

Master's thesis

NTNU  
Norwegian University of  
Science and Technology  
Faculty of Social Sciences  
and Technology Management  
and Department of Psychology

Gunn Kristin Halvorsen

## Gender and cerebral lateralization of audio-visual perception of emotion

Master of Science in Cognitive and Biological Psychology

Trondheim, spring 2014

## Acknowledgements

First of all I want to thank Speech lab for use of lab and equipment, and Dawn Behne for our collaboration in this and pervious projects, and friendship for the last four years. And all connected to the Speech lab, thank you all for good times and tea breaks.

I also would like to thank Elida for helping with the practical running of the experiment and for being brave enough to take part in the project, and doing a good job.

My mother, father and brother also deserves a big thank you for always being there and doing whatever they could to help when needed, like my brother Knut Ingar who got on a plane just to be a participant in the experiment. And who said that when I got my master degree he would like a Quasimodo-like creature follow me saying "Master! master!" I am looking forward to the next time we meet!

And a special thanks to my mother whom I called when my very first university exam was approaching saying that I thought I may not be cut out for this and should drop out and start to work instead, who listened to me and said that that was my choice...after I had taken the exams. So thank you to my family for believing in me when I did not do it myself.

I would also give thanks to the lovely people in Moelv for encouragement and support. And to my best friend Anette who let me use her picture in figure 1, and for always being there. And to Eirik, Marianne and Helle for much fun, great times and a few beers throughout the years!

And finally to my dear Øyvind for sharing my ups and downs, and helping with the project and being supportive in every way he could. Much love to you.

To all a big thank you!

## Content

Acknowledgements.....	I
List of figures.....	IV
List of tables.....	V
List of abbreviations.....	VI
Abstract.....	IIIX
Emotional perception.....	1
Vision and emotion.....	3
Binaural hearing and emotion.....	5
Audio-visual emotion perception.....	6
Gender.....	7
Current study.....	9
Method.....	9
Participants.....	9
Material.....	11
Procedure.....	14
Design.....	17
Results.....	17
Control conditions.....	18
Valence – and the right-hemisphere hypothesis.....	21
Gender.....	30
Discussion.....	30
Gender.....	31
Audio-visual emotion perception.....	32
Binaural hearing and emotion.....	34
Vision and emotion.....	35
Emotion perception.....	38
References.....	41

Appendix A: Informed consent form.....	48
Appendix B: Questionnaire.....	51
Appendix C: Participant receipt.....	54

## List of figures

	Page
1. Picture of experiment setup with calculations and measurements for computer monitor and visual angle .....	13
2. Picture of one of the response pads used in experiment.....	16
3. Mean percentages for correct response of emotion .....	19
4. Mean percentages for correct responses for clips presented in the different hemifields .....	19
5. Means percentages of correct responses to the different clip length.....	20
6. Means percentages of correct responses in different .....	21
7. Means percentages of correct responses in the different emotion and modalities .....	23
8. Interaction emotion and length. Means percentages of correct responses and SD in the different emotion and lengths.....	25
9. Interaction modality, emotion and length. Means percentages of correct responses and SD in the different emotion and lengths.....	28

## List of tables

	Page
1. Number of frames in the MPAFC clips trimmed in AVID before onset of emotion.....	12
2. Means percentages of correct responses and SD in the different emotion and modalities. ....	24
3. Interaction emotion and length. Means percentages of correct responses and SD in different emotion, valence and lengths .....	26
4. Interaction modality and length in RH- and valence ANOVA.....	26
5. Interaction modality, emotion and length. Means percentages of correct responses and SD for valence ANOVA. ....	27
6. Interaction modality, emotion, length. Means percentages of correct responses and SD for RH ANOVA.....	30

## **Abbreviations**

ANOVA	Analysis of variance
AO	Audio-only
AV	Audio-visual
EEG	Electroencephalography
EM	Expectation-Maximization Analysis
IBM SPSS	Statistical package for the social sciences version.
LEA	Left ear advantage
LH	Left hemisphere
LHF	Left hemifield
LVF	Left visual field
MAV	Montréal affective voices
MPAFC	Montréal Pain and Affective Face Clips
NTNU	Norwegian University of Science and Technology
NSD	Norwegian Social Science Data Services
PET	Positron emission tomography
REA	Right ear advantage

