Perceived (in)congruency of audiovisual stimuli consisting of Gabor patches and AM- and FM-tones

ナタリア, ポストノヴァ

出版情報：Kyushu University, 2020, 博士（芸術工学），課程博士
バージョン：
権利関係：

Perceived (in)congruency of audiovisual stimuli consisting of Gabor patches and AM- and FM-tones

Kyushu University

Natalia Postnova

February 2021

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Gerard B. Remijn, for all the help and guidance he has given me over the past three years. I am also extremely grateful to Professor Yoshitaka Nakajima and Professor Kazuo Ueda for the support and advice throughout the doctoral course.

I would also like to extend my gratitude to the Japanese Government for the MEXT scholarship that supported my doctoral study.

And finally, I would like to thank my parents for always believing in me, and my fellow graduate students and my friends for participating in my experiments and giving me friendly support when it was needed, making the completion of this study possible.

Contents

Acknowledgements **i**

Contents **ii**

List of Tables **vi**

List of Figures **vi**

List of Abbreviations **vi**

List of Publications **vii**

Abstract **viii**

1 General Introduction **1**

1.1 General objectives and research plan 11

1.2 Structure of the dissertation 12

2 Experiment 1. Pilot study **14**

2.1 Introduction 14

2.2 Method 15

2.2.1 Participants 15

2.2.2 Apparatus 15

2.2.3 Experiment application and procedure 16

2.2.4 Stimuli 18

2.3 Results 21

2.3.1 Flickering Gabor patches 21

2.3.2 Drifting Gabor patches 27
2.4 Discussion 34
2.5 Conclusions 36

3 Experiment 2. Effect of tone modulation type and temporal (dis)similarity 37

3.1 Introduction 37
3.2 Method 38
3.2.1 Participants 38
3.2.2 Apparatus 39
3.2.3 Experiment application and procedure 39
3.2.4 Stimuli 41
3.3 Results 42
3.3.1 Flickering Gabor patches 42
3.3.2 Drifting Gabor patches 46
3.4 Discussion 52
3.5 Conclusions 55

4 Experiment 3. Effect of a carrier frequency of a tone 56

4.1 Introduction 56
4.2 Method 57
4.2.1 Participants 57
4.2.2 Apparatus 58
4.2.3 Experiment Application and Procedure 58
4.2.4 Stimuli 59
4.3 Results 60
4.3.1 Flickering Gabor patches 60
4.3.2 Drifting Gabor patches 63
4.4 Discussion 68
4.5 Conclusion 69

4.6 Limitations of the study . . . 69

5 Experiment 4. Effect of stimulus duration 70

5.1 Introduction . . . 70

5.2 Method . . . 71

5.2.1 Participants . . . 71

5.2.2 Apparatus . . . 72

5.2.3 Experiment application and procedure . . . 72

5.2.4 Stimuli . . . 73

5.3 Results . . . 75

5.3.1 Flickering Gabor patches . . . 75

5.3.2 Drifting Gabor patches . . . 77

5.4 Discussion . . . 81

5.5 Conclusion . . . 82

6 General Discussion 83

6.1 Summary . . . 83

6.2 Implications . . . 87

6.2.1 Effect of tone-modulation type on perceived congruency . . . 87

6.2.2 Effect of dynamic parameters of Gabor patches and modulated tones . . . 88

6.2.3 Effect of static parameters of Gabor patches and modulated tones . . . 89

6.2.4 Limitations and future studies . . . 91

6.2.5 Conclusion . . . 93

Appendices 94

A Experiment Application . . . 94

B Calibration Procedure . . . 104

B.1 Experiment 1 . . . 105

B.2 Experiment 2 . . . 105

B.3 Experiment 3 106
B.4 Experiment 4 106
C Light and sound levels measurements 107
D Sound generation 108
D.1 Experiment 1 108
D.2 Experiment 2 110
D.3 Experiment 3 110
D.4 Experiment 4 111
E Instruction 114

Bibliography **122**

List of Abbreviations

AM	Amplitude modulation
FM	Frequency modulation
LFR	Logarithm frequency ratio
cpd	Cycle per degree
PSS	Point of subjective simultaneity
TOJ	Temporal order judgement
SJ	Synchrony judgment
PSSA	Point of subjective spatial alignment
lx	Lux
cd/m^2	Candela per square metre
CRT	Cathode-ray tube
D/A	Digital to analog
kHz	kilohertz

List of Publications

Chapter 2 (Experiment 1)

Postnova, N., & Remijn, G. (2019). The effect of sound modulation mode on perceived audiovisual congruency of pure tones and Gabor patches. *35th Annual Meeting of the International Society for Psychophysics. Conference Proceedings*, 101–107

Chapter 3 and 4 (Experiments 2 and 3)

Postnova, N., Nakajima, Y., Ueda, K., & Remijn, G. B. (2020). Perceived congruency in audiovisual stimuli consisting of Gabor patches and AM and FM tones. *Multisensory Research*, 1–21. https://doi.org/10.1163/22134808-bja10041

Abstract

Until now, numerous studies have used stimuli containing Gabor patches and modulated tones to examine how auditory and visual information is combined and processed. The current study uses such stimuli and investigates the roles of their physical parameters in defining perceived congruency.

In the past, the congruency between a Gabor patch and a tone was often determined by matching the spatial frequency of a patch and a carrier or modulation frequency of a tone. However, it is yet unknown whether this crossmodal correspondence has an effect on (in)congruency in the case of dynamic (flickering or drifting) Gabor patches. Besides, in cases when modulated tones are used (AM or FM), the choice of modulation type often remains unsupported by empirical data, as no comparative studies for AM- and FM-tones in the context of their effect on audiovisual (in)congruency were conducted. Therefore, the objectives of the current study were: (1) to investigate the perceived congruency in relatively long stimuli (2 seconds) consisting of a Gabor patch and a modulated tone; (2) to compare the perceived congruency of such stimuli with an AM-tone and an FM-tone; (3) to investigate the effect of static (spatial frequency of the patch and the carrier frequency of the tone) and dynamic or temporal parameters on perceived congruency to define the most prominent factors for congruency in such stimuli.

The present research examined Gabor patches of various spatial frequencies (2-10 cpd) with flickering or drifting gratings in combinations with AM- or FM-tones of 0.5 - 4 Hz modulation, and 500-, 1000- and 2000-Hz carrier frequencies. Data were collected through experiments in which combinations of a Gabor patch and a tone were rated on a scale from 1 (incongruent) to 7 (congruent). The results showed,

first, that stimuli with a flickering Gabor patch and an AM-tone showed significantly higher perceived congruency compared to stimuli with an FM-tone. Besides, the effect was especially strong in stimuli in which the patch-flicker frequency and the tone-modulation frequency were (almost) similar. Second, the dynamic parameters, such as the flickering (temporal) frequency of the patch and the modulation frequency of the tone, played a prominent role in defining perceived congruency, while static parameters had little or no effect. These findings were confirmed for stimuli of different duration (1 - 4 seconds).

The results suggest that the temporal similarity between auditory and visual components plays a prominent role in defining audiovisual (in)congruency. Additionally, the similarity in the dynamics of auditory and visual components can further enhance congruency. The crossmodal matching of the visual component's spatial frequency and the auditory frequencies, on the other hand, has no measurable effect on the (in)congruency of dynamic stimuli.

Chapter 1. General Introduction

We perceive the world around us through different senses, and most of the real-life events are registered by more than one sense at the same time. This multisensory nature of our perception is highly beneficial as it increases perception reliability by compensating the limitations of one sense organ with abilities of the other, or by utilising the redundant or complementary information from two or more senses. One example of such perception enhancement is the improvement of speech intelligibility when the speaker's face (lips) is visible compared to when the face is hidden (Sumby and Pollack, 1954; Summerfield, 1992). In some cases, however, audiovisual perception not only enhances, but alters the perception of the event. Some of the most prominent examples of such crossmodal influence would be the McGurk effect (McGurk and MacDonald, 1976), in which the perception of a spoken syllable changes depending on the perception of visual information regarding lip movements of the speaker. Another example is the stream/bounce illusion (Sekuler et al., 1997), in which the visual movement of two objects is influenced by whether the stimulus is accompanied by a sound or not.

Numerous studies have investigated human audiovisual integration using various methods. The correct choice of stimuli for such experiments could be crucial in getting reliable and meaningful results. One term that is often used to describe how well auditory and visual stimuli work together is *congruency*. Various studies have employed the idea that physically congruent and incongruent stimuli are processed differently. This has often been expressed as a "congruency effect" that can be observed and measured (see Spence, 2007). However, when the term "congruency" is used, different features of audiovisual stimuli can be implied, such as temporal, spatial, or semantic features.

Congruency as a factor for the "Assumption of Unity"

Most of the events in the world are perceived through both vision and hearing, and an ability to correctly match visual information with sound is crucial for successful navigation and survival. Numerous studies investigated how event interpretation and detection are affected by correspondence in stimulus parameters (Sumby and Pollack, 1954; Bolognini et al., 2004; Iordanescu et al., 2010). The important implication of such crossmodal relationships is the "assumption of unity". The "assumption of unity" holds that a particular combination of temporal, spatial and semantic parameters of sound and visual signals can provide clues that these signals pertain to the same multisensory event (Welch and Warren, 1980; Chen and Spence, 2017). "Congruent" stimuli in that case can be interpreted as stimuli that are perceived as referring to one event, while "incongruent" audio and visual signals are perceived as originated by separate events.

Temporal and spatial congruency

In the case of temporal congruency, most of the time when the term "congruency" is used, simultaneity or synchrony is meant. In real life, sound and light from the same event do not reach our senses at the same time, due to different propagation times and due to differences in "central availability", i.e., the differences in processing time of sound and light in the brain. The study of the relation between physical simultaneity and perceptual simultaneity starts by defining the point of subjective simultaneity (PSS; Keetels and Vroomen, 2011). Tasks that are often used to measure PSS are the temporal order judgement task (TOJ) or the synchrony judgment task (SJ). In both cases, participants are asked to judge the audio and visual signal on their simultaneity: "sound first" or "light first" in case of TOJ, "synchronous" or "asynchronous" in case of SJ. Depending on the method and the stimuli, PSS for sound and light can differ and is observer specific (Jaśkowski et al., 1990; Aschersleben and Musseler, 1997; Slutsky and Recanzone, 2001; Stone et al., 2001; Zampini et al., 2003; Vatakis and Spence, 2006; Eijk et al., 2008; Kayser et al., 2008;).

Similar to temporal congruency measurements, spatial congruency can be expressed by obtaining the point of subjective spatial alignment (PSSA; Lewald et al., 2001; Miyauchi et al., 2014). It must be noted, however, that temporal and spatial congruencies are interrelated, and spatial congruency is often studied in the context of the temporal congruency or when temporal simultaneity of the studied stimuli is implied. Commonly, both PSS and PSSA are in the center of a limited range of stimulus characteristics (i.e., a temporal or spatial window) in which simultaneity is perceived. In studies that consider both temporal and spatial factors of sound and light, a spatiotemporal window is identified in which audiovisual integration can happen, depending on both temporal and spatial congruency of the stimuli (Bertelson and Radeau, 1981; Radeau and Bertelson, 1987; Meredith et al., 1987; Stein et al., 1988; Frens et al., 1995; Lewald et al., 2001; Hairston et al., 2003; Alais and Burr, 2004; Zampini et al., 2005; Miyauchi et al., 2014). Congruency in spatiotemporal terms often refers to an assumption that the sound and light come from the same location and originated at the same time. This assumption is closely related to the "assumption of unity" that was described earlier.

Semantic or synaesthetic congruency

Compared to temporal and spatial congruency, semantic congruency is more challenging to define and measure, as it usually refers to stimulus components in terms of their "identity and/or meaning" (Spence, 2011). The differentiation between semantically congruent and semantically incongruent stimuli is quite clear when not abstract, but meaningful stimuli are used - pictures of objects or text for visual components, and environmental sounds or speech for auditory components. The perfect example of such combinations is the Visual-Auditory Stroop task (Donohue et al., 2013). In such tasks, a colour name is presented as a word on the screen with a voiceover saying the same (congruent) or different (incongruent) colour.

The semantic congruency for abstract stimuli is more complex. Unlike temporal and spatial congruency, semantic congruency is not defined by one parameter only. The studies on such congruency often deal with a variety of features. For example, in the study by Lipscomb and Kim (2004) participants were asked to rate how well auditory and visual components matched in audiovisual stimuli. The results showed crossmodal correspondence between vertical location and pitch, size and loudness, and shape and timbre. Similar results were obtained by Evans and Treisman (2010), who used a different method and demonstrated a correspondence between vertical location and perceived pitch, and size and pitch. In the case of abstract stimuli, the term "semantic" could not always be considered suitable. Therefore, in some studies the term "synaesthetic" is used to refer to such crossmodal correspondence (Spence, 2011).

Stimuli with Gabor patches and tones

Among the wide variety of stimuli that are used in studies on audiovisual interactions, a vast number of researchers have used Gabor patches in combination with various sounds to create audiovisual stimuli of certain congruency. Gabor patches are considered to be the building blocks of our visual perception system (Figure 1.1; Gabor, 1947; Watson et al., 1983). They are simple, yet have a range of parameters, which when systematically varied, can create a great variety of visual stimuli with clearly defined properties. These parameters are:

- spatial frequency,
- gratings orientation,
- contrast,
- phase,
- size.

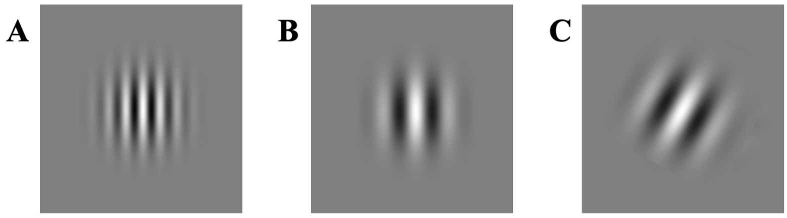

Figure 1.1: Examples of Gabor patches. Patches A and B have different spatial frequencies, patches B and C have different gratings orientations.

As these parameters refer to static Gabor patches, in this dissertation they are referred to as "static" parameters.

Besides static parameters that are described above, various dynamic changes can be applied to the patch by systematically varying static parameters over time:

- gradual change of spatial frequency, creating a pulsating effect;
- gradual change of gratings orientation, creating a rotating effect;
- gradual change of contrast, creating a flickering effect;
- gradual change of phase, creating a drifting effect;
- various combinations of dynamic changes can create more complex effects, for example a gradual change of spatial frequency and size, or spatial frequency and contrast, creates stronger pulsating effects.

Parameters referring to as the dynamic changes of a Gabor patch, such as flickering frequency or drifting speed, are referred as "dynamic" parameters in this dissertation.

When creating an audiovisual stimulus using a Gabor patch, congruency with auditory components can be achieved by matching these parameters with corresponding parameters of a tone or noise, such as its spectral components or modulation frequency.

Research on audiovisual congruency using a Gabor patch and a sound

Studies with short stimuli (<1 second)

Evans and Treisman (2010) specifically examined the effect of combinations of Gabor patches and pure tones in order to answer a more general question concerning the neural mapping between audio and visual features. The results demonstrated a cross-modal correspondence between auditory pitch and visual position and between pitch and spatial frequency. In that study, short stimuli with a duration of 120 ms were used. The pure tones were of 1000 Hz (low) and 1500 Hz (high) frequency and combined with Gabor patches of 2 cycles per degree (cpd), i.e., a low spatial frequency, and 6 cpd, a higher spatial frequency. The combinations with a tone of a high (low) frequency and a patch of a high (low) frequency were considered congruent, while combinations with dissimilar frequencies were considered incongruent. The congruent pairs demonstrated significantly shorter reaction times in a speeded classification task.

Green et al. (2019) also demonstrated a crossmodal correspondence between a spatial frequency of a Gabor patch and the frequency of a tone. However, results demonstrated not an absolute relationship between visual and sound frequencies but rather the correspondence between their ranges. The range of visual frequencies that human eyes are most sensitive to was matched with the range of audio frequencies that the human ears are most sensitive to.

The study of Guzman-Martinez et al. (2012) demonstrated a correspondence between the spatial frequency of a Gabor patch and the modulation frequency of amplitude modulated (AM) white noise. In this experiment, participants adjusted the AM white noise to fit the presented Gabor patch of either 0.5, 2.0 or 4.5 cycles/cm. The results demonstrated that participants "consistently and absolutely" matched a certain modulation frequency to a Gabor patch of a certain spatial frequency. A similar study (Orchard-Mills et al., 2013a, 2013b) confirmed the relations between AM auditory stimuli and the spatial frequency of Gabor patches. However, the exact matching between these two parameters varied in this study and the study by Guzman-Martinez et al. For example, while in the study of Orchard-Mills et al. (2013a) the stimuli with the visual component of 2 cycles/cm were matched with 3-Hz modulated white noise, in the study by Guzman-Martinez et al. the same visual component was matched with 6-Hz modulated white noise. In summary, cross-modal correspondence was investigated mainly between the spatial frequency of a Gabor patch and the frequency or modulation frequency of the tone or modulation frequency of the noise, by using different experimental methods.

Following these studies, multiple researchers based their stimulus design on the above-mentioned findings. For example, Heron et al. (2012), in research on time perception, used short stimuli (20 ms) consisting of a Gabor patch of high (4 cycles/degree) or low (1 cycle/degree) spatial frequency and a pure tone of high (2000 Hz) or low (500 Hz) frequency in order to create “contextually congruent and incongruent” pairs. An effect of congruency, however, was not observed.

In some cases, a similar principle for creating audiovisual pairs was used, i.e., a patch of high (low) spatial frequency was combined with a sound of high (low) frequency, however, no assumption that these stimuli were naturally perceived as congruent was made. A study on learning processes (Altieri et al., 2015) employed 6 pairs of Gabor patches and pure tones of 100 ms, in which Gabor patches of higher spatial frequency were combined with pure tones with a higher frequency, and these pairs were referred to as “previously unassociated”. It can be seen that though the same matching principle was used in both above-mentioned studies (high spatial frequency was matched with the high audio frequency), the utilisation and interpretation of the resulting stimuli were different.

Studies with longer stimuli (> 1 second)

Most of the short (up to a second) stimuli consisting of a Gabor patch and a sound used a pure tone as auditory part with a static Gabor patch, but in the case of longer stimuli, dynamic changes to both visual (Gabor patch) and auditory parts were often introduced. For example, in a study by Schall et al. (2009), congruency between the Gabor patch and the auditory stimulus was created by simultaneous temporal dynamics - the Gabor patch spanned 180 degrees, and the frequency of the tone was modulated between 200 Hz and 300 Hz. In congruent conditions, the so called "trajectory" of these changes were the same, while in incongruent conditions they were different. A similar idea was used in a study by Guttman et al. (2005) where visual rhythm was created by presenting a Gabor patch that reversed in contrast over some random periods of time. This visual rhythm was combined with the auditory rhythm consisting of clicks. Congruent pairs were the combinations of these rhythms where visual contrast reversals and auditory clicks coincided. Another visual rhythm was created by Covic et al. (2017) by introducing a flickering and a pulsating Gabor patch consisting of spatial frequency changes and a frequency modulated (FM) tone. So for relatively long stimuli, the perceived congruency was most often induced by synchronous dynamic changes of the patch and a sound.

Multiple studies have employed a variety of dynamic Gabor patches with modulated tones, however, the empirical data on the perceived congruency of such stimuli are lacking. One of the features that was often used, but rarely tested, is the sound modulation type. The most commonly used types of modulations are amplitude and frequency modulations. In the previously mentioned studies (Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a, 2013b), the frequency of AM white noise was matched with a spatial frequency of a static patch. Interestingly, in a study by Orchard-Mills et al. (2013a) the search benefits persisted even for the stimuli where the spatial frequency of the Gabor patch did not match the tone-modulation frequency. Hence, the matching between the visual and sound frequencies was not the source of the benefit. However, in the study by Guzman-Martinez et al. (2012) the congruency effect was observed for stimuli with the tone-modulation frequency that matched the spatial frequency of the Gabor patch. Evidently, sound modulation affects the perceived congruency of audiovisual stimuli, however, this effect may vary in different contexts and for different stimuli. It might be assumed that in case of FM-tones, similar effects could take place, however, no empirical data on the congruency of stimuli with FM-tones are available.

Suppose there are two stimuli with similar visual components but with auditory components that vary in modulation type (for example, AM and FM). Given that the modulation frequency is the same, their frequency content is still different (Schorer, 1986; Tsumura et al., 1990). In that case, would the stimuli with identical visual components and modulated tones, that differ only in modulation type but not the modulation frequency, be perceived as equally congruent? Fujisaki et al. (2005), in a study on audiovisual synchrony, used both AM- and FM-tones and combined them with a flickering Gaussian blob and a rotating Gabor patch correspondingly to induce temporal congruency. These stimuli showed similar results in visual search tasks, however, the individual effects of AM- and FM-tones could not be conclusively assessed here, as these tones were combined with different visual stimuli.

To summarise, although the AM- and FM-tones were widely used in audiovisual integration studies, no study assessed the effect of modulation type on perceived congruency by comparing the stimuli with an AM- or an FM-tone of the same modulation frequency and identical visual components. Furthermore, the design of relatively long (over 1 second) stimuli with Gabor patches and modulated tones has been often based on creating a temporal synchrony of dynamic changes between the visual and the auditory components of the stimuli. The nature of these changes, however, is somehow vague with no empirical evidence of perceived congruency level in these cases. The available data of perceived congruency of stimuli consisting of a Gabor patch and a tone mostly concern short (less than a second) stimuli. Whether matching principles that are applicable for short stimuli consisting of a Gabor patch and a sound also hold for longer stimuli remains unknown. Besides, the congruency of the stimuli was most often studied by assessing congruency effect, i.e., analysing the changes in response time or temporal window of integration in speed classification or search tasks (Evans and Treisman, 2010; Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a; Green et al., 2019). However, it is unknown whether stimuli examined in these studies are actually perceived as congruent or incongruent by participants, as no (in)congruency judgement was explicitly asked from participants. Also, the level of perceived (in)congruency of such stimuli was never assessed, though it can be assumed that certain stimuli could evoke stronger perceived congruency (or incongruency) than others.

1.1 General objectives and research plan

The stimuli consisting of a Gabor patch and a modulated tone are widely used to investigate the mechanisms of audiovisual integration and congruency perception. For these purposes, such stimuli are often varied on congruency levels, where certain combinations of a Gabor patch and a modulated tone are considered congruent, and others incongruent. The justification of such categorisation is often found in studies of relatively short stimuli ($\approx$100 ms), in which Gabor patches and tones or modulated noises demonstrated certain crossmodal correspondences between parameters of a static Gabor patch and a frequency of a tone. However, the exact effect of these parameters on perceived congruency in other contexts and conditions, i.e., in longer or dynamic stimuli, remains unknown. Besides, in the case of modulated tones, the type of modulation is often chosen arbitrarily, and no comparison studies on amplitude and frequency modulation in such stimuli are available.

The current research aims to investigate the effect of modulation type, specifically amplitude and frequency modulation, on the perceived congruency of stimuli consisting of a modulated tone and a Gabor patch. This study employs non-static visual stimuli consisting of flickering and drifting Gabor patches. A wide variety of crossmodal parameters will be explored and compared, including the comparison of potential effects from physical (dis)similarities between static parameters, i.e., the spatial frequency of the patch and a carrier frequency of the tone, and dynamic (temporal) parameters, such as flickering frequency or drifting speed of the patch and modulation frequency of the tone.

Perceived audiovisual (in)congruency is obtained through rating experiments in which participants are asked to rate the stimuli on an (in)congruency scale. On this scale, higher ratings correspond to higher perceived congruency, while lower ratings correspond to stronger incongruency. The average ratings for different groups of stimuli are evaluated and compared.

The main objective of this dissertation is to define the parameters that affect the perceived congruency of Gabor patches and modulated tones, i.e, (1) to compare the perceived congruency evoked by AM- and FM-tones in stimuli consisting of a Gabor patch and a modulated tone; (2) to confirm whether static parameters have an effect on perceived congruency in the presence of dynamic changes in relatively long audiovisual stimuli (2 seconds), as it was demonstrated for short stimuli with a static Gabor patch (Evans and Treisman, 2010; Green et al., 2019); and (3) to assess the effect of correspondence between the physical frequencies of the auditory and the visual components of the stimuli on perceived audiovisual congruency.

1.2 Structure of the dissertation

Chapter 2 describes a pilot study (Experiment 1) that employs a wide variety of Gabor patches and modulated tones in order to, first, investigate whether the modulation type (AM or FM) had any effect on the perceived congruency, and second, to identify the parameters of Gabor patches and tones that affect perceived congruency. This experiment defined the direction of the current study and the framework of Experiments 2-4.

The goal of Chapter 3 was to confirm and further explore findings described in Chapter 2, namely the advantage of AM-tones over FM-tones in evoking perceived congruency with Gabor patches, and the role of temporal (dis)similarities in establishing congruency between visual and auditory components. For these purposes the initial parameters of stimuli were adjusted to create audiovisual stimuli in which auditory and visual components were of similar temporal frequencies. Besides, the experimental application was modified and a fixation point was introduced.

Chapter 4 investigates the effect of carrier frequency of modulated tones on perceived audiovisual congruency and further describes the effect of modulation type. For the purpose of the study, tones of various carrier frequencies were added to the stimulus pool. The special focus was on the crossmodal correspondence between the spatial frequency of the Gabor patch and the carrier frequency of the tone, which was observed in the previous studies.

Chapter 5 explores whether stimulus duration had an effect on perceived audio-visual congruency. Stimuli with a duration that was shorter and longer than used in Experiments 1-3 were employed and their (in)congruency ratings were compared. This chapter aims to answer the question of whether the stimulus duration could affect the perceived congruency or amplify/weaken the effects of static and dynamic parameters.

Chapter 6 summarizes the findings of the current study and provides a discussion based on individual and cumulative results of the previous chapters, and includes suggestions for future study.

Chapter 2. Experiment 1. Pilot study

2.1 Introduction

To clarify the role of different parameters of visual and auditory components of the audiovisual stimuli on perceived (in)congruency, the following experiment was developed and conducted. The main purpose of the experiment was to investigate the effect of tone modulation (AM and FM) on perceived congruency. It was hypothesised that if modulation type affects the congruency, the difference in average perceived (in)congruency ratings between audiovisual stimuli containing AM and FM tones could be observed.

According to previous studies (Evans and Treisman, 2010; Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a; Orchard-Mills et al., 2013b), the spatial frequency of the Gabor patch can evoke differences in perceived congruency in combinations with different tones. In order to clarify and compare the roles of different parameters on perceived congruency, a wide range of stimulus parameters was used. The choice of parameters was done in a way so that it overlaps with previously mentioned studies. The main focus was on the following features: a) the modulation type of the tones, b) spatial frequency of the Gabor patches, and c) temporal (dis)similarities between these auditory and visual components.

As the interest of the study was perceived congruency, the stimuli were rated directly by participants. Average ratings were calculated for the groups of stimuli aggregated by different parameters or parameters' pairs. Analysis of such groups were conducted to define which parameters played prominent roles in establishing the congruency. The results of this pilot study became the foundation of this dissertation and defined the direction of the subsequent experiments.

2.2 Method

2.2.1 Participants

Initially 15 persons participated in Experiment 1. One participant was excluded prior to the experiment due to a self-reported visual impairment. The other 14 participants reported normal or corrected-to-normal vision and normal hearing. Among them 8 were males, 6 were females. The mean age was 28 years (SD = 6). Twelve out of 14 were Kyushu University students, and 2 were recruited from outside of the university.

Prior to the experiment, all participants were given an explanation about the purpose of the experiment and they received written instructions about the procedure. Each participant provided written informed consent. The experiment was conducted with prior approval of the Ethics Committee of Kyushu University.

2.2.2 Apparatus

The experiment was conducted in a dim-lit and sound-attenuated (less than 27 dBA in quiet) experiment booth. The illuminance at the participant level was measured with a TOPCON IM-1000 lux meter and was as follows: below 0.05 lx for a black screen, 1 lx for the grey background of the experiment's application and 2.97 lx for the white screen.

