Master in Computational Colour and Spectral Imaging (COSI)

Investigation of Universality in Cross-modal Correspondences Between Hue and Pitch

Master Thesis Report

Presented by

Viktoriia Kuznietsova

and defended at the

Norwegian University of Science and Technology

September 2023

Academic Supervisor: Dr. Joni Hyttinen
Host Supervisor: Prof. Shigeki Nakauchi
Jury Committee:

1. Dr. Peter Nussbaum, Norwegian University of Science and Technology, Norway
2. Dr. Rafael Huertas, University of Granada, Spain

Submission of the thesis: 10$^{\text{th}}$ August 2023
Day of the oral defense: 4$^{\text{th}}$ September 2023

Abstract

The combination of different senses is necessary to enhance perception. There exists a hypothesis that correspondences between different modalities are processed automatically in our daily lives and thus seemingly unrelated connections of cross-modal correspondences exist. Cross-modality helps with the efficient processing of information by connecting different modalities. Strong evidence of congruence between shapes and sizes with pitch has been found. Several studies show that with adjustable colour saturation and lightness, the participants will use high values for associations with the high-pitched sound. However, there is still no consensus on the topic of the association between the pitch and hue of the colour. Here we investigate the importance of lightness and show the difference in equiluminant and non-equiluminant stimuli in association to pitch. We found a strong connection between yellow colour with high pitch and blue colour with low pitch in the HSV binary-choice experiments. Contrary to that, the results of the CIELCH binary choice experiment showed a tendency for green hues for high-pitch tones and red for low-pitch. The results showcase a drastic difference in equiluminant and non-equiluminant stimuli cross-modality and support the importance of lightness in the colour-sound correspondences. The use of experiments that test subconscious associations could be a possible way for development since the conscious results correlate heavily to the culture and personal experience than to natural intrinsic cross-modal connections. Based on the results of this work, additional experiments should be conducted with the use of visual stimuli without the influence of lightness. The connection of hue to the sounds that are more complex than pure pitch may be investigated to find further cross-modal associations.

Acknowledgment

This work would not be possible if not for the support and encouragement of many academic members. I would like to show my gratitude to my external supervisor, Shigeki Nakauchi, professor in the Department of Computer Science and Engineering at the Toyohashi University of Technology for his knowledge and enthusiasm which helped me greatly during my work. I also wish to pay my regards to my academic supervisor, Joni Hyttinen, a postdoctoral researcher in the Faculty of Science, Forestry and Technology, School of Computing at the University of Eastern Finland for his valuable advice and attention to detail.

Additionally, I would like to thank the Department of Computer Science and Engineering of the Toyohashi University of Technology for providing me with the technology and space for the work on my thesis. I would like to express my appreciation to Kenta Miyamoto and Michael Makoto Martinsen for providing me with guidance in experimental designs and paper writing.

I am grateful to my family and friends for their eternal encouragement on every step of this journey. I appreciate my COSI groupmates who supported each other as a true family. We banded together such a long way from home, helping each other in difficult times.

Finally, I would like to express my appreciation to the COSI program and all of its members for the wonderful experience that this degree has provided.

Contents

1 Introduction

An expression that translates as "yellow cheering" exists in Japan, which refers to high-pitched cheers by women. Such cross-modal phrases could be found in other languages as well, the Blues music genre is one of many such examples. This indicates that colour and sound associations have been a part of human culture long before modern times. However, hue-pitch correspondences are not established in the field, which means that no consensus has yet been reached as to whether a correspondence between hue and pitch exists.

The term “sensory modalities” in perception and cognition refers to the various ways humans detect and perceive the world, including sight, hearing, touch (Sathian and Stilla, 2010), taste, and smell (Kim, 2013). The ability to receive and combine information from multiple sensory modalities is essential for creating a coherent perceptual experience. For the perception of a coherent world, the brain rebinds input signals through different sensory channels into representations of original multi-attribute objects and events (Driver and Spence, 2000).

1.1 Cross-modality

Cross-modality is the interaction of data among various sensory modalities. For example, movies are perceived as an amalgamation of sounds and images. Multimedia uses auditory and visual signals to improve our ability to perceive and comprehend the message. One media adds context to another, creating a fuller experience (Brown and Hopkins, 1967). Understanding how sensory modalities interact and have an impact on one another can help us understand how the brain integrates and processes sensory data and how this integration affects our ability to perceive and learn. Cross-modality training can reorganise the brain to help with better functioning with loss of other senses (Likova, 2014, 2015). Various studies in psychology, neurology, and cognitive science focus on perception and cognition to discover more about cross-modality.

Modern society is used to seeing cross-modal media all around, from movies to art installations. Cross-modality can also enhance senses in virtual reality to make the simulated experience feel real (Biocca et al., 2001). However, the combination of modalities has always been part of human culture. Artists have always experimented with the integration of several sensory modalities by blending noises, smells, colours, and even temperatures (Ho et al., 2014; Spence, 2020).

Changing usual perception by combining several modalities serves to immerse the observer into a specific feeling. Cultural customs related to cuisine also use cross-modality to enhance the experience. Traditions in different countries vary when it comes to taste and fragrance as well as texture and visual presentation of the dishes (Wan et al., 2014). Songs of different nations, textile choices and rituals all bring diversity when it comes to cross-modality. Hence, even the neighbouring nations could have different crossmodal associations based on their cultural background.

1.1.1 Statistical regularities

Statistical patterns in the environment could be a reason for cross-modal common connections. The cases of such correspondences would be the ability to guess the difference in the sound of the dropped objects based on their size (Grassi, 2005) and recognise smaller animals by the high-pitched voice (Sèbe et al., 2010). Therefore, since particular sensory stimuli are more likely to occur together in the environment, the brain may have developed associations between them.

It could also be that the associations are learned by experience or prolonged exposure to a specific culture. For example, the children depict the sun as yellow in France but would use red to colour it in Japan because of the difference in culture (Yano, 2011). The time of day is often connected to warm colours and loud sounds, while the night is associated with cold colours and silence. It is common to assign colours to seasons and temperatures (Lam et al., 2020).

1.2 Sensory functions

Many scientists discussed the idea that cross-modal associations developed as a result of evolution and adaptation (Spence and Deroy, 2012). It may be that inter-medium connection in the brain developed to help in processing information from our surroundings more efficiently.

For example, when judging the ripeness of the fruit, primates first asses the smell and texture (Dominy, 2004; Sumner and Mollon, 2000; Melin et al., 2019). Because of the modern food production industry, a person is more likely to choose food by appearance first and asses smell and texture later (Spence, 2015; Hutchings, 2011; MacDougall, 2002). However, when judging the edibility of food, humans use multiple sensory cues to tell whether the consumption is safe (Chrea et al., 2004).

Both colour and sound are prevalent in our sensory perception of the world. To accommodate the ever-changing environments, living beings developed physical reactions and psychological adaptations (Robinson and Sloutsky, 2010). The ear collects the sound and directs it to the eardrum, while the retina behind the eyes captures the light. Both types of information are converted to electric signals. The sensations are then transferred to the brain for additional processing and integration with other modalities for the synthesis of a complete scene (Stein and Stanford, 2008).

The stimuli combinations are considered congruent when they integrate, improving behavioural performance (Laurienti et al., 2004; Holmes et al., 2007). Various studies have been successful in probing the consistency in the association between pitch and size (Parise and Spence, 2009, 2008; Evans and Treisman, 2010). However, the correspondence of connection between hue and pitch has been far less consistent. For example, some studies have been investigating the congruence between yellow and blue visual stimuli with the low and high-pitch auditory stimuli (Parise, 2016; Hamilton-Fletcher et al., 2017; Anikin and Johansson, 2019; Orlandatou et al., 2012).

1.2.1 Synesthesia

Because of the nature of multi-modal perception, the connection between different senses is almost impossible to avoid, but there are people for whom this connection is even stronger (Martino and Marks, 2001). Synesthesia is a relatively rare condition where the activation of one sensory pathway results in automatic experience in another sensory pathway. It is a blending of the senses in which people encounter a strong linkage between two or more sensory modalities.

The condition is found in about 2-4% of the population with the most common cross-modal combinations related to symbol-colour, sound-colour and symbol-space pairs. Though this type of perception is unusual, it is not considered a disorder. Individuals with synesthesia might see colours when they hear music, feel emotions when they touch textures (Wheeler, 2013) or imagine numbers and letters in spatial layouts (Brang et al., 2010). It gives a unique perspective on the world and might even be a source of creativity. Individuals with synesthesia are usually tested separately from non-synesthetic in experiments where cross-modal perception matters since their condition might heavily influence the results.

Some people experience vivid colour sensations when listening to music or seeing shapes in connection with sounds. The strong sensation associations and cross-modal connections are considered synesthetic (Parise and Spence, 2009). They are not typical for neurocognitive-normal people. The sensations can be unique from person to person and be rare within communities. While the majority of the population does not experience associations to the same level, it can also experience lighter forms of synesthesia too.

1.3 Research question

In this work, we aim to discuss sound-colour cross-modality and expand on the topic of hue and pitch correspondence. For this purpose, it is necessary to investigate the previous research in the field and design the experiments that would investigate the posed questions.

Our main goals in this work:

- Study the cross-modal connection between colour hue and sound pitch in non-synesthetic observers;
- Investigate the association between yellow with a high pitch and blue with a low pitch;
- Find the importance of lightness in the evaluation of colour and sound congruence;
- Test whether there is a significant difference in responses in groups of different ethnicity.

The occurrence when the information received from different sensory modalities aligns is called cross-modal congruence. It refers to how well information from the various senses is combined. In the case of a mismatch between sensory modalities, for example, if the visual cues present a peaceful scene, but the audio indicates a distressed mood, there will be no cross-modal congruence. Such inconsistency leads to confusion. At the same time, incongruence can help detect discrepancies in a mixture of unnaturally-paired stimuli (Walker et al., 2010).

The thalamus and neocortex activate with the reception of stimulation on the sensory information pathways (Castro-Alamancos, 2009). The cortical networks are active even in the passive state of sleep while they are processing the information at a slower rate. The brain engaged in perceiving multisensory spatial coherence receives fuller information from the surroundings and thus has a higher chance of survival especially when it comes to visual and auditory stimuli (Wallace et al., 1992). Developed for survival, the sensory pathways progressed for better perception of the surrounding world in more detail.

1.3.1 Auditory and Visual stimuli

This thesis focuses on the cross-modal connection between visual and auditory stimuli, thus the topics of the presentation of the stimuli should be discussed.

While loudness refers to the subjective perception of sound volume, the phon scale gives a representation of perceived loudness based on human perception of different frequencies. The loudness should be carefully considered in cases with a variety of different tones (Suzuki and Takeshima, 2004). The phons of all tones are not linear in the range of frequency from 16 Hz to 16 kHz so it must be considered when making experiments since the threshold of hearing could be influencing the results of the test.

The design of the experiments that we developed intends the use of screens to display the stimuli. Different colour spaces could be used, each serving a specific purpose (Wyszecki and Stiles, 2000). CIE XYZ is a linear transformation of colour space that was based on human tristimulus colour perception. It is device independent, perceptually uniform and is often used for colour analysis and processing. CIE RGB is a theoretical colour space which combines three primaries. Is not often used for practical purposes but is a reliable reference for other devices and models.

The CIELAB colour space is also device independent. Even though CIELAB is not a truly perceptually uniform colour space, it should be suitable enough for the purposes of our experiments since our focus is equiluminant hues without a predefined setting for the chroma. It has such channels as L* (lightness), a* (green to red), and b* (blue to yellow).

CIELCH is a colour space that was based on CIELAB. It inherits the qualities of the CIELAB system but is more intuitive to the average user.

HSV colour space is an intuitive colour space that is widely used in program applications by graphic designers. It is composed of Hue, Saturation and Value(Brightness) settings. It is not equiluminant but is preferred in some applications because of its ease of use.

For the purposes of the experiments, we chose HSV and CIELCH. Both spaces have the H channel for colour circles which are necessary for continuous adjustment of the colour stimuli by the design of the experiments.

1.3.2 Experiment aims

Based on our goals, we need to perform psycho-physical experiments that test the presence of an inherent connection between colour and sound cross-modalities.

Some studies showed that there was a regularity in yellow colour and high-pitch association while blue was predominantly connected with a low-pitch sound (Hamilton-Fletcher et al., 2017; Parise, 2016). Basing the experiment on the idea that such correspondence is predetermined would harm the credibility of the final results. Thus, it is necessary to give observers enough control over the stimuli adjustment. At the same time, the freedom of colour choice should not overwhelm the observer. A choice between two hues at a time without the adjustment of lightness and saturation could be a suitable design.

When choosing a colour, a person engages in a thought process. The instinctive decision might be influenced by previous experiences and culture. Since the perceived multi-media information can have a different temporal order, the brain is also efficient in distinguishing which came first for the means of survival. But the task becomes more difficult the closer in time the stimuli are to each other since there is no time to think. An idea was introduced that the temporal order in a pair of synesthetically-matched stimuli is harder to distinguish than a mismatched one (Parise and Spence, 2009). If the stimuli are shown rapidly to the observer with the task to distinguish which came second, the person would not have time to rely on personal experience, thus delivering the result that is inherent to their natural perception.

The contributions of this thesis to the hue and pitch cross-modality correspondence are:

- Design and execution of several experiments to test cross-modal colour-sound correspondences in non-synesthetic observers;
- Investigation of universality in the hue-pitch associations of the Japanese and Norwegian observers;
- Achievement of the quantifiable significance of lightness contribution in colour-pitch correspondences.