RH	Right hemisphere
RHF	Right hemifield
RRH	Right hemisphere hypothesis
RVF	Right visual field
SD	Standard deviation
VH	Valence hypothesis
VO	Video-only

## **Abstract**

Presentation of brief (120ms, 160ms, 520ms) audio prosody-, video- and audio-visual clips containing congruent emotion (anger, fear, happiness, neutral, sadness/ positive- and negative valence) was used in a divided visual field technique/dichotic listening behavioral experiment consisting of 17 males and 17 females to investigate a possible gender difference in cerebral lateralization in perception of emotion. Clips were created from Montréal affective voices and the Montréal Pain and Affective Face Clips. Accuracy percentages of correct recognition of emotion were recorded. Findings showed no support for either the right-hemisphere- or the valence hypothesis. Gender as a between subject factor was non significant. Clips containing both audio and video had the highest accuracy score of all modalities. Audio-only prosody had significant lower accuracy score compared to video-only and audio-visual clips. Positive valence in the short length may have an early accuracy advantaged compared to negative valence in the audio-visual modality that dissipates in 120ms-160ms range, with the accuracy difference disappearing between the categories. The same advantage can be found in anger, while happiness, fear and neutral have no significant differences in accuracy in lengths in the audio-visual modality.

## **Emotional perception**

To recognize the emotion of people around us is an important ability. How do we best recognize emotion? By the sound of laughter, or the sight of a smile, or when the both are combined? And are females better at recognizing emotion in others compared to men?

The ability to perceive facial emotions was studied as early as the 19th century by Darwin who proposed that emotion was closely related to action (Darwin, 1872; Frijda, 1989). The ability to recognize what others feel is a cognitive skill that is fundamental for our social interaction (Collignon, Girard, Gosselin, Saint-Amour, Lepore, & Lassonde, 2010). Both Darwin and Ekman (Ekman, 1992; Vytal, & Hamann, 2010) argue for a limited set of basic emotions consisting of happiness, sadness, anger, fear and disgust. Emotions can be expressed through facial affect, affective prosody, and emotional lexical items. Emotion itself consist of three different components; valence; a range from positive/pleasant to negative/unpleasant, arousal; a range from calm to aroused, and lastly through motor activation with a range from approaching to avoiding, or neither of the two (Gadea, Espert, Salvador & Martí-Bonmatí, 2011). Emotion prosody consists of different pattern of pitch, loudness and length which combined communicates emotion. Laughter may therefore function as emotional prosody for a positive emotion as happiness (Gadea et al., 2011).

The processing of speech and language has through many years of research been found to be lateralized predominantly to the left hemisphere (Studdert-Kennedy, & Shankweiler, 1969; Kimura, 1961; 1987). The first connection between the right hemisphere and processing of emotion was first documented hundred years ago by Mills who connected right-hemisphere lesions to a decrease in emotional expression (Found in Demaree, Everhart, Yongstrom, & Harrison, 2005).

Further study of the lateralization of emotion has produced different hypotheses on the subject. The right-hemisphere hypothesis postulates that the right hemisphere is specialized for emotional expression, experience and perception (Demaree et al., 2005; Erhan et al., 1998; Heilman, & Bowers, 1990; Tucker, 1981). Behavioral studies have also shown evidence that the left side of the face, and therefore the right hemisphere is more emotionally expressive (Sackeim et al., 1987, found in Wager, Phan, Liberzon, & Taylor, 2003). Deficits in emotional speech sounds, or prosody have been found in people with damage to the frontal part of the right hemisphere (Ross & Mesulum, 1979, found in Wager et al., 2003).The ability

to recognize facial expressions of emotion have also been linked to the right hemisphere from studying hemispheric damage (Wager et al., 2003). Support have also been found through experiments using divided visual field technique (Lavadas, Umiltà & Ricci-Britt, 1980)

On the other side many studies have not found evidence to support the right-hemisphere hypothesis in lateralization of emotions (Caltagirone, Ekman, Friesen, Gainotti, Mammucari, Pizzamiglio, & Zoccolotti, 1989; Mammucari, Caltagirone, Ekman, Friesen, Gainotti, Pizzamiglio, & Zoccolotti, 1998; Wager et al., 2003). Wager et al., (2003) found no advantage for the right hemisphere when they performed a meta-analysis consisting of 65 neuroimaging studies looking at lateralization of emotion.

