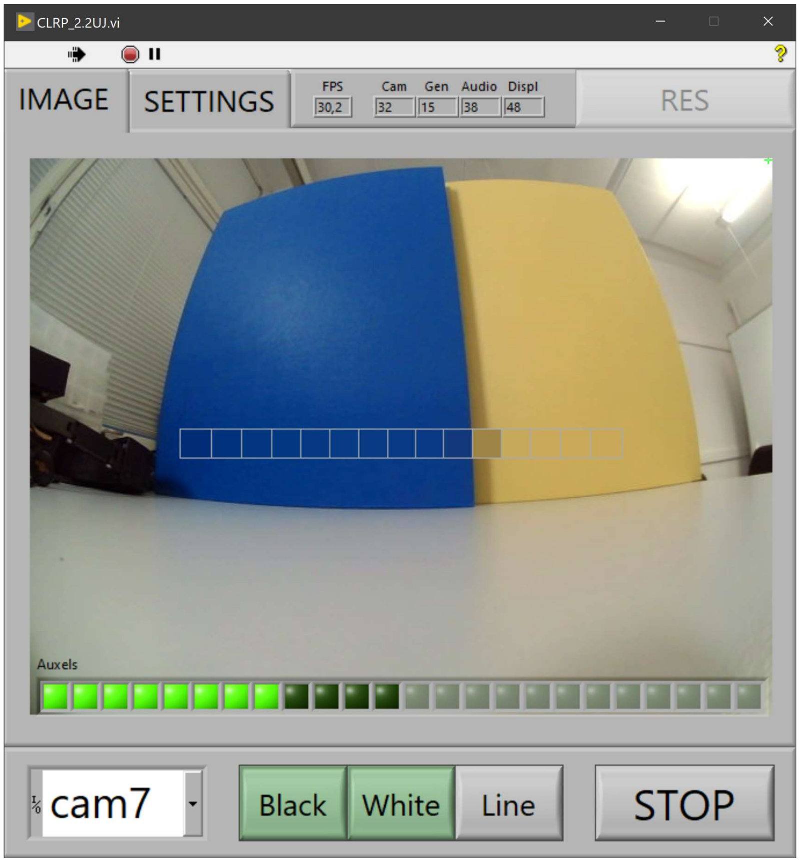


Fig. 15. Colorophone 2.2 GUI in researcher mode.

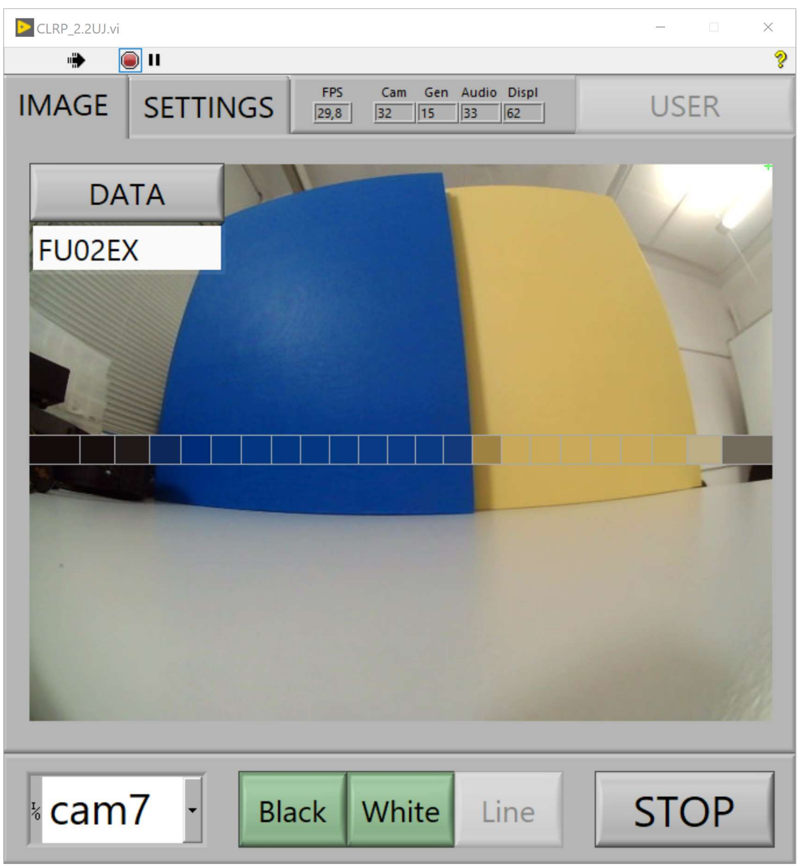


Fig. 16. Colorophone 2.2 GUI in user mode.