The experiment procedure was controlled from outside the booth through a MacBook computer (MacBook Pro Early 2015). The application for the experiment procedure was developed with PsychoPy3 Experiment Builder (Peirce et al., 2019, see Appendix A for the application description). The visual stimuli were presented using a CRT monitor (Iiyama HF703U E, 16 inches). The sound was passed through a D/A converter (Behringer U-Phono UF0202), a filter (NF DF-3BL; cut-off frequency 14 kHz), and a headphone amplifier (STAX SRM-323A), and presented to

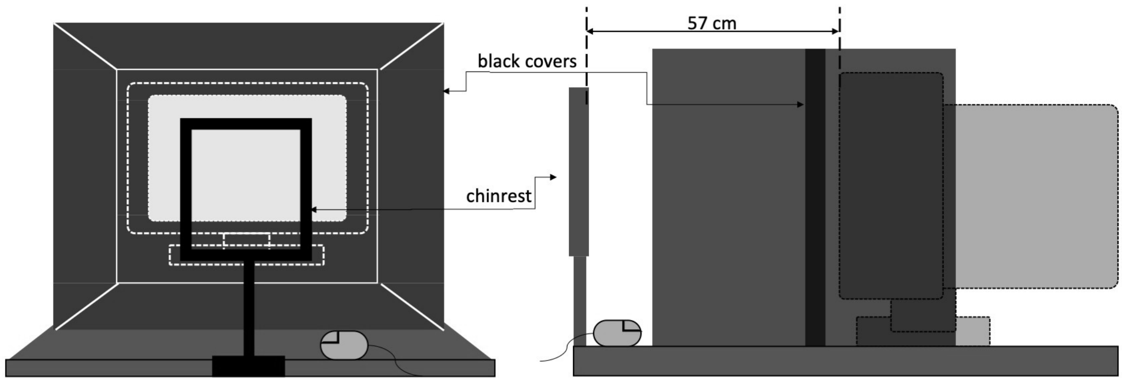

Figure 2.1: Experiment settings

participants binaurally through headphones (STAX SR-307) with maximum sound level of the stimuli set to 57.8 dB (the full information on sound measurements see Appendix C). To ensure simultaneity of the audio and visual signals, the calibration was conducted by using a multi-channel oscilloscope (IWATSU DS-5314). The full calibration procedure is described in Appendix B. Each participant was asked to place their head on a head-and-chin rest in order to fixate the distance from the monitor (57 cm) and to ensure a constant observation angle.

To eliminate distractions during the experiment, the monitor was set inside a tunnel-like black cover and a front black cover for the monitor designed such that all parts of the front side of the monitor, except the display, were covered with black carton paper (Fig. 2.1). Participants were also asked to cover their hands with black cloth, to avoid distractions caused by the reflection of light from their watches or red LED light from the optical mouse.

2.2.3 Experiment application and procedure

The experiment consisted of 500 trials and was divided in two parts conducted on different days. Each part consisted of 5 sessions of 50 trials each with 3-minute breaks between the sessions.

The rating scale that was used in the experiment was designed based on a 7-Points Likert Scale. The number of points on the scale was chosen so that it allowed to collect the ratings with high precision (3 levels of congruency and 3 levels of incongruency) and included a neutral answer option. A value was assigned to every point on the scale from 1 ("incongruent") to 7 ("congruent"). Point 4 represented the neutral response. Thereby, high ratings correspond to stronger perceived congruency, while low ratings correspond to stronger perceived incongruency.

Each participant received an instruction (see Appendix E) with detailed description of the experiment flow, information about pauses and breaks between trials, and expected completion time. The main task was formulated in the instruction as follows: "Please, choose the most appropriate rating for the seen stimulus from 1 to 7, where "1" is completely incongruent and "7" is perfectly congruent." During the experiment, only the rating scale with the word "incongruent" written on the left side of the scale, and the word "congruent" written on the right side of the scale appeared (Fig. 4 in Appendix A). In cases when participants needed a clarification on what congruency meant, the following explanation was given: "how well auditory and visual parts work together".

The flow of the experiment is presented in Figure 2.2. Every trial started with a 2-second pause after which the fixation point appeared, indicating where the stimuli would be located. The fixation point was visible for 1 second followed by 1 second pause, after which the stimulus appeared and lasted for 2 seconds. After the stimulus, there was another 2-second pause, after which the rating scale appeared. Answering time was not limited, but as soon as the response was recorded the trial was over. Participants had to confirm that they were ready for the next trial by clicking the mouse for the second time ("Click for next" message appeared on the screen).

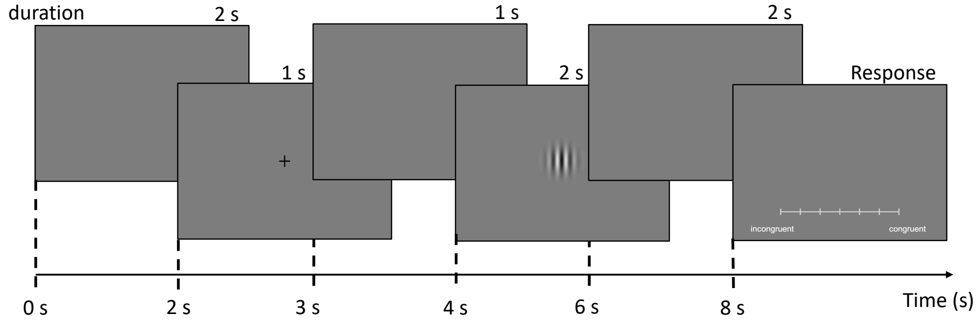

Figure 2.2: Trial flow of Experiment 1.

The experiment started with a practice session during which the experimenter was inside the experiment room together with the participant (the experiment was performed before the COVID-19 outbreak, May 2019). This session consisted of 5 trials, randomly selected for each participant from the pool of all trials. After completion of the practice session, the experimenter left the room.

2.2.4 Stimuli

Every stimulus consisted of a Gabor patch and a tone. The stimuli had a duration of 2 seconds.

Gabor patches

All Gabor patches were 2×2 degrees in visual angle in size with a Gaussian envelope, and consisted of black-and-white sinusoidal gratings with the peak luminance at 9.53 cd/m^2 (white) and the minimum at 0.01 cd/m^2 (black) on a grey background (2.78 cd/m^2). The full measurements done prior to the experiment are presented in Appendix C.

The stimuli consisted of Gabor patches varied across the following parameters:

- spatial frequency (Fig. 2.3), measured in cycles per degree (cpd): 2, 4, 6, 8, and 10 cpd

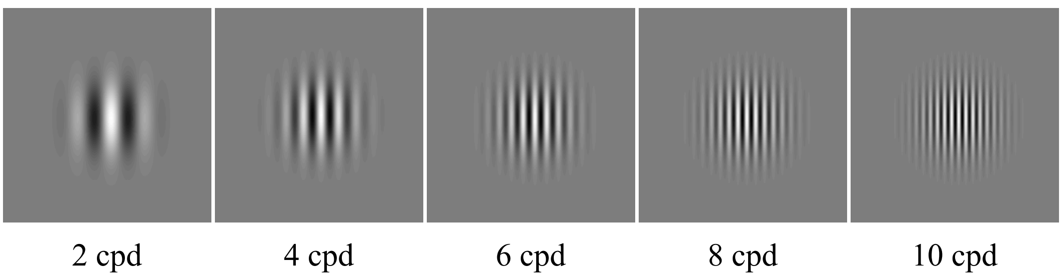

Figure 2.3: Example of the stimuli used in Experiment 1 consisting of Gabor patches with a spatial frequency of 2, 4, 6, 8 and 10 cps.

- dynamic mode:
  - Flickering. A "Flickering" effect was created by gradually changing the luminance contrast of the patch from 0 to 1 and back (Fig. 2.4 A). The phase was changed by π when the patch became invisible (luminance contrast = 0). Flickering Gabor patches varied in flickering frequencies: 0.5, 1, 2, 3 and 4 flickers per second.
  - Drifting. The gratings in the Gabor patches drifted from left to right. This effect was created by gradually changing the phase of the sinusoidal grating (Fig. 2.4 B). Drifting Gabor patches varied in drifting speed: 0.5, 1, 1.5, 2 and 2.5 degrees per second.

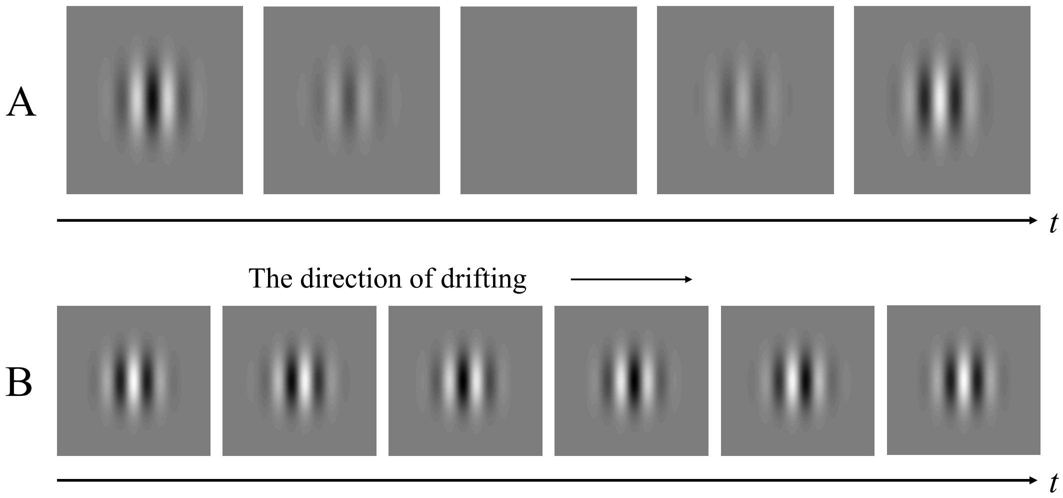

Figure 2.4: Example of a flickering Gabor patch (A) and a drifting Gabor patch (B) as used in Experiment 1.

The Gabor patches were generated by the experiment application in real time during the experiment. The full information about the stimulus generation procedure and a code are presented in Appendix A.

Tones

Amplitude-modulated (AM) and frequency-modulated (FM) tones with the following parameters were used in this experiment:

- carrier frequency: 1000 Hz;
- rise and fall time: 500 ms, with a cosine-shaped ramp;
- modulation:
  - AM,
  - FM;
- modulation depth:
  - for AM: 1 amplitude (the signal changed between the original intensity and 0);
  - for FM: ±80 Hz, half a critical band width for a 1000 Hz tone (Zwicker, 1961);
- modulation frequency (same for AM- and FM-tones): 0.5, 1, 2, 3 and 4 Hz. The set of modulation frequencies was the same as the set of flickering frequencies for flickering Gabor patches.

Auditory stimuli were generated before the experiment (44100 Hz sampling frequency, 16 bit quantization) and called by the application during the experiment. The scripts that were used to generate the audio files are presented in Appendix D.

The consolidated table with all parameters for Experiment 1 and the subsequent experiments is presented in General Discussion on page 86 in Figure 6.1.

All together there were 50 different patches (25 flickering Gabor patches: 5 spatial frequencies × 5 flickering frequencies, and 25 drifting Gabor patches: 5 spatial frequencies × 5 drifting speeds) and 10 different tones (5 modulation frequencies × 2 modulation types, AM and FM), resulting in 500 possible combinations of Gabor patches and tones with the described parameters. Each participant rated each of them once, excluding the 5 random stimuli that appeared in the practice session (practice session ratings were not included in the analysis of the results).

2.3 Results

As the main purpose of the experiment was to define the difference in perception evoked by tones with a different type of modulation, the average perceived (in)congruency ratings were calculated separately for stimuli with AM- and FM-tones, and then compared.

Another purpose was to define which parameters of the stimuli, or more specifically what combinations of a visual parameter and an auditory parameter of the stimuli, influenced perceived congruency. As Gabor patches had two parameters that were varied (spatial frequency and flickering frequency, or spatial frequency and speed), different grouping principles were applied in order to determine which parameter played a more important role in perceived (in)congruency. It can be assumed that if some parameter or a crossmodal combination of parameters affects the perceived congruency, then the average ratings of the stimuli grouped by these parameters will demonstrate some difference between those groups, and the difference will be greater between groups aggregated by parameters with a stronger effect on the congruency.

Flickering and drifting Gabor patches were analysed separately as they had different physical parameters. Nevertheless, the same statistical methods were used for both stimulus types. Also, as the obtained data were not normally distributed, as confirmed with the Kolmogorov-Smirnov test of normality, non-parameteric tests were used to compare the ratings of different groups. Below, the results are first described for flickering Gabor patches, and then for drifting Gabor patches. This structure is kept throughout the dissertation.

2.3.1 Flickering Gabor patches

Tone modulation

The average ratings between the stimuli containing AM-tones and FM-tones were compared using the sign-test. The results demonstrated a significant difference between the stimuli containing AM-tones and corresponding stimuli containing the same visual component and the FM-tone of the same modulation frequency: $N_{plus} = 745$, $N_{minus} = 452$, $Z = -8.440$, $p < 0.001$, where stimuli containing AM-tones evoked higher perceived (in)congruency ratings than stimuli containing FM-tones.

Spatial frequency

According to the previous studies, spatial frequency and tone frequency (perceived pitch) demonstrated cross-modal correspondence (Evans and Treisman, 2010). Therefore, given that we were using only one carrier frequency for all our tones, Gabor patches of certain spatial frequency could potentially have stronger correspondence with these tones than others. In order to check that, we calculated the average ratings across all stimuli with a flickering Gabor patch (Figure 2.5).

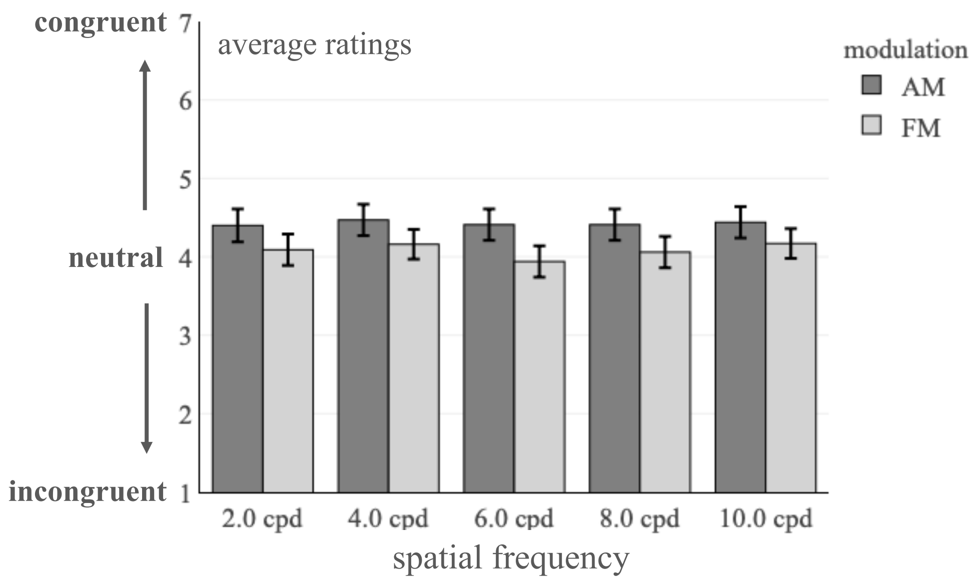

Figure 2.5: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

The average ratings for stimuli with a Gabor patch with different spatial frequencies demonstrated that participants did not have any preferences towards any spatial frequency in terms of perceived congruency. All the average ratings regardless of the spatial frequency occupied roughly the middle of the rating range: 4.38-4.49

for AM-tones, 3.93-4.17 for FM-tones. However, the results in Figure 2.5 clearly demonstrate that stimuli with AM-tones scored higher on the rating scale, and the effect was consistent for stimuli with visual components of all spatial frequencies.

Based on previous findings, we can also assume that Gabor patches of a certain spatial frequency can evoke different perceived congruency when combined with tones of different modulation frequencies (Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a; 2013b). Therefore, (in)congruency ratings for both spatial frequency of the Gabor patch and the modulation frequency of the tone were combined for analysis. As it can be seen in Figure 2.6, some of the combinations scored higher than others, however, no clear rating pattern can be observed and the difference between the groups was not statistically significant (based on 95%-confidence intervals, Figure 2.6). Here too, we can clearly see the difference between AM- and FM-stimuli, and the difference is greater for the stimuli with auditory components of higher modulation frequencies. In summary, spatial frequency of the Gabor patch had no clearly distinguishable effect on perceived (in)congruency.

Flickering frequency

The temporal synchronicity between the visual and the auditory component of a stimulus can play a prominent role in establishing congruency (Munhall et al., 1996; Fujisaki et al., 2005). Similar to previous research, as expected, perceived (in)congruency was strongly affected by similarity or dissimilarity between the temporal frequencies of the auditory (modulation frequency) and the visual (flickering frequency) part of the stimuli. Figure 2.7 shows the average ratings with 95%-confidence intervals around the mean. For a better demonstration of the rating pattern, a grayscale heatmap is also presented in Figure 2.8.

In contrast to groupings by spatial frequency with roughly no difference between the groups' ratings, grouping by flickering and modulation frequencies demonstrated a wide range of average rating values - from significantly congruent (significantly above 4 as shown by the 95%-confidence interval around the mean) to significantly

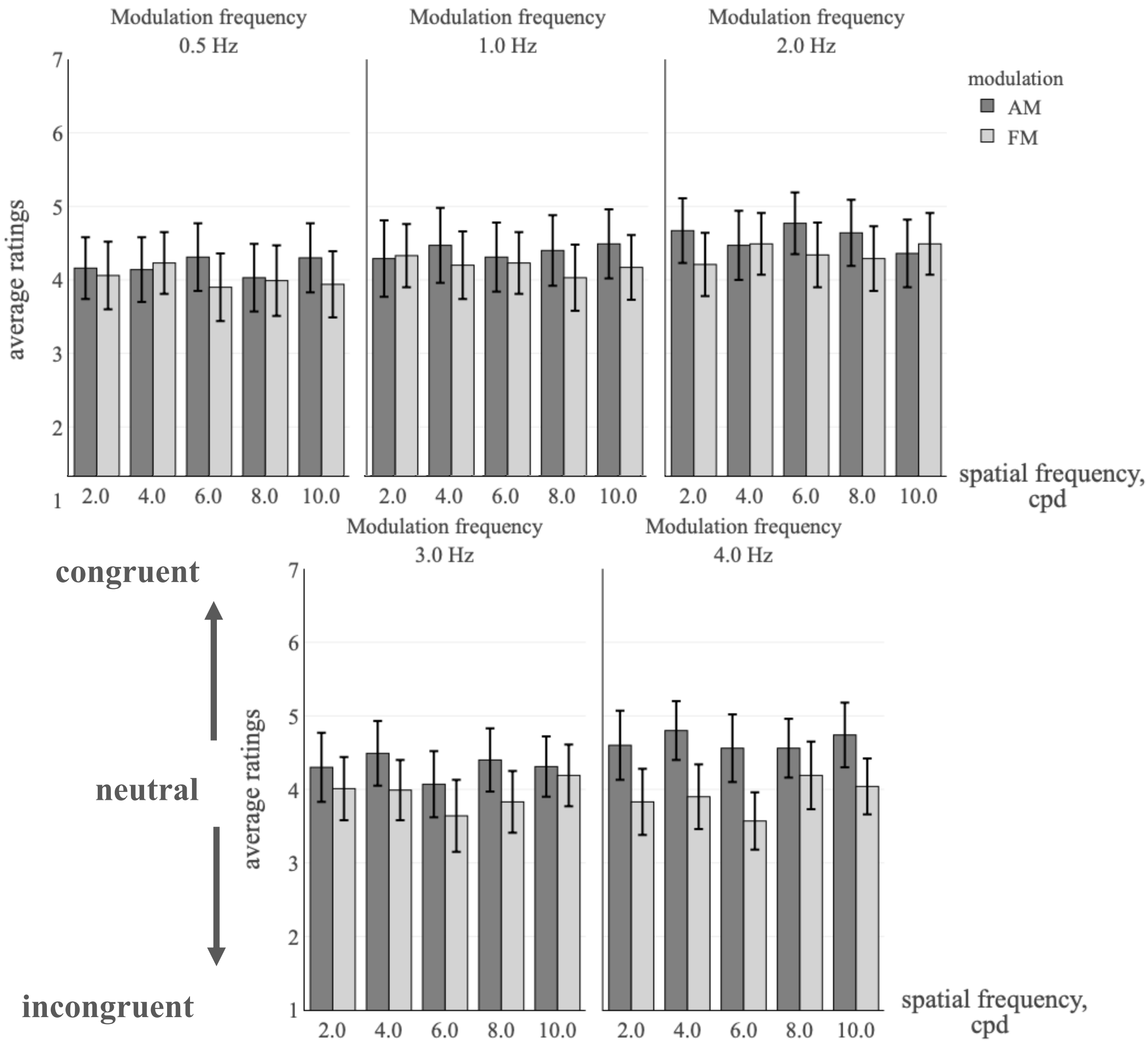

Figure 2.6: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

incongruent (significantly below 4). As expected, the combinations of a Gabor patch and a modulated tone with similar temporal frequencies received significantly higher ratings (see Figure 2.7 and Figure 2.8 for average ratings values and 95%-confidence intervals around the mean) than the combinations of auditory and visual components whose temporal frequencies did not coincide. This is true for both AM- and FM-tone stimuli.

The maximum ratings for stimuli with certain modulation frequencies were always achieved when these stimuli contained a Gabor patch of the flickering frequency equal to this modulation frequency. However, it can be noticed that the

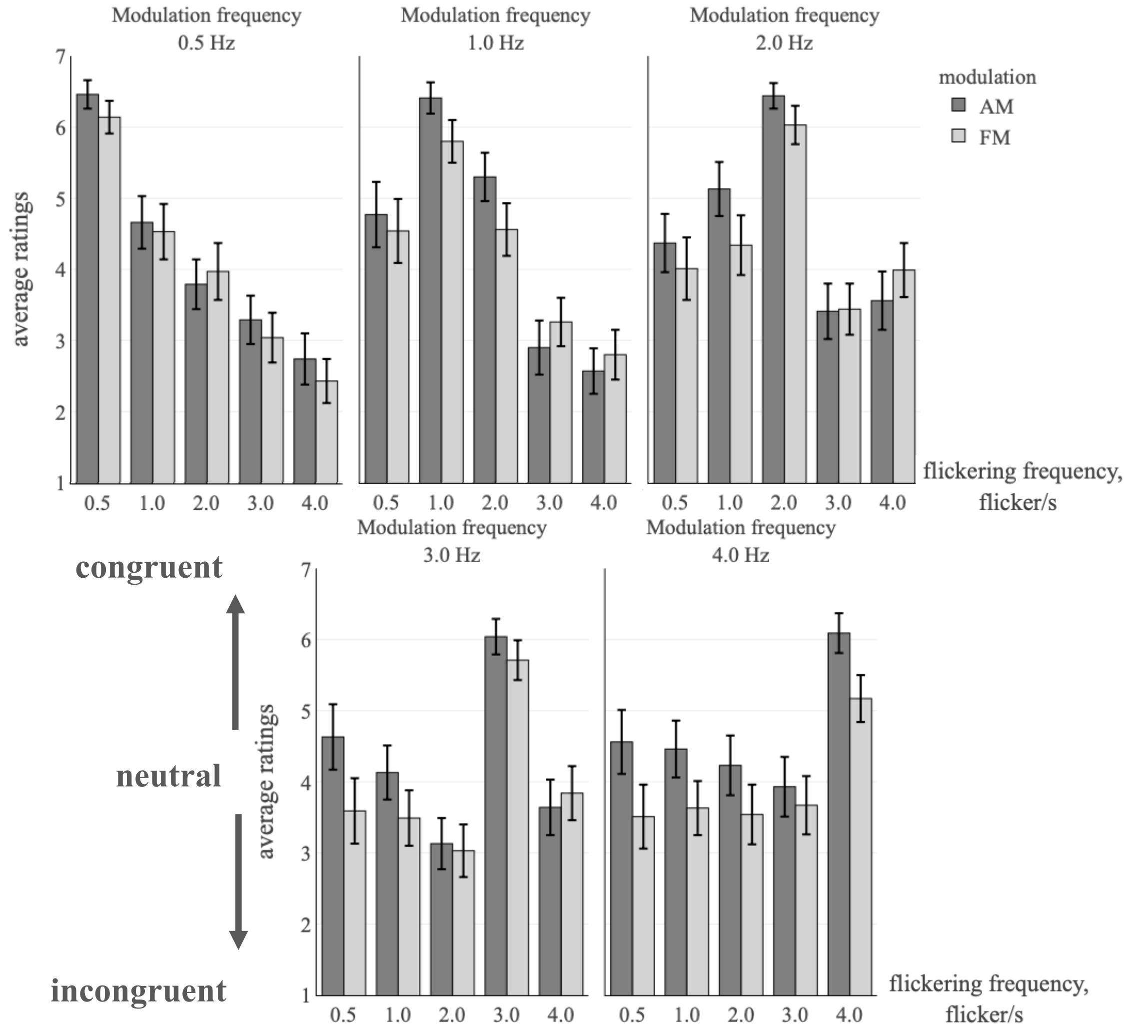

Figure 2.7: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

flickering frequency		AM, Hz					FM, Hz				
		0.5	1	2	3	4	0.5	1	2	3	4
flicker/s	0.5	6.46	4.77	4.37	4.63	4.56	6.14	4.54	4.01	3.59	3.51
	1	4.66	6.41	5.13	4.13	4.46	4.53	5.8	4.34	3.49	3.63
	2	3.79	5.3	6.44	3.13	4.23	3.97	4.56	6.03	3.03	3.54
	3	3.29	2.9	3.41	6.04	3.93	3.04	3.26	3.44	5.71	3.67
	4	2.74	2.57	3.56	3.64	6.09	2.43	2.8	3.99	3.84	5.17

Figure 2.8: Results of Experiment 1, presented as a grayscale heatmap. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The colour demonstrates the degree of (in)congruency: grey - high (in)congruency rating, white - low (in)congruency rating.

stimuli whose flickering frequency and modulation frequency were not equal, but whose patch-flickering versus tone-modulation ratio could be expressed as a whole number, for example, 1 flicker/s and 0.5 Hz, or 1 flicker/s and 2 Hz, got still higher ratings than the stimuli who's flickering-modulation ratio was a fraction, for example 2 flicker/s and 3 Hz. This relation can be expressed and illustrated using the formula for "logarithm of the frequencies ratio" (LFR):

$$LFR = \log_2 \frac{tone\, modulation\, frequency}{flickering\, frequency} \tag{2.1}$$

Using this formula, the stimuli of equal LFR were aggregated and the average ratings were calculated. The resulted function is presented in Figure 2.9.

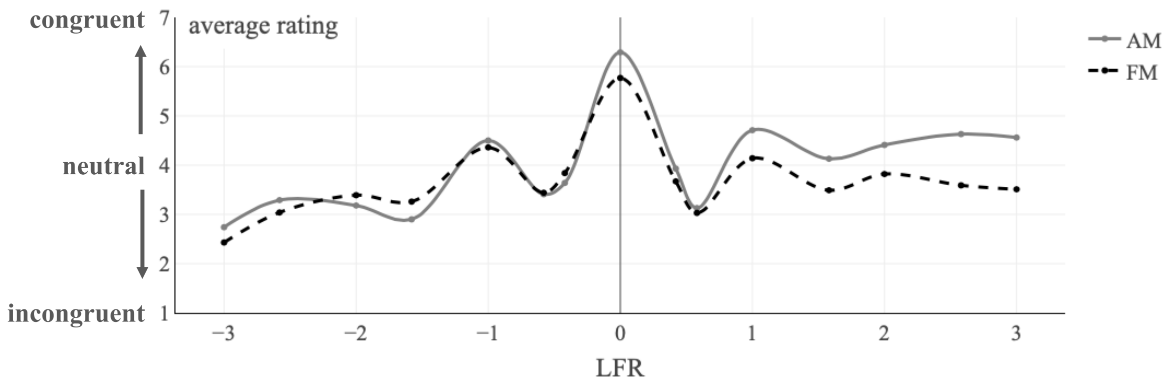

Figure 2.9: Results of Experiment 1. Average ratings of stimuli consisting of a flickering Gabor patch in combination with an AM- or an FM-tone aggregated per LFR (spline interpolation is applied).

The highest average rating was as expected at the point of LFR = 0, i.e., where the tone-modulation frequency and the Gabor patch-flickering frequency were equal. Two roughly symmetrical spikes can be also observed at the point where LFR = 1 and -1, thus where the tone-modulation frequency was exactly double or half the flickering frequency. The local minimums are located at the points where the logarithm is a fraction, for example, for a stimulus with a Gabor patch of 3 flicker/s and a modulated tone of 2 Hz. For these combinations, the temporal mismatch was the strongest. That is, apart from the simultaneous beginning and the end of the signal, none of the other changes in the tone and the patch coincided.