The valence hypothesis has its inspiration from the Wada test experiments in which barbiturates is injected into arteries leading to only one hemisphere inhibiting neural activity in said hemisphere for a said amount of time (e.g Tan, 2013). Wada experiments found that when the left hemisphere was inhibited depressive symptoms like crying, pessimism and guilt were found. When the right hemisphere was inhibited symptoms of euphoria were evident (Goldstein, 1939 found in Tan, 2013). These Wada findings concluded that the right hemisphere (RH) was specialized for negative valence emotions (anger, sadness, disgust and fear) and the left hemisphere (LH) for positive valence emotions (happiness, interest, pleasant surprise) (Tan, 2013; Erhan, et al., 1998; Wager et al., 2003). The valence hypothesis divides the experience and expression of positive emotion to the left hemisphere, while the right hemisphere holds the negative emotions (Erhan, et al., 1998; Silberman & Weingartner, 1986). Pourtois, de Gelder, Bol, & Crommelinck (2005) found evidence for the valence hypothesis with the use of positron emission tomography (PET) activating structures in the left hemisphere when audio-visual clips with the emotion happiness were presented. A connection between the emotion fear and activation in the right hemisphere were also found in the same study supporting the valence hypothesis.

Evidence from other studies done in regards to the valence hypothesis have not all been in support for the hypothesis. Borod (1992) when looking at lesion data found support for the right hemisphere dominance and not the valence hypothesis when comparing the two approaches. And several studies have not found support for the valence hypothesis with use of EEG (Wager et al., 2003).

The approach-withdrawal model connects emotions to approach or withdrawal behaviors and locating this process with the left- and right-anterior regions of the brain (Demaree et al., 2005). This hypothesis uses evolutionary concepts arguing that emotions are closely linked to behavior of an individual in its environment (Alves, Fukusima, & Aznar-Casanova, 2008; Demaree et al., 2005). Negative emotions elicit a withdrawal response while positive emotion elicits an approach response (Gadea et al., 2011). Emotions that therefore elicit the withdrawal response are fear, sadness or disgust, trying to lead the individual away from the source of the emotion. Emotions as anger, happiness and surprise elicit the approach response, triggering a drive toward the source of the emotion (Alves et al., 2008). Anger stands out in this model being a negative emotion eliciting the approach- and not the withdrawal response and correlation to activation in the left frontal part of the brain. This evidence goes against the right-hemisphere- and valence theory which puts anger in the right hemisphere (Carver & Hamon-Jones, 2009). Even with this divide the valence hypothesis and approach-withdrawal model share many similarities and are closely linked (Alves et al., 2008).

In the research of emotion perception both the right-hemisphere hypothesis (Demaree et al., 2005; Erhan et al., 1998; Heilman, & Bowers, 1990; Tucker, 1981) and the valence hypothesis (Erhan, et al., 1998; Silberman & Weingartner, 1986) are the most used approaches. Still new views as the approach-withdrawal model (Demaree et al., 2005) also try to explain emotion perception. To take a closer look at emotion perception vision, binaural hearing, the use of audio-visual clips and gender differences are all important aspects of understanding more about emotion perception.

### **Vision and emotion**

According to the divided visual field theory, clips that are perceived in one hemifield are received and processed by the contra lateral hemisphere. Visual information in the left hemifield (LHF) are therefore received and first processed by the right hemisphere (RH), and a stimulus in the right hemifield (RHF) first received and processed by the left hemisphere (LH) (e.g. Bourne, 2006).

When reading, looking or searching for an object, our eyes make natural eye movements, known as saccades. Between saccades our eyes can remain fixated for a duration of 200-300ms. The duration of one saccade depends on the distance. For example, a two degree

saccade when reading takes 30ms, whereas a five degree saccade will take 40-50ms (Rayner, 1998).

The visual field, when looking straight ahead, can be divided into the foveal-, parafoveal-, and peripheral region. The foveal region stretches from the center, straight ahead to 2 degrees to both sides from center. Accuracy is best in the foveal region. When our eyes fix straight on an object the object is located in the foveal region of the visual field. Looking at an object in the foveal region allows the best possible perception of the object. The parafoveal region stretches from the foveal region to 5 degrees from center vision, and is not as good of an area to perceive objects compared to the foveal region. Beyond the parafoveal region lays the peripheral region extending from 5 degrees and beyond. This part of the visual field results in the poorest perception of an object (Rayner, 1998). The movement of the eyes and the visual angle together produce the concept of eccentricity. These facts need to take into consideration when using divided visual field technique to research lateralization. If both eyes perceive the clip at all times one cannot say anything about lateralization because both eyes/hemifields, and by that, both hemispheres will receive the information (Jordan & Thomas, 2007; Rayner, 1998).