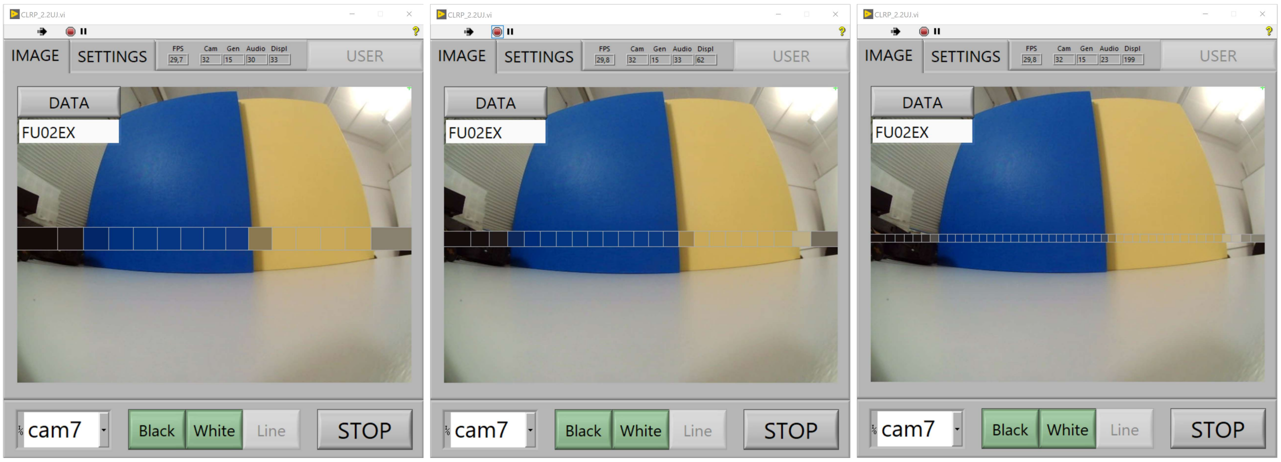


Fig. 17. Colorophone 2.2 comparison of different auxel resolutions

Fig. 18. Colorophone 2.2 settings tab

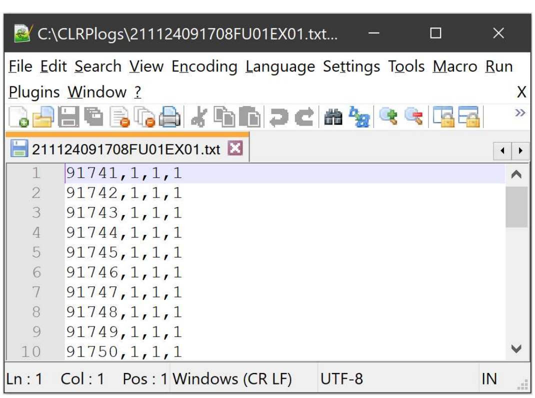


Fig.19. Colorophone 2.2 user mode - log example.

Datalogging every second.
Filename: YYMMDDhhmmsslogstring.txt
Loginfo:

- Timestamp (hhmmss)
- Non-spatial/Spatial
- Black auditory color component
- White auditory color component

3) Colorophone 2.2 assistant

The Colorophone 2.2 Assistant program features the Polish GUI, automatic camera choice, no access to the settings menu, constant data logging and a fixed 23 auxel resolution.