We can also see that AM-tones had higher ratings compared to stimuli with FM-tones, especially at the points of a local maximum (-1, 0, 1). Almost no difference is seen at points where LFR is a fraction, i.e., in between -1 and 0, and 0 and 1 at local minimum points.

To summarise, flickering Gabor patches in combination with modulated tones received higher ratings when combined with AM-tones. The most prominent factor that defined the congruency of these audiovisual stimuli was the similarity of their temporal frequencies: similarity in the flickering frequency of the Gabor patch and the modulation frequency of the tone. The difference between AM- and FM-tone stimuli was especially profound in perceptually congruent combinations of a Gabor patch and a tone.

2.3.2 Drifting Gabor patches

Tone modulation

Similarly to the previous analysis, firstly the ratings for audiovisual stimuli with either an AM- or FM-tone were compared using a sign test. The results showed that stimuli containing a drifting Gabor patch also evoked higher perceived congruency when in combination with an AM-tone rather than with an FM-tone (sign test: $N_{plus} = 701$, $N_{minus}=540$, Z = -4.542, $p < 0.001$).

The analysis of the stimuli with modulated tones of different frequencies demonstrates that stimuli with higher tone-modulation frequencies (AM or FM) received higher average (in)congruency ratings. The only exceptions to that tendency were the stimuli containing tones with a modulation frequency of 0.5 Hz, which were rated as high as the stimuli containing the highest modulation frequency tones of 4 Hz. The average ratings for stimuli with tones of different modulation frequencies are presented in Figure 2.10.

Figure 2.10 also shows that similarly to the stimuli with a flickering Gabor patch, stimuli with a drifting Gabor patch with higher ratings demonstrate greater difference between AM- and FM-tones. Table 2.1 shows a positive correlation between the average ratings (AM and FM together) and the difference between average ratings of stimuli containing AM- and FM-tones.

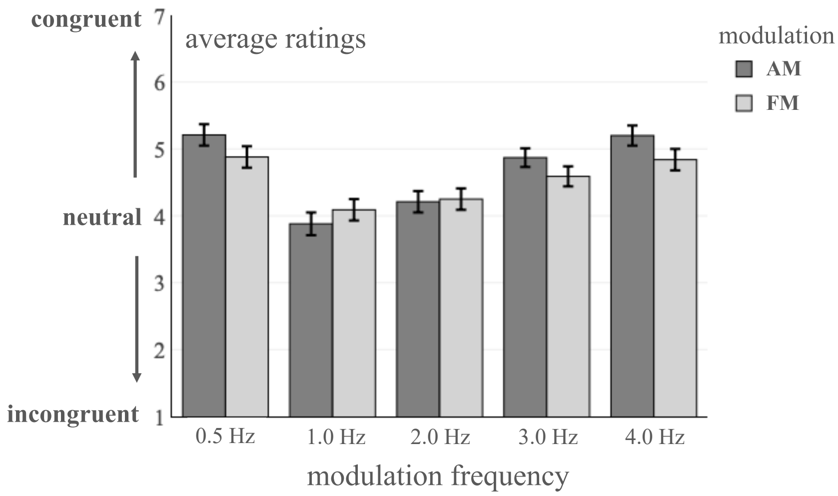

Figure 2.10: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Table 2.1: Results of Experiment 1. The difference between average ratings of stimuli with an AM- and FM-tone and a drifting Gabor patch.

Modulation frequency	Average ratings (AM and FM)	Difference between ratings of stimuli with AM- and FM-tones
1.0 Hz	3.985	-0.21
2.0 Hz	4.23	-0.04
3.0 Hz	4.73	0.28
4.0 Hz	5.02	0.36
0.5 Hz	5.045	0.33

Spatial frequency

Next, as the second objective of this study, the analysis of the average ratings for the groups of stimuli containing Gabor patches of different spatial frequencies was conducted (Fig. 2.11). In contrast with flickering Gabor patches, we can see a tendency to rate stimuli with a drifting Gabor patch slightly different depending on their spatial frequency. The ratings were higher for the patches with a lower spatial frequency and lower for the patches with a higher spatial frequency This tendency can be observed for both stimuli with AM- and FM-tones, but FM-stimuli were rated a bit lower than AM-stimuli for most of the spatial frequencies, demonstrating the difference between AM- and FM-tones that was described in the beginning of this section (sign test for stimuli with AM- and FM-tones: N_{plus} = 701 , N_{minus}=540, Z = -4.542, $p < 0.001$).

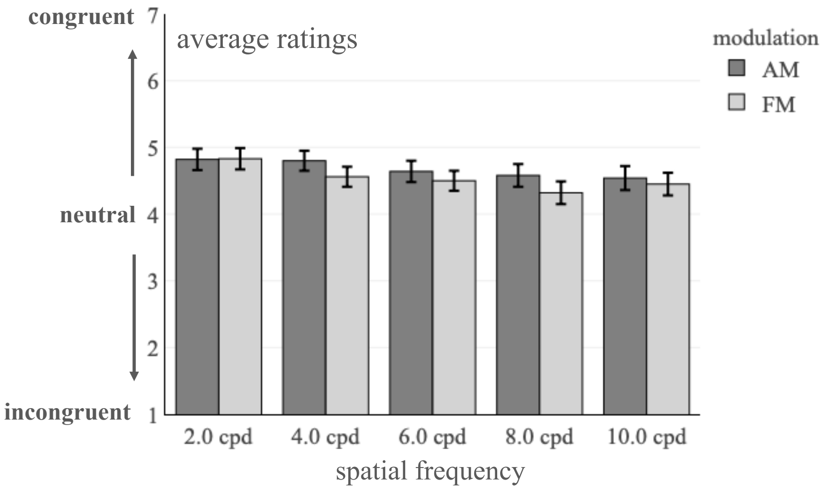

Figure 2.11: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

The average ratings for the stimuli grouped by their spatial frequency and modulation frequency were also calculated. The results are presented in Figure 2.12 and Figure 2.13. The tendency to rate low-spatial frequency stimuli with higher perceived (in)congruency can be also seen here for all modulation frequency-tones except those with a 0.5-Hz modulation.

Speed

Next it was checked if drifting speed affected the perceived congruency. Figure 2.14 demonstrates that stimuli with different drifting speeds were rated quite equally overall. The 95%-confidence intervals suggest that there were no significant differences between stimuli.

As a third objective of the experiment, similarly to the analyses of flickering patches, we calculated the average ratings for the stimuli aggregated by temporal parameters of both the visual and the auditory component, thus for drifting speed and modulation frequency. As can be seen based on the 95%-confidence intervals around the mean in Figures 2.15, no clear rating pattern demonstrating the cor-

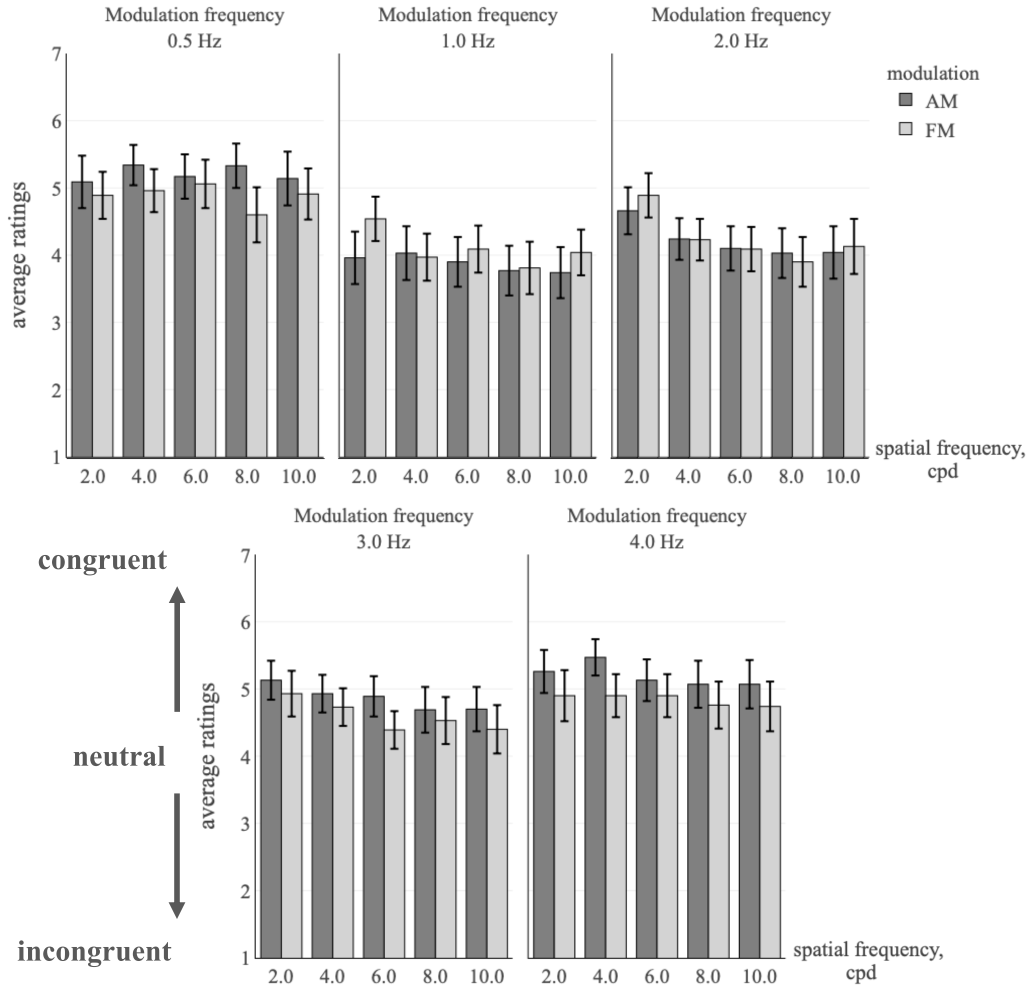

Figure 2.12: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

spatial frequency		AM, Hz					FM, Hz				
		0.5	1	2	3	4	0.5	1	2	3	4
cpd	2	5.09	3.96	4.66	5.13	5.26	4.89	4.54	4.89	4.93	4.9
	4	5.34	4.03	4.24	4.93	5.47	4.96	3.97	4.23	4.73	4.9
	6	5.17	3.9	4.1	4.89	5.13	5.06	4.09	4.09	4.39	4.9
	8	5.33	3.77	4.03	4.69	5.07	4.6	3.81	3.9	4.53	4.76
	10	5.14	3.74	4.04	4.7	5.07	4.91	4.04	4.13	4.4	4.74

Figure 2.13: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The colour demonstrates the degree of (in)congruency: grey - high (in)congruency rating, white - low (in)congruency rating.

respondence between the speed and the modulation frequency or any significant difference in ratings between stimuli containing patches of different speed can be observed.

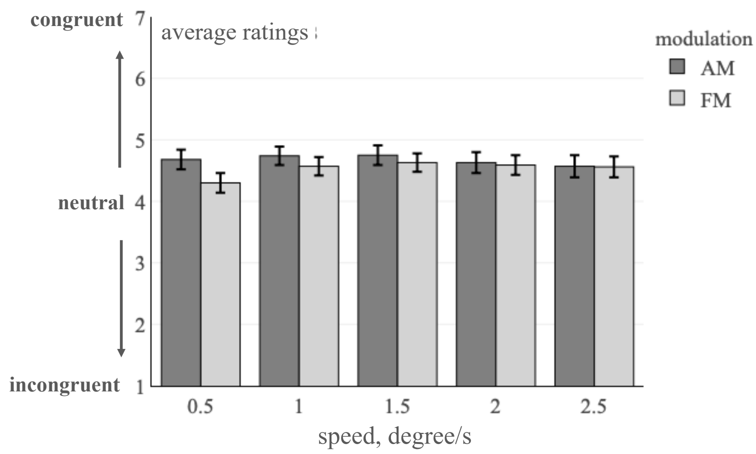

Figure 2.14: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

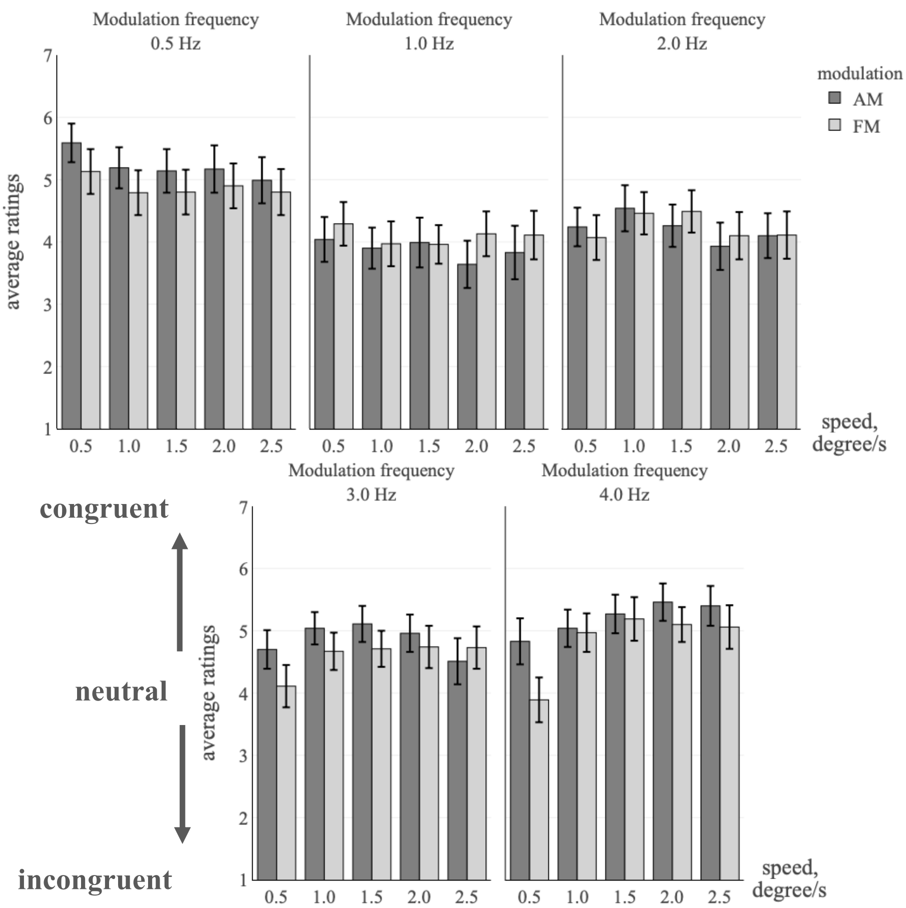

Figure 2.15: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Temporal frequency

Additionally, in the case of drifting stimuli, one derivative parameter should be considered: temporal frequency. In the case of flickering Gabor patches, temporal frequency is the frequency of the contrast change of the whole patch, where every point of a patch changes its contrast simultaneously with others. In the case of drifting Gabor patches, the frequency of luminance fluctuations at one point caused by drifting of black and white gratings should be considered. The temporal (or local flickering) frequency of one such point can be expressed using the following formula (Ashida and Osaka, 1995):

$$Temporal frequency = spatial frequency \times speed \tag{2.2}$$

Every point of the Gabor patch retinal image flickers with such temporal frequency, therefore it might be assumed, that this local flickering frequency is also registered.

The average ratings for the groups of stimuli aggregated by temporal frequency of the drifting patch and the modulation frequency of the tone are presented below in Figure 2.16.

The analysis of data for stimuli aggregated by temporal frequency of a Gabor patch and a modulation frequency of a tone showed a potential pattern similar to the one we observed with flickering Gabor patches. Upon a closer view, the following can be noticed: first, for a modulation frequency of 1 Hz, the maximum rating was achieved for stimuli with a temporal frequency of 1 flicker per second for both AM- and FM-tones. Second, for stimuli with a modulation frequency of 2 Hz, the maximum (in)congruency rating was achieved for stimuli consisting of 2 flicker/s for AM and 3 flicker/s for FM (second high score for 2 flicker/s). Third, for a modulation frequency of 3 Hz the maximum (in)congruency rating was achieved for stimuli consisting of 6 flicker/s for AM, with the second highest for stimuli consisting of 3 flicker/s. Finally, for a modulation frequency of 4 Hz the maximum

spatial frequency		AM, Hz					FM, Hz				
		0.5	1	2	3	4	0.5	1	2	3	4
cpd	1	5.5	4.43	4.29	4.5	4.36	5	4.64	4.43	3.36	3.29
	2	5.5	4.18	4.86	4.82	5.29	5.21	4.36	4.71	4.71	4.21
	3	5.36	4.29	4.61	5.21	5	5.04	4.5	4.82	4.71	4.61
	4	5.29	3.57	4.36	5.17	5.14	4.79	4.21	4.43	4.86	4.93
	5	4.93	3.46	4.25	4.86	5.29	4.89	4.29	4.18	4.82	5
	6	5.21	3.93	4.04	5.25	5.36	5.04	3.86	4.21	4.79	5.29
	8	5.43	4.29	4.04	4.86	5.59	4.68	3.57	4.04	4.75	5.04
	9	5	4.07	4.21	4.79	5.07	4.93	3.64	4.36	4.14	5.43
	10	5.11	4	4.46	4.86	5.21	4.86	4	3.96	4.68	5.07
	12	5.04	3.54	3.96	5.11	5.29	4.79	4.18	4	4.61	4.89
	15	5.18	3.89	4.07	4.64	5.29	4.86	3.93	4.04	4.5	5
	16	5.5	3.79	4	4.43	4.71	4.79	3.71	3.64	4.43	4.79
	20	4.89	3.29	3.64	4.54	5.46	4.61	4.11	4.07	4.39	5.07
	25	5.21	4.29	3.93	4.14	4.93	5.07	4.07	4.36	4.43	4.36

Figure 2.16: Results of Experiment 1. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The colour demonstrates the degree of (in)congruency: grey - high (in)congruency rating, white - low (in)congruency rating.

(in)congruency rating was achieved for stimuli with 8 flicker/s for AM and 9 flicker/s for FM. As can be noticed, similarity of the frequencies seems to play an important role here: the highest congruency is obtained with stimuli containing an auditory and visual component of similar frequencies, i.e., high (low) modulation frequency of the tone and high (low) temporal frequency of the patch. Besides, the numerical values of these auditory and visual frequencies are often (almost) the same or their ratio is a whole number.

Based on these results we can assume that temporal frequency could also play a role in perceived (in)congruency for drifting patches. Similarly to the flickering Gabor patches, stimuli with a drifting Gabor patch received higher ratings when the modulation frequency of the tone and the temporal frequency of the patch were the same or their ratio was a whole number. This analysis, however, cannot lead to a definitive conclusion as the number of cases for some frequencies was too low to produce statistically valid results. This flaw was fixed in the next experiments.

To summarise, stimuli with a drifting Gabor patch also demonstrated higher perceived (in)congruency ratings for the stimuli containing an AM-tone compared to stimuli containing an FM-tone. Also, the analyses showed that spatial frequency of the Gabor patch affected the perceived (in)congruency ratings. However, this effect could be related to the effect of temporal frequency of a drifting Gabor patch and its' (dis)similarity to the modulation frequency of the tone. This should be investigated further in the next experiments.

2.4 Discussion

The first objective of the experiment was to to compare the perceived (in)congruency evoked by AM- and FM-tones. The results demonstrated that the stimuli containing an AM-tone evoked higher perceived congruency than stimuli with FM-tones, and the difference was significant for both stimuli with a flickering patch and stimuli with a drifting patch. The more detailed analysis also showed that AM- and FM-tones demonstrated similar rating patterns for groups of stimuli aggregated by visual or auditory parameters. Furthermore, stimuli with AM-tones showed higher ratings compared to FM-tones especially for congruent stimuli, with ratings significantly above 4 (the rating-scale midpoint). This was also observed in the analysis of the LFR parameter for stimuli with a flickering Gabor patch (Figure 2.1), where points of local and global maximums showed higher ratings for stimuli with AM-tones than for ratings with FM-tones, while local minimums showed almost no difference between the ratings.

Second, the roles of stimulus parameters in evoking perceived congruency were investigated, and specifically if crossmodal correspondence that was reported in the previous studies (Evans and Treisman, 2010; Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a; 2013b) had any effect on perceived (in)congruency ratings. In case of stimuli with a flickering Gabor patch, analysis of visual parameters showed that spatial frequency had no observable effect on the average ratings in the presence of dynamic changes. In case of stimuli with a drifting Gabor patch, the spatial frequency of the patch demonstrated an effect on perceived congruency where stimuli with a Gabor patch with low spatial frequency were rated higher than stimuli with a patch with high spatial frequency. Regardless of the parameters of the tones, participants preferred the stimuli with low spatial frequencies.

Third, the dynamic parameters of the Gabor patch (flickering frequency and speed) and their effect on perceived congruency were analysed. In case of stimuli with a flickering patch, it was observed, that temporal similarity, i.e., similarity between flickering frequency of the patch and the modulation frequency of the tone, played a prominent role in establishing congruency. Besides, when peaks and troughs of dynamic changes coincided partially, for example, when the flickering frequency of the patch was twice as high as the modulation frequency of the tone and vice versa, an increase in average ratings could also be seen (Figure 2.7, 2.8 and 2.9). The analysis of the stimuli with a drifting Gabor patch demonstrated that similarities between the temporal frequency of the patch and the modulation frequency of the tone yielded ratings higher than for other stimuli, however, the effect did not reach significance (according to 95% confidence intervals). It should be taken into account, though, that in this experiment the temporal frequencies of Gabor patches were in the range of 1 - 25 flickers/s with medium 8, while modulation frequencies of tones were in the range of 0.5 - 4 Hz. Therefore, there were no "perfect" combinations for patches with high spatial frequency in terms of the auditory component with an equal temporal (modulation) frequency in this experiment, resulting in lower average ratings for such drifting stimuli. Similarly, as there were more combinations with higher temporal frequencies (see Figure 2.16), the higher ratings were also shifted towards higher modulation frequencies. This was clarified in the following experiments.

Considering these findings, it can be concluded that the effect of crossmodal correspondence was not confirmed here.

Limitations of the study

First, though this pilot study employed a wide range of stimuli, the analysis showed that the range that was used might not have fully satisfied the purposes of the study. In the analysis of the drifting Gabor patches it was noticed that the temporal frequencies of most patches were much higher than the modulation frequency of the tones. The range of temporal frequencies of Gabor patches was from 1 flicker/s to 25 flicker/s, while the tones were modulated only by 0.5 Hz to 4 Hz.

Second, another issue related to the analysis of temporal frequencies concerns the presentation of the stimuli. All the stimuli were presented with no fixation point. Therefore it might be assumed that when the stimuli with a drifting Gabor patch were presented, participants moved their gaze to follow the grating's drift. This stimulus tracking potentially could decrease the effect of temporal frequency of the drifting Gabor patch on perceived congruency as on the retina this patch would appear as a static image. The presentation of the fixation point would fix the gaze and enhance the local flickering effect of the Gabor patch, because for a given point on the retina the temporal frequency of the Gabor patch would be equal to the flickering frequency at this particular point.

Finally, it should be noted that this study mainly focused on the effect of parameters such as the spatial and temporal frequency of the patch and modulation type and frequency of the tone.

2.5 Conclusions

- AM-tones evoked significantly higher perceived congruency with Gabor patches compared to FM-tones, and the difference between ratings was higher for stimuli rated as "congruent" (significantly higher than "4", the scale midpoint).
- Stimuli containing a flickering Gabor patch demonstrated significantly higher ratings when temporal frequencies, i.e., the flickering frequency of the Gabor patch and the modulation frequency of the tone, were similar than when they were dissimilar.
- Stimuli containing a drifting Gabor patch of low spatial frequency received higher (in)congruency ratings than stimuli with a patch of high spatial frequency.

Chapter 3. Experiment 2. Effect of tone modulation type and temporal (dis)similarity

3.1 Introduction

To further clarify perceived (in)congruency of Gabor patches and AM- and FM-tones, Experiment 2 was conducted. The results of the previous experiment demonstrated that AM-tones in combination with Gabor patches evoked higher perceived congruency than FM-tones. Besides, the temporal (dis)similarity between a tone and a patch showed a greater effect on perceived (in)congruency than static features such as the spatial frequency of a patch in case of flickering Gabor patches. In case of drifting Gabor patches, however, some effect of spatial frequency of the Gabor patch on perceived congruency ratings was observed. Therefore, the objectives of Experiment 2 were the same as the objectives of Experiment 1: first, to compare the perceived congruency evoked by AM- and FM-tones; second, to confirm whether static parameters (spatial frequency) have an effect on perceived congruency in the presence of dynamic changes; and finally, to assess the effect of dynamic parameters of the stimuli (flickering or temporal frequency of a patch and modulation frequency of a tone) on perceived congruency.

In the previous experiment, the analyses of temporal frequencies showed that temporal (dis)similarities had a strong effect on perceived congruency in case of flickering Gabor patches. For drifting Gabor patches conclusive results were not found, due to incompatibility between the visual and auditory parameters: the temporal frequency range of the patches did not coincide with the temporal frequency range of the tones. This issue was addressed in this experiment, and the parameters of the Gabor patches and the tones were adjusted so as to match their ranges.

Furthermore, a fixation point was introduced below the stimuli. In the previous experiment no fixation point was used. Therefore, it can be assumed that drifting gratings of the patch caused involuntary tracking of the gratings. In that case the reflection of gratings on the retina is static with only a small possible effect of temporal frequency. When participants use the fixation point, the gaze is fixed and the image on the retina is changing so that one particular point on it is flickering with a frequency equal to the temporal frequency, i.e., speed $\times$ spatial frequency. It was hypothesised that in that case the effect of temporal frequency of the drifting Gabor patch and its (dis)similarity with the modulation frequency of the tone should be stronger.

3.2 Method

The method was the same as in Experiment 1 with some minor differences, which are described below.

3.2.1 Participants

Fourteen participants (8 males, 6 females) took part in the experiment. The mean age was 28 years old, SD = 6 years. All reported normal or corrected-to-normal vision and normal hearing. Thirteen out of 14 participants were university students and 1 was recruited from outside of the university. Prior to the experiment each participant received instructions and an explanation about the purpose of the experiment and the procedure. All agreed to participate and provided written informed consent. The experiment was conducted with prior approval of the Ethics Committee of Kyushu University.

3.2.2 Apparatus

Experiment 2 was conducted in the same sound-attenuated experiment booth as Experiment 1 and employed the same apparatus (see section 2.2.2 for detailed description). The illuminance at the participant level was measured with a TOPCON IM-1000 lux meter and was 2.7 lx for the white screen, 0.9 lx for the grey background, and less than 0.1 lx for the black screen. The application used in Experiment 2 was based on the original application developed for Experiment 1 with some necessary changes, in order to meet the criteria for the new experiment and to eliminate some of the limitations of the previous experiment. The main change was for the inclusion of the fixation point during stimuli presentation (see the full description in Appendix A).

3.2.3 Experiment application and procedure

Experiment 2 consisted of 108 trials divided in 4 sessions conducted on the same day with 3-minute breaks in between.

Instruction was given to the participants prior to the experiment with the detailed description of the experiment flow and explanations on how to navigate the application (see Appendix E for the instruction example). One of the main points in the instruction was the request to keep the eyes on the fixation point throughout the trial. The main task was expressed as follows: "Every stimulus consists of a visual pattern (Gabor patch) and an auditory signal of 2 seconds long. There are pauses before and after the stimuli. 1 second before the stimulus the fixation point will appear, please, fixate your gaze on it. The stimuli will appear above the fixation point, however, do not move your gaze from the fixation point. The gaze should be at the fixation point through the whole time of the trial - before and during appearance of the stimuli . . . After each stimulus, a rating scale will appear. Please, choose the most appropriate rate for the seen stimulus from 1 to 7, where "1" is completely incongruent and "7" is perfectly congruent".

The flow of Experiment 2 is presented in Figure 3.1. The trial started with a 3-second pause during which nothing appeared on the grey screen, giving participants time to prepare for the stimuli. One second before the stimuli the fixation point appeared. The stimulus was located 2 visual degrees above the fixation point and lasted for 2 seconds, after which both the stimulus and the fixation point disappeared. The rating scale appeared 2 seconds after the stimulus. As soon as the response was registered, the scale disappeared. Participants were offered to click the mouse one more time to confirm they were ready for the next trial. As soon as the click was received, the next trial started.