Since the two hemispheres are connected through the corpus callosum the length of clips presented is critical. One might control the placement of the visual clips, but ballistic eye movements like saccades can give the information to both visual fields and thereby both hemispheres, one have to therefore control the degrees from the eye to the fixation point and clip and ensure that the clip is no longer than 180ms for the divided visual field technique to work correctly (Bourne, 2006; Jordan & Thomas, 2007; Rayner, 1998).

Behne, Wang, Belsby, Kaasa, Simonsen, & Back, (2008) used divided visual field technique to try to answer if hemifield in which visual speech was presented affected the audio-visual perception. 20 participants (8 male, 12 female) were set to answer a forced-choice speeded identification task. Participants were presented with McGurk audio, visual or audio-visual clips. The clips were presented at three different locations on the monitor, left from fixation point, center on the monitor or to the right, and in six different lengths: 80, 120, 160, 200, 240 and 280ms. The result showed increased audio-visual fusion in the right hemifield after excluding that the difference was a result of facial asymmetry. The results from the different lengths used in the experiment showed results indicating early audio-visual processing in the right hemifield, corresponding with left hemisphere dominance for speech segments. All clip

positions on the monitor gave findings corresponding with the visual component presented with short visual clips lengths. When visual clips lengths reached 160-200 ms the trend shifted from visual responses to audio responses (Behne et al., 2008).

Research on emotion using divided visual field technique has shown that visual clips presented in the left hemifield were judged by the participants as more emotional and gave greater autonomic responses (Levine, & Levy, 1986; Wager et al., 2003). Önal-Hartmann, Paulio, Ockleburg, & Güntütkün (2012) used divided visual field technique combined with flexion and extensor response with the use of a joystick to try to find support for the valence hypothesis. They hypothesized that positive visual clips would elicit flexion responses “towards action” to the visual clips when presented to the left hemisphere, while extensor “away action” response would occur when the right hemisphere were presented with negative visual clips. Findings showed that flexion responses were faster for positive visual clips, and negative visual clips elicited faster extensor responses. They also found a right hemisphere advantage for negative visual clips, and the shortest response time when presenting visual clips in the left visual field. The response times showed no hemispheric advantage for positive visual clips. The mixed findings showing no clear support for either hypothesis and it is still needed more research on lateralization of emotion though the use of visual field technique.

### **Binaural hearing and emotion**

Dichotic listening is a reliable, non-invasive way to investigate lateralization in the brain and therefore a frequently used tool used in research (Gadea et al., 2011). In dichotic listening the participant is presented with two different sounds at once, one going to one ear the other sound to the other ear. One can also use only one sound, presenting the sound to only one ear while the other are presented with noise. Dichotic listening exploits the brains incapacity to handle two things at the same time. Audio presented to the right ear is received and processed by the left hemisphere, and audio presented to the left ear are received and processed first by the right hemisphere (Hugdahl, 2000). The dynamic dual pathway model of auditory language comprehension, syntactic and information with semantic qualities is primary processed in the left hemisphere temporal-frontal pathway, and prosody is processed in the right hemisphere temporal-frontal pathway via the posterior part of the corpus callosum (Gadea et al., 2011).

When using emotion prosody, an intonational pattern used to convey an emotion, researchers have found a left ear advantage (LEA) supporting the right hemisphere hypothesis (Erhan et

al., 1998). When using nonverbal sounds, neutral sentences stated with emotional intonation (as mentioned in Erhan et al., 1998) or nonsense and emotionally intoned neutral sentences where white noise or babble were presented to the right ear supports the right-hemisphere hypothesis (Haggard & Parkinson, 1971; Hatta & Ayetani, 1985; Henry, 1983; Safer & Leventhal, 1977). Erhan et al. (1998) found a LEA independent of emotion in recognition accuracy for all types of emotion when presented as audio indicating a right hemisphere advantage.

Gadea et al. (2011) reviewed several dichotic studies and suggested that the right hemisphere has an advantage for the prosody of sadness. Anger were suggested to be processed more bilaterally or more in the left hemisphere following in the lines of the valence hypotheses indicating that emotion perception of prosody is in need of more research.