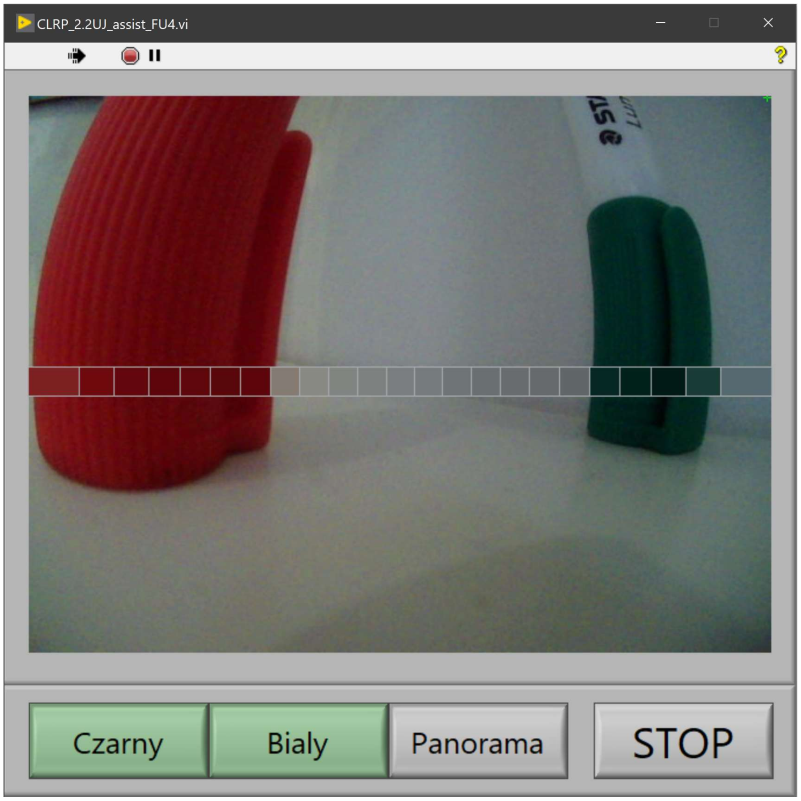
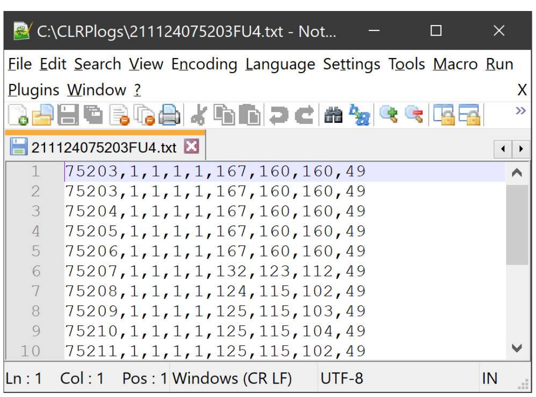


Fig. 20. Colorophone 2.2 assistant GUI.




```
75203,1,1,1,1,167,160,160,49
75203,1,1,1,1,167,160,160,49
75204,1,1,1,1,167,160,160,49
75205,1,1,1,1,167,160,160,49
75206,1,1,1,1,167,160,160,49
75207,1,1,1,1,132,123,112,49
75208,1,1,1,1,124,115,102,49
75209,1,1,1,1,125,115,103,49
75210,1,1,1,1,125,115,104,49
75211,1,1,1,1,125,115,102,49
```

Fig. 21. Colorophone 2.2 assistant log example.

Datalogging every second.
Filename: YYMMDDhhmmsslogstring.txt
Loginfo:

- Timestamp (hhmmss)
- Non-spatial/Spatial

- Black auditory color component
- White auditory color component
- Movement
- RGB values for the central superpixel
- Volume

4) Image hearer

The Image hearer program enables image sonification that allows for the presentation of every auditory color transition by moving the finger on the chosen color change trajectory on the tablet. The system features similar functions to the previously described programs, however the processed image comes from a file and the sonification area can be chosen by finger movements.