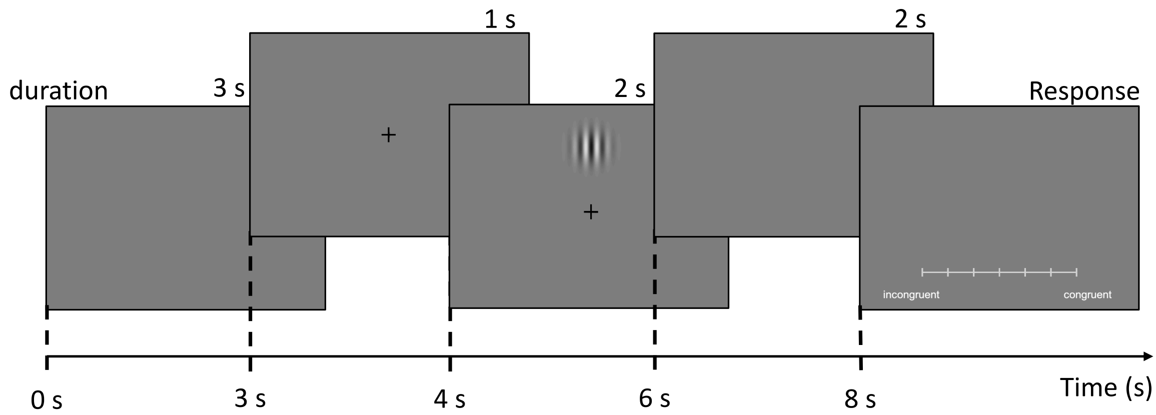

Figure 3.1: Trial flow of Experiment 2

Every experiment session started with a practice session, during which the experimenter was in the same room with the participant to ensure everything was understood and to answer questions in case they appeared (the experiment was performed before the COVID-19 situation, November 2019). Also, for each participant the importance of keeping the eyes on the fixation point during the trial was emphasised during the practice session. Some of the participants initially reported that it was challenging not to move their gaze along with patch gratings, but after 5 practice trails everyone confirmed that they could do it with no excessive efforts.

3.2.4 Stimuli

Similar to the previous experiment, every stimulus of Experiment 2 consisted of a Gabor patch and an AM- or FM-tone (see table in Figure 6.1 on page 86 of General Discussion section for the consolidated table on the parameters in all Experiments).

Gabor patch

The Gabor patches were 2×2 degrees in size with a Gaussian envelope. The luminance of the sinusoidal gratings ranged from 9.47 cd/m^2 (white) to 2.87 cd/m^2 (the same as the grey background) and 0.02 cd/m^2 (black).

Gabor patches varied across the following parameters:

- spatial frequency: 2 cpd, 3 cpd and 4 cpd;
- dynamic mode:
  - flickering: 2, 3 and 4 flickers/s;
  - drifting: 0.5, 1.0 and 1.5 degrees/s.

Gabor patches were generated by the application in a real time during the experiment.

Tones

The AM- and FM-tones used in Experiment 2 were similar to the ones used in Experiment 1 with the following parameters:

- 1000-Hz carrier frequency;
- 500-ms rise and fall time with a cosine-shaped ramp;
- modulation: AM and FM;
- modulation depth:
  - for AM: 1 amplitude;

- for FM: ±80 Hz, half a critical band width for a 1000-Hz tone (Zwicker, 1961);

- modulation frequency: 2 Hz, 3 Hz and 4 Hz. The modulation frequency was the same for AM- and FM-tones, and they were similar to the flickering frequency of flickering Gabor patches.

The tones were generated prior to the experiment using J-language script (see appendix D for the script details).

3.3 Results

3.3.1 Flickering Gabor patches

Tone modulation

With regard to the first objective of this experiment, a sign test for all corresponding pairs of stimuli containing AM- and FM-tones, confirmed the findings of Experiment 1: the perceived congruency of a flickering Gabor patch and a modulated tone was higher for stimuli containing an AM-tone, than for stimuli containing an FM-tone (N_{plus} = 163, N_{minus} = 113, Z = -2.950, p = 0.003). The detailed analysis of this effect in different groups of stimuli, and the effects of other parameters on perceived congruency is presented below.

Spatial frequency

With regard to the second objective, namely the clarification of the roles of static parameters on perceived congruency, the analysis showed that in Experiment 2, spatial frequency also did not affect the perceived congruency ratings of Gabor patches and modulated tones (Fig. 3.2) [χ^2 (df=2, N= 252) = 0.127, p = 0.94].

The analyses of the stimuli aggregated by the spatial frequency of a Gabor patch and modulation frequency of a tone is presented in Figure 3.3. Based on previous findings, it can be assumed that Gabor patches of a certain spatial frequency can

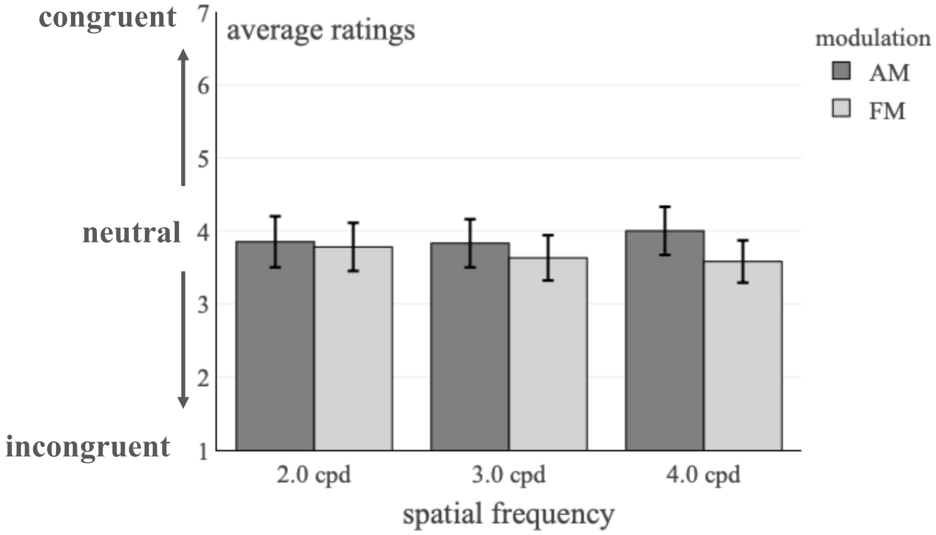

Figure 3.2: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

evoke differences in perceived congruency in combination with tones of different modulation frequencies (Guzman-Martinez et al., 2012). However, no significant effect of spatial frequency of the flickering Gabor patch was found [2-Hz tone modulation frequency: χ^2 (df=2, N= 84) = 0.696, p = 0.71; 3-Hz tone modulation frequency: χ^2 (df=2, N= 84) = 0.095, p = 0.95; 4-Hz tone modulation frequency: χ^2 (df=2, N= 84) = 0.577, p = 0.75].

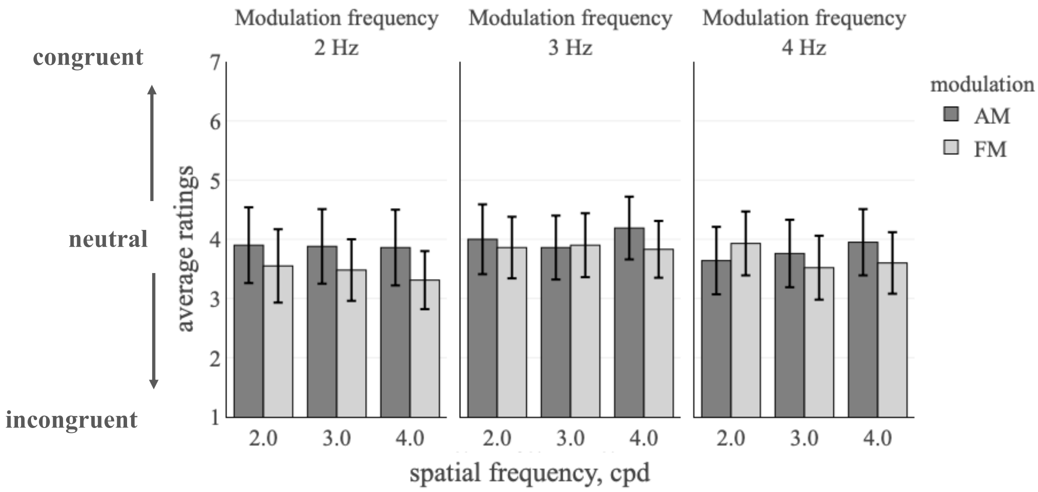

Figure 3.3: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Flickering frequency

With regard to the third objective of this experiment, i.e., the clarification of the role of the dynamic parameters on perceived congruency, the analyses of stimuli aggregated by their visual and auditory temporal frequencies, according to the flickering frequency of a Gabor patch and the modulation frequency of a tone, demonstrated the same rating pattern that was observed in Experiment 1. The 95%-confidence intervals around the mean in Figure 3.4 demonstrate that groups of stimuli with similar temporal frequencies received significantly higher congruency ratings than stimuli with dissimilar visual and auditory temporal frequencies. The highest ratings for all modulation frequencies were given to the stimuli where modulation frequency equals the flickering frequency of the Gabor patch, for example in a stimulus with a 2-Hz tone and 2-flicker/s Gabor patch. The lowest ratings were given to the stimuli in which the difference between the visual and the auditory temporal frequencies were maximum: such as in stimuli with a 2-Hz tone and 4-flicker/s Gabor patch, and 4-Hz tone and 2-flicker/s Gabor patch.

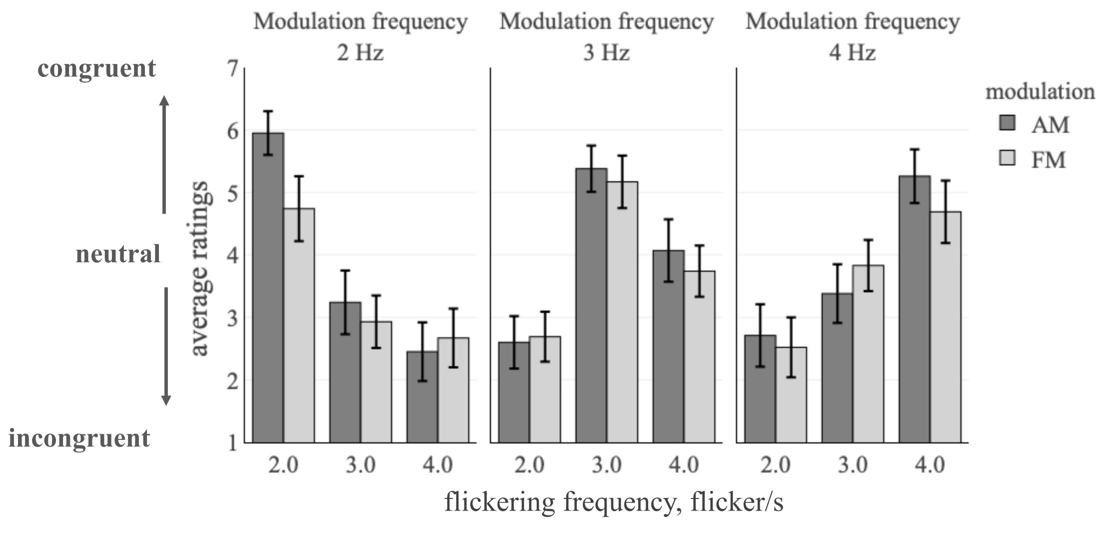

Figure 3.4: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Furthermore, here too the stimuli with high congruency ratings demonstrated the largest difference between stimuli with AM-tones and stimuli with FM-tones.

A prominent effect of tone modulation (AM or FM) could be observed especially for the stimuli containing 2-Hz tones (Figure 3.4). A sign test performed for these stimuli confirmed the statistically significant advantage of amplitude over frequency modulation: $N_{plus} = 58$, $N_{minus} = 35$, Z= -2.281, p = 0.02. A more detailed analysis demonstrated an even stronger effect for the group of stimuli containing only stimuli in which the flickering frequency of the patch and the modulation frequency of the tone were the same: 2-flicker/s Gabor patch and 2-Hz tone: $N_{plus} = 29$, $N_{minus} = 7$, Z = -3.500, $p < 0.001$. Similar analyses conducted for stimuli containing tones of 3-Hz and 4-Hz modulation frequency demonstrated a similar tendency, however the effect was not statistically significant: for stimuli with a 3-Hz tone: $N_{plus} = 54$, $N_{minus} = 40$, Z = -1.341, p = 0.18; for all stimuli with a patch of 3-flicker/s and a 3-Hz tone: $N_{plus} = 19$, $N_{minus} = 12$, Z = -1.078, p = 0.28; for all stimuli with a 4-Hz tone: $N_{plus} = 51$, $N_{minus} = 38$, Z = -1.272, p = 0.20; for stimuli with a patch of 4-flicker/s and a 4-Hz tone: $N_{plus} = 19$, N- = 9, Z = -1.701, p = 0.09.

Figure 3.5 demonstrates the ratings of stimuli grouped by their LFR (logarithm of the frequency ratio, formula 2.1). The roughly symmetrical function with the peak at the point LFR = 0 and lower values on the sides (LFR = -1 and LFR = 1) is similar to the function observed in Experiment 1. It can be also noticed that the points with the highest difference between ratings of AM and FM stimuli are located in the center, while on the sides the average ratings are almost equal. The sign test over the stimuli with LFR = 0 showed a highly significant difference between stimuli with AM-tones and stimuli with FM-tones: $N_{plus} = 67$, $N_{minus} = 28$, Z = -3.899, $p < 0.001$.

The results demonstrated in Figure 3.5 slightly differed from the results in Experiment 1 (Figure 2.9). In Experiment 1 the stimuli with LFR = 1 or -1, i.e, the stimuli whose patch-flickering frequency was exactly half or double of the tone-modulation frequency, received higher ratings than the stimuli with LFR in between -1 and 0, and 0 and 1. We see a different tendency here, where congruency ratings of stimuli with LFR between -1 and 0, and between 0 and 1, are not lower than the ratings of the stimuli with LFR = 1 or -1.

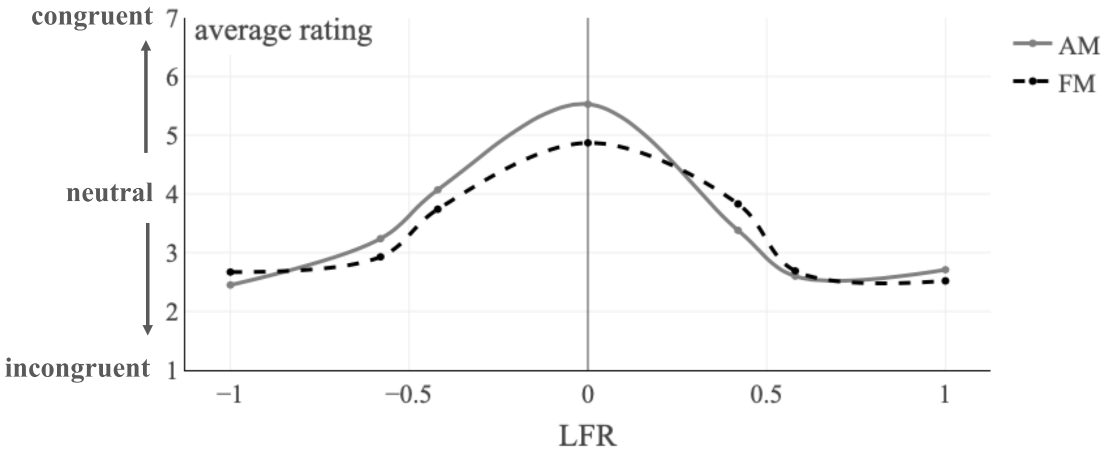

Figure 3.5: Results of Experiment 2. Average ratings of stimuli with a flickering Gabor patch in combination with an AM- or an FM-tone aggregated per LFR (logarithm of the frequencies ratio).

To summarise, Experiment 2 confirmed the main findings of Experiment 1: most of the stimuli containing an AM-tone received higher perceived congruency ratings compared to stimuli containing an FM-tone and the effect was especially strong for stimuli containing a 2-Hz modulated tone. Also, stimuli with an equal auditory and visual temporal frequency received the highest ratings and demonstrated the strongest effect of AM- over FM-modulation on perceived audiovisual congruency.

3.3.2 Drifting Gabor patches

Tone modulation

With regard to the first objective of the study, similarly to the analysis in Experiment 1, for stimuli containing a drifting Gabor patch, the difference between stimuli with AM- and FM-tones was assessed. The average ratings for stimuli with a drifting Gabor patch and an AM-tone were higher than for stimuli with an FM-tone, however, here the effect of modulation type on the perceived congruency ratings did not reach statistical significance: $N_{plus} = 152$, $N_{minus} = 143$, $Z = -0.466$, $P = 0.64$.

Spatial frequency

With regard to the second objective, i.e., the clarification of the roles of static parameters on perceived congruency, the average ratings over spatial frequency of the drifting Gabor patch are presented in Figure 3.6. Contrary to the results of Experiment 1, here there was no tendency that stimuli with a Gabor patch with a low spatial frequency were rated higher.

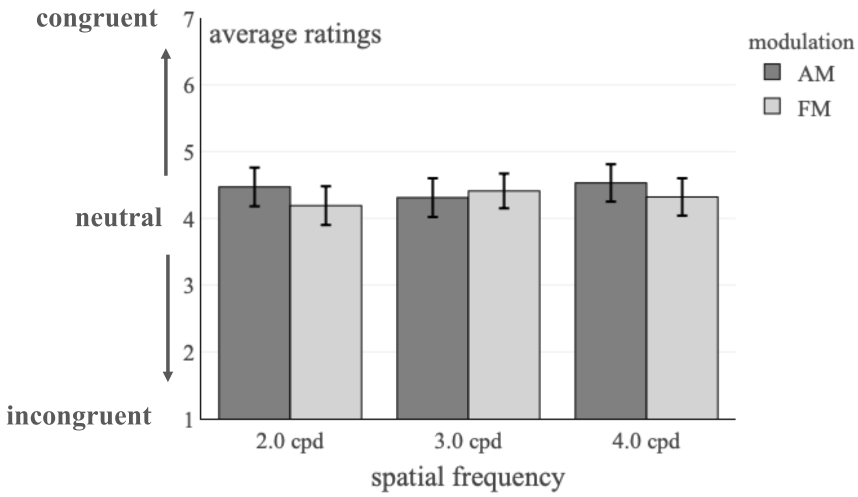

Figure 3.6: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

As it can be seen in Figures 3.6 and 3.7, there were no wide variations in the average (in)congruency ratings between stimuli in which the spatial frequency of a patch and the modulation frequency of a tone were different. However, stimuli in which the Gabor patch had a relatively high spatial frequency received higher congruency ratings in combinations with a tone with a higher modulation frequency, both AM and FM. This effect reached significance in stimuli with modulated tones of 4 Hz [χ^2 (df=2, N=84) = 11.970, p = 0.003], but not in stimuli with modulated tones of 2 Hz [χ^2 (df=2, N=84) = 3.024, p = 0.22] or 3 Hz [χ^2 (df=2, N=84) = 0.185, p = 0.91]. Separate analyses of the stimuli with AM- and FM-tones showed that the effect of spatial frequency on congruency ratings was stronger in stimulus pairs containing 4-Hz FM tones than 4-Hz AM tones, although the effect tended towards significance for the latter as well [FM 4 Hz: χ^2 (df=2, N=42) = 7.75, p =

0.02; AM 4 Hz: χ^2 (df=2, N=42) = 4.762, p = 0.09].

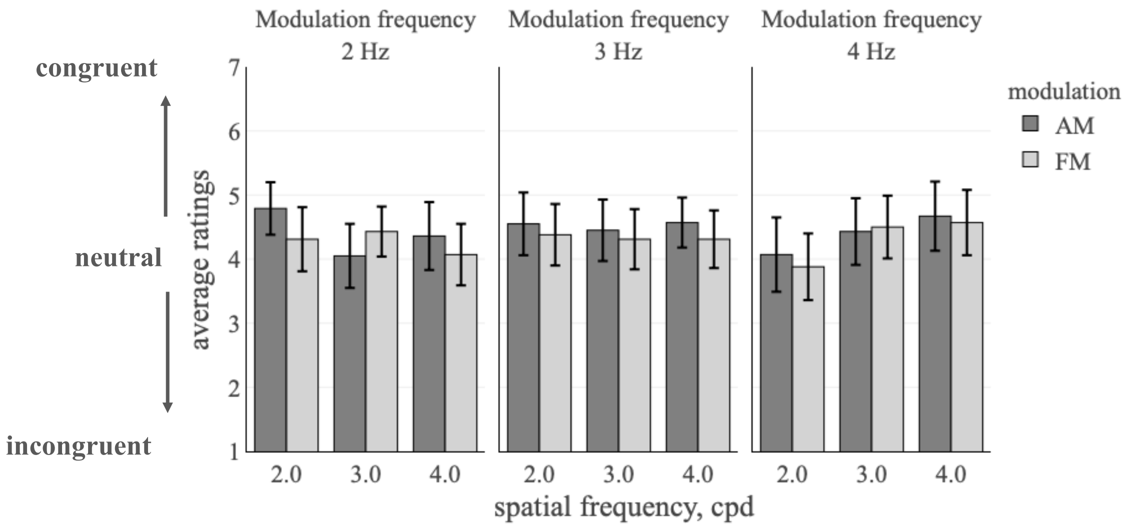

Figure 3.7: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Speed

Next, the effect of dynamic parameters of the Gabor patch, namely the effect of speed and temporal frequency, was assessed. Figure 3.8 demonstrates the average ratings over drifting speed of a Gabor patch, with 95%-confidence intervals around the mean. A statistically significant effect of drifting speed can be clearly observed: stimuli with a higher patch-speed received higher ratings (χ^2 (df = 2, N = 252)= 29.728, $p < 0.001$).

More detailed analysis of the effect of the drifting speed and tone-modulation frequency showed that the (dis)similarities in temporal characteristics of the auditory and visual components, affected perceived congruency (Figure 3.9). The results for stimuli with a modulation frequency of 3 Hz and of 4 Hz demonstrate a rating pattern similar to the one in Figure 3.8: when combined with a Gabor patch with a higher drifting speed, the perceived congruency was higher. For the stimuli containing a 4-Hz modulated tone, based on the 95%-confidence intervals, the stimuli containing 1.5-degree/s Gabor patch were rated significantly higher than the stimuli with 0.5-degree/s Gabor patch.

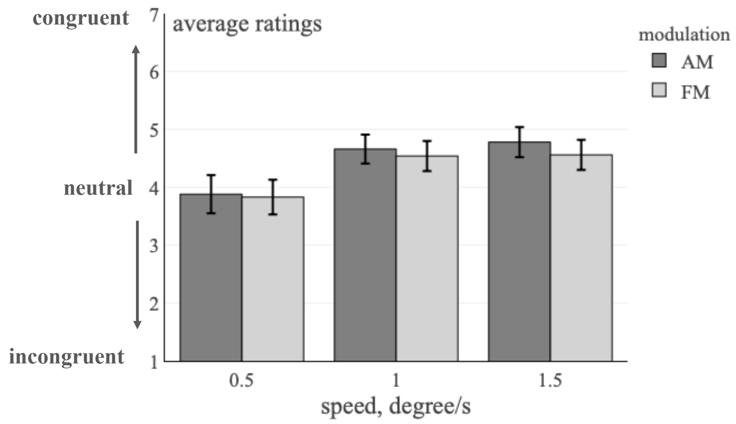

Figure 3.8: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

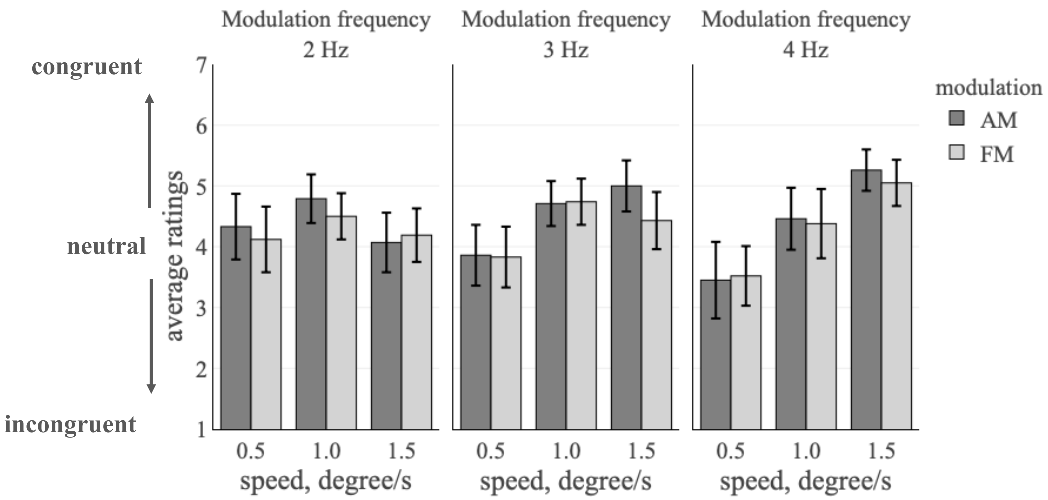

Figure 3.9: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

The Friedman test results comparing stimuli containing different modulated tones demonstrated a significant difference between (in)congruency ratings for stimuli with a 1.5-degree/s Gabor patch, where stimuli containing a 2-Hz tone were rated lower than stimuli with other tones: (χ^2 (df = 2, N = 84) = 15.643, p < 0.001), and for stimuli with a 0.5-degree/s Gabor patch, where stimuli containing a 4-Hz tone were rated higher than stimuli containing a 2-Hz tone: χ^2 (df = 2, N = 84) = 11.006, p = 0.004). No significant difference was found for the groups of stimuli containing a 1.0 degree/s Gabor patch: χ^2 (df = 2, N = 84) = 0.768, p = 0.68.

Therefore it might be concluded that compared to static parameters of the patch (spatial frequency), the dynamic parameters, i.e., speed, had a stronger effect on perceived congruency for stimuli used in Experiment 2.

Temporal frequency

As it was described in section 2.3.2 (subsection "Temporal frequency"), the drifting Gabor patch can be considered as a set of separated points that flicker with the same flickering frequency, but have a gradual phase shift between them. The local flickering frequency is calculated using Formula 2.2 (page 32) and referred to as temporal frequency (Ashida and Osaka, 1995). The average ratings of the groups of stimuli aggregated by temporal frequencies of the drifting Gabor patches and modulation frequencies of the tones are presented in Figures 3.10 and 3.11. These results demonstrate rating tendencies that are similar to those observed for the stimuli with a flickering Gabor patch.

The stimuli with a patch with a low temporal frequency and a tone with a high modulation frequency received the lowest average (in)congruency ratings. However, when the temporal frequency of the patch and the modulation frequency of the tone were approximately similar, (in)congruency scores were higher.