** Addressing AI usage for Thesis Report Writing: No AI tools were used to generate code or text for use during the work on this research. Aiding the writing process was the only purpose for the utilisation of such tools as ChatGPT and Grammarly.*

2 | Previous work

Cross-modal correspondence assesses how humans match different, seemingly unrelated senses, and whether neurocognitive regularities exist in the participant's sensations. However, though the field has been studied for decades, there is no clear definition of what constitutes a crossmodal correspondence. The modalities have been considered separate by some scientists, while others claim that there is no harsh separation (Shimojo and Shams, 2001). In fact, there are so many minor factors that could influence the cross-modal associations that it is not possible to take them all into account. However, there is strong evidence that cross-modality is closely related to the semantics and context of the measured subjects (Spence, 2011).

Spence (2022) conducted a series of tests examining crossmodal correspondences in groups of people with common criteria such as culture, disorders and specific abilities. Notably, he studied those with synaesthesia and autistic spectrum disorder to study potential group distinctions. He discovered that people with synesthesia tended to cross-modal associations more than others. However, Ward et al. (2006) states that people with synesthesia do not use different but instead use the same pathways between visual and auditory stimuli that a non-synesthetic brain would.

2.1 Cross-modal correspondences

The fact that most auditory and visual sensory signals are often temporarily simultaneous leads to their high integration as stated in the work by Koelewijn et al. (2010). The authors conclude that the timing of the auditory and visual signals played a more important role in cross-modality association than their positioning which implies a high temporal and a lower spatial connection. Multisensory input draws more attention than unisensory stimuli which is shown in animal as well as human brain function. The research discussed that not all circumstances are suited for cross-modal correlation because the connection between the senses is not automatic.

Nevertheless, when the association between two sensory feelings takes place, the impact is much greater than it would have been with just one stimulus. The attention to the stimuli, the positioning in central peripheral zones, and the level of thresholding were important to evaluate the results. The authors suggested that the sound and the visual stimuli are processed in a parallel integration framework. This means the two stimulus types could be processed in conjunction or separately depending on the stimuli pair.

2.1.1 Synesthesic influence

Spence and Deroy (2012) challenged the claim that animals have innate crossmodal correspondences. The authors suggest that finding the natural origin of acquired modal correlation is better than calling animals innately synesthetic. He reviewed several direct and indirect experiments with fast discrimination tasks. Direct experiments tested observers by giving them specific instructions on judging stimuli, while indirect tested for the results incidentally. Both versions showed that the effects are indeed subconscious.

Itoh et al. (2017) explored the pitch class and colour correspondence in people with synesthesia. The participants stated that they received sensations, not from the pitch itself, but from the musical notes. The results of all participants were relatively consistent between the sequence of syllables and the colours of the rainbow. It was reported to not be a taught behaviour since the music teachers of the participants did not use colour designation during the learning process. Other than that, the responses of the participants seemed to be based on verbal representations of the pitches which lead to the idea that the pitch-colour labelling heavily relies on the participant's language and vocabulary.

2.1.2 Cultural influence

Parise (2016) addresses the need for concrete guidelines when assessing cross-modality. According to him, the mechanisms generating crossmodal correspondences should be investigated in future studies, including whether they are based on similar sensory processing pathways or more complex cognitive associations. Additionally, he advised that academics look at the cultural and developmental aspects that affect crossmodal correspondences.

It is common for cross-modality influence to come from a culture, especially in cases where the stimuli are between colour and odour since cuisine is quite distinct in different places around the world. Fragrances are especially important when it comes to food traditions, the surrounding environment and local flora. It is reasonable to assume that specific differences in colour and smell associations depend on the participant's ethnicity. However, the study by Levitan et al. (2014) showed that the geographical origin of the participants was not of essential consequence when comparing the countries. The set of 14 odours was presented to participants from different geographic locations. The participants were asked to assign congruent and incongruent colours to each of the odours and the results showed high agreement between the national groups such as US and German. It was notable that German and Malay participants had more agreement in their answers than Malay and Netherland-residence Chinese.

In his paper, Spence (2015) discussed the correlation of geographical cuisine with colours. He reviewed several papers that studied food and colour correspondence. It was concluded that the connection was heavily reliant on several cultural factors such as age, geographical location and genetic taste status. One of the papers he reviewed was a study by Shankar et al. (2010). In this study, participants were presented with seven drinks, six of which were coloured. The participants were asked to identify the flavour of the drinks. The authors proposed that the investigation of higher-level factors such as expectations should be studied more to comprehend such associations better. Additionally, Wan et al. (2014) examined cross-modality in various populations, including participants from Western and non-Western cultures. On the basis of previous studies, they concluded that people from non-Western cultures were less likely to associate sweet tastes with the colour red.

2.2 Audio-visual correspondence

The hue-pitch cross-modal correspondence is heavily reliant on experiences, culture, and native language as it has been stated by Johansson et al. (2020). In their work, the authors discuss the connection between different sounds from various languages to the colours. They found that the sounds were highly correlated with the luminance and were more reliable than for saturation. They discussed that the brightness was more associated with the sonority of vowels, while the saturation was influenced by the consonants to a lesser degree.

Fu et al. (2020) performed an extensive analysis regarding selective attention in cross-modality with the spatial perception of auditory and visual stimuli. The discussion explains one of the most popular theories about cross-modal integration. The flow of the interaction between visual and auditory processing is represented in Figure 2.1. The main idea of the interaction is that it could work as a bottom-up, as well as a top-down process, meaning that the information is processed either as buildup or decomposition respectively. At its core, attention or lack thereof is a mechanism for humans to efficiently learn and sort important information from the massive amounts of incoming stimuli. For example, several studies conducted by Parisi et al. (2017, 2018) and Fu et al. (2018) had a cross-modal experiment

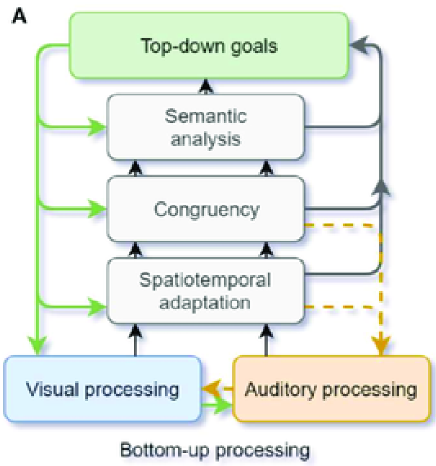


Figure 2.1: *The visualisation of crossmodal integration and attentional control in the human brain. The bottom-up processing is depicted as black and grey arrows, the green arrow is used to show top-down attention adjustment, and the yellow arrow means recurrency of adjustment (Adapted from Fu et al. (2020), under terms of Creative Commons Attribution 4.0 International license).*

with four speakers placed behind each of four virtual avatars displayed on a single monitor. The setup was testing the attention of the participants by asking them to guess which speaker is working when the lips of the avatars were moving with the task for the participants to point to the active auditory source. This would involve a person simultaneously perceiving information from both visual and auditory pathways.

In the study of Palmer et al. (2013), participants listened to music and reported a correspondence of colour and emotion. The results showed that fast-paced happy music was associated with brighter colours like yellow, while slow sad melodies were marked as desaturated, dark, and blue colours. The emotions influenced the response of the participants which means that they are significantly correlated with cross-modal associations. The combination of colour and emotion connection was discussed in the work by Nakajima et al. (2017). In this study, the observers were shown pictures of faces with different emotional gradations. The faces were coloured shades of blue and red and the participants had to identify the emotion in the facial expression in the picture. The results showed that the participants were more likely to mark blue-coloured faces as showing sad emotions. Other combinations of facial expressions and colours did not grant as much significance, implying that the correspondence is disproportionate in its effects.

2.2.1 Congruence in auditory-visual pairing

Müller et al. (2011) discussed the role of incongruence in the processing of cross-modal information. The authors showcased that the amygdala response of the participants was enhanced when a picture with emotion joined with a congruent sound (fearful face and scream). The amygdala response from the participants was weakened when the incongruent stimuli were paired, while other parts of the brain became more active. This means that the brain reacts to the inconsistencies in order to process the information and find a resolution.

Anikin and Johansson (2019) conducted a study to find a connection between colour and sound. The visual stimuli could be changed in hue, lightness and saturation. The sound stimuli settings had five dimensions with loudness, spectral centroid, and pitch among them. It was noted in the work that, unlike what other studies have shown, the saturation of the colours was connected to the spectral centroid, loudness and pitch, while only pitch was associated with the lightness of the colours. Additionally, the study concluded that the use of synthetic vowels could have significance in association with the colour because of low- and high-frequency energy and not the vowels themselves.

2.2.2 Sound-shape associations

The case of the Kiki-Bouba effect first introduced by Köhler (1947) is a typical example of sound-shape cross-modality. In the study, the participants of the experiment made connections between the sound of the words and visual shape stimuli. Later studies confirmed the connection with tests among different ethnic groups with vastly different cultures. It was found that even though the languages from different parts of the world were so different, the participants connected the sharp “Kiki” sound with the pointed shape while assigning the soft “Bouba” - with the dull shape. Similar studies were conducted in other regions with variations of similar sounds and shapes. The results of those studies were consistent with the original one, confirming the cross-modal connection.

A study by Lin et al. (2021) adapted the concept of Kiki-Bouba to 3D printed shapes. The results showed that observers associated shape angularity with red colours and low brightness. Similarly, the round shapes were connected to blue colours with high brightness. Interestingly, the study concluded that the concepts did not directly extrapolate from the 2D plane to the 3D. This suggests that the connection between the dimensional perception is not linear. Thus, there should be a more compound way to understand the connection between the two mediums.

The study by Parise and Spence (2009) discussed the time-order experiment which texted congruence. The sound stimuli consisted of three pairs of visual stimuli and a pair of pure tone frequencies (300 Hz and 4500 Hz sine waves) with seven stimulus onset asynchrony variables. The visual stimuli differed by type (circle, waveform, space) and size (small, big). Each of the types of visual stimuli was tested in a time-order experiment with the two sine frequencies. The experiment used the effects of audiovisual temporal asynchrony to test whether the congruent stimulus pairs proved harder to discern in the time-order judgement. The results showed that the harder it was for the observer to judge which stimuli from type-frequency pairs came first, the stronger the congruence of that pair was. This proved that multi-sensory integration was connected to synesthetic congruency.

2.3 Hue-pitch correspondence

The auditory sound pitch was found to be robust in cross-modal correspondence in studies of Parise (2016); Hamilton-Fletcher et al. (2017); Anikin and Johansson (2019) and Orlandatou et al. (2012). It is still not clear whether this correspondence is relative or absolute. Various experiments suggest that pitch cross-modality is mostly relative. However, judging by the experiments in which the pitch was used as one of two dimensions for mapping the participant's responses, the responses are also connected with linguistics. In the study by Parise (2016), it was also discussed that the decisions of an observer are heavily influenced not only by the language but also, by the daily environment and adaptivity to it. The paper also suggests taking multiple types of approaches to encompass a wider analysis of the reasoning behind cross-modal correspondences.

2.3.1 Yellow-blue correspondence

A study by Hamilton-Fletcher et al. (2017) investigated the topic of colour-sound cross-modality. The visual stimuli that they used were CIELUV colour space with $L* = 65$ and adjustment by radius and azimuth with chroma constrained to 60. The two types of sound stimuli that they used were pure pitch and vocal timbres which were combined in different sets. During the experiment, the participants assigned yellow to the higher pitch and blue to the lower pitch. The results did not show any particular tendency towards other hues. The authors suggest that the equiluminance of the stimuli influenced the dichotomy of the participant results. Notably, the hue adjustment changed linearly in the case of pure tone trials with both changes in frequency bands and loudness function. Contrary to that, the hue association stayed constant in sounds that varied in the centre of gravity.

In the survey by Parise (2016), various cross-modal associations were tested against each other. A little less than 80% of the participants responded with a connection of yellow colour to high pitch and blue colour to low pitch. Though the survey had some methodological shortcomings, it was significant to find a persistent association between colour and sound outside of psychophysical tests.

The study by Orlandatou et al. (2012) compares the three types of sound stimuli in connection to colour in synesthetic observers: pure tones, noise bursts and complex sine tones. In the results yellow and green were consistently assigned to high tones.

3 Material and methods

Instructions and stimuli were presented on a Microsoft Surface Pro 4 CR3-0014 laptop tablet. The device has a 12.3-inch display with a resolution of 2736 x 1824 pixels, a pixel density of 267 PPI and a refresh rate of 60 Hz. The colourimetric properties of the display were measured with a TOPCON SR-3AR and processed with the measurement software Colorimetry Program CS-900A. We used Sennheiser HD650 headphones for sound stimuli.

The experiments were conducted in a shielded darkroom. The setup is visualised in Figure 3.1. We estimated the distance between the positions of the observer's head and the display to be 44-48 cm during the experiments. Additionally, a headrest was used during the Time-order experiment.

The software for visualizing and interacting with the stimuli was implemented in Python 3.7, PsychoPy and PsychoPy Builder 2023.1.1 packages. The integrated development environment used was JetBrains PyCharm (Community) 2023.1.1. The PsychoPy framework is available at `https://www.psychopy.org/`. We used JASP 0.17.2.1 and R 4.4.2 with anovakun 4.8.6 (analysis of variance function) for statistical analysis of the results.