### **Audio-visual emotion perception**

The use of audio-visual clips, also called bimodal clips, is relatively new practice in research on perception of emotion. Previous studies on perception of emotion have been concentrated on face perception using still pictures of facial emotions going back to Ekman's (1992) basic emotion studies that showed differences in perception of emotion in regards to age or culture. De Gelder & Vroomen (2000a) proposed that since we use eyes and ears in speech perception the same bimodality can be said about perception of emotion, expressed in the voice as well as in the face, giving bimodal clips strong ecological validity (Collignin et al., 2008; 2010; Kreifelts, Ethofer, Grodd, Erb & Wildgruber, 2007; de Gelder & Vroomen, 2000a; 2000b; Simon, Craig, Gosselin, Belin, & Rainville, 2008). The modality of emotion clips used evolved from Ekman's (1992) classic emotion still pictures to bimodal clips with facial expression video clips and audio clips with emotional uttering's or prosody. This use of video and or audio clips in researching perception of emotion has only been seen in research the last years (Collignin et al., 2008; 2010, Simon et al., 2008). In Pourtois et al., (2005) the use of positron emission tomography (PET) showed that a greater regional cerebral blood flow to middle temporal gyrus during audio-visual clips than only visual- or audio only clips, indicating a neurobiological difference between the modalities. Behavioral audio-visual presentation of non-verbal emotional information have produced a higher accuracy recognition score than video- or audio clips alone (Collignin, Girard, Gosselin, Roy, Saint-Amour, Lassode, & Lepore, 2008; Kreifelts et al., 2007) Collignin et al. (2010) found in their study that audio-visual clips resulted in a significant advantage in a speed-accuracy composite

score compared to the score produced from the audio-only or video-only clips. Audio-visual clips are therefore a modality here to stay in regards to research on emotion perception and will continued to be used in research years to come.

## **Gender**

There have been found a difference between the genders when it comes to emotion. Women have in these studies been found to be more emotionally expressive compared to men (Kring & Gordon, 1998; Wager et al., 2003). Some have argued that this may be a result from a difference in socialization between the genders (Grossman & Wood, 1993, found in Wager et al., 2003). Women have also been found to have stronger psychophysiological responses to clips of an emotional character compared to men (Kring & Gordon, 1998; Orozco & Ehlers, 1998; Wager et al., 2003).

From an anatomically view women exhibit more gray-matter volume in the cingulated cortex, a part connected to the limbic system, compared to men (Good, Johnsrude, Ashburner, Henson, Fritson, & Frackowiak, 2001). Schneider, Habel, Kessler, Salloum, & Posse (2000) found that males show activity in the amygdala during negative affect from being exposed to pictures portraying sadness, while the females did not show this activation pattern during neuroimaging, indicating neurobiological differences between the genders. Studies have also revealed that men show more lateralization of brain function than women (Hines, Chiu, McAdams, & Bentler, 1992; Russo, Persegani, Papersch, Nicoloni, & Trimarchi, 2000; Wager et al., 2003). Good et al. (2003) showed also that men had more asymmetry in the temporal cortex than women.

Only a few studies have taken a closer look at gender differences in the processing of prosodic audio see Belin, Fillion-Bilodeau, & Gosselin (2008) for review. An experiment done by Collignin et al. (2010) showed results indicating a sex difference when processing and expressing emotion. In this experiment the participants were to discriminate the emotions of fear and disgust expressed through audio clips, video clips and audio-visual clips. The women in the experiment outperformed men in processing audio clips, video clips and bimodal clips by answering faster and more accurate on which emotion that was presented. The results also suggested that women not only processed emotional information more efficiently, but were also better at integrating vocal and facial expressions. Collignin et al.

(2010) suggests that these findings support the hypothesis that men and women process emotions differently.