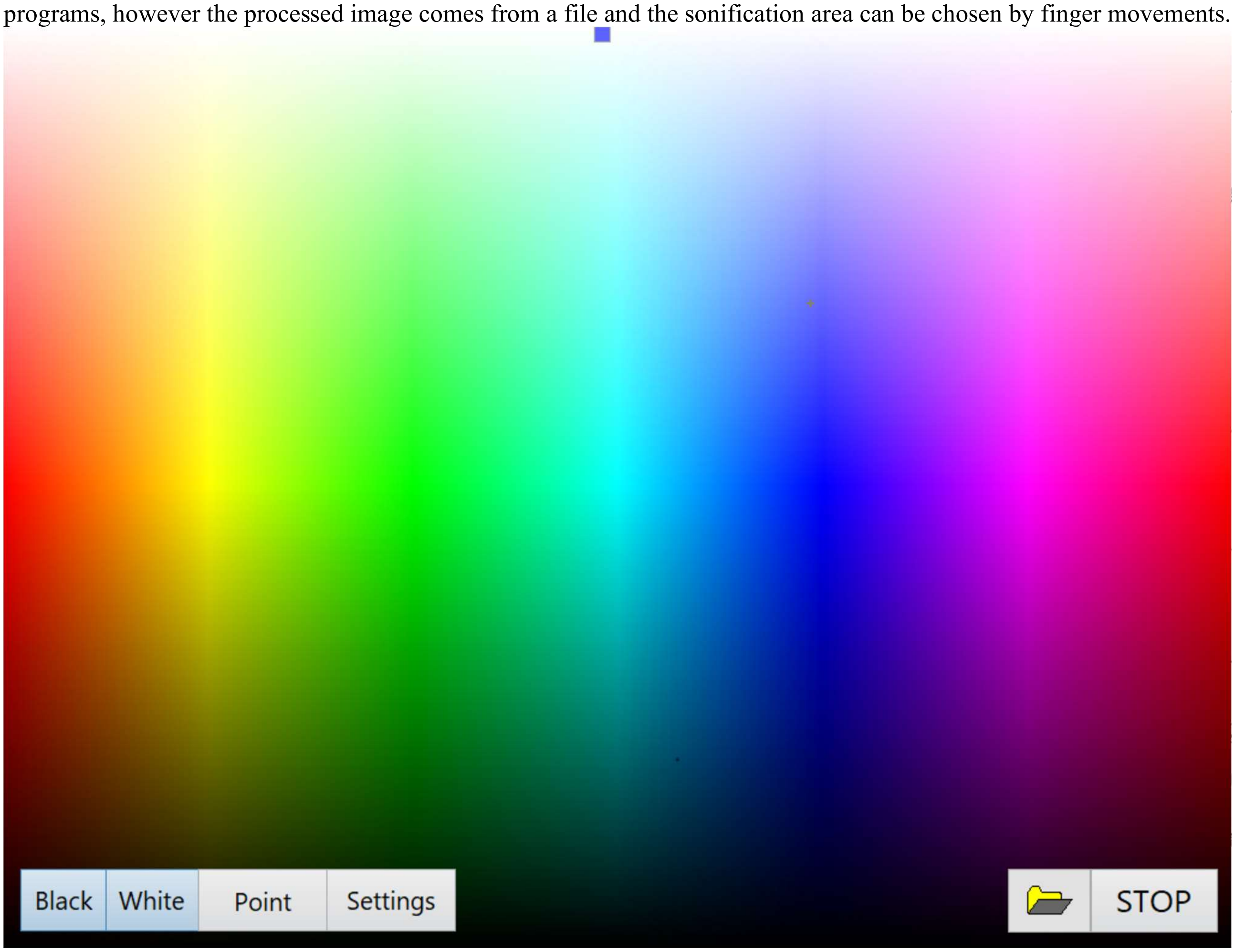


Fig. 22. Colorophone image hearer in non-spatial mode

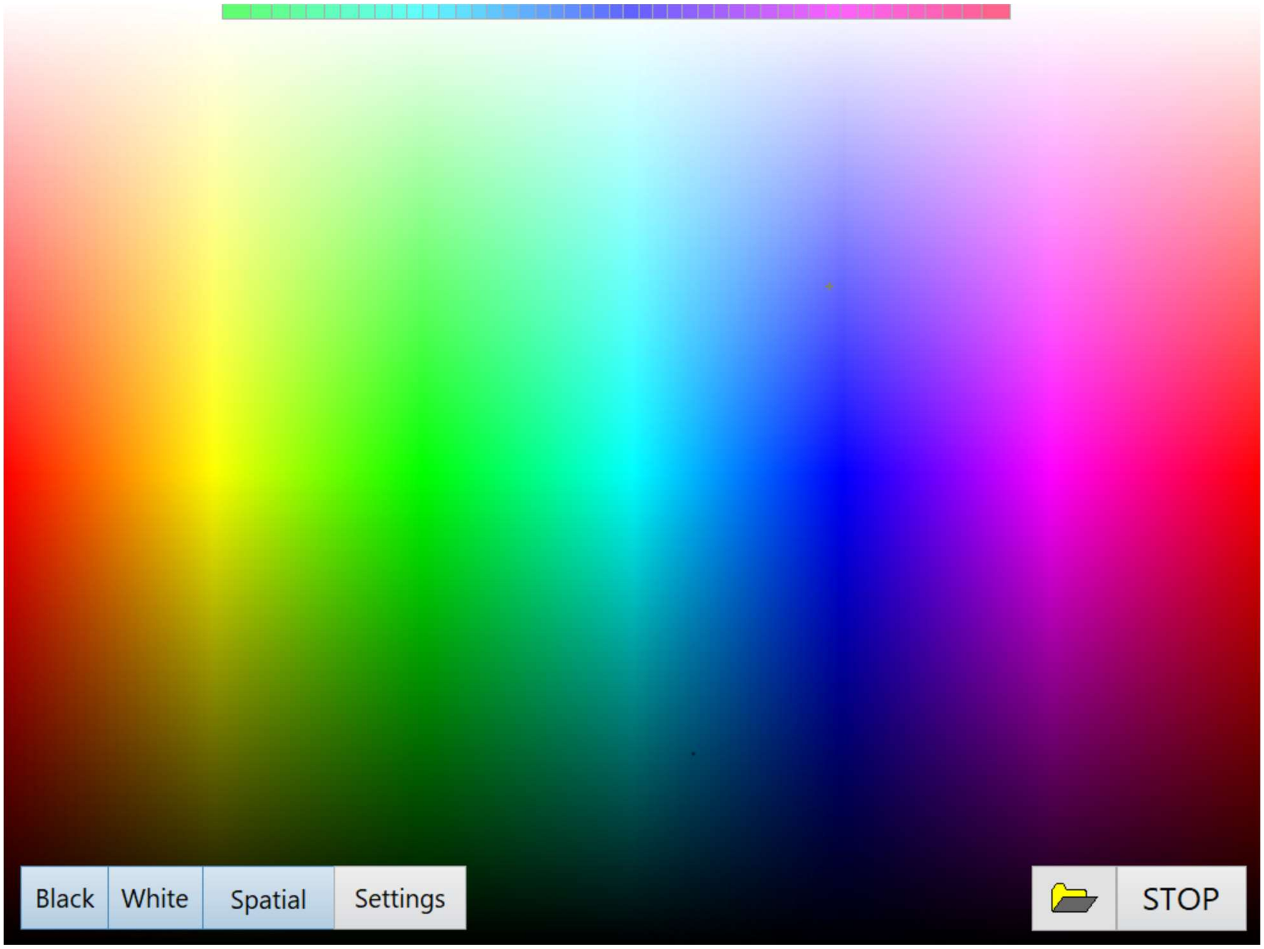


Fig. 23. Colorophone image hearer in spatial mode

III. System control

A. Colorophone 2.1

The Colorophone 2.1 user version of the system can be controlled by the volume changes that are applied through the tablet or Bose glasses. Turning the volume up will switch the program into the non-spatial mode, while turning the volume down will turn on the spatial mode. Activating inversion, i.e., changing between white and black color components to be represented by the sound of the rainfall, can be done by pressing the button on the interface or by adjusting the volume to zero.

B. Colorophone 2.2

The 2.2 version of the system in user mode can be controlled by pressing buttons on the Bluetooth remote controller depicted in Fig. 24. The VOL+ and VOL- buttons are used for the volume control. The button normally used for playing the previous song, placed on the left side of the controller, is mapped to work as a toggle switch for the black auditory color component. The button normally used for playing the next song located on the right side on the controller is mapped to work as a toggle switch for the white auditory color component. The central button normally used for playing and pausing, placed in the center of the controller, is mapped to switch between non-spatial and spatial modes.

Fig. 24. Bluetooth remote controller used in Colorophone 2.2.