In this work, we present three experiments:

- HSV binary-choice - a cross-modal experiment with HSV stimuli for both Japan and Norway;
- CIELCH binary-choice - a binary-choice cross-modal with CIELCH stimuli experiment with Japan;
- Hue-sound time-order - experiment with HSV yellow-blue and light-dark grey stimuli pairs in Japan.

The same display and headphones were used for all experiments. We chose a setup that could be easily transported for consistent data acquisition in different locations.

Figure 3.1: *The experimental setup consists of the table, chair, laptop and headphones. The experiment was conducted in a shielded darkroom in a quiet environment to have control over the lighting conditions and surrounding sounds. The same computer and headphones were used in different rooms with similar conditions during international testing.*

3.1 Sound stimuli

Sound stimuli were created by changing the frequency of the sine wave. We used a total of 6 stimuli with different frequencies within the human hearing range (20 Hz - 20 kHz). The specific frequencies chosen were 100 Hz, 200 Hz, 400 Hz, 800 Hz, 1600 Hz, and 3200 Hz sine waves which were based on the pitch stimuli from the study by Hamilton-Fletcher et al. (2017). During the experiments, we used 400 Hz sound as the reference for the loudness adjustment because of its middle placement in the scale of frequencies.

The sound volume of the laptop was kept unchanged to have consistent conditions for all trials of each test. The initial sound volume for each stimulus was set based on the preliminary test (1.66, 0.52, 0.3, 0.41, 0.81, 0.42). We used the

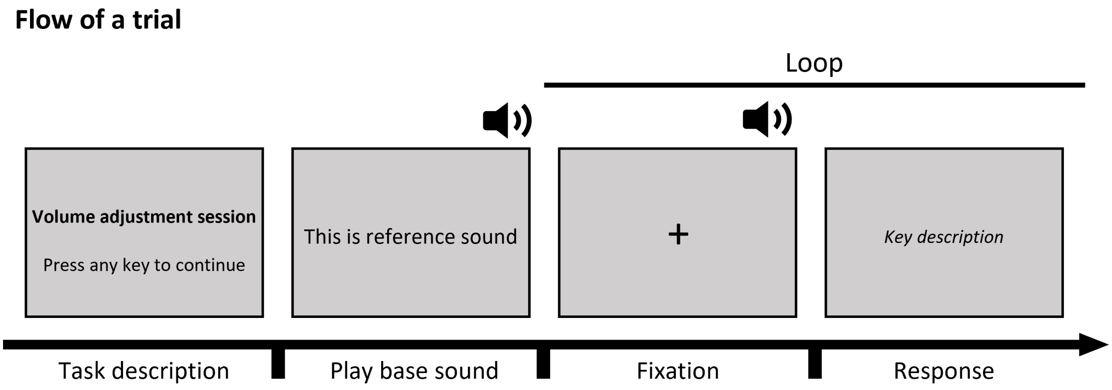


Figure 3.2: *The flow of the preliminary volume adjustment experiment. The first screen introduced the observer to the experiment with a short task description. The target sounds were played next with the task of comparing them to the reference sound (400 Hz) one by one.*

volume scale from 0 to 2 instead of the default PsychoPy settings. The adjustment was performed by the keyboard with the possibility to decrease/increase volume with the step of 0.2 and 0.05. Observers were asked to adjust further the loudness of each sound stimulus to match the reference sample (400 Hz) for a better fit to their personal sound perception. The volume-adjusted sounds were then used as auditory stimuli during the experiment. A number of four sets of six stimuli total was performed during the experiment.

The binary-choice experiments (HSV, CIELCH) used all six frequencies as stimuli. The scales for the volume of each sound could be adjusted by 0.2 and 0.05 both up and down for more control over sound loudness. One sound from the 6 available frequencies was presented while the binary colour choice was displayed on the screen. We randomised the sequence of the sounds to ensure the absence of any bias.

The turn-order experiment had only one pair of sounds (400 Hz and 1600 Hz) since the colour stimuli were also paired. The 400 Hz sound had a constant volume of 0.3 and was used as a reference to adjust the 1600 Hz sound loudness with an initial volume of 0.81. We set the step for volume adjustment as 0.2.

3.2 Colour stimuli

In both binary choice experiments (HSV and CIELCH) we presented colour stimuli as two circles of 2 cm in diameter and 4 cm distance from each other. In the time-order experiment, we only had one circle with a diameter of 8 cm. We set the visual angle to be 10-degree (44-48 cm from the display). We used an achromatic background with constant luminance to keep pupillary response consistent. We calibrated the display to ensure that the colour stimuli was shown as it was defined.

In both HSV and CIELCH binary-choice experiments the observers were supposed to choose a pair of colours that would be used to specify a tone between the two hues. We decided to use five stimuli since it was a small enough quantity to make the time of the experiment short to not exhaust the observers. Furthermore, the selection of the colours was enough to showcase most regions of the colour circle. It was important to select an odd number of stimuli to have a specific direction for colour narrowing. If the chosen number of colours was four and the observer chose two opposing hues, the results would be heavily affected by the decision of which side of the colour circle the hues average on.

The stimuli colours were selected to be situated 72 degrees from each other, dividing the 360-degree colour circle evenly into five hues. This allowed us to change the colours by only changing the hue value in both HSV (Figure 3.5) and CIELCH (Figure 3.6) colour spaces. Since hue was the only value that was changing, we could display hue change without the loss of saturation and chroma values.

The colours for binary-choice HSV and time-order experiments were sampled from an isoluminant plane in HSV colour space. The saturation and brightness of the HSV colours were set to maximum values during the experiment. HSV hue selection allowed us to display the colours typically as how they are used on the screen in most computer applications. If we changed the lightness of the HSV stimuli to be equal, the typical HSV yellow and blue could not be displayed.

The colours for binary-choice CIELCH and time-order experiments were sampled from an isoluminant plane in CIELCH colour space. The results of the colour range search in the CIELAB colour space are visualised in Figure 3.3. We set the L* as 80 to keep the yellow stimuli light enough to be classified as yellow by the observers. The limitation of the consistent CIELCH stimuli was that it is not possible to display equally saturated yellow and blue of the same lightness on the display because of its fit within the sRGB gamut. To keep the chroma equal, we took the colours on the same radius from the middle of the colour range map. The yellow from the available radius in $L* = 70$ was too dark to be recognised by the observers as yellow instead of brown. We decided to use desaturated colours because the main presentation medium that we used was a screen instead of other display methods with a larger gamut (e.g. lasers). Choosing a low value for the chromaticity of the stimuli allowed us to keep equal lightness for all of the hues while having the same saturation.

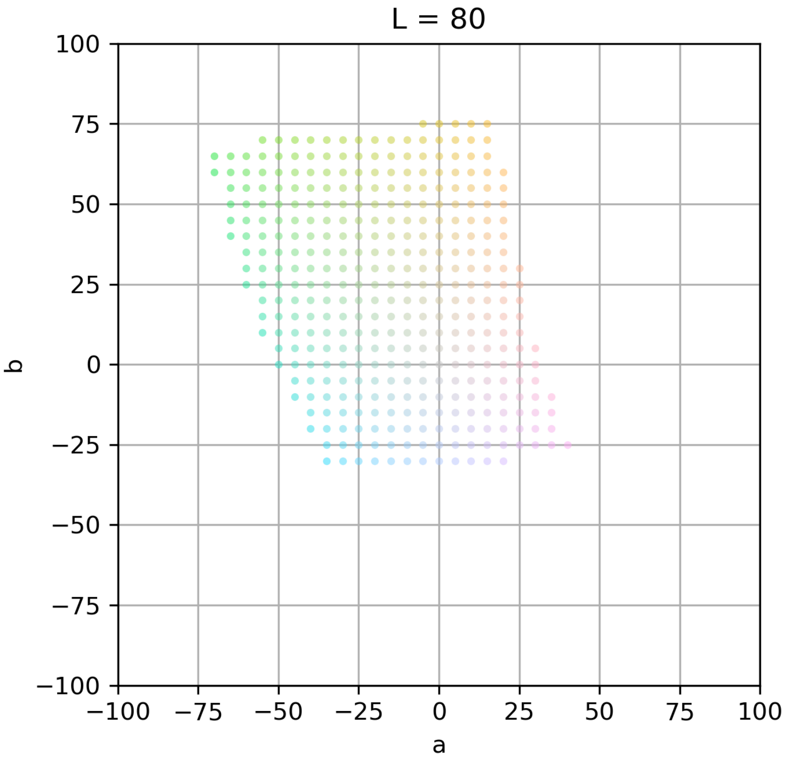


Figure 3.3: *Results of colour range search in CIELAB colour space with L* set as 80.*

3.3 Observers

We aimed for 18 observers for each of the experiments. The number of necessary observers was determined by performing repeated ANOVA within-between interactions with an effect size based on a preliminary test experiment. For the first experiment, we aimed to investigate, whether the language and culture of the observer will play a significant role in the results. Thus, the first experiment has a trial with both Norwegian and Japanese observers. All of the observers from Norway were students of the Norwegian University of Science and Technology in Gjøvik. All Japanese observers were students from the Toyohashi University of Technology. All observers had normal colour vision which was confirmed with the Ishihara Test for Color Blindness. Observers filled out a consent form that explained the use of their experimental results and personal information. You can see the information on the sex and average age of the observers in Table 3.1.

Table 3.1: *Observer sex and age information for all of the performed experiments. The range, mean value and standard deviation of ages are listed.*

Experiment	Female	Male	Ages [a]	μ_{age} [a]	σ_{age} [a]
Binary choice HSV Norway	7	11	20-32	22.8	2.8
Binary choice HSV Japan	9	9	19-24	22.3	1.5
Binary choice CIELCH Japan	5	13	21-25	22.8	1.2
Time-order Japan	5	13	21-25	22.3	1.2

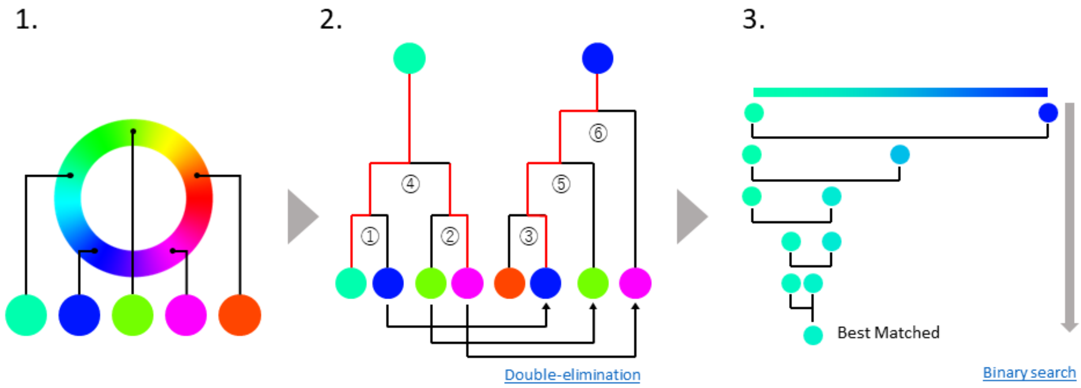


Figure 3.4: *The algorithm for the binary-choice experiment. The five hues from the colour circle were used in the double-elimination competition. A binary search was implemented for the two winning hues to determine the final hue.*

3.4 Binary choice experiment

We displayed the colour stimuli as two circles equally spaced from the centre of the display. They are shown to the observer in a randomised order using double-elimination, making six elimination pairs in total. In each of the elimination pairs, the observer chooses one of the two colours. The colour that the observer picked was saved for comparison with the colour that the observer picked next. The colour that was not picked initially was shown a second time in another comparison. As a result of the pair elimination, a pair of colours was chosen. The final shade was then specified using a four-step binary search on the two selected colours. The binary search results in the choice of a ratio between the two colours. This can either be one of the colours from the pair or a mix between the two. The algorithm for the hue determination is shown in Figure 3.4.

Before the experiment, observers indicated that they had normal or corrected-to-normal visual acuity and no hearing or colour vision deficiencies by filling out a consent form. The volume adjustment session started after the observer answered

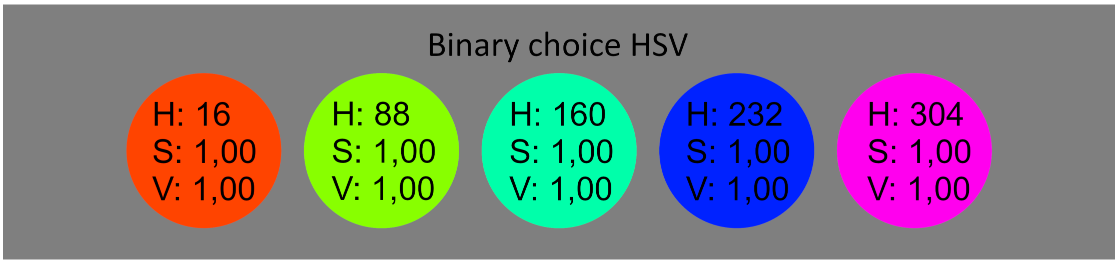


Figure 3.5: *The HSV stimuli for the binary choice experiment.*

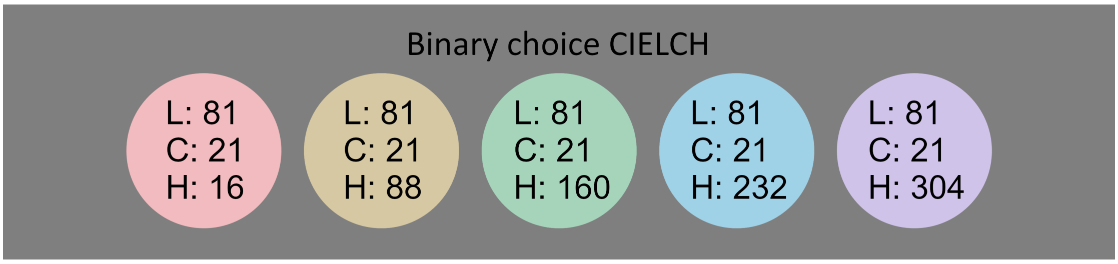


Figure 3.6: *The CIELCH stimuli for the binary choice experiment.*

the questions in the form. During the task, the six sound stimuli were given in random order. observers were required to adjust the sound stimulus (100 Hz, 200 Hz, 800 Hz, 1600 Hz, 3200 Hz) to be the same loudness level as the reference sound (400 Hz sample). The flow of volume adjustment is visualised in Figure 3.2. During the trial, observers could listen to the sound stimulus and reference sound anytime to compare the loudness and adjust depending on their perception. observers were forwarded to the colour adjustment section after they finished the volume adjustment.