For temporal frequencies of 4.5 flicker/s and 6 flicker/s the differences between ratings of stimuli in combinations with different modulation frequencies was statistically significant as shown by Friedman test: for 4.5 flicker/s χ^2 (df = 2, N = 28) = 13.196, p = 0.001; for 6 flicker/s χ^2 (df = 2, N = 28) = 8.696, p = 0.01. In both cases (in)congruency ratings were significantly higher for stimuli containing a 4-Hz modulated tone compare to stimuli with a 2-Hz and a 3-Hz tone. The differences between ratings of the stimuli with Gabor patches of 1 flicker/s temporal frequency bordered on significance: χ^2 (df = 2, N = 28) = 5.839, p = 0.05 (the stimuli containing a 4-Hz modulated tone received lower ratings compare to stimuli with other tones). The results for other temporal frequencies demonstrated similar tendencies, but did not reach statistical significance: for 1.5 flicker/s χ^2 (df = 2, N = 28) = 3.554, p = 0.17; for 2 flicker/s χ^2 (df = 2, N = 56) = 3.813, p = 0.15; for 3 flicker/s

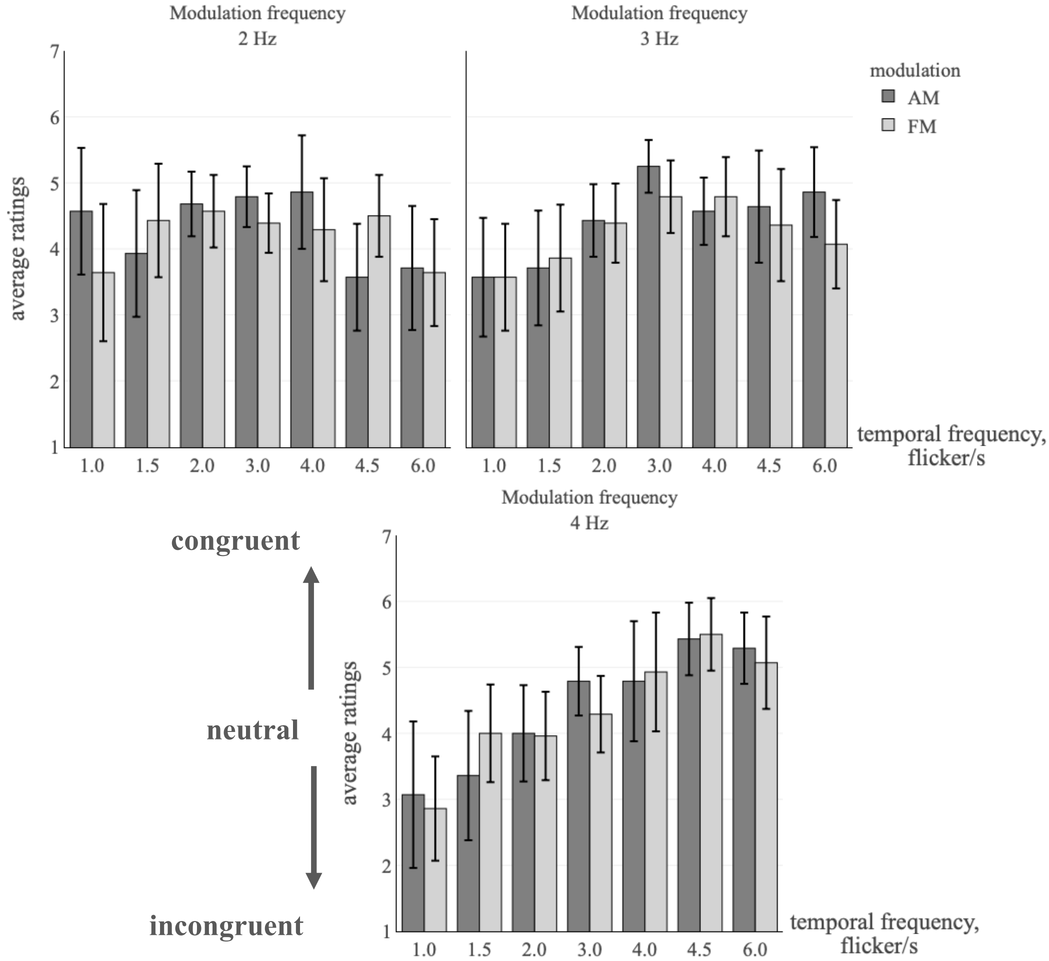

Figure 3.10: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

temporal frequency		AM, Hz			FM, Hz		
		2	3	4	2	3	4
flicker/s	1.0	4.57	3.57	3.07	3.64	3.57	2.86
	1.5	3.93	3.71	3.36	4.43	3.86	4
	2.0	4.68	4.43	4	4.57	4.39	3.96
	3.0	4.79	5.25	4.79	4.39	4.79	4.29
	4.0	4.86	4.57	4.79	4.29	4.79	4.93
	4.5	3.57	4.64	5.43	4.5	4.36	5.5
	6.0	3.71	4.86	5.29	3.64	4.07	5.07

Figure 3.11: Results of Experiment 2. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The colour demonstrates the degree of (in)congruency: grey - high (in)congruency rating, white - low (in)congruency rating.

χ^2 (df = 2, N = 56) = 3.857, p = 0.15; for 4 flicker/s χ^2 (df = 2, N = 28) = 0.375, p = 0.83.

To summarise, the results could not confirm the advantages of AM-tones over FM-tones in evoking perceived congruency in stimuli with a drifting Gabor patch and a modulated tone. However, the importance of (dis)similarity in the temporal frequencies between the patch and the tone was demonstrated clearer in Experiment 2 compared to Experiment 1. Also, contrary to the results of Experiment 1, here the spatial frequency demonstrated no effect on perceived (in)congruency ratings, while drifting speed showed a significant effect on ratings as a stand-alone parameter as well as in combination with modulation frequency.

3.4 Discussion

The method and the objectives of Experiment 2 were the same as the objectives of Experiment 1.

The first objective of Experiment 2 was to compare the perceived congruency of stimuli with an AM-tone and stimuli with an FM-tone. The results of Experiment 2 demonstrated that AM-tones evoked higher perceived congruency in stimuli with a flickering Gabor patch compared to FM-tones in most of the cases. Furthermore, here too the strongest effect of modulation type was observed for the congruent groups of stimuli, i.e., stimuli with the average ratings significantly higher than 4 (midscale) based on 95%-confidence intervals. This can be clearly seen for stimuli aggregated by flickering frequency of the patch and modulation frequency of the tone (Figure 3.4), as well as in Figure 3.5 that demonstrates the LFR function for AM- and FM-tones, and where the highest difference between average ratings of stimuli with AM- and FM-tones can be seen at the point of LFR = 0, i.e., where modulation frequency of the tone is equal to the flickering frequency of the patch. It is logical to assume that while modulation creates temporal congruency with a flickering Gabor patch, amplitude modulation enhances it. Regardless of the modulation type, stimuli with (dis)similar auditory and visual temporal frequencies were often given a low rating (below 4), and modulation type had no significant effect on the ratings in that case. On the other hand, in the case of stimuli that were mostly rated as “congruent” (stimuli with similar patch-flickering and tone-modulation frequencies), the ratings were even higher in cases where stimuli contained an AM-tone compared to stimuli with an FM-tone.

In the case of drifting Gabor patches, the difference between ratings of stimuli with AM-tones and FM-tones was not confirmed. The ratings for stimuli containing AM-tones were on average higher than the ratings of the stimuli containing FM-tones, however, the difference did not reach statistical significance in that case.

The second objective of Experiment 2 was to confirm whether spatial frequency (a static parameter) had an effect on perceived congruency. In Experiment 1 the stimuli with a Gabor patch of a lower spatial frequency were rated higher than stimuli with a higher spatial-frequency patch. However, in Experiment 2 the analysis demonstrated no difference between groups of stimuli with a different spatial-frequency patch. The tendency that was observed in Experiment 1 might have been related to the temporal frequency effect, and the fact that stimuli in Experiment 1 were not well balanced in terms of temporal frequencies of drifting Gabor patches and modulated tones. In Experiment 1 only stimuli with a low-frequency Gabor patch could be combined with a tone of a similar frequency. Therefore, these stimuli had an advantage, and as a result they received on average higher ratings. In Experiment 2 the parameters of the Gabor patches and modulated tones were adjusted in a way that made temporal frequencies of Gabor patches and modulated frequency of the tones approximately in the same range (1 - 6 flicker/s for drifting Gabor patches and 2 - 4 Hz for modulated tones), and the analysis of these stimuli showed no difference between average ratings of stimuli with a different spatial frequency patch.

The third objective of Experiment 2 was to assess the effect of dynamic parameters on perceived audiovisual congruency. The comparative analysis of stimuli aggregated by parameters of the patch and/or parameters of the tone, showed that the most important factor that affects the congruency between the modulated tone and the Gabor patch was the similarity between the flickering or temporal frequency of the patch and modulation frequency of the tone. This confirms the findings of Experiment 1. The effect is especially prominent in case of stimuli with a flickering Gabor patch, where stimuli with similar patch-flickering frequency and tone-modulation frequency received ratings significantly higher than the stimuli with dissimilar frequencies between the patch and a tone. In case of a stimulus with a drifting patch, the analysis of dynamic parameters, speed and modulation frequency, demonstrated a significant effect of speed on perceived (in)congruency ratings. The stimuli containing modulated tones of higher frequencies were rated significantly higher when they contained a Gabor patch of high speed than when they contained a Gabor patch of low speed (Figure 3.9). Furthermore, analysis of the temporal frequency of a patch demonstrated a clearer effect of temporal (dis)similarities in Experiment 2 than was observed in Experiment 1. The analysis of these stimuli demonstrated a rating pattern where stimuli with a drifting Gabor patch and a tone with modulation frequency similar to the temporal frequency of the patch received higher ratings. Therefore, it might be concluded that the dynamic parameters of Gabor patches (flickering, temporal frequency or speed) had a stronger effect on perceived congruency than static parameters (spatial frequency).

Limitations of the study

First, though the importance of dynamic (temporal) parameters of stimuli over static (spatial frequency) was demonstrated in Experiment 1 and 2, in both experiments tones of only one carrier frequency were used. Previous studies demonstrated cross-modal correspondence between spatial frequency of a Gabor patch and perceived pitch of the tone (Evans and Treisman, 2010). Hence the lack of evidence of an effect of crossmodal modulation between the Gabor patch and the tone might be related to the insufficient range of stimuli, i.e., the lack of variation in the carrier frequency of the tone. Therefore, for further research, various carrier frequencies of modulated tones should be used.

Second, in drifting Gabor patches only one drifting direction was used. Though the auditory signal is presented through the headphones with no lateralization, other drifting directions should be balanced.

3.5 Conclusions

- Stimuli with an AM-tone demonstrated on average higher perceived (in)congruency ratings than stimuli with an FM-tone in case of flickering Gabor patches. Stimuli with a drifting Gabor patch demonstrated that tendency in Experiment 1, however, in Experiment 2 the difference between the ratings of stimuli with an AM-tone and an FM-tone did not reach significance.

- The results of Experiment 1 and Experiment 2 suggest that the perceived congruency of stimuli consisting of a Gabor patch and a modulated tone is defined mostly by the dynamic parameters of a patch (flickering and temporal frequencies, and drifting speed) and modulation frequency of a tone. A static property such as spatial frequency has less or no impact on perceived congruency.

Chapter 4. Experiment 3. Effect of a carrier frequency of a tone

4.1 Introduction

In the previous experiments it was observed that AM-tones in combinations with Gabor patches evoked higher perceived congruency than FM-tones. This effect reached significance for both flickering and drifting Gabor patches in Experiment 1, and only for flickering Gabor patches in Experiment 2. Furthermore, the analysis of static (spatial frequency) and dynamic (flickering frequency, speed and temporal frequency) parameters of Gabor patches demonstrated that dynamic rather than static parameters played a prominent role in defining perceived congruency in dynamic audiovisual stimuli. The analysis of flickering Gabor patches showed the same results in both experiments: first, there was no effect of spatial frequency on perceived congruency, and, second, there was a significant effect of (dis)similarity between flickering frequency of the patch and modulating frequency of the tone on perceived (in)congruency ratings. In case of drifting Gabor patches the results varied, but Experiment 2 demonstrated that similar to flickering Gabor patches, here too dynamic parameters as well as the temporal (dis)similarities had a stronger influence on perceived congruency than static parameters.

Evans and Treisman (2010) demonstrated crossmodal correspondence between the spatial frequency of a Gabor patch and a perceived pitch of a tone. Following this research, various studies used the principal of combining high (low) spatial frequency Gabor patches with high (low) tones to create (assumingly) congruent audiovisual stimuli and high (low) spatial frequency Gabor patches with low (high) tones to create (assumingly) incongruent audiovisual stimuli (Heron et al., 2012; Altieri et al., 2015; Green et al., 2019). However, the results of these studies are ambiguous. Besides, there are no conclusive data on whether the above-mentioned principle holds for dynamic stimuli longer than a second. In Experiment 1 and 2 the effect of spatial frequency was not convincingly observed, however, tones of only one carrier frequency were used. To test the above-mentioned principle further and to investigate the effect of carrier frequency on perceived congruency, new auditory stimuli with various carrier frequencies were implemented in Experiment 3.

The first purpose of Experiment 3 therefore was to clarify whether AM-tones evoke higher perceived congruency than FM-tones in case of stimuli with a drifting Gabor patch, as the analysis of Experiment 1 and Experiment 2 did not yield the same results. The second purpose was to investigate the effect of a carrier frequency of the tone on perceived congruency. Similar to Experiment 1 and Experiment 2, participants were asked to rate the stimuli on the scale from 1 (incongruent) to 7 (congruent). In accordance with the previous research (Evans and Treisman, 2010) that demonstrated crossmodal correspondence between the carrier frequency of the tone and spatial frequency of a patch, it was expected to see an effect of carrier frequency on perceived audiovisual congruency.

4.2 Method

The same method and experimental settings were used in Experiment 3 as in the two previous experiments. The minor adjustments that were made in order to reach the objectives of Experiment 3, are described below.

4.2.1 Participants

Thirteen participants took part in Experiment 3. The average age of the participants was 30 years old (SD = 6), 9 were females, 4 were males. Nine of them were students at Kyushu University, 4 were recruited from outside of the university. All reported normal or corrected to normal vision and normal hearing, and provided written informed consent. The experiment was conducted with prior approval of the Ethics Committee of Kyushu University.

4.2.2 Apparatus

The same settings, including the use of the experimental room and the equipment, were used in Experiment 3 as in Experiments 1 and 2 (see section 2.2.2 for the detailed description). The illuminance at the participant level was measured using a TOPCON IM-1000 lux meter and was as follows: 2.7 lx for the white screen, 0.9 lx for the grey background and less than 0.1 lx for the black screen.

4.2.3 Experiment Application and Procedure

The same application was used for Experiment 3 as in Experiment 2 with minor changes regarding the stimuli. There were 144 trials in this experiment divided in 4 sessions with three-minute breaks in between. Instructions were given to the participants prior to the experiment (see Appendix E for the instruction example). Similar to Experiment 2 the importance of keeping the eyes on the fixation point was expressed together with the main task: “Every stimulus consists of a visual pattern (Gabor patch) and an auditory signal of 2 seconds long. There are pauses before and after the stimuli. 1 second before the stimulus the fixation point will appear, please, fixate your gaze on it. The stimuli will appear above the fixation point, however do not move your gaze from the fixation point. The gaze should be at the fixation point through the whole time of the trial - before and during appearance of the stimuli . . . After each stimulus, a rating scale will appear. Please, choose the most appropriate rate for the seen stimulus from 1 to 7, where “1” is completely incongruent and “7” is perfectly congruent”. The flow of the experiment was exactly the same as in Experiment 2 (Figure 3.1). Here too, each experiment started with a practice session consisting of 5 randomly chosen trials.

4.2.4 Stimuli

Stimuli consisted of a Gabor patch and a modulated tone with the following parameters.

Gabor patches

All Gabor patches were 2×2 degrees in size with a Gaussian envelope and black-and-white vertical sinusoidal gratings ranging from 9.8 cd/m^2 (white) to 0.02 cd/m^2 (black) presented on a grey background of 2.9 cd/m^2 (see table in Figure 6.1 on page 86 of General Discussion section for the consolidated table on the parameters in all Experiments).

Gabor patches used in Experiment 3 had the following parameters:

- length: 2 seconds;
- spatial frequency: 2 cpd, 4 cpd, 6 cpd;
- dynamic mode:
  - Flickering with flickering frequencies of 2 flicker/s and 4 flicker/s;
  - Drifting with the speed of 1 degree/s (left to right drift) and -1 degree/s (right to left drift).

The Gabor patches were generated by the experimental application in real time during the experiment.

Tones

AM- and FM-tones with the following parameters were used for Experiment 3:

- length: 2 seconds (same as the Gabor patches);
- rise and fall time: 500 ms, with a cosine-shaped ramp;
- carrier frequency: 500 Hz, 1000 Hz and 2000 Hz;

- modulation:
  - AM;
  - FM;
- modulation depth:
  - for AM: 1 amplitude;
  - for FM: varied for tones of different carrier frequency. The exact values were calculated using the formula for equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990) and were as follows: 39 Hz for 500 Hz tones, 66 Hz for 1000 Hz tones, 120 Hz for 2000 Hz tones.

All modulated tones were generated prior to the experiment (see Appendix D for the J-language script information). All together there were 12 Gabor patches (3 spatial frequencies × 2 flickering frequencies for flickering Gabor patches and 3 spatial frequencies × 2 drifting speeds for drifting Gabor patches) and 12 tones (3 carrier frequencies × 2 modulation frequencies × 2 modulation type - AM and FM) resulting in 144 stimuli.

4.3 Results

4.3.1 Flickering Gabor patches

Tone modulation

Firstly, the effect of modulation frequency that was observed in the previous two experiments was checked here. Stimuli with AM-tones received significantly higher ratings, than stimuli with FM-tones ($N_{plus} = 208$, $N_{minus} = 118$, $Z = -4.929$, $p < 0.001$), confirming the results of Experiment 1 and 2.

Carrier frequency

With regard to the first objective of the experiment, the results demonstrated that participants did not rate stimuli with modulated tones of different carrier frequencies differently. Figure 4.1 shows that groups of stimuli with tones of 500 Hz, 1000 Hz, and 2000 Hz received similar ratings. It also can be noticed that for all carrier frequencies, stimuli with AM-tones showed higher ratings than stimuli with FM-tones.

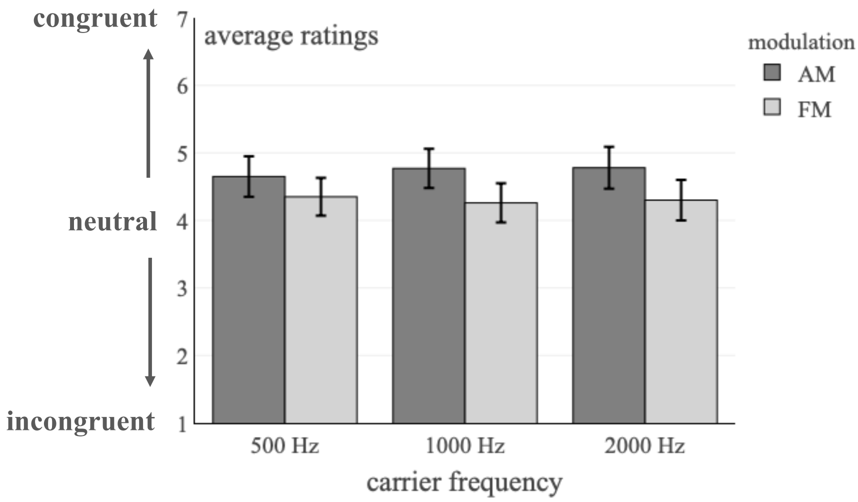

Figure 4.1: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

With regard to the second objective of this experiment, according to the previous studies (Evans and Treisman, 2010), there is a crossmodal correspondence between the spatial frequency of the Gabor patch and the frequency (perceived pitch) of the tone. Here, the perceived (in)congruency of the groups of stimuli aggregated by these two parameters was analysed. Figure 4.2 shows that no effect of carrier frequency on perceived congruency was obtained.

Similarly, Figure 4.3 also demonstrates no effect of carrier frequency on the perceived congruency of Gabor patch stimuli with different flickering frequency.

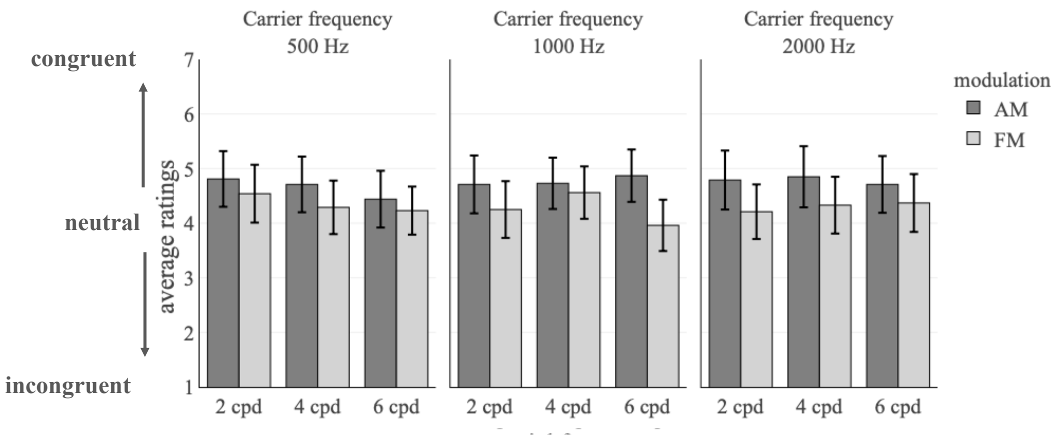

Figure 4.2: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

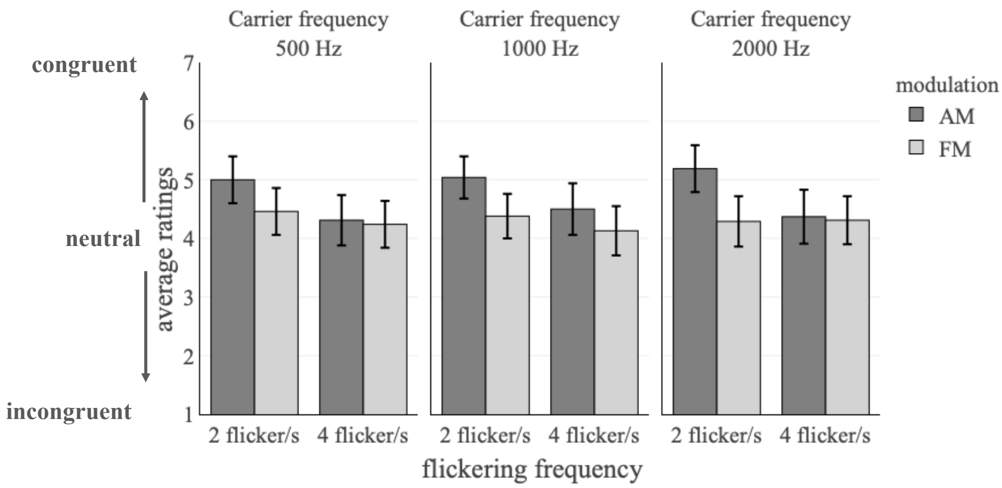

Figure 4.3: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Flickering frequency and modulation frequency

Similar to previous experiments, the groups of stimuli aggregated by dynamic parameters of the auditory and visual component were analysed. Figure 4.4 shows that temporal (dis)similarities had a prominent effect on perceived congruency, as was also observed in Experiment 1 and Experiment 2. In Experiment 2, the stimuli containing tones with a modulation frequency of 2 Hz contributed the most to the modulation-frequency effect as they showed the biggest difference between stimuli with AM- and FM-tones. Here, however, stimuli with a modulation frequency of 4 Hz showed greater difference according to Figure 4.4. This was confirmed by the sign test: for stimuli with 2-Hz tones $N_{plus} = 91$, $N_{minus} = 74$, Z = -1.246, p = 0.21; for stimuli with 4 Hz tones $N_{plus} = 117$, $N_{minus} = 44$, Z =-5.674, $p < 0.001$. Besides, in case of stimuli containing a Gabor patch of 2 flicker/s and a modulated tone of 4 Hz the difference was significant as can be seen by the 95%-confidence intervals. Futhermore, the stimuli with AM-tones were rated significantly above the midpoint, i.e, as congruent, while stimuli with FM-tones were rated significantly below 4, i.e, as incongruent, demonstrating that in some cases stimuli with the same visual component and modulated tones of the same frequency can be perceived as congruent or incongruent depending on the modulation type.

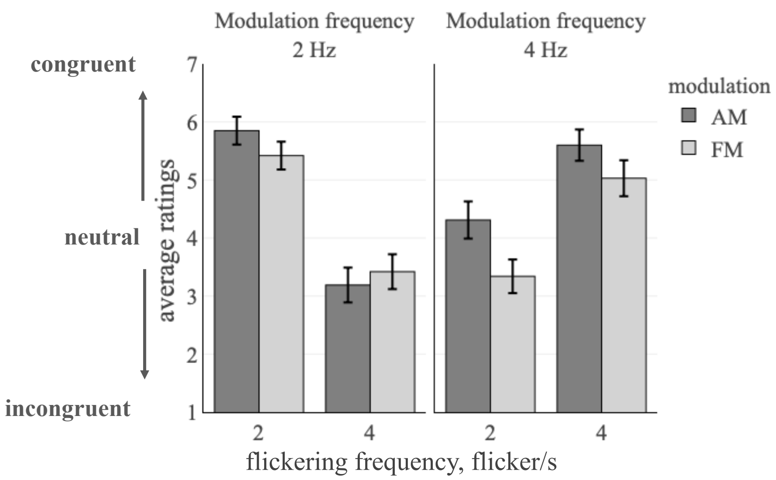

Figure 4.4: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

4.3.2 Drifting Gabor patches

Tone modulation

First, the effect of modulation type was assessed. The effect that was observed in Experiment 1 for drifting Gabor patches and in both experiments (1 and 2) for flickering Gabor patches, where stimuli with AM-tones were rated higher than stimuli with FM-tones, was not confirmed here. The results of a sign test were as follows: $N_{plus} = 155$, $N_{minus} = 180$, Z = -1.311, p = 0.19.

The detailed analyses in this experiment, as it was already mentioned earlier, was focused on parameters of the tones in Experiment 3. Also, as only one drifting speed was used for all Gabor patches (1 degree/s), the analyses of the spatial frequency and temporal frequency would yield the same result. Therefore, no separate analysis of the spatial frequency and temporal frequency in this experiment was conducted.

Carrier frequency

The average perceived (in)congruency ratings for stimuli with tones of a different carrier frequency are presented in Figure 4.5. No difference in perceived congruency between groups of stimuli containing a tone with a different carrier frequency was observed.

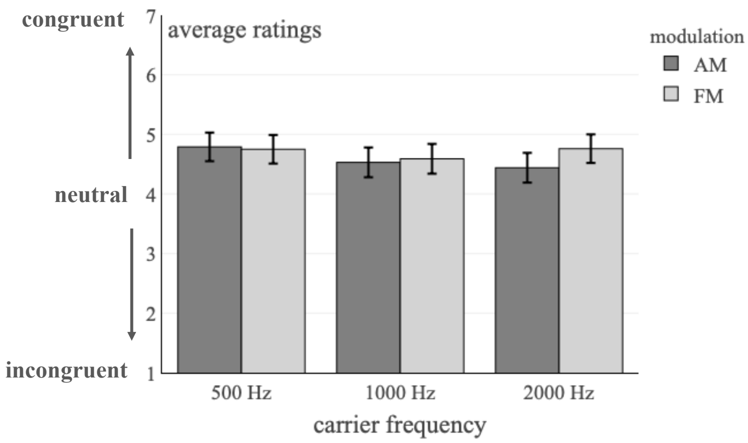

Figure 4.5: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

The groups of stimuli aggregated by carrier frequency of the tone and spatial frequency of the patch are presented in Figure 4.6. As can be seen from these figures, there is a weak rating pattern where combinations of Gabor patches with a high (low) spatial frequency and tones with a high (low) carrier frequency received higher perceived (in)congruency ratings than other stimuli. This tendency, however, did not reach statistical significance in most cases. The Friedman tests were run row-wise (Figure 4.7), with comparison of groups of stimuli with different carrier frequency tones, but the same spatial frequency Gabor patches; and column-wise with comparison of groups of stimuli with different spatial frequency Gabor patches, but the same carrier frequency tones. It can be assumed, that if crossmodal correspondence affects perceived congruency, the difference between the tested groups can be observed in both or either directions.

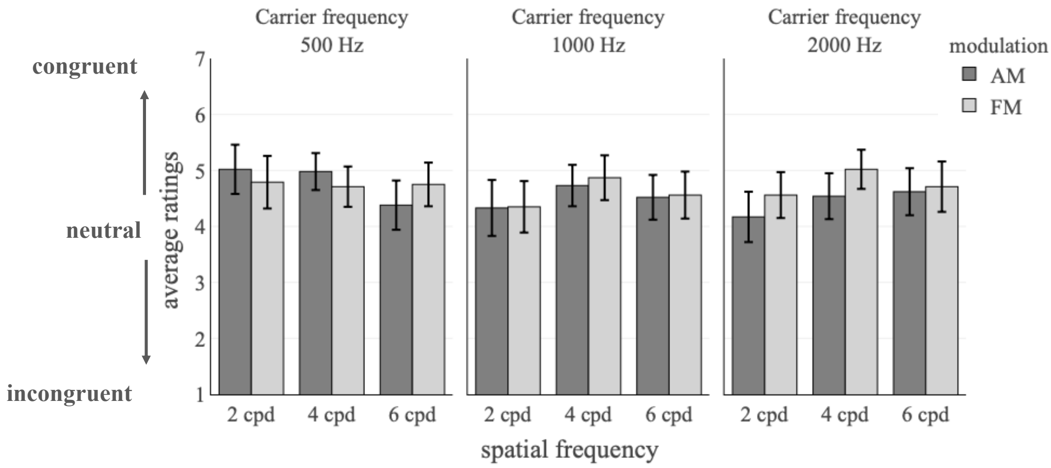

Figure 4.6: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

spatial frequency		AM, Hz			FM, Hz		
		500	1000	2000	500	1000	2000
cpd	2	5.02	4.33	4.17	4.79	4.35	4.56
	4	4.98	4.73	4.54	4.71	4.87	5.02
	6	4.38	4.52	4.62	4.75	4.56	4.71

Figure 4.7: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The colour demonstrates the degree of (in)congruency: grey - high (in)congruency rating; white - low (in)congruency rating.