IV. System's Evaluation

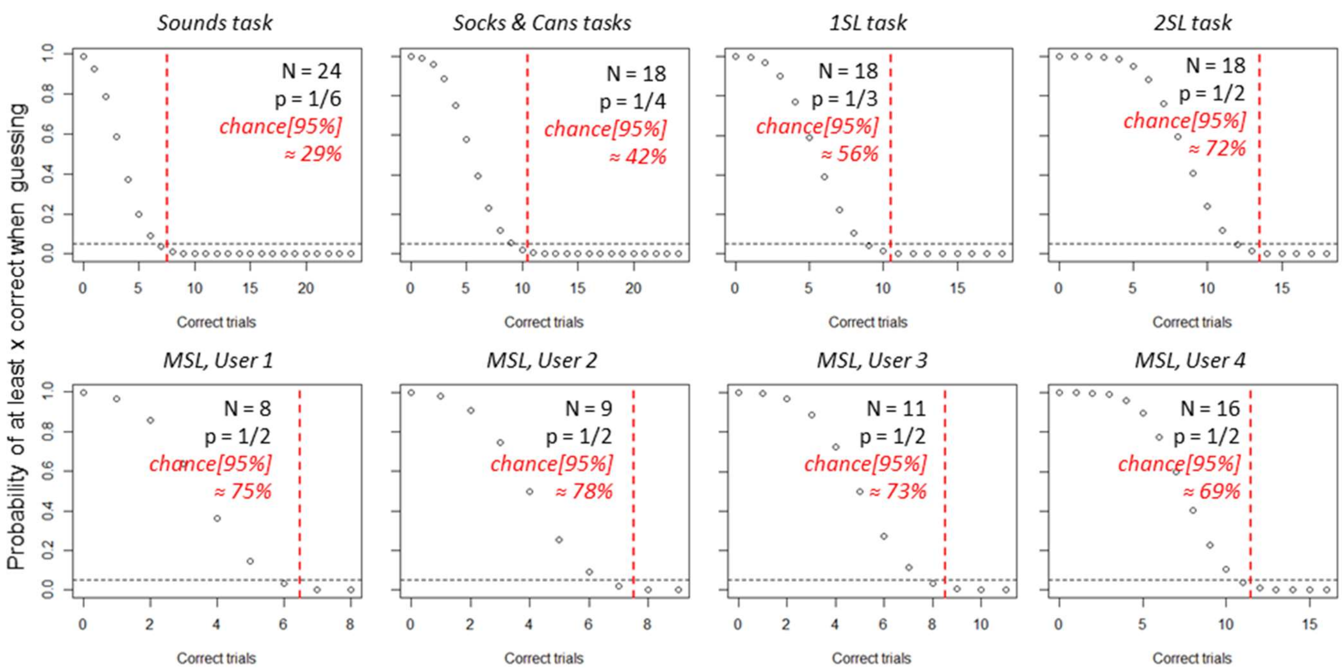


Fig. 25. Chance performance levels for auditory color recognition, object identification and source localization tasks. For each task, we calculated the 95th percentile accuracy that would be observed from a random guessing using the Binomial Distribution. For MSL task, the chance level was calculated for each user individually due to differences in number of underwent trials. Abbreviations: N – number of underwent trials per task or user; p – probability of a correct answer in a given task; chance [95%] – accuracy for the threshold of 5% of guessing.