Two colours were presented with one of the sound stimuli playing in the background. Observers were required to select the colour that better matched the stimulus sound. After choosing the colour, the next colour set was presented. This process was repeated until the best-matched colour was determined. The task took approximately 30 minutes to complete on average. By repeating the selection 11 times (six selections for determining two better-matched colours, and five selections for narrowing the colour range), observers get the best-matched colour with the sound. This flow is regarded as one trial as shown in Figure 3.7. The experiment consisted of five sets, which had six trials each. In total, observers select the colour in thirty trials (5 sets x 6 sounds = 30 trials). The sequence of colours and sounds was randomised to prevent accidental bias. Each trial ended in a break session.

We have decided to use the same design for both HSV and CIELCH binary-choice experiments. The main difference between the two experiments is the stimuli

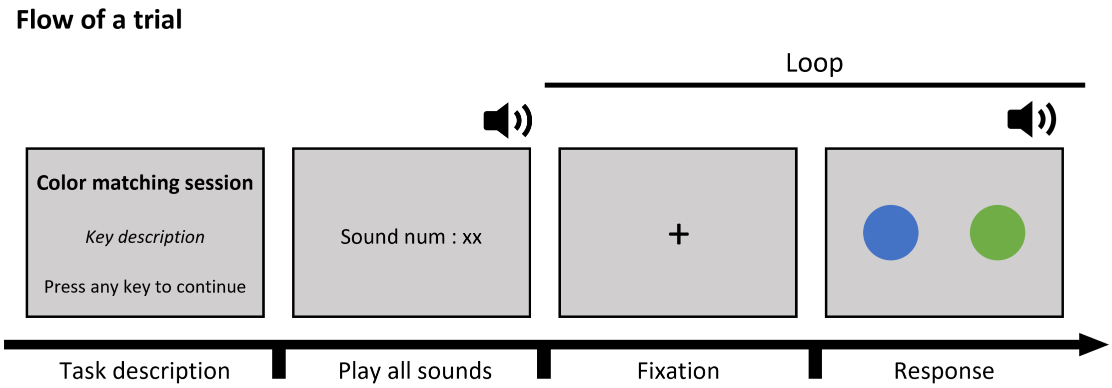


Figure 3.7: *Flow of the hue binary-choice experiments. This model was used for both HSV and CIELCH binary-choice experiments. The keys and the task were explained to the observer and then the randomised sequence of 5 sounds was presented. Each of the sounds was compared to the reference sound (400 Hz) like it was in the preliminary sound adjustment experiment (Figure 3.2). After the volume adjustment, the observer .*

set.

3.5 Time-order experiment

In this experiment, we wanted to see how closely the cross-modality between colour and sound is connected. The previous experiment had no time constraints thus giving time to observers to think and make a judgement. We know that the low pitch is associated with big objects (Parise and Spence, 2009) but the link between colours and sounds is not as clear. We designed a time-order experiment to check whether there is a real connection in the subconscious of the observers between the colour and pitch. To test the synesthetic congruence between the colours and sounds, the stimuli were shown rapidly so the observers did not have time to make a conscious judgement. Contrary to the binary-choice experiments, we used a headrest in the time order trials. This was important to eliminate unnecessary movement and display distance change to enhance focus since the experiment relied on the attention of the observers.

We based the stimulus onset asynchrony (SOA) of the time-order experiment on the paper by Parise and Spence (2009). The values for the SOA are 0 ms, 75 ms, 133 ms, 200 ms, 267 ms, 333 ms and 467 ms. The same values were used as negative for the SOA condition where the visual stimulus was presented after the audio and positive for the audio presented second. The trial was run 10 times for each of the SOA, except for 0 ms where we used 20 runs to account for both conditions. The sequence of the stimuli pairs and SOA was randomized to avoid accidental observer bias. On average, the experiment took 20 minutes to complete.

The time-order experiment took the visual and sound stimuli from the HSV binary-choice experiment. We used 400 Hz and 1600 Hz sound samples for auditory stimuli with an initial volume scale set as 0.3 and 0.81 respectively. After that, a small section with volume adjustment was added. We asked observers to match 1600 Hz volume to the constant volume of 400 Hz to accommodate for individual sound perception. The 1600 Hz was used as a high pitch sound while the 400 Hz was presented as a low pitch. We chose 400 Hz because it was different enough from the 1600 Hz sound but was not drastically different to distract the attention of the observer from the flow of the experiment. The experiment flow is shown in Figure 3.9.

We made two pairs of stimuli for the time-order experiment: chromatic and achromatic. Using the reasoning from binary-choice HSV experiment results, we picked a typical blue and yellow stimuli pair. We transformed the colours from HSV into CIELCH and set Chroma setting to zero to get the achromatic stimuli pair. In this way, we matched the perceived lightness between the chromatic (blue and yellow) and achromatic (dark and light grey) stimuli, leaving chroma as the main difference between the two sets of stimuli. We displayed the HSV and CIELCH parameters for both resulting pairs in Figure 3.8.

For this experiment, we label a pair of low-pitch and blue and dark grey stimuli as a synesthesically congruent cross-modal connection. A pair with high-pitch and yellow or light grey was also labelled as congruent. Consequently, the connection between the two groups was labelled as incongruent. The full list of all stimuli combinations can be found in Table 3.2.

Table 3.2: *Combinations of the visual and auditory stimuli in the time-order experiment. The display of the listing of conditions that are considered synesthetically congruent and incongruent in the experiment.*

	Chromatic		Achromatic	
Condition	Yellow	Blue	Light-grey	Dark-grey
400 Hz	Incongruent	Congruent	Incongruent	Congruent
1600 Hz	Congruent	Incongruent	Congruent	Incongruent

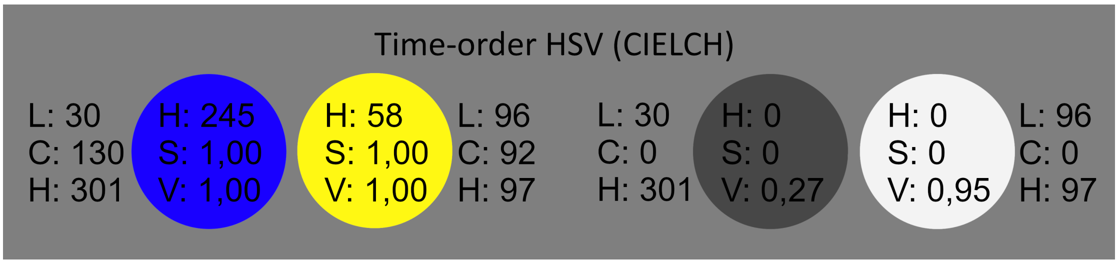


Figure 3.8: *The stimuli colours for chromatic (left) and achromatic (right) time-order experiments.*

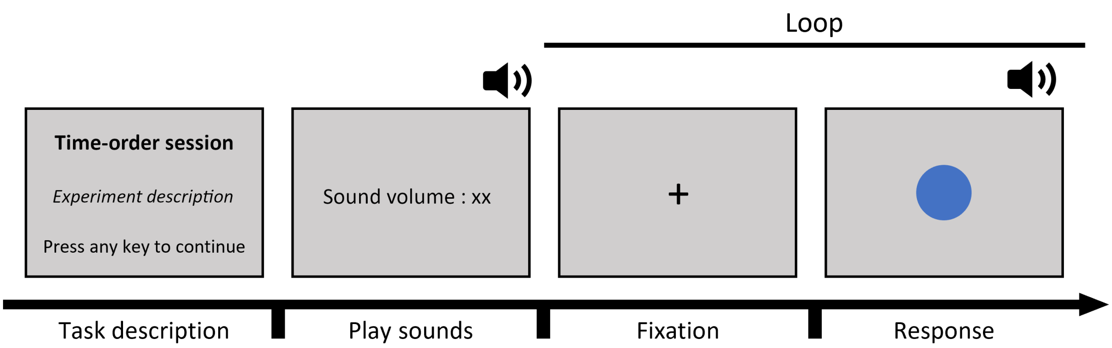


Figure 3.9: *Flow of the hue time-order part of the experiment. The keys and the description were shown first. This is followed by the adjustment of the 1600 Hz pitch to match the 400 Hz volume by the observer. Next, the loop is set to show visual and auditory stimuli with alternate order and timing.*

4 Results

All of the experiments presented in this work start from the volume adjustment section. The initial values of volume levels were adjusted by every observer individually before the start of each experiment, only having 400 Hz as a common fixed value for the purposes of having a reference sound. All six sounds were used in the two binary-choice experiments and only 400 Hz and 1600 Hz were used in the time-order experiment. The results for the volume adjustment were moderately consistent throughout all of the experiments. You can see the total average of all volume adjustment session results in Table 4.1. The results of the volume adjustment had a high degree of consistency, only having a major inter-observer disagreement in the 100 Hz pitch adjustment.

The noticeable jump in the volume in the 1600 Hz pitch frequency adjusted by observers is accordant with ISO 226:2003 equal-loudness contour curves (Suzuki and Takeshima, 2004), where the loudness in the region of 1000-2000 Hz is perceived as lower.

Table 4.1: *The volume of equal loudness levels averaged from the experiments.*

Frequency	Volume level
100 Hz	1.5931
200 Hz	0.5390
400 Hz	0.3000
800 Hz	0.3715
1600 Hz	0.6230
3200 Hz	0.3380

For each of the experiments, we performed the analysis of variance to determine the significance of the results. Analysis of the colour choices of the observers for all of the sound stimuli was performed for the two binary-choice experiments. The main focus of binary-choice experiments was to determine the average hues for each pitch and test the effect of observer consistency. The time-order experiment relies on the observers to choose in which order a cross-modal stimuli pair was presented. We compared the statistical results of the stimuli combinations to find the significance of the data. The p-value determines and tests the null hypothesis. The smaller the p-value, the more evident regularities are observed in the data. The value of $p < 0.05$ is generally considered statistically significant. We used the F-value to calculate the ratio of explained variance to unexplained variance. To have statistical significance, the value should be $F > 2.5$. The η_p^2 value is used in ANOVA models to measure the effect size of different variables. η_p^2 has a medium effect size when it is 0.06.

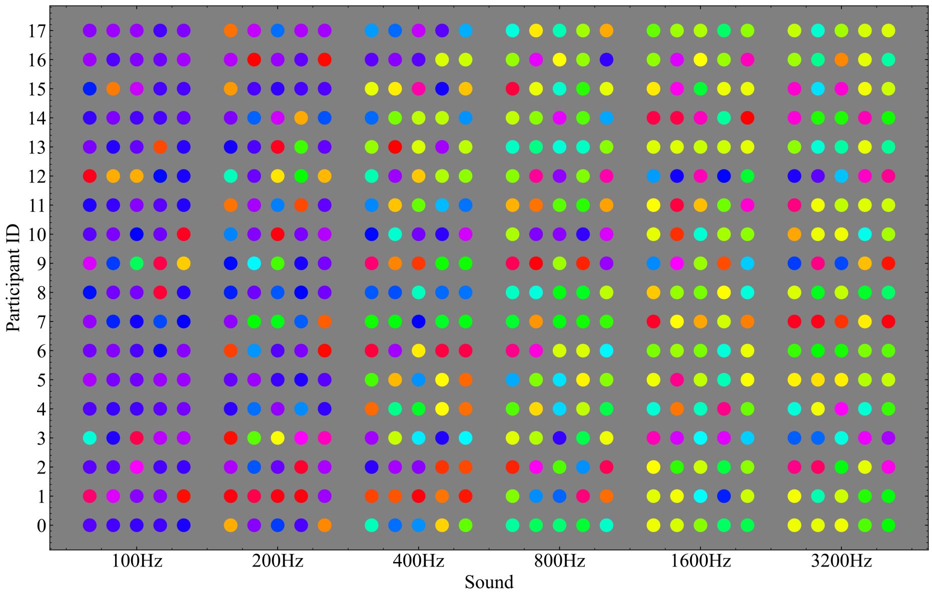
4.1 HSV Binary choice results

The results of the first binary choice experiment can be compared between two ethnic groups: Japanese and Norwegians. We visualised the results of all 18 observers by plotting the chosen HSV stimuli for each of the six sound stimuli. Each pitch had 5 repetitions to check for consistency in the observer's judgement. Both Japanese and Norwegian observer groups mainly chose blue-violet in 100-200 Hz and yellow-green in 800-3200 Hz which is visualised in Figure 4.1. It is hard to judge the predominant hue in the 400 Hz. The choice of red hue is inconsistent and scarcely exists in the whole range of the sound stimuli. In the oral survey, the observers reported that some pitches reminded them of an alarm that inclined them to choose a red hue.