The tests, however, demonstrated only one significant case. For the groups of stimuli containing a Gabor patch of 2 cpd spatial frequency the highest perceived congruency was obtained in stimuli with a lower carrier frequency tone: χ^2 (df = 2, N = 104) = 10.058, p = 0.007. For other spatial frequencies no significant difference between the groups of different carrier frequency was observed: spatial frequency 4 cpd χ^2 (df = 2, N = 104) = 0.188, p = 0.91; spatial frequency 6 cpd χ^2 (df = 2, N = 104) = 1.139, p = 0.57. No significant difference was observed either when comparing groups of Gabor patches with a different spatial frequency with one carrier frequency: 500-Hz tones χ^2 (df = 2, N = 104) = 2.457, p = 0.29; 1000-Hz tones χ^2 (df = 2, N = 104) = 3.692, p = 0.16; 2000-Hz tones χ^2 (df = 2, N = 104) = 1.688, p = 0.43.

Modulation frequency

Similar to the previous experiments, the effect of dynamic parameters was assessed. As it was mentioned in the beginning of this section, as only one speed was used the analysis of the effect of temporal frequency and spatial frequency yields the same results:

$$Speed = 1\ degree/s$$

$$Temporal frequency = spatial frequency \times speed$$

Figure 4.8 demonstrates the average perceived (in)congruency ratings of drifting Gabor patches aggregated by their spatial frequency (temporal frequency) and a modulation frequency.

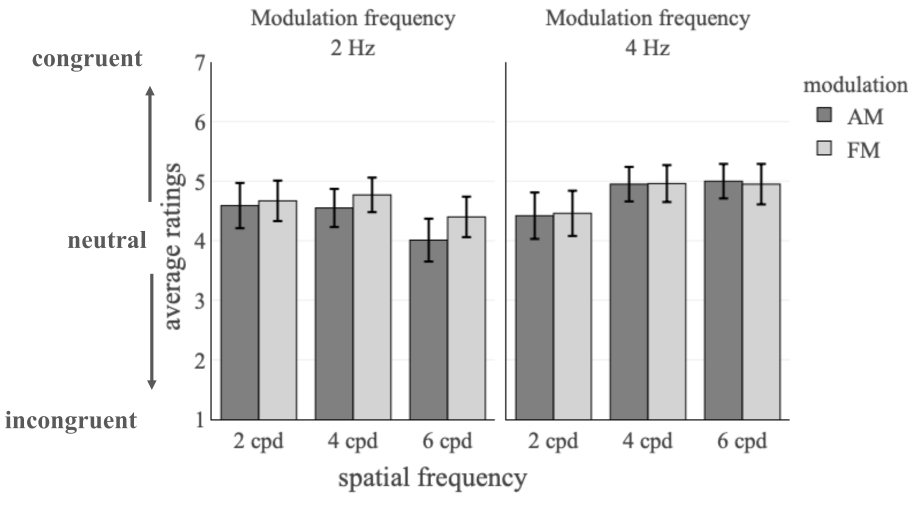

Figure 4.8: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

It can be noticed that for stimuli with a modulation frequency of 2 Hz the lowest ratings falls on the highest spatial frequency patches of 6 cpd (temporal frequency of 6 flicker/s), while for the stimuli with a modulation frequency of 4 Hz the lowest rating falls on the lowest modulation frequency patch of 2 cpd (2 flicker/s). Friedman test showed that the difference is significant for stimuli containing 2-Hz modulated tones (χ^2 (df = 2, N = 156) = 7.022, p = 0.03), but not for the stimuli with 4-Hz modulated tones (χ^2 (df = 2, N = 156) = 4.106, p = 0.13).

Drifting direction

Lastly, as one of the objectives of the current research was to investigate the role of dynamic parameters on perceived congruency, the influence of the drifting speed direction was clarified. In Experiment 3 two drifting directions were used - left to right and right to left. Figure 4.9 demonstrates the average ratings for the stimuli with Gabor patches of different drifting directions. It can be seen, drifting direction had no effect on perceived congruency.

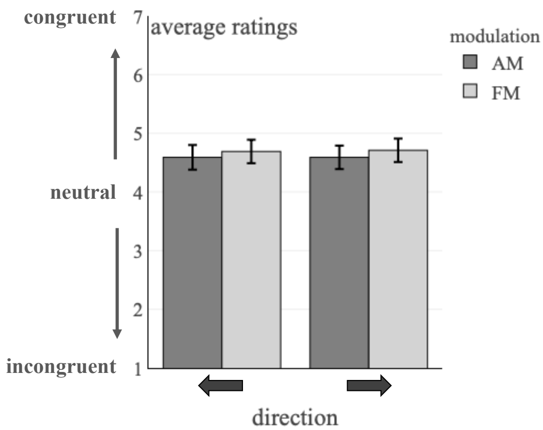

Figure 4.9: Results of Experiment 3. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

4.4 Discussion

The main purpose of the current experiment was to investigate whether carrier frequency of the tones had an effect on perceived congruency of the stimuli consisting of a Gabor patch and AM- and FM- tones. Similar to the previous experiments participants rated the stimuli on the scale from 1 ("incongruent") to 7 ("congruent").

The analysis of stimuli containing a flickering Gabor patch showed no effect of carrier frequency on perceived (in)congruency ratings. The tests conducted over groups of stimuli with certain spatial frequencies also did not yield any significant results and demonstrated no effect of crossmodal correspondence between spatial frequencies of Gabor patches and carrier frequencies of tones (Evans and Treisman, 2010). Similar analyses of stimuli containing a drifting Gabor patch demonstrated that stimuli with Gabor patches of 2-cpd spatial frequency had significantly different ratings when combined with tones of different carrier frequencies. However, this effect did not reach significance in any other cases, suggesting that there is no strong effect of carrier frequency, or that the effect of carrier frequency is masked by a stronger effect of dynamic properties.

Similar to the previous experiments, here too (dis)similarity in temporal frequencies of auditory and visual components of the stimuli, namely the flickering or temporal frequency of the patch and modulation frequency of the tone, played a prominent role in establishing perceived congruency. In might be suggested, that in the presence of dynamic changes, such as the flickering or drifting effect of Gabor patches or tone modulation, the effects of static parameters such as spatial or carrier frequencies are outmuscled by the stronger effect of temporal (dis)similarities.

Additionally, the comparison of the ratings of the stimuli containing AM- and FM-tones confirmed once again the advantage of AM-tones over FM-tones in evoking perceived congruency in stimuli with a flickering Gabor patch. In stimuli with a drifting Gabor patch no significant difference between ratings of stimuli with AM-tones and FM-tones was observed.

4.5 Conclusion

- Althouth the tone-carrier frequency has been considered as a factor that could influence perceived congruency of audiovisual stimuli in previous research, it had no effect on perceived congruency in the current experiment.
- The advantage of amplitude modulation over frequency modulation in evoking perceived congruency was confirmed one more time for stimuli containing a flickering Gabor patch and a modulated tone. In case of drifting Gabor patches, similar to the results of Experiment 2, no such advantage was observed.
- The groups of stimuli containing Gabor patches of different drifting directions showed the same average perceived (in)congruency ratings. In other words, drifting direction had no effect on perceived congruency.

4.6 Limitations of the study

Though the results of Experiment 3 did not demonstrate any effect of carrier frequency on perceived audiovisual congruency, as it was said earlier, that might indicate not the absence of the effect, but rather the presence of a stronger effect of dynamic changes such as patch flicker or patch-gratings drifting. In that case, a similar experiment with pure (not modulated) tones might clarify if there is any effect of carrier frequency in such stimuli.

Although the parameters for the current experiment were chosen in a way that they overlap with parameters of stimuli used in previous studies, the parameters' range was still limited. One of the serious limitations is the duration of the stimuli, that was constant throughout Experiments 1-3, i.e., 2 seconds. As the main findings of the experiments suggest that temporal similarities play a prominent role in evoking perceived congruency, the amount of temporal information or the length of the stimuli might affect the congruency.

Chapter 5. Experiment 4. Effect of stimulus duration

5.1 Introduction

In the previous Experiments, the effects of different auditory and visual parameters of stimuli consisting of a Gabor patch and a modulated tone on their perceived congruency were assessed. It was demonstrated that dynamic parameters and specifically (dis)similarities between flickering or temporal frequencies of the patch and modulation frequency of the tone play a prominent role in perceived (in)congruency of such stimuli. Static parameters, such as spatial frequency of a Gabor patch and carrier frequency of a tone demonstrated no observable effect. Besides, it was shown that, in case of stimuli with a flickering Gabor patch, AM-tones evoked higher perceived congruency compared to FM-tones.

All mentioned results were demonstrated for stimuli of one length of 2 seconds. Previous studies in this area, however, used stimuli of different duration. Evans and Treisman (2010), for example, studied short stimuli (120 ms) and demonstrated crossmodal correspondence between the spatial frequency of the patch and perceived pitch of the tone. Green et al. (2019) also used short (50 ms) stimuli and showed multisensory integration based on correspondence between spatial frequency of Gabor patches and sound frequency (pure tones). As the current study did not observe any evidence suggesting that perceived congruency of audiovisual stimuli might be significantly affected by such correspondence, it might be assumed, that the role of static parameters (i.e., spatial frequency of the patch and carrier frequency of the tone) is different in stimuli of different durations. The results of Experiments 1-3 demonstrated the importance of temporal information on congruency judgements. However, the amount of temporal information is less in shorter stimuli. It might be suggested that the longer the stimulus, the more clues to (in)congruency can be perceived.

To check this hypothesis, stimuli of 3 different lengths - 1 second, 2 seconds and 4 seconds, were used in Experiment 4. If the amount of temporal information influences perceived congruency, stimuli of different duration would demonstrate different ratings.

5.2 Method

The settings of Experiment 4 were identical to the settings used in Experiments 1-3. However, several adjustments to the settings had to be implemented in order to meet the objectives of Experiment 4 and enable prevention measures related to the COVID-19 pandemic. The full information on these and other adjustments is presented in the following sections.

5.2.1 Participants

Sixteen participants (11 females, 5 males) took part in Experiment 4. The average age of the participants was 30 years old (SD = 5). Nine participants were students of Kyushu University at the time of the experiment, 7 were recruited from outside of the university. Every participant received a written instruction and explanation about the purpose of the experiment and the procedure, and gave a written informed consent to participate. The COVID-19 prevention measures were also explained to every participant. These measures included wearing a mask for both participant and experimenter, hand sanitising before and after the experiment, and extra ventilation of the experimental booth. Additionally, a questionnaire was taken in which participants stated their health condition at the time of the experiment and signed that they had not been in contact with a person diagnosed with COVID-19 in the past two weeks.

The experiment was conducted with prior approval of the Ethics Committee of Kyushu University.

5.2.2 Apparatus

The settings for this experiment were the same as previously used for Experiments 1-3 (see section 2.2.2 for detailed description) with the following adjustments.

First, the infection prevention measures were implemented. These measures, however, did not affect the flow of the experiment and were mainly concerned with extra disinfection of the equipment that was used, and additional ventilation of the room. A fan was set in the experimental booth, that provided extra air circulation before the experiment and during the breaks. The participants were asked to leave the experimental booth during the breaks as an extra precaution. During the trials the fan was off to keep the room quiet and secure, similar to previous experimental conditions.

Second, the computer that controlled the experiment had to be replaced with a similar model (MacBook Pro Early 2015) due to calibration issues that arose with the computer that was used for Experiments 1-3. After the replacement the calibration procedure was successfully conducted. The final calibration measurements can be seen in Appendix B.

5.2.3 Experiment application and procedure

Experiment 4 consisted of 96 trials that were conducted over 3 sessions with 3-minute breaks. Similar instruction to the ones used in the previous experiments were provided to all participants (see Appendix E). An additional health check questionnaire was also performed. After the completion of all the forms, the experiment started with a practice session. Unlike previous experiments, the practice session in Experiment 4 was conducted with the door of the experimental booth opened and the experimenter guiding the participant from outside the booth to ensure recommended social distance between the participants and the experimenter. These measures did not cause any issues and participants reported no difficulties with experimental application navigation or any other confusions with the experiment after the practice session was over. The practice session consisted of 5 randomly chosen trails. The results of these trials were not included in the final analysis.

The flow of the experiment was identical to the one used in Experiments 2-3 and can be seen in Figure 3.1.

5.2.4 Stimuli

Stimuli consisting of a Gabor patch and a tone of various lengths were used in this experiment (see table in Figure 6.1 on page 86 of General Discussion section for the consolidated table on all parameters).

Gabor patches

All Gabor patches were of 2×2 degrees in visual angle with a Gaussian envelop and vertical sinusoidal gratings. The gratings had the peak luminance at 9.49 cd/m^2 (white) and the minimum at 0.02 cd/m^2 (black) on a grey background (2.83 cd/m^2). Gabor patches varied across the following parameters:

- spatial frequency: 2 cpd and 4 cpd;
- length: 1 second, 2 seconds and 4 seconds;
- dynamic mode:
  - Flickering with flickering frequencies of 2 Hz and 4 Hz;
  - Drifting with drifting speeds of 0.5 degree/s and 2.0 degrees/s.

Gabor patches were generated by the application in real time during the experiment.

Tones

AM- and FM-tones varied across the following parameters:

- length: 1 second, 2 seconds and 4 seconds (always combined with a Gabor patch of a similar length);
- rise and fall time: 20 ms, with a cosine-shaped ramp. This parameter was changed in this experiment (500 ms in Experiments 1-3) to secure adequate level of the tone for the shortest stimuli;
- carrier frequency: 1000 Hz;
- modulation type:
  - AM,
  - FM;
- modulation depth:
  - for AM: 1 amplitude;
  - for FM: ±80 Hz;
- modulation frequency: 2 Hz and 4 Hz.

Tones were generated prior to the experiment (see Appendix D for the generating scripts).

All together there were 96 stimuli: 4 flickering Gabor patches (2 spatial frequencies × 2 flickering frequencies) + 4 drifting Gabor patches (2 spatial frequencies × 2 drifting speeds) combined with 4 modulated tones (2 AM- and 2 FM- tones of different modulation frequencies) × 3 lengths of the stimuli. Each participant rated each stimulus once, excluding the 5 random practice trials.

5.3 Results

As the purpose of this experiment was to investigate the perceived congruency of Gabor patches and modulated tones in stimuli of different lengths, the analysis was performed for all stimuli to confirm the tendencies that were observed previously, as well as on groups of stimuli with different lengths separately, to investigate if the same tendencies were observed in stimuli with shorter or longer stimuli.

5.3.1 Flickering Gabor patches

Tone modulation

First, the effect of modulation type was confirmed. As in previous experiments, here too stimuli with AM-tones received significantly higher (in)congruency ratings than stimuli with FM-tones (N_{plus} = 145, N_{minus}, Z = -2.129, p = 0.03).

Stimulus duration

Second, as the main purpose of the experiment, the effect of stimulus length was analysed. Figure 5.1 demonstrates the average ratings of the stimuli aggregated by duration. As can be seen in the figure, no difference between ratings of stimuli with different durations can be observed (χ^2 (df = 2, N = 309) = 1.345). However, it can be noticed, that shorter stimuli (1 seconds and 2 seconds) demonstrated the highest difference between stimuli with AM- and FM-tones (AM-tones higher), and that was confirmed by the sign tests: stimuli of 1 second length N_{plus} = 57, N_{minus} = 34, Z = -2.306, p = 0.02; 2 seconds length N_{plus} = 47, N_{minus} = 31, Z = -1.698, p = 0.09; 4 seconds length N_{plus} = 41, N_{minus} = 45, Z = -0.323, p = 0.75.

In the previous experiments, the importance of physical (dis)similarity between flickering frequency of a Gabor patch and a modulation frequency of a tone was demonstrated. Here too, Figure 5.2 shows the same rating pattern where stimuli with a similar flickering frequency of a Gabor patch and a modulation frequency of a tone were rated higher than stimuli with dissimilar frequencies.

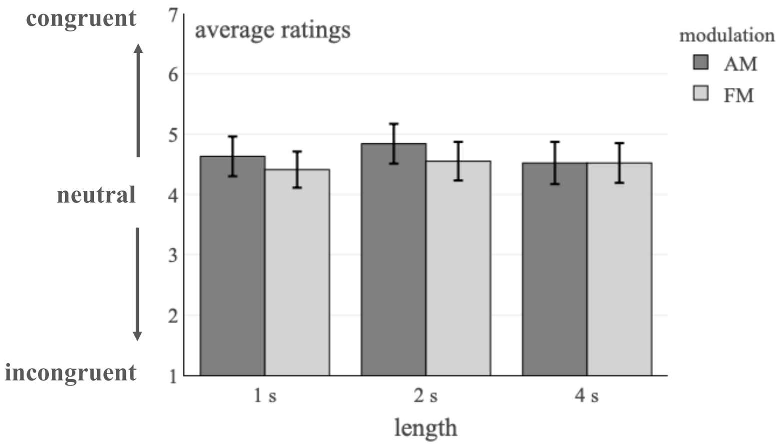

Figure 5.1: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

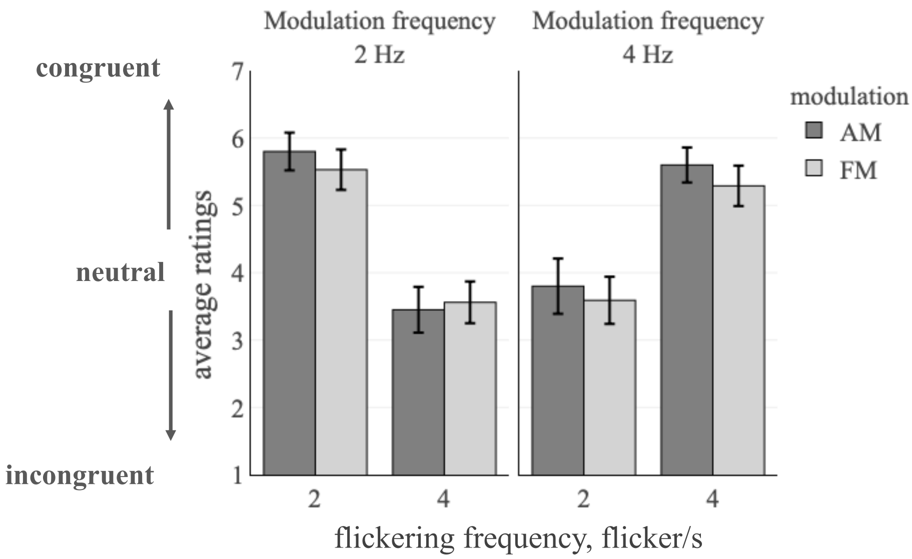

Figure 5.2: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

Figure 5.3 demonstrated the rating pattern for stimuli of different lengths. It can be seen that regardless of the length, perceived (in)congruency is rated similarly overall. Besides, it can be noticed that the effect of modulation type is also similar for all the lengths. For stimuli 1, 2 or 4 seconds length, all groups of congruent stimuli with AM-tones were rated higher than congruent stimuli with FM-tones (Figure 5.3).

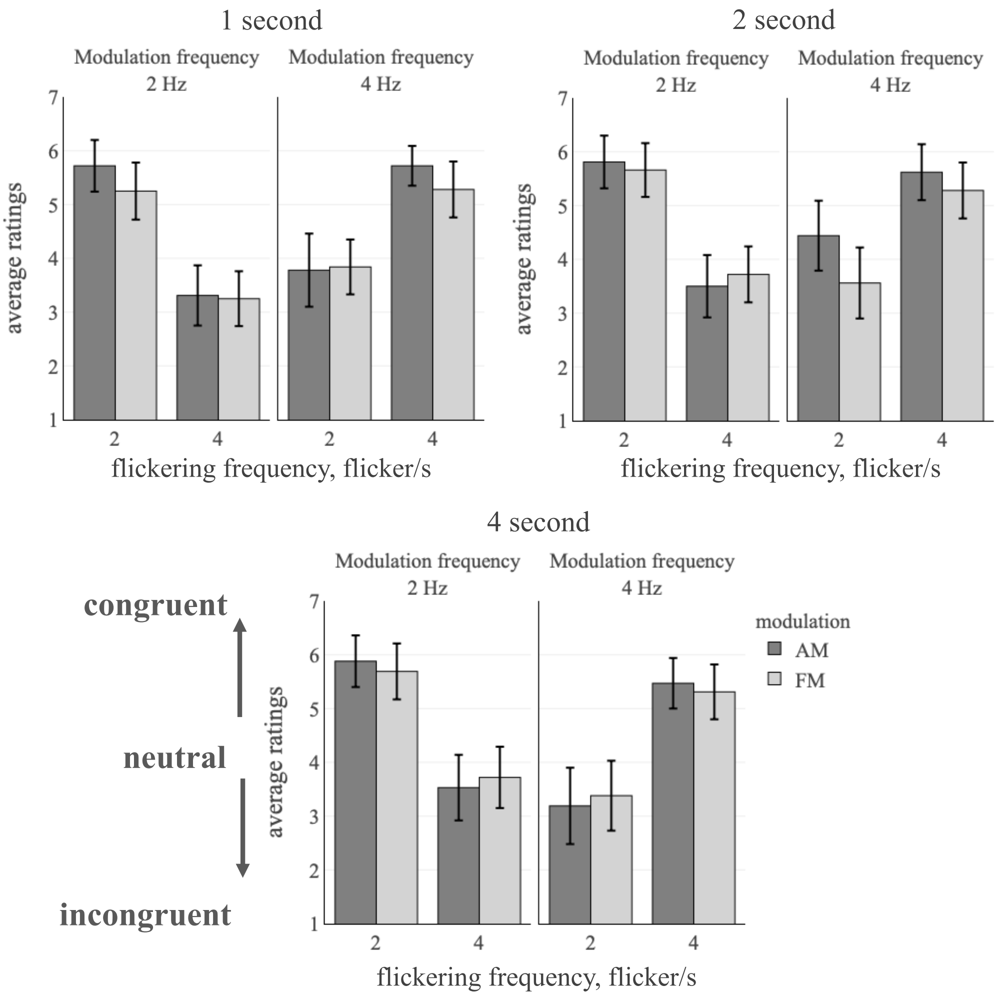

Figure 5.3: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a flickering Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

5.3.2 Drifting Gabor patches

Tone modulation

First, similar to the stimuli with a flickering Gabor patch, the effect of modulation type (AM or FM) was checked. Here too, as in Experiments 2 and 3, drifting Gabor patches demonstrated no effect of modulation type on (in)congruency ratings (N_{plus} = 132, N_{minus} = 138, Z = -0.304, p = 0.76).

Stimulus duration

Second, as the main purpose of the experiment, the effect of the length was checked for all stimuli with a drifting Gabor patch. Figure 5.4 demonstrates the average (in)congruency ratings of stimuli with a different length. It can be noticed than the average ratings for the shorter stimuli were lower than the average ratings for the longer stimuli. This difference, however, was not significant (χ^2 (d = 2, N = 256) = 3.377, p = 0.18).

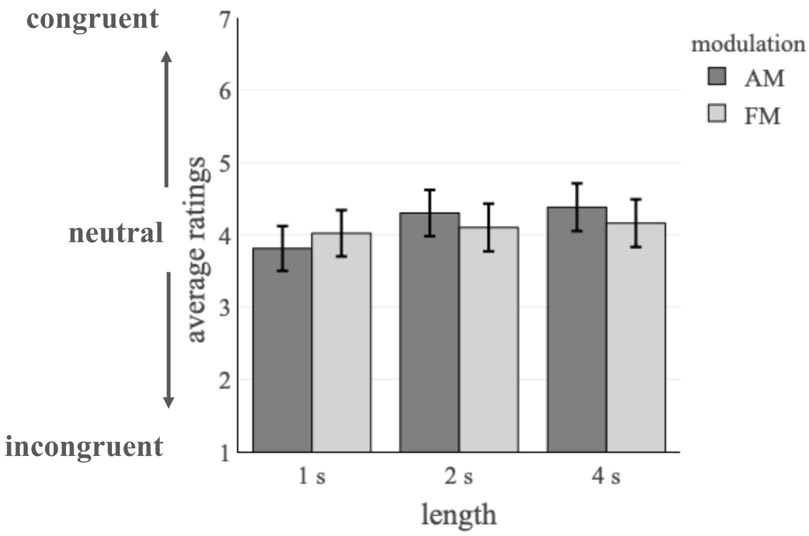

Figure 5.4: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

The analysis of temporal parameters demonstrated a significant difference between ratings of stimuli containing Gabor patches of different speeds and a 4-Hz modulated tone (statistical significance is in accordance with 95%-confidence intervals around the mean, Figure 5.5). Stimuli with these tones and with Gabor patches with a drifting speed of 2 degrees/s were rated as congruent (significantly higher than 4), while stimuli with a Gabor patch of 0.5-degree/s drifting speed were rated as incongruent (significantly below 4). Also, it can be seen in the figure, that the stimuli with a patch-speed of 2 degrees/s received higher ratings when combined with 4-Hz tones, than when combined with 2-Hz modulation frequency tones (according to 95%-confidence intervals), and for FM-tones this difference is significant.

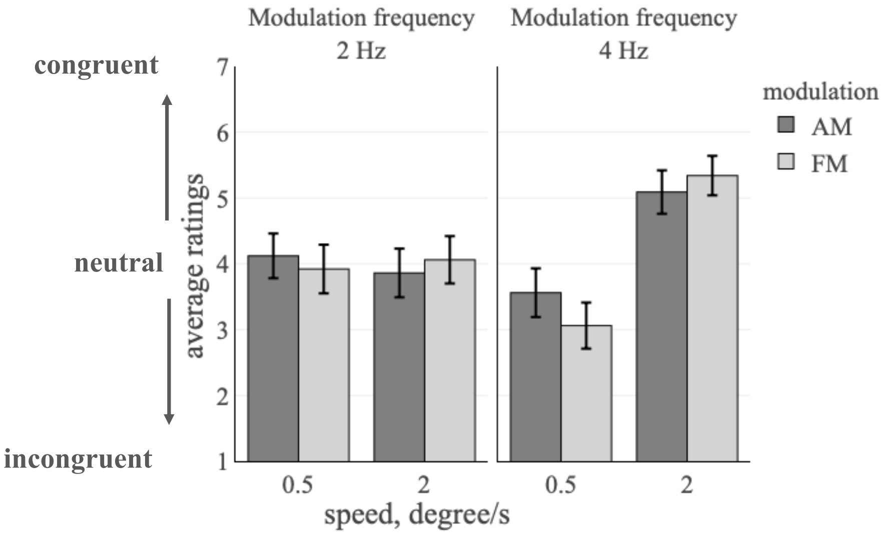

Figure 5.5: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

The analysis of the temporal frequency (Formula 2.2) showed a similar pattern (Figure 5.6). The stimuli with 4-Hz modulated tones demonstrated a clear rating pattern where temporal frequency (dis)similarity had a prominent effect on ratings. For stimuli containing 4-Hz modulated tones the highest rating was given to the stimuli containing a drifting Gabor patch of 4 flicker/s, while the lowest ratings were given to the stimuli with a drifting Gabor patch of 1 flicker/s (the difference is significant according to 95%-confidence intervals around the mean, Figure 5.6).

A more detailed analyses of temporal frequencies conducted for stimuli of different length separately is presented in Figure 5.7. It can be seen that a similar rating pattern can be observed for stimuli of all durations. For stimuli containing a 4-Hz modulated tone, the highest rating was given to stimuli with a patch of 4 flicker/s and the lowest to stimuli with 1 flicker/s, for stimuli of all durations. The (in)congruency ratings for the stimuli containing 2-Hz modulated tones, however, were on average higher for the longer stimuli. The Friedman test comparing stimuli of different lengths containing only 2-Hz modulated tones showed that the difference between them was statistically significant: χ^2 (df = 2, N = 128) = 6.691, p = 0.04. No difference was observed for such groups of stimuli containing only 4-Hz modulated tones: χ^2 (df = 2, N = 128) = 0.520, p = 0.77.