An early experiment done by Lavadas, Umiltà & Ricci-Britt (1980) found a gender difference where the female participants answered fastest when clips was shown to the right hemisphere through presenting the emotional picture in the left visual field. This experiment used 36 photographs from three female- and three male actors and six different emotions: happiness, sadness, fear, anger, disgust and surprise. The experiments result indicates a female advantage in discriminating emotion compared to the male participants, and give support to the right-hemisphere hypothesis. Thayer & Johnsen (2000) found support for a female advantage producing more accurate discriminations than the male sample. Here 28 females and 16 males watched still pictures expressing disgust, fear, happy, surprise, sad and angry from Ekman and Friesen standardized facial expression set, and answered how much they had experienced the a array of different emotions with the six emotion used included. The females sample classified the emotion right 45.7% of the time compared to 39.7% in the male sample. The same study also argued that females may base their judgment on the emotional content in the clips and males having more difficulty extracting the emotional information from an emotionally expressing face. The male sample had the most difficulty in recognizing negative emotions expressed by a female, supporting this. The results from this study showed a gender difference where females compared to males may have a differentiated representation of emotion making them better at accurately determining the correct emotion shown independently of gender expressing the emotion.

Wager et al. (2003) through performing a meta-analyse of 65 neuroimaging studies on emotion found that emotional processing showed that men exhibit more lateralized activation of emotion in the brain than women.

The study of gender difference in perception of emotion are of importance and can give information about gender differences found in mood anxiety disorders and depressive disorders that have a higher frequency in women compared to men (Gohier, Senior, Brittain, Lounes, Law, Phillips, & Surguladze, in press; Rose,2011) or why there are a higher frequency of men compared to women that suffers from disorders that effect their ability of understanding and recognizing emotion as in autism (Collignin et al, 2010) .

## **Current study**

This study investigates gender differences in cerebral lateralization in perception of emotion by presenting video and audio-visual clips of emotional facial expressions to the left, center and right hemifields, and by presenting clips containing emotion prosody dichotic to only one or both ears as behavioral tests of laterality.

The study will first examine if there exists support for the right-hemisphere hypothesis seen by having a higher percentage of correct answers when audio-, video- and audio-video clips are presented in the left hemifield/ear (e.g Demaree et al., 2005; Erhan et al., 1998; Heilman, & Bowers, 1990; Tucker, 1981), or for the valence hypothesis where emotions with positive valence are expected to have a higher percentage of correct answers when presented in the right hemifield/ear, and negative valence when presented in the left hemifield/ear (e.g Erhan, et al., 1998; Silberman & Weingartner, 1986; Pourtois, de Gelder, Bol, & Crommelinck, 2005).

The study will further examine if there is any support for a gender difference. The female sample is hypothesized to have higher percentage of correct answers independently of modality compared to the male sample (Collignin, 2010). Men is hypothesized to have the highest accuracy rate when clips are presented in the left hemifield giving support that men are more right hemisphere lateralized in processing emotion (Hines, Chiu, McAdams, & Bentler, 1992; Russo, Persegani, Papersch, Nicoloni, & Trimarchi, 2000; Wager et al., 2003).

## **Method**

### **Participants**

The final sample consisted of 34 participants ranging from 19 -29 years, with an mean age of 22,79 (SD=2,69). In this sample 17 were males with an mean age of 22,94(SD=2,82), and 17 females with a mean age of 22,65(SD=2,64).

The decrease in participants from 38 to 34 was due to 2 failed audiometric tests, one failed Snellen vision test and 1 from questionnaire due to not being a Norwegian native.

The participants were predominately students located in Trondheim, Norway. The sample was recruited through information given to students at NTNU via their intranet and word of mouth.

Before the experiment participants took an audiometric test to ensure normal hearing. The requirements of normal hearing were an audiometric score over 25dB from 250-4000 Hz. The audio used in this experiment was presented dichotic, such that one ear were presented audio while the other ear were presented with pink noise, this made it especially important to ensure that the hearing was not only normal, but that the ears have comparable acuity. The audiometric test was given to one ear at a time, and the participant had to have hearing within the requirements on both ears to pass the audiometric test.

The participants completed a Snellen vision test to make sure that they had the vision requirements needed for this experiment. To pass the Snellen test the participants had to read the correct letters up to line 8 from a distance of 4,15 meters, giving them at least 20/25 to 20/20 vision. Participants were allowed to use glasses or lenses as long as they passed the Snellen test and used them during the experiment.

If a participant did not pass the audiometric- or Snellen vision test they were excluded from the sample and did not take part in the experiment. They also received a letter stating that the person performing these test where not a audiometrician or optometrist by profession, and that the result itself did not conclude that there was something wrong with their hearing or vision, but merely that they did not fit in for the specific criteria's for this experiment.