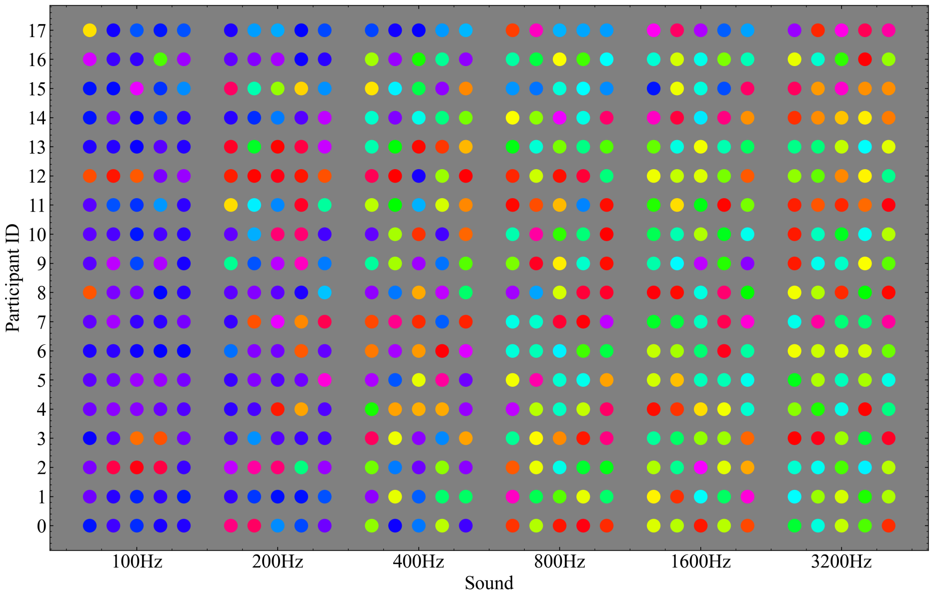
There are some intra-observer inconsistencies within the same pitch results, however, the overall inter-observer comparison shows a high level of agreement. We converted the HSV hue results to CIELAB to check for the consistencies within L*, a* and b*.

We compared the L* values for results from both Japan (Figure 4.2) and Norway (Figure 4.3) first. It should be noted that the results of the observers with 100 Hz were the most consistent in both countries, having a small fluctuation. Since L* is used for expressing perceptual lightness, we can see that the lowest chosen lightness is at 100 Hz with the L* value in the average value of 40-50. 300 Hz had less consistency then 100 Hz with the results in $L^* < 60$. The observers had the least consistency with the L* value in 400 Hz. The results of 800 Hz, 1600 Hz and 3200 Hz did not differ considerably, however, the observers in Japan had less consistency with the result of 3200 Hz than the Norwegian. It should be noted that there are no stimuli below the value of 30 in L* since we used the HSV colours with the setting of 100 in both saturation and value. Therefore, even though the L* varies in the provided stimuli, the value does not become lower than 32.

The results of a* analysis which is displayed in Figures 4.4 and 4.5 show an even bigger similarity between the Japanese and Norwegian observers. The 100 Hz and 200 Hz are the highest in a* value. 400 Hz marks a rapid drop in both the average value and consistency which means that starting from 400 Hz. This means



(a) *Japanese results.*



(b) *Norwegian results.*

Figure 4.1: *Japanese (a) and Norwegian (b) results of the 18 participants in the HSV binary-choice experiment for the range of 100-3200 Hz frequencies. In both visualisations the divide between blue-violet and yellow-green in the 400 Hz is noticeable.*

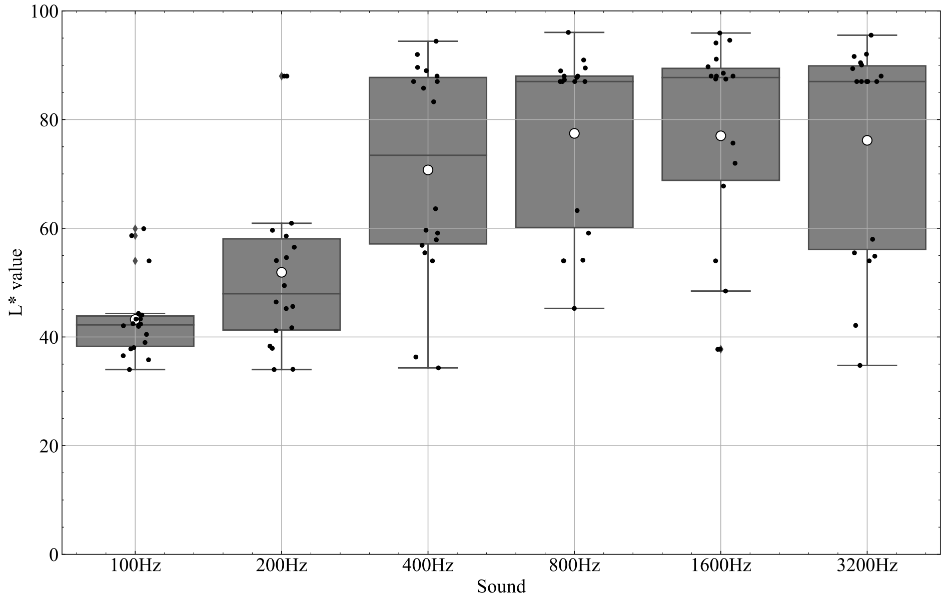


Figure 4.2: *Japanese L* values of binary-choice HSV experiment results.*

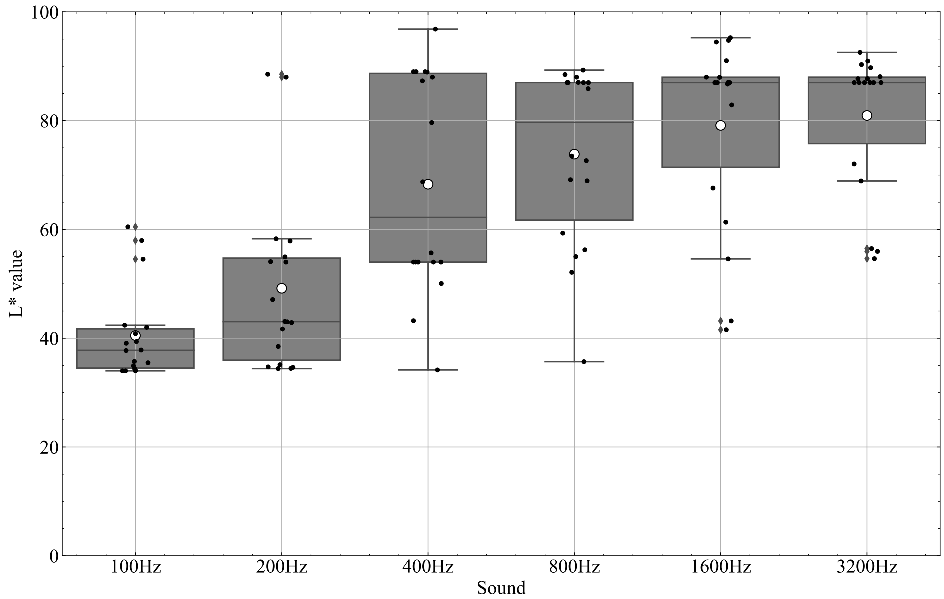


Figure 4.3: *Norwegian L* values of binary-choice HSV experiment results.*

that the observer response in the 100-200 Hz had the least a* value, meaning the stimuli that were chosen had more red than green. Similar to the L*, the chosen stimuli in the 3200 Hz are less consistent in a* with Japanese than Norwegians.

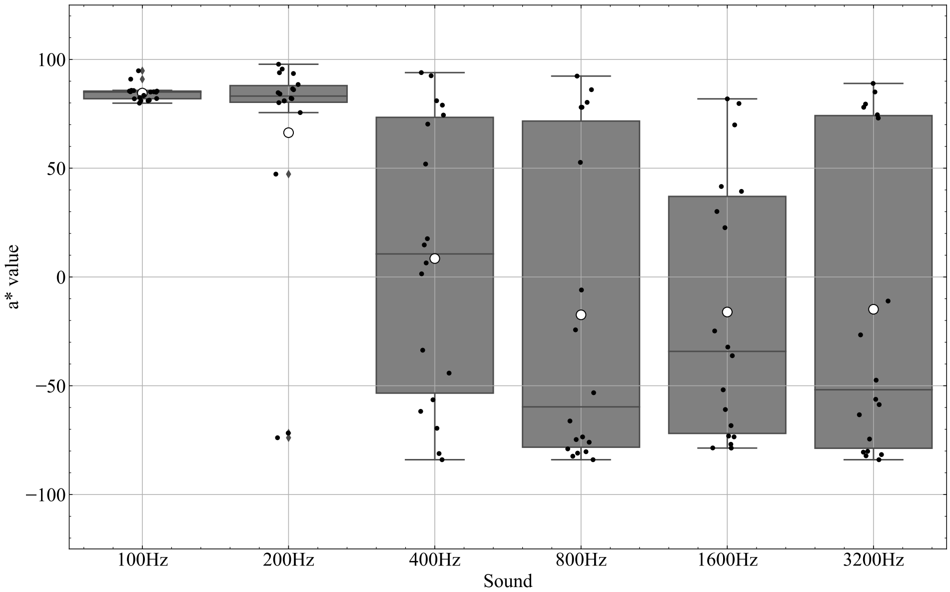


Figure 4.4: *Japanese a* values of binary-choice HSV experiment results.*

Averaged b* values show us more information about the violet-blue and green-yellow dichotomy which can be seen in Figures 4.6 and 4.7. There are several outliers, but the average results of 100 Hz and 200 Hz are under -50 in b*, showing a strong incline towards blue-coloured stimuli. We measured that 400 Hz stimuli are the least consistent with b*, however, the majority of the responses place the average over 0, which shows a tendency for yellow. The stimuli at 800 Hz, 1600 Hz and 3200 Hz show the average results to be higher than 50 which means that the stimuli have a high value in yellow. The pitch of 200 Hz in b* shows more consistent answers with the Japanese observers. However, in the 400 Hz the Norwegian observers are more consistent. The average values for both countries are otherwise similar within the b* analysis.

The analysis of variance showed the results for Japanese (Table 4.2) and Norwegian (Table 4.3) binary-choice HSV data adjusted by Greenhouse-Geisser's Epsilon. The F-test for Japanese data resulted in L* having the most significance and certainty with p-value ($F_{L*}(2.3, 39.11) = 33.1354$, $p < 0.005$, $\eta_p^2 = 0.6609$). The effect size (η_p^2) for the L* value in Japanese data is also quite high. In the Norwegian data, the L* and b* values are very close in significance. The L* main effect ($F_{L*}(2.81, 47.75) = 47.2295$, $p < 0.005$, $\eta_p^2 = 0.7353$) is only comparable to

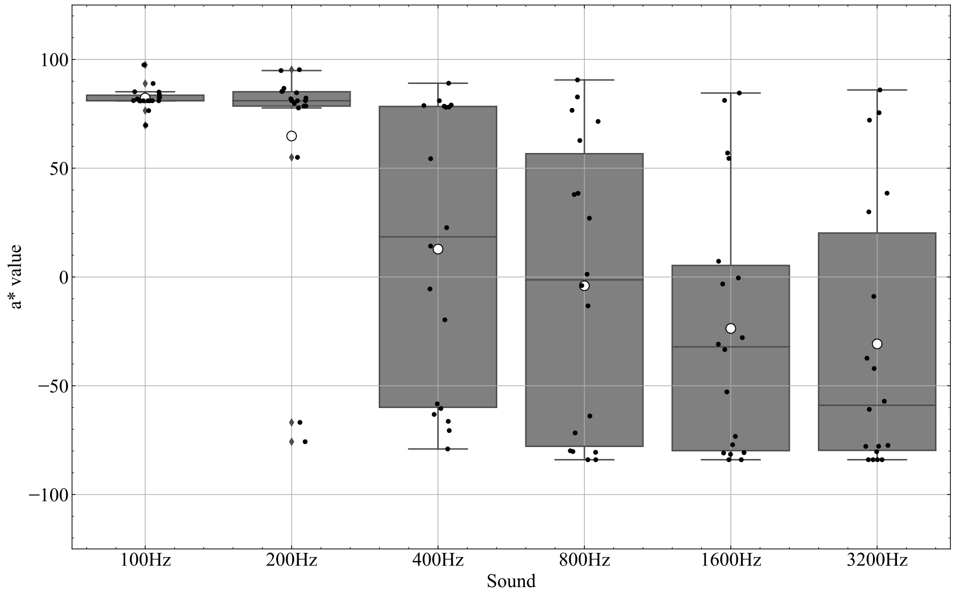


Figure 4.5: *Norwegian a* values of binary-choice HSV experiment results.*

b* ($F_{L*}(2.58, 43.94) = 45.8519$, $p < 0.005$, $\eta_p^2 = 0.7295$). This is quite a notable difference in the main effect values between the two countries, but they both show strong significance in the results.