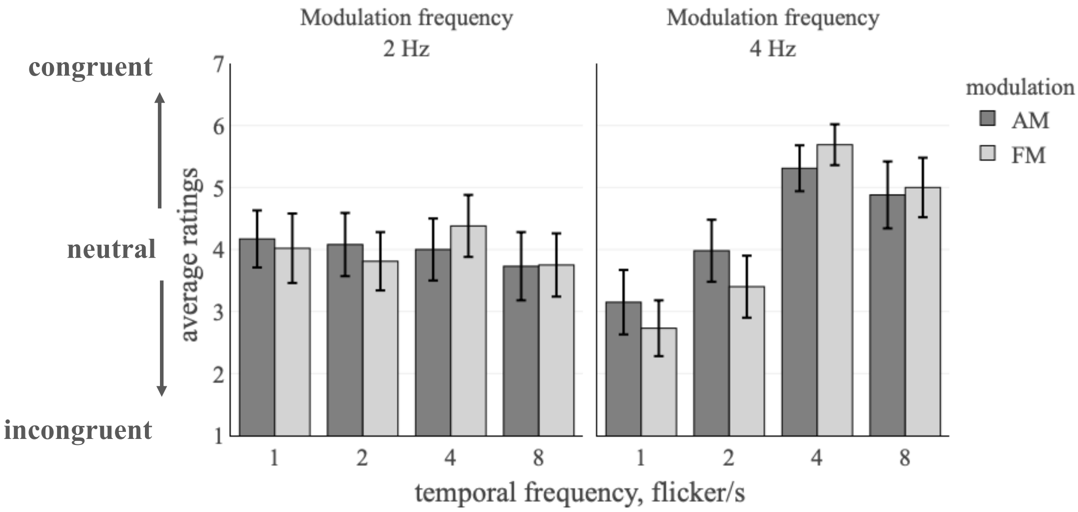

Figure 5.6: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

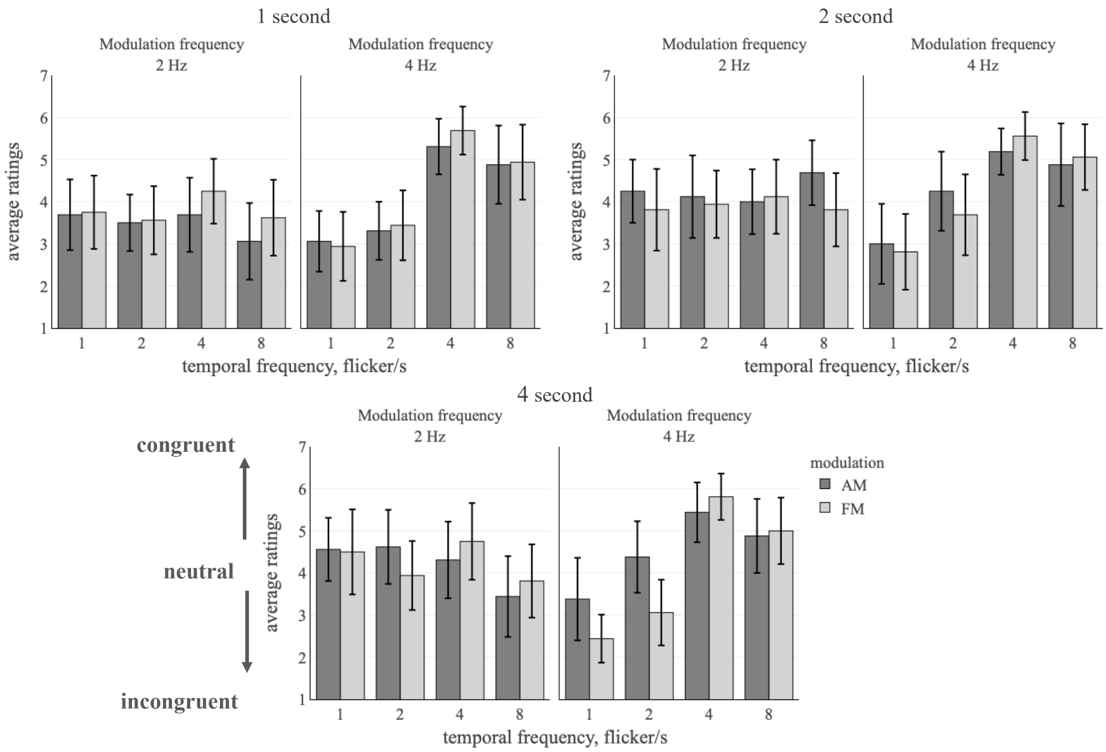

Figure 5.7: Results of Experiment 4. Average perceived (in)congruency ratings of stimuli consisting of a drifting Gabor patch and an AM- or FM-tone. The error bars indicate 95%-confidence intervals.

5.4 Discussion

The main purpose of Experiment 4 was to clarify the role of stimulus duration on perceived congruency of stimuli consisting of a Gabor patch and a modulated tone. The stimuli were rated on the scale from 1 (congruent) to 7 (incongruent), and the average ratings were analysed. The results demonstrated the following. First, that stimuli with a flickering Gabor patch and an AM-tone received significantly higher ratings than stimuli with an FM-tone. All stimuli with a flickering Gabor patch regardless of their duration demonstrated the advantage of AM-tones over FM-tones in congruent stimuli. Furthermore, stimuli with a similar flickering frequency of the patch and modulation frequency of the tone were rated relatively high, while stimuli with dissimilar patch-flicker and tone-modulation frequencies were perceived as relatively incongruent. This effect was observed in stimuli of all lengths.

Based on this finding, it might be suggested that a certain minimum of temporal information is required for a given parameter to influence audiovisual congruency. In stimuli of higher temporal frequencies, this parameter would reach this threshold faster than in stimuli with lower temporal frequencies, and therefore their effect becomes prominent in shorter stimuli. The results for stimuli consisting of a drifting Gabor patch and a modulated tone of 2 Hz demonstrated higher (in)congruency ratings for longer stimuli. Furthermore, stimuli with 2-Hz modulated tones showed no significant congruency difference when paired with a Gabor patch with a different temporal frequency. However, in stimuli with a 4-Hz modulated tone, a clear congruency difference was found between stimuli with a temporally (dis)similar patch and a tone. Stimuli with a 4-Hz modulated tone have twice the amount of temporal information than stimuli with a 2-Hz modulated tone, and that might be the reason of such difference in results. That theory should be investigated in future experiments.

5.5 Conclusion

The perceived congruency of stimuli of different duration was rated and the results were as follows:

- No effect of the stimulus length on perceived congruency of audiovisual stimuli was observed for stimuli consisting of a Gabor patch and a modulated tone of 1, 2, or 4 seconds duration.
- In case of flickering Gabor patches, stimuli with AM-tones were rated higher than stimuli with FM-tones, and the (dis)similarity between flickering frequency of a patch and a modulation frequency of a tone played a prominent role in establishing perceived (in)congruency in all such stimuli regardless of duration.
- Stimuli with a drifting Gabor patch also demonstrated that (dis)similarities between the temporal frequency of the patch and the modulation frequency of the tone played an important role in perceived (in)congruency, regardless of stimulus duration. However, the effect of temporal (dis)similarity was stronger in stimuli with a higher tone-modulation frequency.

Chapter 6. General Discussion

6.1 Summary

Multiple studies utilise stimuli consisting of a Gabor patch and a modulated tone. Often these stimuli are classified as "congruent" or "incongruent" based on principles that correlate certain parameters of the visual component to the corresponding parameters of the auditory component. Often these parameters are the spatial frequency of the patch and the frequency or modulation frequency of the tone (e.g., Evans and Treisman, 2010), as well as the spatial frequency of the patch and AM-frequency of a noise (e.g., Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a, 2013b). Numerous studies employed longer stimuli of over a second duration and applied dynamic variations to Gabor patches that were matched to the frequency or amplitude modulation of the tones. However, no empirical data are available on perceived congruency in dynamic stimuli, i.e, stimuli consisting of a non-static Gabor patch (flickering, drifting or with any other dynamic changes applied to a patch) and a modulated tone, though such stimuli are also widely used in multisensory research. This dissertation investigated the perceived congruency of such stimuli. The main objective of this study was to clarify the effect of different visual and auditory parameters, and specifically the modulation type of the tone and static and dynamic parameters of a Gabor patch, on perceived audiovisual congruency.

Chapter 2 describes a pilot study (Experiment 1), that employed a wide range of parameters, and its results helped to finalise the framework of this research, as well as the direction of subsequent experiments. The main objective of this experiment was to investigate which parameters have the strongest impact on perceived (in)congruency of stimuli consisted of a flickering or drifting Gabor patch and a modulated tone. The parameters that were checked are the following: the spatial and flickering frequency or speed of the patch and the modulation frequency and modulation type (AM and FM) of the tone. Results demonstrated that stimuli with an AM-tone received significantly higher perceived (in)congruency ratings than stimuli with an FM-tone. Also, it was observed that temporal (dis)similarity was a prominent factor of establishing perceived (in)congruency, while spatial frequency of a patch had no or little effect on congruency. The effect of temporal (dis)similarities was clearly shown in flickering Gabor patches, where stimuli with a similar flickering frequency of the patch and modulation frequency of the tone were rated significantly higher than stimuli with dissimilar patch-flicker and tone-modulation frequencies. A similar analysis for stimuli with a drifting Gabor patch, however, did not yield any conclusive results, as parameters of drifting Gabor patches and the tones that were used in Experiment 1 did not fully match. This was corrected in the subsequent experiments.

Chapter 3 describes the results of Experiment 2. The objective for Experiment 2 was to confirm and clarify the findings of Experiment 1, and specifically to confirm the advantage of AM-tones over FM-tones in evoking perceived congruency with stimuli consisting of a Gabor patch and a modulated tone, and to further investigate the effect of static (spatial frequency) and dynamic (flickering and temporal frequency) parameters of a patch on perceived congruency. Also in Experiment 2, stimuli with an AM-tone received significantly higher (in)congruency ratings compared to stimuli with an FM-tone in combinations with a flickering Gabor patch, however, the advantage of AM-tones did not reach significance for stimuli with a drifting Gabor patch. Similar to the results of Experiment 1, the spatial frequency of the Gabor patch had no effect on perceived audiovisual congruency. However, dynamic parameters of the patch, such as flickering frequency or temporal frequency, had a strong effect on perceived congruency. Stimuli with a similar flickering (temporal) frequency of the Gabor patch and modulation frequency of the tone received significantly higher ratings than stimuli with dissimilar frequencies.

Chapter 4 focused on the effect of carrier frequency of modulated tones on perceived audiovisual congruency. Modulated tones of three different carrier frequencies were employed in Experiment 3. Previous studies reported crossmodal correspondence between the spatial frequency of Gabor patches and perceived pitch (frequency) of the tones (Evans and Treisman, 2010, Green et al., 2019), however, no clear effect of carrier frequency or crossmodal correspondence between carrier frequency and spatial frequency was observed. Dynamic parameters demonstrated similar effects on perceived congruency as observed in Experiment 1 and 2. These results suggested that static parameters such as carrier frequency of the tone or spatial frequency of the Gabor patch had no effect on perceived congruency of dynamic stimuli.

Chapter 5 confirms the results observed in previous experiments in stimuli of different lengths of 1, 2 and 4 seconds. As in previous experiments, each stimuli contained a flickering or drifting Gabor patch and an AM- or FM-tone. The comparison showed no difference in perceived congruency between these stimuli of shorter or longer duration. The analysis demonstrated a strong influence of dynamic parameters on perceived congruency and demonstrated no significant effect of static stimuli on perceived congruency regardless of the stimulus duration.

The stimuli parameters used in the current thesis are presented in Figure 6.1.

Spatial frequency of Gabor patches				Flickering mode frequencies				Drifting mode speed (two directions)			
Exp 1	Exp 2	Exp 3	Exp 4	Exp 1	Exp 2	Exp 3	Exp 4	Exp 1	Exp2	Exp 3	Exp 4
2	2	2	2	0.5	-			-	-	-1.0	
-	3			1	-			0.5	0.5		0.5
4	4	4	4	2	2	2	2	1.0	1.0	1.0	
6	-	6		3	3			1.5	1.5		
8	-			4	4	4	4	2.0	-		2.0
10	-							2.5	-		

Stimuli length:
Exp 1-3: 2 seconds
Exp 4: 1, 2, 4 seconds

Auditory stimuli:
Modulation Frequencies

Carrying frequencies, Hz			
-	-	500	
1000	1000	1000	1000
-	-	2000	

Figure 6.1: Stimuli parameters used in Experiments 1-4.

6.2 Implications

The main findings can be summarised and classified into these key points, which will be individually discussed in detail in sections 6.2.1, 6.2.2, and 6.2.3 below:

- effect of tone-modulation type on perceived audiovisual congruency;
- effect of static parameters (spatial frequency of the patch and carrier frequency of the tone) on perceived audiovisual congruency;
- effect of dynamic parameters (flickering and temporal frequency of a patch and modulation frequency of a tone) on perceived audiovisual congruency.

6.2.1 Effect of tone-modulation type on perceived congruency

One of the main objectives of the current study was to compare the perceived congruency in stimuli with an AM- and FM-tone. The presented results demonstrated that average perceived congruency of stimuli consisting of a flickering Gabor patch and a modulated tone was in general significantly higher when an AM-tone was used compared to an FM-tone. This difference was the most prominent in stimuli with a similar flickering frequency of the patch and modulation frequency of the tone. Although most of such stimuli were rated as "congruent" (i.e., received a significantly higher rating than the rating midpoint) in both cases, stimuli with an AM-tone received even higher ratings than stimuli with an FM-tone. One of the explanations for such preference could be that in case of stimuli with a flickering patch and an AM-tone, the fluctuations that happen within the visual and auditory component of a stimulus are similar. As the AM-tones in this study had a full amplitude modulation depth, and the flickering Gabor patch had the same dynamics with full disappearance/appearance of the patch, both auditory and visual components of the stimuli demonstrated the same dynamics, and in case of "congruent" stimuli it happened simultaneously. An FM-tone on the other hand, had the same temporal structure as the changes in the Gabor patch, however, it always fluctuated above threshold, demonstrating a different dynamic compared to the Gabor patch flickers. Similarly, in stimuli with a drifting Gabor patch, where the patch was always above the detection/visibility threshold, the advantage of AM-tones disappeared.

Whether the sub-threshold fluctuations or the simultaneous change of energy in both modalities plays a crucial role here should be investigated further. A similar experiment with AM-tones of various modulation depths, including fully supra-threshold fluctuating tones, can clarify this effect further.

6.2.2 Effect of dynamic parameters of Gabor patches and modulated tones

The second objective of the study was to investigate what parameters affect the perceived congruency of stimuli consisting of a Gabor patch and a modulated tone. In the current study, dynamic Gabor patches were used: flickering and drifting. For both types of Gabor patches the most convincing effect on perceived congruency was demonstrated by similarity between the flickering or the temporal frequency of the Gabor patch and the modulation frequency of the tone.

In case of stimuli with a flickering Gabor patch, the results were very clear. Similar to previous findings (e.g., Schall et al., 2009; Frings and Spence, 2010; Covic et al., 2017), the stimuli with a similar flickering frequency of the patch and modulation frequency of the tone were rated as "congruent", while most of the stimuli with dissimilar patch-flicker and tone-modulation frequencies were rated as "incongruent". In Experiment 1 it was observed that if the ratio of the tone-modulation frequency and the flickering frequency could be expressed as a whole number (for example when the tone-modulation frequency was half or double of the patch-flickering frequency), then stimuli with such parameters were also rated higher than stimuli which parameters that did not satisfy this condition. It can be assumed that if the peaks and troughs of the auditory and visual signal coincide even partially, perceived congruency might increase. This, however, was clearly demonstrated only in Experiment 1, and needs to be investigated further.

Stimuli with a drifting Gabor patch demonstrated a more complex relationship between visual and auditory components. As the dynamic parameters of a Gabor patch are not independent, i.e., its temporal frequency is related to spatial frequency through drifting speed, various interrelated tendencies were observed in Experiments 1-4. The most convincing audiovisual congruency, however, was still demonstrated when the temporal frequency of the patch and the modulation frequency of the tone were similar.

6.2.3 Effect of static parameters of Gabor patches and modulated tones

Various previous studies employed a principle where a Gabor patch with high (low) spatial frequency in combination with high (low) frequency tone or modulated noise constituted a congruent audiovisual stimulus, while a Gabor patch with low (high) spatial frequency in combination with high (low) frequency tone or modulated noise constituted an incongruent audiovisual stimulus. This principle was based on studies that demonstrated a crossmodal correspondence between the spatial frequency of the Gabor patch and the frequency (perceived pitch) of the tone (Evans and Treisman, 2010), as well as the spatial frequency of the patch and AM-frequency of the noise (Guzman-Martinez et al., 2012; Orchard-Mills et al., 2013a, 2013b).

As a third objective, in the current study this principle was tested using dynamic Gabor patches and modulated tones, however, no relation between (dis)similarity in such frequencies and stimulus (in)congruency was convincingly observed. Spatial frequency of the patch and carrier frequency of the tone were tested as stand-alone parameters, as well as in crossmodal combinations with each other and other parameters, but in stimuli with both a flickering or a drifting Gabor patch, not the static but the dynamic parameters affected the congruency the most.

In can be argued that as in this study relatively long stimuli were used (1-4 seconds) the effect of static parameters was outmuscled by the dynamic parameters of such stimuli. Indeed, most of the mentioned previous studies used stimuli shorter than a second. In such a short stimulus, the amount of temporal information might not be sufficient for a congruency judgment based on dynamic (temporal) parameters, and then correspondence between static parameters might become crucial. In Experiment 4, stimuli of different duration were used, however, no changes in perceived congruency that would indicate the stronger effect of static parameters in shorter stimuli were observed. Further study with a wider range of lengths (from less than a second) and a wider (lower) range of temporal frequencies can clarify if the amount of temporal information needs to surpass a certain minimum that is necessary to become crucial in defining congruency.

To summarise and to put the current study in a more general context, the presented findings demonstrate that the similarities in dynamic changes of sound and the visual information we perceive play the primary role in audiovisual perception. Indeed, most of the real-life events are represented in prolonged and often continuous sound or visual signals. In that case the temporal similarity in dynamic changes of the signals (changes of intensity or frequency contents), rather than the actual instantaneous signal values, should be the key for audiovisual integration. From the evolutionary perspective, the efficient identification and analysis of physical changes in the environment around us is crucial for survival, as usually the changes rather than the permanent state of the environment represent immediate danger. The correct and quick identification of the moving object from the array of similar static ones in that case could be a good example of prominence of the dynamic characteristics of the sound and light signals in contrast to the features of the steady signals from the static objects.

6.2.4 Limitations and future studies

First, the current study demonstrated the advantage of AM-tones over FM-tones in stimuli containing a flickering Gabor patch and a modulated tone. However, as only one AM-modulation depth was used, the question remains whether the observed advantage was caused by a synchronous change in energy, or synchronous fluctuations above and below auditory and visual detection thresholds. Similar tests with various modulation depths can clarify this. Furthermore, in the present study in stimuli with similar flickering frequency of a patch and modulation frequency of a tone the peaks and troughs always coincided. A study of stimuli with a phase shift between the auditory and visual fluctuations might help to clarify the importance of synchronous (peaks and trough coincide) versus parallel (with a phase shift) changes in the Gabor patch and the tone.

Second, the importance of (dis)similarity between the temporal frequency of the patch and the modulation frequency of the tone was demonstrated in stimuli of different durations. A wider range of flickering frequencies and stimulus lengths should be investigated in order to clarify the importance of the amount of information and to show whether the role of such (dis)similarity remains in a shorter stimulus or in stimuli with lower patch-flicker frequency and/or lower tone-modulation frequency. That can also help to clarify whether static parameters, such as carrier frequency of a tone and spatial frequency of a patch, become more important when temporal information lacks.

Third, although in the present study the most common physical parameters of Gabor patches and tones were used, other physical parameters that were not investigated should also be considered in future studies, such as the loudness of the tone and the contrast, size or gratings orientation of the Gabor patch.

Fourth, the concept of incongruency should be investigated. While the principle of matching a patch and a tone to ensure high perceived congruency was discussed and assessed, a similar investigation of incongruency is lacking. Unlike high (congruent) ratings, low ratings indicating incongruency were not consistently observed throughout the study - certain stimuli were rated significantly incongruent in one experiment, but got higher ratings in another experiment making it impossible to define the most incongruent pair of a Gabor patch and a tone. Besides, the rating distributions in general were skewed towards "congruent" values, and "incongruent" average ratings were rare. This can be the result of the stimuli parameter choice - for example, for stimuli with flickering Gabor patches, the parameters were always chosen in a way to secure the presence of congruent audio-visual pairs (a patch and a tone with similar flickering and modulation frequencies), while the level of incongruency was not considered. Another reason for skewed distribution and lack of significantly incongruent ratings might be the presence of some minimal level of perceived congruency provided by the simultaneous onset and offset of a patch and a tone. This should be examined in future experiments by a careful selection of parameters accounting for both congruency and incongruency levels with special focus on incongruent stimuli.

Fifth, some improvements in the rating scale design should be implemented, as labelling of the rating scale can affect the results (Weijters et al., 2010). In the case of the current study only the extreme values were labelled. These extreme values represented two opposite judgments - congruent and incongruent. Values between them corresponded to various levels of (in)congruency, with incongruency levels on the left (values 1, 2, 3), neutral in the middle (value 4), and congruency on the right (values 5, 6, 7). However, the numerical values that were assigned to the scale ticks (1-7) and were quoted in the written instruction could have given the participants the idea that the scale represents levels of *congruency* from 1 to 7. Full verbal labelling of the scale with clear identification of congruency and incongruency levels and the neutral point, as well as the omission of the scale-values (1 to 7) from the written instruction should resolve that issue.

Finally, the current study utilises only one experimental method - the rating experiment. Other methods, for example, the method of adjustment or force-choice tasks, could be also considered and used in order to confirm or clarify the current findings.

6.2.5 Conclusion

This dissertation focused on parameters of audiovisual stimuli consisting of a Gabor patch and a modulated tone (AM or FM) that affected perceived (in)congruency of such stimuli. It was demonstrated that AM-tones in combinations with flickering Gabor patches evoked higher perceived congruency than FM-tones. Furthermore, static parameters of audiovisual stimuli, such as spatial frequency of the Gabor patch and the carrier frequency of the tone had no or little effect on perceived congruency in the presence of dynamic changes. Dynamic parameters, i.e., the flickering and temporal frequency of the Gabor patch and the modulation frequency of the tone, however, played a prominent role in perceived (in)congruency. Stimuli with auditory and visual components of similar temporal frequencies (flickering frequency of the patch and modulated frequency of the tone) received significantly higher ratings than stimuli with dissimilar components. Additionally, it was confirmed, that above mentioned findings hold true for stimuli of shorter and longer durations.

Appendices

A Experiment Application

The application was developed using PsychoPy3 Experiment Builder. The same application was used for all 4 experiments with some minor changes. In cases when a certain component of the application was changed to accommodate the differences in experiments, this component is described for different experiments separately. In order to run the experiment, the application has to be used with two files - "trials_practice.csv" which sets parameters for 5 practice trials, and "trials.csv" which sets the parameters and the order of the experimental trails. These files are generated using R Studio prior to the experiment for each participant.

	A	B	C	D	E	F	G	H	I	J	K
1	spacial_fr	flicker_fr	f	speed	m	sound	phase_rate	trialN	text	time	min
42	8	2	1	0	0	sounds/am2	0	41	.	0	0
43	8	0	0	1	1	sounds/fm3	8	42	.	0	0
44	6	2	1	0	0	sounds/am3	0	43	.	0	0
45	8	0	0	0.5	1	sounds/fm3	4	44	.	0	0
46	2	2	1	0	0	sounds/am2	0	45	.	0	0
47	6	0	0	1.5	1	sounds/fm3	9	46	.	0	0
48	4	0	0	2	1	sounds/am4	8	47	.	0	0
49	6	1	1	0	0	sounds/am1	0	48	.	0	0
50	2	0	0	2	1	sounds/fm4	4	49	.	0	0
51	6	0	0	0.5	1	sounds/fm3	3	0	1 out of 5 is done! Let's have a break:	180	3
52	10	1	1	0	0	sounds/am4	0	1	.	0	0
53	4	0	0	2.5	1	sounds/am3	10	2	.	0	0
54	6	0	0	1.5	1	sounds/am1	9	3	.	0	0
55	4	0	0	2.5	1	sounds/fm1	10	4	.	0	0
56	10	0	0	2	1	sounds/fm5	20	5	.	0	0
57	4	2	1	0	0	sounds/fm4	0	6	.	0	0
58	8	1	1	0	0	sounds/am5	0	7	.	0	0
59	6	0	0	1	1	sounds/fm4	6	8	.	0	0

Figure 2: An example of the trails' information table "table.csv".

The files contain tables where each row corresponds to one trial with the following parameters stored in the columns (see Figure 2 for the example of this table for Experiment 1):

- "spatial_fr" - spatial frequency for Gabor patches;

- “flicker_fr” - flickering frequency for flickering Gabor patches (set to 0 for drifting);
- “speed” for drifting Gabor patches (set to 0 for flickering);
- “phase rate” = spatial_fr * speed;
- “sound” - sound filename and it’s location;
- “trialN” - trial order number;
- “text” - text for the breaks notifications;
- “time” - time in seconds for the breaks between trials (180 for the trials followed by a break, 0 for all other trials);
- “f” - 0 for trials with drifting Gabor patches, and 1 for trials with flickering Gabor patches;
- “m” - 0 for trials with flickering Gabor patches, and 1 for trials with drifting Gabor patches.

Additional parameter that was used in Experiment 3:

- “dir” - drifting direction for drifting Gabor patches (-1 or 1 for drifting patches, 0 for flickering).

Additional parameter that was used in Experiment 4:

- “length” - length of the stimuli in seconds.

The flow of the experiment as it was developed in the Builder View (PsychoPy3) is presented in Figure 3. The developed application consisted of 10 original routines (Fig.3).

Routine 1: “Instruction”. On the grey background (constant thoughout the experiment) the message appears: “Please, read the provided printed instruction and then click the mouse to start the practice session”. This message stays on the screen till the mouse is clicked. Routine 1 was created using two components:

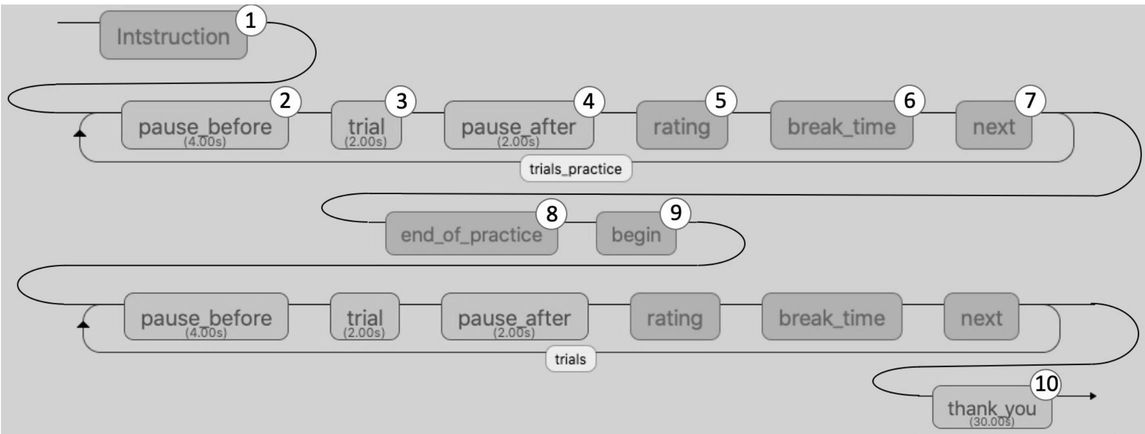

Figure 3: Flow of the experiment as in Builder View. The 10 original routines are marked each with the number. In cases when the same routine was used twice, the number only marks the first occasion.

- "text". Displayed the message mentioned above. The duration is set to "conditional" (condition was a mouse click);
- "mouse" set to "End Routine on any click";
- "code". "Begin Routine": `win.mouseVisible = False` # makes cursor invisible till another command is passed.

From here on, only the components' parameters that were set differently from the default are described.

After this Routine is over, the practice trial loop begins (Fig. 3). It repeats 5 times (5 practice trails) and the parameters of the stimuli are defined by the "trials_practice.csv" file. It is followed by the experiment loop "trial" that consists of the same routines, but uses a different parameters' file - "trials.csv". All routines described further apply to both loops.

Routine 2: "pause before". This routine provides a 4-second pause before the beginning of the trial and the appearance of the stimuli. The routine was created by using the following components:

Experiment 1:

- "grating" component with the parameters set as following: "opacity" = 0 (fully transparent), time duration = 4 seconds. This resulted in the grey empty screen for 4 seconds;

- two “polygon” components set to “line” option - one horizontal and one vertical line forming the fixation cross. The parameters were set so the cross is located in the middle of the screen with 0.4 visual degree width and height. Time: start at 2 seconds from the beginning of the routine, duration - 1 second. As a result, the fixation point appeared at the spot where the center of the stimuli would be located 2 seconds before the stimulus, lasting 1 second and disappearing 1 second before the stimuli appeared.