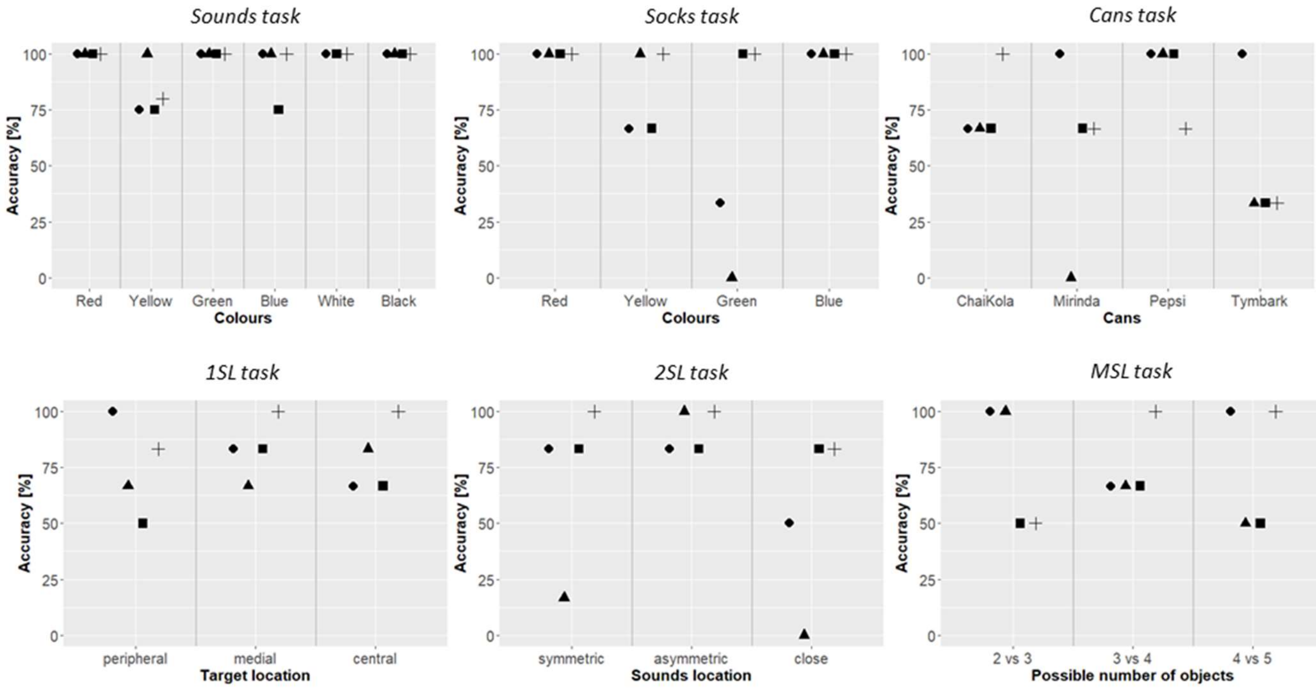


Fig. 26. Accuracy in the auditory color recognition, object identification and source localization tasks with respect to blocks and users (coded with shapes). Users' markers: User 2 – circle, User 3 – triangle, User 4 – square, User 5 – cross.

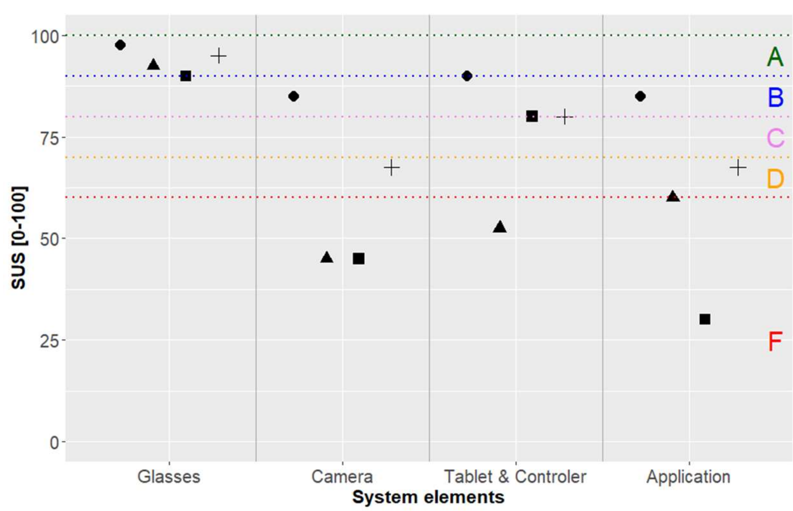


Fig. 27. System Usability Scale results with respect to the system's parts and users (coded by shapes). Dotted lines delineate the boundaries of SUS scores [2]. Users' markers: User 2 – circle, User 3 – triangle, User 4 – square, User 5 – cross.