Table 4.2: *The ANOVA results of the Japanese observer data.*

Main effect	F	p	η_p^2
L* value	33.1354	0.0000	0.6609
a* value	22.1114	0.0000	0.5653
b* value	23.3953	0.0000	0.5792

Table 4.3: *The ANOVA results of the Norwegian observer data.*

Main effect	F	p	η_p^2
L* value	47.2295	0.0000	0.7353
a* value	23.3810	0.0000	0.5790
b* value	45.8519	0.0000	0.7295

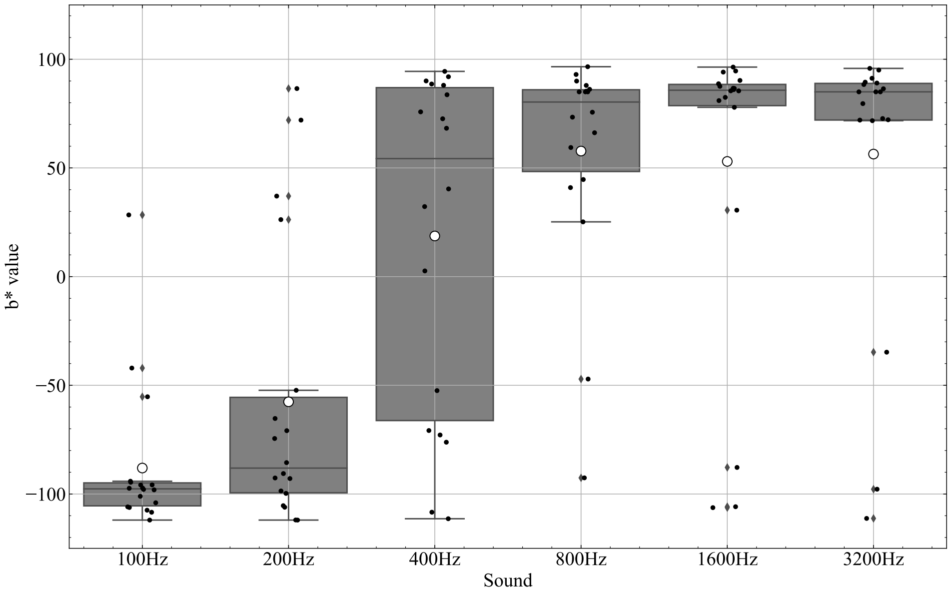


Figure 4.6: *Japanese b* values of binary-choice HSV experiment results.*

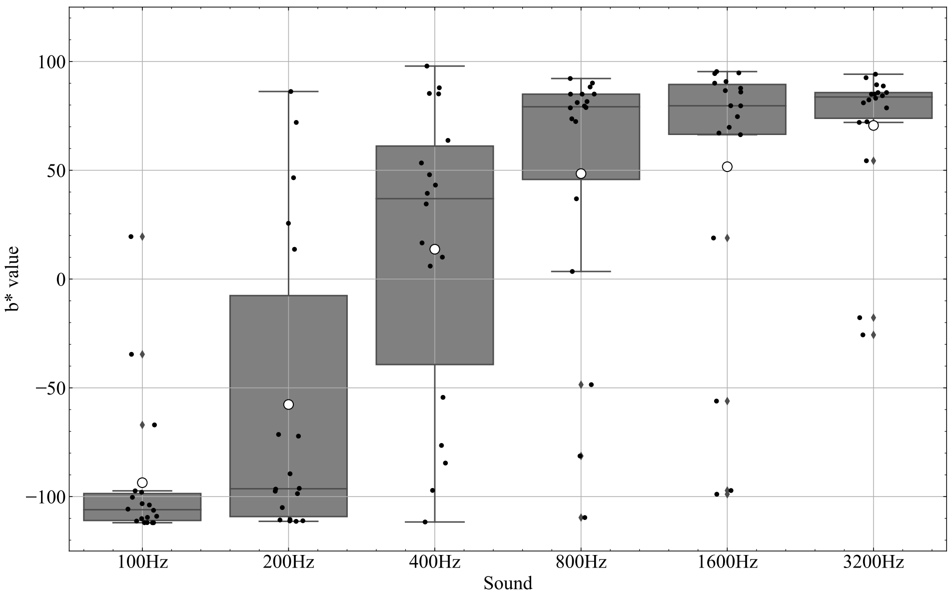


Figure 4.7: *Norvegian b* values of binary-choice HSV experiment results.*

4.2 CIELCH Binary-choice results

The stimuli change in the CIELCH binary-choice experiment significantly influenced the results (Figure 4.8). Unlike the results in the HSV experiment, in CIELCH we do not observe any universal tendencies towards choosing specific hues for a range of pitches. Some of the observers choose mostly blue hues for 100-200 Hz while the others choose blue for 800 Hz 1600 Hz and 3200 Hz. Since we kept L* consistent for all of the stimuli in the CIELCH experiment, is necessary to see examine a* and b* value statistics.

It should be noted that since the stimuli are kept consistent in L* value and are fixed in saturation, the values for both a* and b* are kept under 21 in both positive and negative values. The y-axis of the graphs was changed to have a better illustration of the difference in the observer response.

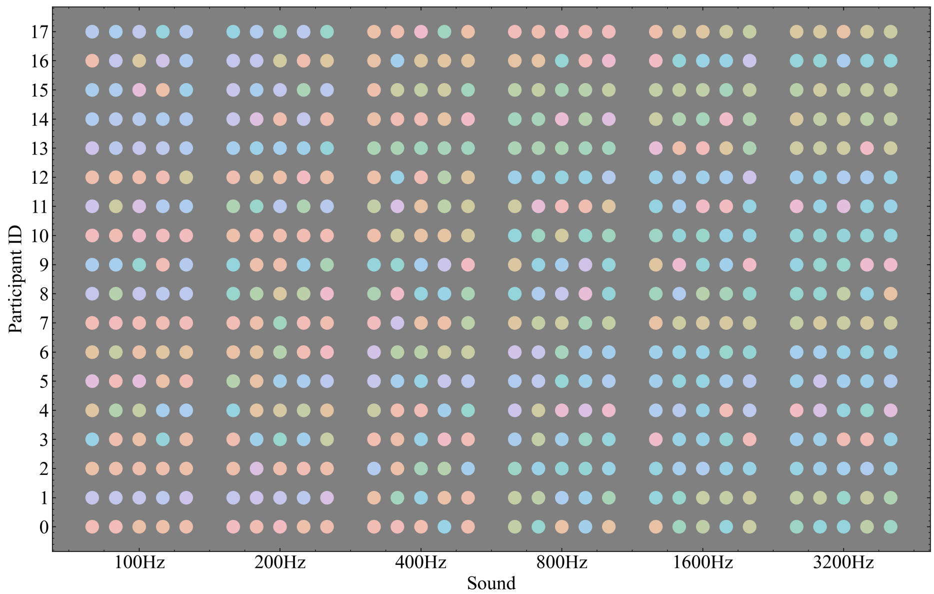


Figure 4.8: *Cross-modal CIELCH binary choice experiment colour results. The most noticeable is that observers chose red and violet hues for 100-200 Hz with blue and green hues in 800-3200 Hz.*

In the a* (Figure 4.9), the observers chose mostly positive values in the 100-400 Hz resulting in more red-tinted hues, while the chosen colours for 800-3200 Hz were in the negative a* resulting in green and yellow hues. We found the least inter-observer agreement in 200 Hz hue choice.

The observers mostly chose the stimuli with the b* (Figure 4.10) value of 10 and more in the 400 Hz pitch which is the most yellow stimuli among all of the sounds. The 800 Hz average hue was situated on 0 b*, but 1600 Hz and 3200 Hz tended to have negative values. All of the pitches except for the 400 Hz have inconsistent results which implies a low significance of b* in the total hue choice in the experiments with consistent L*.

We analysed the CIELCH data with ANOVA and adjusted it by Greenhouse-Geisser's Epsilon (Table 4.4). This experiment does not rely on the L* results since the lightness is constant for all of the hue stimuli in the CIELCH experiment. Contrary to the HSV experiment result, the main effect was found in the a* ($F_{L*}(2.43, 41.35) = 6.9148$, $p < 0.005$, $\eta_p^2 = 0.2891$). Though it is drastically smaller than the effect in the previous experiment, it is still significant.

Between the a* and b* for the CIELCH experiment, the a* has more influence on the stimuli choice. Unlike the experiment with HSV stimuli, b* does not have a significant enough influence and L* analysis is not applicable since it is used as a constant.

Table 4.4: *The ANOVA results of CIELCH observer data.*

Main effect	F	p	η_p^2
a* value	6.9148	0.0014	0.2891
b* value	1.0516	0.3477	0.0583

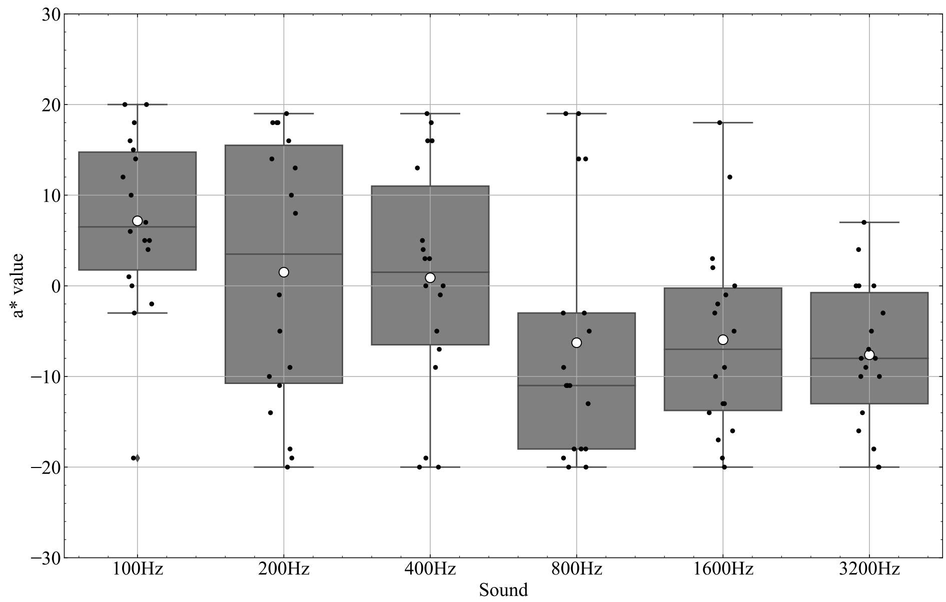


Figure 4.9: *CIELCH Binary choice a* results.*

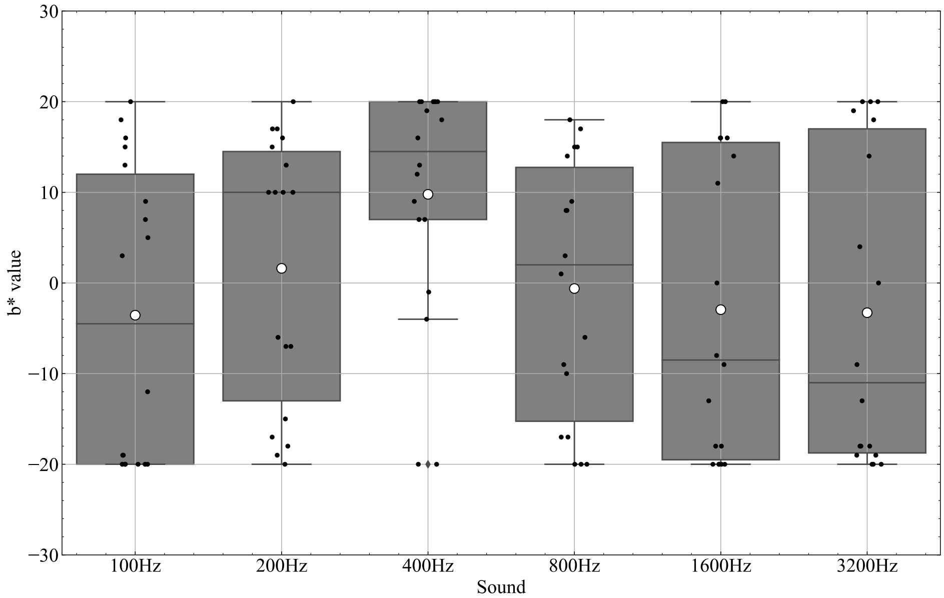


Figure 4.10: *CIELCH Binary choice b* results.*

4.3 Time-order results

The time-order experiment is different from the binary-choice HSV and CIELCH experiments because we analyse the success rate of the observers in guessing the order of the two types of stimuli instead of the observer consciously choosing a colour that fits the sound the most in their subjective opinion. We can extract information on the hue influence by comparing the chromatic and achromatic stimuli pair order assignment success rate. This means that even when the observer is guessing randomly, there typically should be a 50% chance of being right unless there were other factors that influenced the results.

We have conducted an extensive analysis of the time-order experiment data to find the significance by means of the ANOVA. We used different combinations of the data to test every significance that could be important 4.5. The main focus was to use repeated measures to find the Point of Subjective Equality (PSE) and Just-Noticeable Difference (JND) for the observer results. PSE corresponds to the SOA in which auditory and visual stimuli are perceived at the same time.