Experiments 2-4:

- “grating” (same as described above)
- two “polygon” components similar to the ones described above. The parameters were set so the cross is located in the middle of the screen with 0.4 visual degree width and height. Time: start at 3 seconds from the beginning of the routine, duration - 1 second. The fixation point appeared 1 second before the stimuli below the spot where the stimulus would appear.

Routine 3: “trial”. Duting this routine the stimuli were demonstrated. The visual part of the stimulus was generated, while for the sound part the external sound file was called.

Experiment 1:

- “grating” component named “moving”. The parameters were set as following:
  - Time: start 0, duration: 2 seconds;
  - Opacity:

    `(sin((trialClock.getTime()-0.5)*pi)+1)/2*m`,

    set every frame;
  - Size: (1,1). That would be equivalent to 2 visual degree diameter on the experimental monitor due to the MacBook Retina Display conversion (one pixel for two);

- Texture: sin;
- Mas: gauss;
- Phase (in cycles): `trialClock.getTime()*phase_rate`, set every frame;
- Spatial frequency: spatial_fr, set every repeat.

- "grating" component named "flickering". The parameters were set as following:
  - Time: start 0, duration: 2 seconds;
  - Opacity:
    `(sin((2*trialClock.getTime()*flicker_fr-0.5)*pi)+1)/2*f`,
    set every frame;
  - Size: (1,1). (2 visual degrees);
  - Texture: sin;
  - Mas: gauss;
  - Phase (in cycles):
    `round((sin((trialClock.getTime()*flicker_fr)*pi)+1)/2)/2`,
    set every repeat;
  - Spatial frequency: spatial_fr, set every repeat.
- "sound"
  - Time: start 0, duration: 2 seconds;
  - Sound: sound (sound filename), set every repeat.

Experiment 2 and 3:

- all the same components that were described for Experiment 1 with the minor change: both "grating" components ("flicker" and "moving") are moved two visual degrees up from the center of the screen;

- two "polygon" components identical to the ones used in the previous routine (Routine 2, Experiments 2-4). From the participant perspective the fixation cross (these two polygons) appears one second before the stimulus (during previous routine) and stays on the screen through the stimulus presentation. Time settings: start 0 s, duration 2 s.

Experiment 4 used different settings for all components, as the length of the stimuli were varying from trial to trial:

- "grating" component named "moving". The parameters were set as following:
  - Time. Start: 0 s, duration: length;
  - Opacity: m;
  - Position: (0,1) (2 visual degrees from the center);
  - Size: (1,1) (2 visual degrees);
  - Texture: sin;
  - Mas: gauss;
  - Phase: `trialClock.getTime()*phase_rate`, set every frame;
  - Spatial frequency: spacial_fr, set every repeat.
- "grating" component named "flickering". The parameters were set as following:
  - Time. Start: 0 s, duration: length;
  - Opacity:

    `(sin((2*(trialClock.getTime())*flicker_fr-0.5)*pi)+1)/2*f`;
  - Position: (0,1) (2 visual degrees from the center);
  - Size: (1,1) (2 visual degrees);
  - Texture: sin;
  - Mas: gauss;

  - Phase:

    `round((sin((trialClock.getTime()*flicker_fr)*pi)+1)/2)/2`,

    set every frame;

  - Spatial frequency: spacial_fr, set every repeat.

- "polygon" set to "rectangle" mode. The purpose of this component is to create an envelope for the Gabor patches and make them appear gradually. The parameters were as following:

  - Time. Start: 0 s, duration: 0.2 s;

  - Opacity: `1-sin(trialClock.getTime()*2.5*pi)`, set every repeat;

  - Size: (1.5, 1.5).

- "polygon" set to "rectangle" mode. The purpose of this component is similar to that previously described with the difference that this component makes Gabor patches disappear gradually. Most of the components are the same with the previous polygon with these two differences:

  - Time. Start: length - 0.2, duration: 0.2 s;

  - Opacity: `sin((trialClock.getTime()-length-0.2)*2.5*pi)`.

- two "polygons" forming fixation cross as described above (Experiment 2 and 3).

Routine 4: "pause after". It provides 2-seconds pause between a stimulus and the rating. The routine consists of one component:

- "grating" with opacity set to 0 (empty screen), time duration: 2 seconds.

Routine 5: "rating". During this routine a rating scale appeared (Fig. 4) and participants gave ratings to the stimuli. It consisted of the following components:

- "slider" with the following parameters:

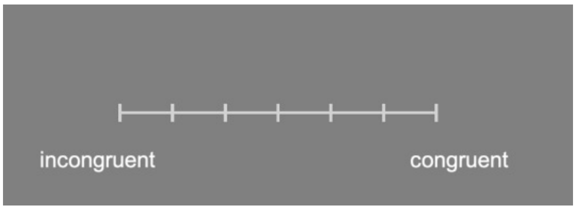

Figure 4: Rating scale

  - Size: (7, 0.4);
  - Position: (0. -3.0);
  - Ticks: (1,2,3,4,5,6,7);
  - Granularity: 1 (integer scale);
  - Style: rating.

- two "text" components located on the left side of the slider (text "incongruent") and on the right side of the slider (text "congruent).
- "code":
  - Begin Routine: `win.mouseVisible = True` # makes cursor appear;
  - End Routine: `win.mouseVisible = False` # makes cursor disappear.

Routine 6: "break time". This routine provided occasional breaks for participants. Components:

- text:
  - Time: duration "time" (0 for most of the trials, 180 seconds for trials that are followed by a break);
  - Tex: text, set every repeat.

Routine 7: "next". This routine prevented the next trial from starting straight away after rating without participants confirming their readiness. Components:

- text: "Thanks! Click for next";

- mouse: end routine on any click;
- two polygons forming a progress bar (Fig. 5). One of them with the constant parameters (external frame of the progress bar), while parameter of the other (width) were varying to reflect the progress:
  - Position: (0, -4) for the steady polygon, `(-(36-trialN)/20, -4)` for the "growing" polygon;
  - Size: (3.6, 0.7) for the steady, `(trialN/10, 0.5)` for the growing.

Figure 5: Progress bar

If more trials were scheduled for this loop, the Routine 2 follows (4-second pause break before the stimuli). In case of the last scheduled trial in the practice session Routine 8 followed, in case of the last scheduled trial in the experiment session, Routine 10 followed.

Routine 8: "end of practice". During this routine the following message was displayed: "The practice sessions is finished. Do you have any questions? If yes, please, ask them now. If no, click the mouse to begin the experiment."

- "text": the text mentioned above;
- "mouse": end routine on any click.

Routine 8: "begin". The message announcing the beginning of the experiment: "The experiment is about to begin. Make sure your head is resting on the chinrest and the door is properly closed. If you are ready, click the mouse to start the first trial."

- "text": the text mentioned above;

- “mouse”: end routine on any click.

Routine 10: “thank you”. The message announcing the end of the experiment: “This is the end of the experiment. Thank you!”

- “text”: the text mentioned above. Duration set to 30 seconds.

After Routine 10 is over, the application automatically closes. All the data is saved in “.csv” file containing the name of the experiment, participant’s id number and the time of completion.

B Calibration Procedure

Calibration of the visual and audio components' congruency was conducted using Iwatsu Digital Oscilloscope DS-5314, a luminance meter and a sound level meter. The luminance meter was connected to the first channel of the oscilloscope, and the sound meter (RION NL-42) was connected to the second channel through the AC output. The luminance meter was located in close proximity to the monitor specifically where the stimuli appeared. The sound level meter was directed at one of the headphones. The changes in the visual picture on the screen or the appearance of the sound were reflected on the oscilloscope monitor. In case of the auditory signal, a small 1.3-ms delay should have been accounted for as the microphone input signal was converted into digital format for further processing and then returned back to analog format for AC output (according to RION NL-42/NL-52 Instruction Manual).

The original experiment design had to be modified for the purpose of calibration. First of all, the grey background should have been turned to a black one, so that the appearance of the stimuli on the screen would show more contrast and therefore would be more clearly seen on the waveform. The stimuli themselves were also modified in a following way. As the stimuli had a long rise and fall time for both the auditory and visual part, the depiction of their appearance on the oscilloscope's screen was a gradual rise, therefore the exact beginning of the stimuli was impossible to determine. To overcome this issue the rise time for the auditory part and the gradual opacity change for the visual part were temporarily removed. This ensured the contrast change in both modalities and therefore a clear depiction of it on a waveform.

Using the described method several measurements were done. In situations when the resulting synchronisation level was not satisfactory, certain delays were introduced. After that another set of measurements was conducted. The process was repeated till the required level of synchronisity was reached. The results of the final measurements are presented in the tables below.

B.1 Experiment 1

Introduced delay: no delay was needed.

Table 1: Experiment 1. Calibration measurements.

N	Sound delay ms	Sound meter ms	Resulting delay ms
1	4	-1.3	2.7
2	-7	-1.3	-8.3
3	4	-1.3	2.7
4	3	-1.3	1.7
5	-5	-1.3	-6.3
6	6	-1.3	4.7
7	3	-1.3	1.7
Average delay			-0.16

B.2 Experiment 2

Introduced delay: 4-ms video delay

Table 2: Experiment 2. Calibration measurements.

N	Sound delay ms	Sound meter ms	Resulting delay ms
1	-10	-1.3	-8.7
2	8	-1.3	9.3
3	-5	-1.3	-3.7
4	-3	-1.3	-1.7
5	5	-1.3	6.3
Average delay			0.3

B.3 Experiment 3

Introduced delay: 4-ms video delay

Table 3: Experiment 3. Calibration measurements.

N	Sound delay ms	Sound meter ms	Resulting delay ms
1	0	-1.3	-1.3
2	5	-1.3	3.7
3	2	-1.3	0.7
4	0	-1.3	-1.3
5	0	-1.3	-1.3
Average delay			0.3

B.4 Experiment 4

Introduced delay: 32-ms audio delay

Table 4: Experiment 4. Calibration measurements.

N	Sound delay ms	Sound meter ms	Resulting delay ms
1	0	-1.3	-1.3
2	1	-1.3	-0.3
3	2	-1.3	0.7
4	0	-1.3	-1.3
5	2	-1.3	0.7
Average delay			-0.3

C Light and sound levels measurements

The average values for luminance and illuminance are presented in the table below.

Table 5: Light measurements.

		White	Grey	Black
Experiment 1	Illuminance	2.97	1	less than 0.05
	Luminance	9.53	2.78	0.01
Experiment 2	Illuminance	2.7	0.9	less than 0.1
	Luminance	9.47	2.87	0.02
Experiment 3	Illuminance	2.7	0.9	less than 0.1
	Luminance	9.8	2.9	0.02
Experiment 4	Illuminance	3	1	less than 0.1
	Luminance	9.49	2.83	0.02

Sound measurements were done differently for the first (pilot) experiment and for Experiments 2-4:

- Experiment 1 (Pilot).

  The sound level was set to a comfortable level and was as follows: 1000 Hz continuously played sound on average at the level of 50 dB measured at the position of the participant (A-weighted, Fast). The maximum sound level of the stimuli was at 57.8 dB.

- Experiment 2-4.

  The level was set in a way, so the maximum sound level for all stimuli was 60 dB. It corresponded to a slightly higher, but still comfortable sound level compared to Experiment 1.

D Sound generation

The sounds for the experiment were generated using Melo Demo Home package based on J language. Two scripts (for AM- and FM-sounds) were created for each experiment. The texts of the scripts are presented below.

D.1 Experiment 1

Amplitude modulated tones

```
directory=:'my_dir'
  pw=:0 NB. fall and rise are cosine curve
  frs=:  44100 NB. sampling frequency
  lag=:0.000
  offtime=:0.000
  rtime =:  0.500
  ftime =:  0.500
  dur=:  2.000 NB. duration
  a=:  tone 1000
  am1=:(0.5 1 _1.57 0.000 0.000) amvib a
  am2=:(1 1 _1.57 0.000 0.000) amvib a
  am3=:(2 1 _1.57 0.000 0.000) amvib a
  am4=:(3 1 _1.57 0.000 0.000) amvib a
  am5=:(4 1 _1.57 0.000 0.000) amvib a
  (am1) savemn44 'am1.wav'
  (am2) savemn44 'am2.wav'
  (am3) savemn44 'am3.wav'
  (am4) savemn44 'am4.wav'
  (am5) savemn44 'am5.wav'
```

Frequency modulated tones

```
directory =: 'my_dir'
  pw=:0
  frs=:44100
  lag=:0.000
  offtime=:0.000
  dur=:6.000
  rtime=:0.500 ftime=:0.500
  NB.iph=:_1.57
  ifreq=:1000
  ilevel=:0
  f1 =: 1000 + 80*lft 0.5
  fm1 =: fluct f1
  f2 =: 1000 + 80*lft 1
  fm2 =: fluct f2
  f3 =: 1000 + 80*lft 1.5
  fm3 =: fluct f3
  f4 =: 1000 + 80*lft 2
  fm4 =: fluct f4
  f5 =: 1000 + 80*lft 2.5
  fm5 =: fluct f5
  (fm1,a) savemn44 'fm1.wav'
  (fm2,a) savemn44 'fm2.wav'
  (fm3,a) savemn44 'fm3.wav'
  (fm4,a) savemn44 'fm4.wav'
  (fm5,a) savemn44 'fm5.wav'
```

D.2 Experiment 2

Same as in Experiment 1, but only the following files were used: ”am3.wav”, ”am4.wav”, ”am5.wav” and ”fm3.wav”, ”fm4.wav”, ”fm5.wav”.

D.3 Experiment 3

Some changes to the script were introduced, as in Experiment 3 carrier frequency varied. Also separate scripts for AM- and FM-tones were combined into one script:

```
directory =:  'my_dir'
pw=:0
frs=:44100
lag=:0.000
offtime=:0.000
rtime=:0.500
ftime=:0.500
dur=:2.000
iph=:0
ilevel=:0
tone500=:tone 500
tone1000=:tone 1000
tone2000=:tone 2000
tone500am2=:(2 1 _1.57 0.000 0.000) amvib tone500
tone500am4=:(4 1 _1.57 0.000 0.000) amvib tone500
tone1000am2=:(2 1 _1.57 0.000 0.000) amvib tone1000
tone1000am4=:(4 1 _1.57 0.000 0.000) amvib tone1000
tone2000am2=:(2 1 _1.57 0.000 0.000) amvib tone2000
tone2000am4=:(4 1 _1.57 0.000 0.000) amvib tone2000
(tone500am2) savemn44 'tone500am2.wav'
(tone500am4) savemn44 'tone500am4.wav'
```

```
(tone1000am2) savemn44 'tone1000am2.wav'
(tone1000am4) savemn44 'tone1000am4.wav'
(tone2000am2) savemn44 'tone2000am2.wav'
(tone2000am4) savemn44 'tone2000am4.wav'
iph=:_1.57
f1 =: 500 + 39*lft 2
tone500fm2 =: fluct f1
f2 =: 500 + 39*lft 4
tone500fm4 =: fluct f2
f3 =: 1000 + 66*lft 2
tone1000fm2 =: fluct f3
f4 =: 1000 + 66*lft 4
tone1000fm4 =: fluct f4
f5 =: 2000 + 120*lft 2
tone2000fm2 =: fluct f5
f6 =: 2000 + 120*lft 4
tone2000fm4 =: fluct f6
(tone500fm2) savemn44 'tone500fm2.wav'
(tone500fm4) savemn44 'tone500fm4.wav'
(tone1000fm2) savemn44 'tone1000fm2.wav'
(tone1000fm4) savemn44 'tone1000fm4.wav'
(tone2000fm2) savemn44 'tone2000fm2.wav'
(tone2000fm4) savemn44 'tone2000fm4.wav'
```

D.4 Experiment 4

Separate scripts for AM- and FM-tones were created. Rise and fall time were changed.

Amplitude modulated tones

```
directory=:'my_dir'
   pw=:0 NB. fall and rise are cosine curve
   frs=:  44100 NB. sampling frequency
   rtime =:  0.020
   ftime =:  0.020
   dur=:  1.000 NB. duration
   a=:  tone 1000
   am21=:(2 1 _1.57 0.000 0.000) amvib a
   am41=:(4 1 _1.57 0.000 0.000) amvib a
   (am21) savemn44 'am21.wav'
   (am41) savemn44 'am41.wav'
   dur=:  2.000 NB. duration
   a=:  tone 1000
   am22=:(2 1 _1.57 0.000 0.000) amvib a
   am42=:(4 1 _1.57 0.000 0.000) amvib a
   (am22) savemn44 'am22.wav'
   (am42) savemn44 'am42.wav'
   dur=:  4.000 NB. duration
   a=:  tone 1000
   am24=:(2 1 _1.57 0.000 0.000) amvib a
   am44=:(4 1 _1.57 0.000 0.000) amvib a
   (am24) savemn44 'am24.wav'
   (am44) savemn44 'am44.wav'
```

Frequency modulated tones

```
directory=:'my_dir'
   NB. 1 second files
```

```
pw=:0 NB. fall and rise are cosine curve
frs=:  44100 NB. sampling frequency
rtime =:  0.020
ftime =:  0.020
dur=:  1.000 NB. duration
f21 =:  1000 + 80*lft 2
fm21 =:  fluct f21
f41 =:  1000 + 80*lft 4
fm41 =:  fluct f41
(fm21) savemn44 'fm21.wav'
(fm41) savemn44 'fm41.wav'
NB. 2 seconds files
dur=:  2.000 NB. duration
f22 =:  1000 + 80*lft 2
fm22 =:  fluct f22
f42 =:  1000 + 80*lft 4
fm42 =:  fluct f42
(fm22) savemn44 'fm22.wav'
(fm42) savemn44 'fm42.wav'
NB. 4 seconds files
dur=:  4.000 NB. duration
f24 =:  1000 + 80*lft 2
fm24 =:  fluct f24
f44 =:  1000 + 80*lft 4
fm44 =:  fluct f44
(fm24) savemn44 'fm24.wav'
(fm44) savemn44 'fm44.wav'
```

E Instruction

On the next page the instruction for Experiment 1 is presented. The instructions for other experiments were similar to this one with the following differences:

- Overall length (time) of the experiment and the number of breaks was changed in accordance with the number of trials for each experiment

- In Experiments 2-4 the fixation point was used differently from the Experiment 1, therefore the point 2 of "Experiment instructions" was changed to the following text: *The experiment will start with a short trial section for practicing only. You will be shown audio-visual stimuli one by one. Every stimulus consists of a visual pattern (Gabor patch) and an auditory signal 2 seconds long. There are pauses before and after the stimuli. 1 second before the stimulus the fixation point will appear, please, fixate your gaze on it. The stimuli will appear above the fixation point, however do not move your gaze from the fixation point. The gaze should be at the fixation point through the whole time of the trial - before and during appearance of the stimuli. Please, report to experimenter if you find it very difficult.*

Perceived (in)congruency experiment on 8th floor building 3, at Kyushu University

Participant's ID: _______

Dear participant,

Thank you for joining our perceived (in)congruency experiment. In this experiment, we would like you to rate the congruency of audio-visual stimuli.

We expect individual differences based on personal aptitudes and physical and mental state at the time of the experiment. There are no (health) risks involved in joining the experiment and bear in mind that you can opt out of the experiment any time - participation is on voluntary basis. The experiment consist of two parts and one part will take approximately 1 hour. The time between these parts is at least 12 hours.

To process the data accurately, we would like to ask you some information. We will use the information to analyze our data and, possibly, for data publication of group means. However, we guarantee your privacy: your data will be numbered and we will not disclose data of single individuals.

Here are our questions:
1. Do you have normal or corrected-to-normal vision? [yes / no]
2. Do you have normal hearing? [yes / no]
3. What is your age? ...... years old
4. What is your gender? [male / female]
5. What is your nationality? ____
6. How tired are you now? Please, rate from 0 to 10 (0 – not tired at all, 10 – extremely tired): before the experiment ____________; after the experiment______________
7. How sleepy are you now? Please, rate from 0 to 10 (0 – not sleepy at all, 10 – extremely sleepy): before the experiment ____________; after the experiment______________

Experiment instructions:

1. We will first set you in front of the display so the distance from you to the screen is 57 cm. Please, follow the instructions of the experimenter.
2. The experiment will start with a short trial section for practicing only. You will be shown audio-visual stimuli one by one. Every stimulus consists of a visual pattern (Gabor patch) and an auditory signal 2 seconds long. There are pauses before and after the stimuli. 1 second before the stimulus the fixation point will appear, please, look at it – it will guide you to where the stimulus will be on the screen. After each stimulus, a rating scale will appear. Please, choose the most appropriate rate for the seen stimulus from 1 to 7, where 1 is completely incongruent and 7 is perfectly congruent. After the rate is chosen, the progress bar will appear and you will be asked to click the mouse one more time to initiate the beginning of the next trial. Experimenter will be in the room during practice session, so feel free to ask any questions or give comments.
3. After the practice section is over, please, confirm that the task is clear. If anything is unclear, please, ask the experimenter about it. Experimenter will leave the room after the practice session.
4. We then will start the 1st session of the experiment. There will be 10 sessions overall, 5 in a first part and 5 in a second. There will be breaks between the sessions. Each session takes about 8 minutes, break – 3 minutes.

Thank you for your participation!

Natalia Postnova, Gerard B. Remijn

Bibliography

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Altieri, N., Stevenson, R. A., Wallace, M. T., & Wenger, M. J. (2015). Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. *Brain topography*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4043950/

Aschersleben, G., & Musseler, J. (1997). Dissociations in the timing of stationary and moving stimuli. *PsycEXTRA Dataset*. https://doi.org/10.1037/e536982012-582

Ashida, H., & Osaka, N. (1995). Motion aftereffect with flickering test stimuli depends on adapting velocity. *Vision Research*, *35*(13), 1825–1833. https://doi.org/10.1016/0042-6989(94)00270-v

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception Psychophysics*, *29*(6), 578–584. https://doi.org/10.3758/bf03207374

Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2004). Acoustical vision of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282. https://doi.org/10.1007/s00221-004-2005-z

Chen, Y.-C., & Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: A review. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00445

Covic, A., Keitel, C., Porcu, E., Schröger, E., & Müller, M. M. (2017). Audio-visual synchrony and spatial attention enhance processing of dynamic visual stimulation independently and in parallel: A frequency-tagging study. *NeuroImage*. https://www.sciencedirect.com/science/article/pii/S1053811917306699

Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., & Woldorff, M. G. (2013). Cross-modal stimulus conflict: The behavioral effects of stimulus input timing in a visual-auditory stroop task. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0062802

Eijk, R. L. J. V., Kohlrausch, A., Juola, J. F., & Par, S. V. D. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception Psychophysics*, *70*(6), 955–968. https://doi.org/10.3758/pp.70.6.955

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*. https://jov.arvojournals.org/article.aspx?articleid=2121078

Frens, M. A., Opstal, A. J. V., & Willigen, R. F. V. D. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception Psychophysics*, *57*(6), 802–816. https://doi.org/10.3758/bf03206796

Frings, C., & Spence, C. (2010). Crossmodal congruency effects based on stimulus identity. *Brain Research*, *1354*, 113–122. https://doi.org/10.1016/j.brainres.2010.07.058

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2005). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1588), 865–874. https://doi.org/10.1098/rspb.2005.3327

Gabor, D. (1947). Theory of communication. *Journal of the Institution of Electrical Engineers - Part I: General*, *94*(73), 58–58. https://doi.org/10.1049/ji-1.1947.0015

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1-2), 103–138. https://doi.org/10.1016/0378-5955(90)90170-t

Green, J. J., Pierce, A. M., & Adams, S. L. M. (2019). Multisensory integration is modulated by auditory sound frequency and visual spatial frequency. *Multisensory Research*, *32*(7), 589–611. https://doi.org/10.1163/22134808-20191402

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological science*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1431611/

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, *22*(5), 383–388. https://doi.org/10.1016/j.cub.2012.01.004

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, *15*(1), 20–29. https://doi.org/10.1162/089892903321107792

Heron, J., Roach, N. W., Hanson, J. V. M., McGraw, P. V., & Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Experimental brain research*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3324684/

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception Psychophysics*, *72*(7), 1736–1741. https://doi.org/10.3758/app.72.7.1736

Jaśkowski, P., Jaroszyk, F., & Hojan-Jezierska, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, *52*(1), 35–38. https://doi.org/10.1007/bf00867209

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, *18*(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187

Keetels, M., & Vroomen, J. (2011). Perception of synchrony between the senses. *Frontiers in Neuroscience The Neural Bases of Multisensory Processes*, 147–178. https://doi.org/10.1201/9781439812174-12

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research*, *121*(1-2), 69–79. https://doi.org/10.1016/s0166-4328(00)00386-7

Lipscomb, S. D., & Kim, E. M. (2004). Perceived match between visual parameters and auditory correlates: An experimental multimedia investigation. *Proceedings of the 8th International Conference on Music Perception and Cognition*, 72–75.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meredith, M., Nemitz, J., & Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. i. temporal factors. *The Journal of Neuroscience*, *7*(10), 3215–3229. https://doi.org/10.1523/jneurosci.07-10-03215.1987

Miyauchi, R., Kang, D.-G., Iwaya, Y., & Suzuki, Y. (2014). Relative localization of auditory and visual events presented in peripheral visual field. *Multisensory Research*, *27*(1), 1–16. https://doi.org/10.1163/22134808-00002442

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the mcgurk effect. *Perception Psychophysics*, *58*(3), 351–362. https://doi.org/10.3758/bf03206811

Orchard-Mills, E., Alais, D., & Burg, E. V. d. (2013b). Cross-modal associations between vision, touch, and audition influence visual search through top-down attention, not bottom-up capture. *SpringerLink*. https://link.springer.com/article/10.3758/s13414-013-0535-9

Orchard-Mills, E., Burg, E. V. d., & Alais, D. (2013a). Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies. *Journal of Vision*. https://jov.arvojournals.org/article.aspx?articleid=2194122

Peirce, J., Gray, J. R., Simpson, S., Macaskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Postnova, N., & Remijn, G. (2019). The effect of sound modulation mode on perceived audiovisual congruency of pure tones and Gabor patches. *35th Annual Meeting of the International Society for Psychophysics. Conference Proceedings*, 101–107.

Postnova, N., Nakajima, Y., Ueda, K., & Remijn, G. B. (2020). Perceived congruency in audiovisual stimuli consisting of Gabor patches and AM and FM tones. *Multisensory Research*, 1–21. https://doi.org/10.1163/22134808-bja10041

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. *Psychological Research*, *49*(1), 17–22. https://doi.org/10.1007/bf00309198

Schall, S., Quigley, C., Onat, S., & König, P. (2009). Visual stimulus locking of eeg is modulated by temporal congruency of auditory stimuli. *Experimental Brain Research*, *198*(2-3), 137–151. https://doi.org/10.1007/s00221-009-1867-5

Schorer, E. (1986). Critical modulation frequency based on detection of am versus fm tones. *The Journal of the Acoustical Society of America*, *79*(4), 1054–1057. https://doi.org/10.1121/1.393377

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308–308. https://doi.org/10.1038/385308a0

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10. https://doi.org/10.1097/00001756-200101220-00009

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61–70. https://doi.org/10.1250/ast.28.61

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: The same rules of multisensory integration apply. *Brain Research*, *448*(2), 355–358. https://doi.org/10.1016/0006-8993(88)91276-0

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., & Porter, N. R. (2001). When is now? perception of simultaneity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1462), 31–38. https://doi.org/10.1098/rspb.2000.1326

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 71–78. https://doi.org/10.1098/rstb.1992.0009

Tsumura, T., Ura, A., Nakajima, Y., & Teranishi, R. (1990). The auditory processing of frequency and amplitude changes included in frequency- and amplitude-modulated tones. *Journal of the Acoustical Society of Japan (E)*, *11*(5), 299–301. https://doi.org/10.1250/ast.11.299

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, *1111*(1), 134–142. https://doi.org/10.1016/j.brainres.2006.05.078

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature*, *302*(5907), 419–422. https://doi.org/10.1038/302419a0

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception Psychophysics*, *67*(3), 531–544. https://doi.org/10.3758/bf03193329

Zampini, M., Shore, D. I., & Spence, C. (2003). Multisensory temporal order judgments: The role of hemispheric redundancy. *International Journal of Psychophysiology*, *50*(1-2), 165–180. https://doi.org/10.1016/s0167-8760(03)00132-6

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, *33*(2), 248–248. https://doi.org/10.1121/1.1908630