TABLE IV

RESULTS OF PAIRWISE POST-HOC COMPARISONS FOLLOWING THE TWO-WAY REPEATED MEASURES ANOVA TESTING EFFECTS OF ITEM AND DAY ON NASA-TLX RATINGS. BONFERONNI CORRECTION WAS APPLIED TO CALCULATED ADJUSTED P-VALUES.

		t	*df*	*p.adj*	
Mental Demand	Physical Demand	5.33	25	<.001	***
Mental Demand	Temporal Demand	6.12	25	<.001	****
Mental Demand	Performance	-1.17	25	1	ns
Mental Demand	Effort	-4.24	25	.004	**
Mental Demand	Frustration	6.97	25	<.001	****
Physical Demand	Temporal Demand	1.79	25	1	ns
Physical Demand	Performance	-7.39	25	<.001	****

Physical Demand	Effort	-7.03	25	<.001	****
Physical Demand	Frustration	-.90	25	1.00	ns
Temporal Demand	Performance	-8.29	25	<.001	****
Temporal Demand	Effort	-7.86	25	<.001	****
Temporal Demand	Frustration	-2.38	25	.38	ns
Performance	Effort	-.45	25	1.00	ns
Performance	Frustration	5.45	25	<.001	***
Effort	Frustration	7.64	25	<.001	****

V. IDENTIFIED USABILITY ISSUES

Here, we enumerate selected usability issues identified in the interviews with the users collected after the whole training. The findings, grouped in categories used in SUS investigation, are presented together with the example quotes from the users. The user IDs are U2-U5 because the first user was a part of the pilot study with the Colorophone 2.1. The dialogues are coded as [I] – investigator and [P] – participant.

TABLE V
SELECTED USABILITY ISSUES REPORTED BY PARTICIPANTS GROUPED BY CATEGORIES USED IN SUS

System part	**Usability issue**	**Quote**
Glasses	Absence of tangible buttons for volume control	U2 'Just, little buttons, and necessarily tactile! Like a switch!'
Camera	Cable	U2 'You certainly have to take into account the issue that the cables will deteriorate and that you'll have to replace the cables with new ones. I think that on average every two months there will be a new cable.' U5 'I have to pull the cable with my hand in order to turn my head to the right. It's just that the cable gets in the way a bit. Ok, it definitely has to be there and you can't avoid it, but still, if you could do without the cable, it would be great.'
Camera	Color constancy	U2 'Maybe a professional would know that something is wrong, I don't know, some color distortion, but I don't notice such things.' U5 [I] '[…] I would like to ask you what was the most difficult part of your training and why?' [P] 'Well first of all the differences of light. First of all, if there was little light for example, or just the color of the light, right? Well, because for sure in Biedronka there's a different color, there's some ads and there's something there, some different lights than at home when there's a regular light bulb, right? So, it just mattered. The color of the green light is certainly different than the incandescent light bulb.'
Camera	Fixed, hard to adjust camera angle	U5 'If I look to the sides, it's ok, but for example, if I'm walking, well, I had the camera positioned a bit upwards, right? When I was walking down the street, I had to keep my head like this [the respondent tilts his head forward/down], and for me it was not only uncomfortable, but I also felt strange.'
Tablet and controller	Controller – small buttons	U2 'Well, and I think if it was very cold, these buttons so tiny, here like this, in a circle [the subject draws a small circle with his finger on the glass he's holding], the distances between these buttons with fingers that are so moderately efficient in a cold environment, well they might blend together a little bit. But you know, under certain conditions only.'
Tablet and controller	Controller – sleep mode	U2

		‘Because the issue is that you would have to extend the time of that...what do you call it? This state of pause. Because I’m pressing it and I have to press it a couple of times to turn something off, because it pauses right away.’
Tablet and controller	Controller – unnecessary additional device	U2 ‘…the issue of changing modes I would give to the glasses. I wouldn’t play with additional devices like remote control, but everything in the glasses.’
Tablet and controller	Tablet size	U2 ‘I imagine I’m going to be using alone and not a tablet, but a phone and it’s my own phone because it’s hard to carry two phones.’ U5 ‘It would be the coolest thing if you could have an app like that on your phone. Then... Well, because you always take your phone with you anyway, right? […] It would solve the problem of replacing the tablet with the phone.’
Application	Processing delay	U5 ‘The delays interfere a lot and the small viewing angle.’
Application	Low auditory image resolution	U5 ‘…finding a small object precisely, well sometimes it was already a problem for me I would say. It wasn’t always so good.’

REFERENCES

[1] R. Algazi, C. Avendano, and R. O. Duda, “Estimation of a spherical-head model from anthropometry,” *J. Aud. Eng. Soc.*, vol. 49, pp. 472–479, 2001.

[2] J. Brook, “SUS: a ‘quick and dirty’ usability scale,” *Usability evaluation in industry*, 1996.