We performed repeated measures ANOVA 2x4 to find the PSE of the two chromacity variations, achromatic and chromatic, for four conditions: incongruent-high, incongruent-low, congruent-low, and congruent-high. We tested all of the combinations of the comparisons possible, with the adjustment of the p-value for comparing a family of 28. Interaction in the chromatic combinations was significant

$p = 0.006$. However, the post hoc tests revealed no significance in this case. The smallest p-value was for "blue-yellow, incongruent-low" and "dark-light, incongruent-low" interactions, which was $p = 0.164$. This means that the interaction of yellow and light visual stimuli with the low pitch was $p = 0.164$ which is not significant.

For the repeated measures ANOVA just for blue and yellow, using 2 chromatic visual stimuli (blue, yellow) and 2 sounds (low, high) there was no interaction, and no main effect was found. Repeated measures ANOVA for dark and light only, 2 achromatic visual stimuli (light, dark) and 2 sounds (low, high). In this case, the interaction was as significant as $p < 0.007$. The interaction between "dark, low" and "light, low" was $p < 0.022$

Additional repeated measures were performed with ANOVA 2x4 to find the JND of the two chromaticity variations (achromatic, chromatic) for four conditions (incongruent-high, incongruent-low, congruent-low, congruent-high).

Post Hoc Tests P-value adjusted for comparing a family of 6 revealed no significance The smallest p was for "congruent-low" vs "congruent-high", which was p=0.065. This means that if a stimulus was Congruent, it may have a chance to enhance the discrimination between low- and high-paired stimuli.

We analysed the repeated measures ANOVA just for the case with two chromatic colours (blue, yellow) and two pitch conditions (low, high). The main effect was found in Color, $p = 0.041$. Yellow had reduced JND compared to the Blue condition.

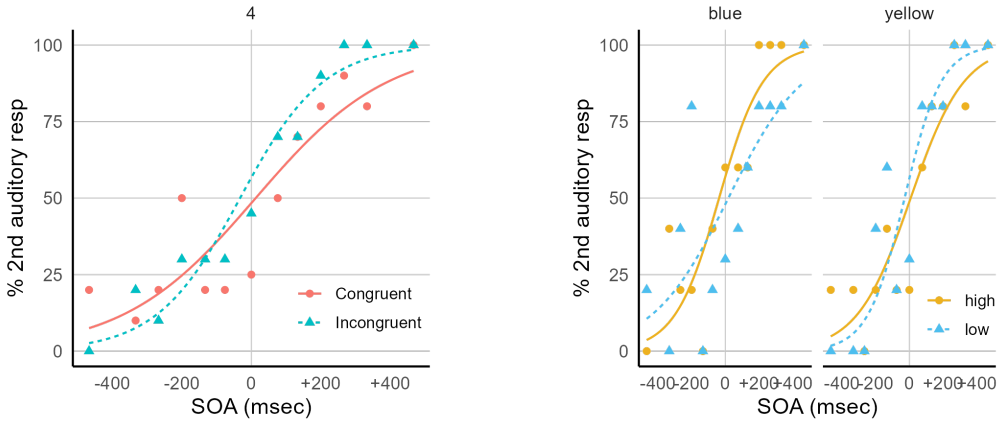
The main effect was found in Color, $p = 0.036$ was found in repeated measures ANOVA just for the case with two achromatic colours (light, dark) and two pitch conditions (low, high). For the achromatic visual stimuli, the light condition JND was reduced compared to the dark.

Table 4.5: *Time-order results of the p-values for the two-way and three-way ANOVA.*

ANOVA Combinations	p
PSE 2 Chromaticity x 4 Conditions	0.006
JND 2 Chromaticity x 4 Conditions	0.344
PSE 2 Chromaticity x 2 Congruence	0.008
JND 2 Chromaticity x 2 Congruence	0.191
JND 2 Chromaticity x 2 Lightness x 2 Sound	0.191

The observers tend to perceive the visual stimuli to be second when the actual chance of it being second is 50% from the PSE curves (Figure 4.11) of the participants. This means that the observers perceive the sound coming second when the stimuli are presented simultaneously.

In the repeated measures ANOVA for two experiment chromaticities (achromatic, chromatic) and congruence labels (congruent, incongruent) the PSE has



(a) *Congruent-Incongruent PSE curve.*

(b) *Blue and yellow PSE curves.*

Figure 4.11: *The results of the observer four with congruent-incongruent curve (b) and separated blue-yellow curves (a). The SOA* $= 0$ *ms in the graph is where the auditory and visual stimuli are shown simultaneously. The SOA in* $[-400; 0]$ *is for the visual stimuli coming first, whereas* $[0; 400]$ *is when the audio is presented second. In Figure (a), the blue-low and yellow-high conditions are labelled as congruent.*

an interactive. Post hoc tests for light-dark with congruent and light-dark with incongruent conditions in Figure 4.12. The $p = 0.043$ shows significant interaction in the light-dark grey congruent-incongruent conditions.

We performed a three-way JND ANOVA (Figure 4.13) for two experiment chromaticities (chromatic, achromatic), with two lightness settings (dark, light) and two pitches (high, low).

Within Subjects Effects, lightness was found to be $p = 0.009$. It was found significant in the Post Hoc Comparisons as the lightness was set as $p < 0.01$.

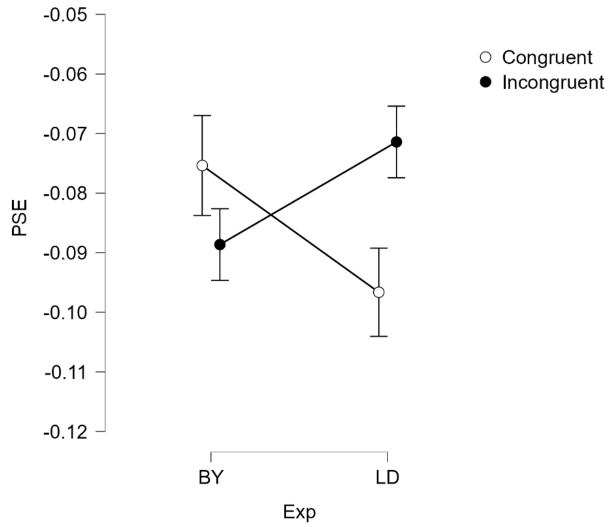
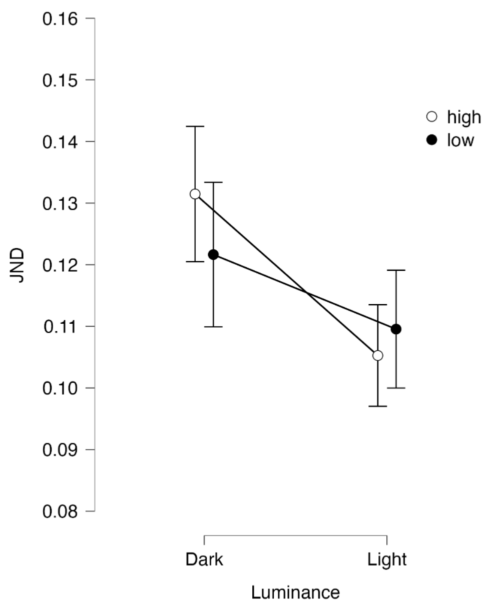
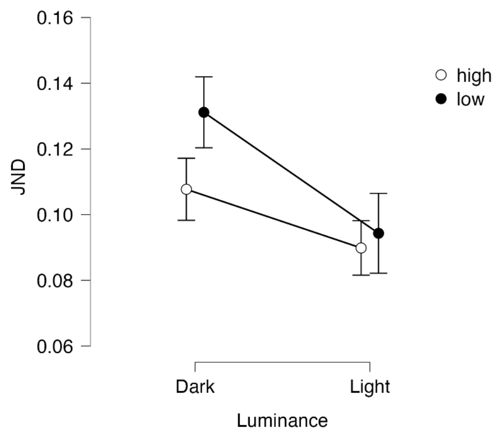


Figure 4.12: *PSE Descriptive mean values for 2 experiment chromaticities (achromatic, chromatic) and congruence labels (congruent, incongruent). The interaction is significant between light-dark congruent and light-dark incongruent conditions*



(a) *chromatic x 2 lightness (dark, light) x 2 sound (high, low).*



(b) *achromatic x 2 lightness (dark, light) x 2 sounds (high, low).*

Figure 4.13: *JND Descriptive mean values for 2 chromaticities (Chromatic(a), Achromatic(b)), with two Lightness settings (dark, light) and two pitches (high, low).*

5 | Discussion

In this study, we investigated how hue and pitch correspond in the field of cross-modality. We conducted three experiments and analysed the data using ANOVA. The first experiment we conducted was the HSV binary choice with participants from Japan and Norway. Since the first experiment stimuli relied on the lightness of the stimuli, the second experiment was designed to use CIELCH stimuli with a similar design as the first one. The average results of all 18 observers in each experiment are presented in Figure 5.1. The third experiment was a time-order one. We used a pair of stimuli based on the first HSV experiment. The observer results of the sound adjustment in all three experiments were consistent with the ISO 226:2003 equal-loudness contour curves (Suzuki and Takeshima, 2004).

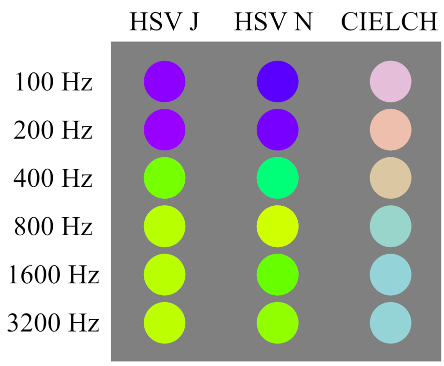


Figure 5.1: *The averaged results for all three binary choice experiments: HSV Japan, HSV Norway, and CIELCH. It should be noted that the hue results for HSV Norway look the least certain while the statistical results of the experiment show the most significance.*

5.1 Binary-choice discussion

The initial design of the HSV binary-choice experiment built on the previously-established idea of blue colour association with a low-pitch, and yellow - to high-pitch sound. Additionally, we wanted to investigate whether two groups of students from Japanese and Norwegian universities would have different results.

The colour division in the HSV binary-choice experiment is distinct with the 100-200 Hz frequencies having mostly blue-violet stimuli while 800-3200 Hz was classified by the observers as yellow-green. An important reason for this result is the difference in the perceived lightness of the HSV colour. The response of observers from both countries seems to be in consensus on the blue-yellow dichotomy.

The choice of red throughout the pitch range was explained by observers in the questionnaire answers as being associated with alarm, though the colour does not only appear in the higher frequencies as it would be expected.

It should be noted that 400 Hz has the least response consistency out of all sound stimuli in the binary-choice experiments. One reason could be because it is a sound that is situated around the middle of the given range (100-3200 Hz), not classifiable as either high or low pitch. Another idea is that the use of the pitch as a reference sound influenced the way observers perceive the auditory stimuli. Observers adjusted the loudness of the pitch stimuli having a constant volume of 400 Hz which could have influenced the inconsistency in the chosen hues.

The difference in lightness of the stimuli had a considerable impact on the results of the HSV binary-choice experiment. It is necessary to compare the two experiments since the results of binary choice HSV are more influenced by the lightness of the visual stimuli than the hue itself. When compared to the results of the binary-choice CIELAB experiment where the lightness was a constant, L* in the HSV experiment was the most important value when deciding the hue-pitch choices of the observers. Judging by the HSV experiment, the b* value should be influential as well. The a* and b* carried less significance than L*, but the b* was the second-most significant value. However, in the case of equiluminant stimuli of the CIELCH experiment where the L* was not influencing the decision, a* had the main significance ($F_{L*}(2.43, 41.35) = 6.9148$, $p < 0.005$, $\eta_p^2 = 0.2891$).

5.2 Time-order discussion

The results of the chromatic and achromatic time-order experiments do not show a strong connection between the luminosity of the stimuli and the pitch of the presented sound. This suggests that the connection in the colour-pitch pairs of blue-low and yellow-high is not an unconscious mechanism, but instead, a learned behaviour. Instead, the data had more interaction for the dark and light grey stimuli which implies the congruence for dark with low pitch and light with high pitch. Additionally, the observers had a tendency to perceive the sound coming second when the stimuli were presented simultaneously. This result is consistent with the find in the study by Uno and Yokosawa (2022).

5.3 General discussion

According to the questionnaire results that the observers had to complete after the test, the bright frequencies were labelled as "loud". Some of the consistent associations were pairs of lightness and loudness, darkness and quietness. This corresponds with both HSV binary-choice and time-order experiments since they both used HSV colour space where the variable of lightness is changing. The perception of yellow as lighter is naturally dependent on the luminous efficiency functions.

We tested the observers from different cultures during the HSV binary-choice experiment. The results showed consistency between the two observer groups, even in the inconsistent incorporation of red stimuli across the frequencies. Since the CIELCH binary-choice experiment had completely different colour stimuli, a straightforward comparison is impossible. On the other hand, the difference in observer choice of hues is almost the opposite of the first experiment. The difference between the chosen colours of the HSV and CIELCH binary-choice experiments seems like a shift on a colour circle in terms of hue (Figure 5.1). The results of the HSV experiment could be explained by the apparent difference in the luminance of the stimuli, but the CIELCH results require further discussion.

It should be noted that the results in Figure 5.1 are averaged and do not show the full interaction of the observer data because of small statistical significance. The time-order experiment confirmed that there is little interaction between hue and pitch on an unconscious level. Subsequently, it can be inferred that the hue-pitch correspondence could be a cultural phenomenon which needs further investigation. The congruence in lightness and pitch was confirmed to be the determining factor of the cross-modal hue-pitch association.

5.4 Limitations

Firstly, the binary-choice experiments do not account for the fact that the results can be influenced by personal bias: associations, experiences and culture. Since the trials do not rely on reaction and do not apply harsh time constraints, the observer could instead of choosing the stimuli based on the inherent congruence of the cross-modal pair.

Second, the environment could not be strictly controlled for international trials of the binary-choice HSV experiment due to the change in locations (Norway and Japan). The change in the environment might influence the results inconspicuously which would not be possible with a stationary setup.

Third, it is possible that the attention levels of the observers were influenced by external factors (e.g. tiredness, incidental distraction) which could drastically impact the time-order experiment results. The experiments were specifically designed to be as short as possible not to tire the observers, but the factors outside of the experiment scope could also drastically influence observer performance during the trials.

6 Conclusions

We tested both HSV and CIELCH colour spaces in the binary-choice experiment to compare the influences of the lightness on the correspondence with 6 pure-pitch frequencies. In the paired choice test, we asked observers to choose a hue that corresponded the most to the presented sound stimuli.

The results concluded that lightness played the primary role in the HSV experiment based on the L* statistics. Moreover, because of the inherent connection of the HSV colours to the luminance, it resulted in a strong blue-yellow dichotomy on the b* axis. The data for Japanese and Norwegian observer groups showed similar results.

The CIELCH binary-choice experiment, where the stimuli were isoluminant, shows opposite results. Due to the lack of lightness effect, the response in the dichotomy of the answers shifted to the a* value. The ANOVA for CIELCH results showed a tendency in observers to choose a red hue in the low frequencies and a green hue in the high frequencies.

Since the binary-choice experiment provided time for deciding what colour suited the sound most, we designed an experiment that does not rely on the personal judgement of the observer. The results from the time-order experiment imply that lightness has the most significance when it comes to underlying cross-modal congruence. If there was any influence from chroma, it was not big enough to be significant.

6.1 Future work

Since only the HSV binary-choice experiment was performed in both Japan and Norway, it is suggested to investigate the cultural differences in the observers of different ethnicity with equiluminant colour stimuli.

The previous discussion of colour-pitch cross-modality mainly depends on yellow-blue correspondences with high and low pitch. Based on the results of the CIELCH binary-choice experiment, additional investigation of a* influence in equiluminant green-red stimuli cross-modality should be looked into.

The stimuli used in the current work were simple sine wave pitches of different frequencies. Such sounds are hard to find in nature so that might be influential in getting more varied results. The natural sounds of voices, bells, birds and thunder could be used as stimuli to have more complex auditory stimuli compared to simplistic pitch frequencies. Similarly, sets of colour hues could be substituted with pictures of trees, animals and objects to serve as visual stimuli for testing cross-modality. The visual stimuli could be changed to another display method that has a larger gamut than one of a screen. Moreover, the cross-modal temporal perception of the auditory and visual stimuli could be explored further based on the delay in sound perception.

In this work, we tested observers with normal colour vision observers and the absence of synesthesia. There could be different results in testing different groups of observers. In the future, similar experiments could be conducted with musically-inclined observers or individuals with synesthesia.

Bibliography

Anikin, A. and Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception, & Psychophysics*, 81:764–777. (cited on pages 3, 11, and 12)

Biocca, F., Kim, J., and Choi, Y. (2001). Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators & Virtual Environments*, 10(3):247–265. (cited on page 2)

Brang, D., Teuscher, U., Ramachandran, V. S., and Coulson, S. (2010). Temporal sequences, synesthetic mappings, and cultural biases: The geography of time. *Consciousness and Cognition*, 19(1):311–320. (cited on page 3)

Brown, A. and Hopkins, H. (1967). Interaction of the auditory and visual sensory modalities. *The Journal of the Acoustical Society of America*, 41(1):1–6. (cited on page 1)

Castro-Alamancos, M. A. (2009). Cortical up and activated states: implications for sensory information processing. *The Neuroscientist*, 15(6):625–634. (cited on page 4)

Chrea, C., Valentin, D., Sulmont-Rossé, C., Mai, H. L., Nguyen, D. H., and Abdi, H. (2004). Culture and odor categorization: agreement between cultures depends upon the odors. *Food quality and preference*, 15(7-8):669–679. (cited on page 2)

Dominy, N. J. (2004). Fruits, fingers, and fermentation: the sensory cues available to foraging primates. *Integrative and comparative biology*, 44(4):295–303. (cited on page 2)

Driver, J. and Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current biology*, 10(20):R731–R735. (cited on page 1)

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of vision*, 10(1):6–6. (cited on page 3)

Fu, D., Barros, P., Parisi, G. I., Wu, H., Magg, S., Liu, X., and Wermter, S. (2018). Assessing the contribution of semantic congruency to multisensory integration and conflict resolution. *arXiv preprint arXiv:1810.06748*. (cited on page 9)

Fu, D., Weber, C., Yang, G., Kerzel, M., Nan, W., Barros, P., Wu, H., Liu, X., and Wermter, S. (2020). What can computational models learn from human selective attention? a review from an audiovisual unimodal and crossmodal perspective. *Frontiers in Integrative Neuroscience*, 14:10. (cited on pages 9 and 10)

Grassi, M. (2005). Do we hear size or sound? balls dropped on plates. *Perception & psychophysics*, 67:274–284. (cited on page 2)

Hamilton-Fletcher, G., Witzel, C., Reby, D., and Ward, J. (2017). Sound properties associated with equiluminant colours. *Multisensory Research*, 30(3-5):337–362. (cited on pages 3, 5, 12, and 16)

Ho, H.-N., Van Doorn, G. H., Kawabe, T., Watanabe, J., and Spence, C. (2014). Colour-temperature correspondences: when reactions to thermal stimuli are influenced by colour. *PloS one*, 9(3):e91854. (cited on page 2)

Holmes, N. P., Calvert, G. A., and Spence, C. (2007). Tool use changes multisensory interactions in seconds: evidence from the crossmodal congruency task. *Experimental Brain Research*, 183:465–476. (cited on page 3)

Hutchings, J. B. (2011). *Food colour and appearance*. Springer Science & Business Media. (cited on page 2)

Itoh, K., Sakata, H., Kwee, I. L., and Nakada, T. (2017). Musical pitch classes have rainbow hues in pitch class-color synesthesia. *Scientific reports*, 7(1):17781. (cited on page 8)

Johansson, N., Anikin, A., and Aseyev, N. (2020). Color sound symbolism in natural languages. *Language and Cognition*, 12(1):56–83. (cited on page 9)

Kim, Y.-J. (2013). Can eyes smell? cross-modal correspondences between color hue-tone and fragrance family. *Color Research & Application*, 38(2):139–156. (cited on page 1)

Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134(3):372–384. (cited on page 7)

Köhler, W. (1947). Gestalt psychology (1929) new york. *NY: Liveright*. (cited on page 11)

Lam, C. K. C., Yang, H., Yang, X., Liu, J., Ou, C., Cui, S., Kong, X., and Hang, J. (2020). Cross-modal effects of thermal and visual conditions on outdoor thermal and visual comfort perception. *Building and Environment*, 186:107297. (cited on page 2)

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental brain research*, 158:405–414. (cited on page 3)

Levitan, C. A., Ren, J., Woods, A. T., Boesveldt, S., Chan, J. S., McKenzie, K. J., Dodson, M., Levin, J. A., Leong, C. X., and Van den Bosch, J. J. (2014). Cross-cultural color-odor associations. *PloS one*, 9(7):e101651. (cited on page 8)

Likova, L. (2014). Learning-based cross-modal plasticity in the human brain: insights from visual deprivation fmri. *Advanced Brain Neuroimaging Topics in Health and Disease-Methods and Applications*, 2014:327–58. (cited on page 1)

Likova, L. T. (2015). Temporal evolution of brain reorganization under cross-modal training: insights into the functional architecture of encoding and retrieval networks. In *Human Vision and Electronic Imaging XX*, volume 9394, pages 406–420. SPIE. (cited on page 1)

Lin, A., Scheller, M., Feng, F., Proulx, M. J., and Metatla, O. (2021). Feeling colours: Crossmodal correspondences between tangible 3d objects, colours and emotions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12. (cited on page 11)

MacDougall, D. (2002). *Colour in food: improving quality*. Woodhead Publishing. (cited on page 2)

Martino, G. and Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, 10(2):61–65. (cited on page 3)

Melin, A. D., Nevo, O., Shirasu, M., Williamson, R. E., Garrett, E. C., Endo, M., Sakurai, K., Matsushita, Y., Touhara, K., and Kawamura, S. (2019). Fruit scent and observer colour vision shape food-selection strategies in wild capuchin monkeys. *Nature communications*, 10(1):2407. (cited on page 2)

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., and Eickhoff, S. B. (2011). Incongruence effects in crossmodal emotional integration. *Neuroimage*, 54(3):2257–2266. (cited on page 11)

Nakajima, K., Minami, T., and Nakauchi, S. (2017). Interaction between facial expression and color. *Scientific reports*, 7(1):41019. (cited on page 10)

Orlandatou, K., Cambouropoulos, E., Tsougras, C., Mavromatis, P., and Pastiadis, K. (2012). The role of pitch and timbre in the synaesthetic experience. In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece*, pages 751–758. (cited on pages 3, 12, and 13)

Palmer, S. E., Schloss, K. B., Xu, Z., and Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, 110(22):8836–8841. (cited on page 10)

Parise, C. and Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, 442(3):257–261. (cited on page 3)

Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29(1-3):7–28. (cited on pages 3, 5, 8, and 12)

Parise, C. V. and Spence, C. (2009). 'when birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PloS one*, 4(5):e5664. (cited on pages 3, 6, 11, and 22)

Parisi, G. I., Barros, P., Fu, D., Magg, S., Wu, H., Liu, X., and Wermter, S. (2018). A neurorobotic experiment for crossmodal conflict resolution in complex environments. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2330–2335. IEEE. (cited on page 9)

Parisi, G. I., Barros, P., Kerzel, M., Wu, H., Yang, G., Li, Z., Liu, X., and Wermter, S. (2017). A computational model of crossmodal processing for conflict resolution. In *2017 joint ieee international conference on development and learning and epigenetic robotics (icdl-epirob)*, pages 33–38. IEEE. (cited on page 9)

Robinson, C. W. and Sloutsky, V. M. (2010). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(1):135–141. (cited on page 3)

Sathian, K. and Stilla, R. (2010). Cross-modal plasticity of tactile perception in blindness. *Restorative neurology and neuroscience*, 28(2):271–281. (cited on page 1)

Sèbe, F., Duboscq, J., Aubin, T., Ligout, S., and Poindron, P. (2010). Early vocal recognition of mother by lambs: contribution of low-and high-frequency vocalizations. *Animal behaviour*, 79(5):1055–1066. (cited on page 2)

Shankar, M. U., Levitan, C. A., and Spence, C. (2010). Grape expectations: The role of cognitive influences in color–flavor interactions. *Consciousness and cognition*, 19(1):380–390. (cited on page 9)

Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current opinion in neurobiology*, 11(4):505–509. (cited on page 7)

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73:971–995. (cited on page 7)

Spence, C. (2015). On the psychological impact of food colour. *Flavour*, 4(1):1–16. (cited on pages 2 and 9)

Spence, C. (2020). Temperature-based crossmodal correspondences: Causes and consequences. *Multisensory Research*, 33(6):645–682. (cited on page 2)

Spence, C. (2022). Exploring group differences in the crossmodal correspondences. *Multisensory Research*, 35(6):495–536. (cited on page 7)

Spence, C. and Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception*, 3(5):316–318. (cited on pages 2 and 8)

Stein, B. E. and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature reviews neuroscience*, 9(4):255–266. (cited on page 3)

Sumner, P. and Mollon, J. D. (2000). Chromaticity as a signal of ripeness in fruits taken by primates. *Journal of Experimental Biology*, 203(13):1987–2000. (cited on page 2)

Suzuki, Y. and Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, 116(2):918–933. (cited on pages 5, 25, and 39)

Uno, K. and Yokosawa, K. (2022). Cross-modal correspondence between auditory pitch and visual elevation modulates audiovisual temporal recalibration. *Scientific Reports*, 12(1):21308. (cited on page 40)

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., and Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1):21–25. (cited on page 4)

Wallace, M. T., Meredith, M. A., and Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, 91:484–488. (cited on page 4)

Wan, X., Woods, A. T., van den Bosch, J. J., McKenzie, K. J., Velasco, C., and Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between basic tastes and visual features. *Frontiers in psychology*, 5:1365. (cited on pages 2 and 9)

Ward, J., Huckstep, B., and Tsakanikos, E. (2006). Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2):264–280. (cited on page 7)

Wheeler, K. L. (2013). Understanding synesthesia and impact for learning. *Online Submission*. (cited on page 3)

Wyszecki, G. and Stiles, W. S. (2000). *Color science: concepts and methods, quantitative data and formulae*, volume 40. John wiley & sons. (cited on page 5)

Yano, Y. (2011). English as an international language and 'japanese english'. In *English in Japan in the era of globalization*, pages 125–142. Springer. (cited on page 2)

List of Figures

List of Tables