The participants were also given a questionnaire including questions about general health, age, occupation and possible medical conditions that may impact the results of the experiment. (See appendix B). To ensure that the participants all where right-handed they completed a revised version of the Edinburgh Handedness Inventory to determine handedness (Williams, 2008).

Handedness of participants is important since neurocognitive organization of those who are left-handed may differ from those who are right-handed. Cognitive processes that usually are located to the right can in some cases be found in the left or both hemispheres. The hemispheric processing in perception of emotion that normally takes place in the right

hemisphere may not take place in the right hemisphere in a left-handed participant making it best to exclude them from the sample (Kimura, 1961).

## **Material**

The audio clips used in the experiment came from “Montreal affective voices” (MAV) an auditory equivalent to the affective faces picture by Ekman & Friesen (Belin, Fillion-Bilodeau, & Gosselin, 2008). The standardized set consist of burst of affect equivalent to emotions of anger, fear, disgust, pain, sadness, happiness, pleasure and a neutral auditory expression. For this study the prosody for anger, fear, sadness, happiness and neutral was chosen from a male (audio 55) and a female speaker (audio 60) of the ten (five men, five female) available for free download at (<http://vnl.psy.gla.ac.uk/resources.php>) and described in detail in Belin, Fillion-Bilodeau, & Gosselin, (2008).

In Praat version 5.3 (Boersma, & Weenink, 1995), available for free download at (<http://www.fon.hum.uva.nl/praat/>), the audio clips were all adjusted to 68dB and edited to the three desired lengths of 120, 160 and full 520ms and matched with pink noise at a level of 58 dB to make the dichotic audio clips. The full audio clips for fear from the female speaker contained 440 and not 520ms of prosody. The decision to keep the 440ms audio clip was based on the fact that the 520ms clips would function as the “full” expression and the 440ms prosody of fear meet that criterion. The audio-only condition with only audio presented dichotic had three different positions: audio presented to the right-, audio to the left ear or audio presented to both ears.

The visual video clips used came from “Montréal Pain and Affective Face Clips” (MPAFC) where video clips of a female (F1)- and a male speaker (M2) was chosen. The clips of the same emotions as chosen from the MAV set were downloaded from (<http://amor.cms.hu-berlin.de/~simondan/>) after contacting and receiving permission from Daniela Simon (Simon, Craig, Gosselin, Belin, & Rainville, 2008). The “Montréal Pain and Affective Face Clips” are standardized video clips of dynamic, prototypical facial expression of emotions. The emotional expression where modeled after the “Facial Action Coding System” FACS created by Ekman & Friesen and the production are described in detail in Simon et al. (2008).

To ensure onset of emotional cues started at a similar time across emotion the five emotions were viewed in the editing program AVID and noted when the first frame in the clip showed

any form of change from a neutral expression. This was set as onset for emotion and frames before this trimmed.

Table 1

Number of frames in the MPAFC clips trimmed in AVID before onset of emotion in video.

Emotion	Female	Male
Anger	2	3
Fear	6	3
Happiness	3	2
Sadness	11	1
Neutral	0	0

The study depended on the length of the clips, the shortest 120ms; this made it crucial that the onset of emotional cues was similar in time. The onset of emotions varied. It should also be noted that in some of the clips the speaker did not have the mouth closed at the beginning of the clip making it hard to determine onset of emotion.

In the video-only condition participants was presented with only video with no sound at the three different positions on the screen

The audio-visual congruent clips of the five emotions (anger, fear, happiness, neutral, sadness) were created by combined the audio (MAV) and the visual clips (MPAFC) and editing them in AVID to the three length (120, 160, 520ms) and exported as 682 x 522 pixels.

The audio-visual clips had audio to both ears when the video where presented in the center location of the screen. When the audio-visual clips where presented to the left of the fixation point the left ear received audio while the right ear pink noise. When presented in the right of fixation the right ear received audio while the left pink noise.

The clips presented in the left/right hemifields had a distance of 3cm from the speakers ear, and 6 cm to the middle of the speakers face from the fixation, a 1x1cm red cross. In the center location the clips where presented in both hemifields and on top of the fixation.

The distance from the clips center to the left/right of the fixation point was calculated using Pythagoras formula using 76 cm as the distance from the screen, and an angle of  $2,2^\circ$  to the speakers ear and  $4,5^\circ$  corresponding to the middle of the speakers face in the video (see figure 1), corresponding with knowledge of eccentricity (Jordan & Thomas, 2007; Rayner, 1998).

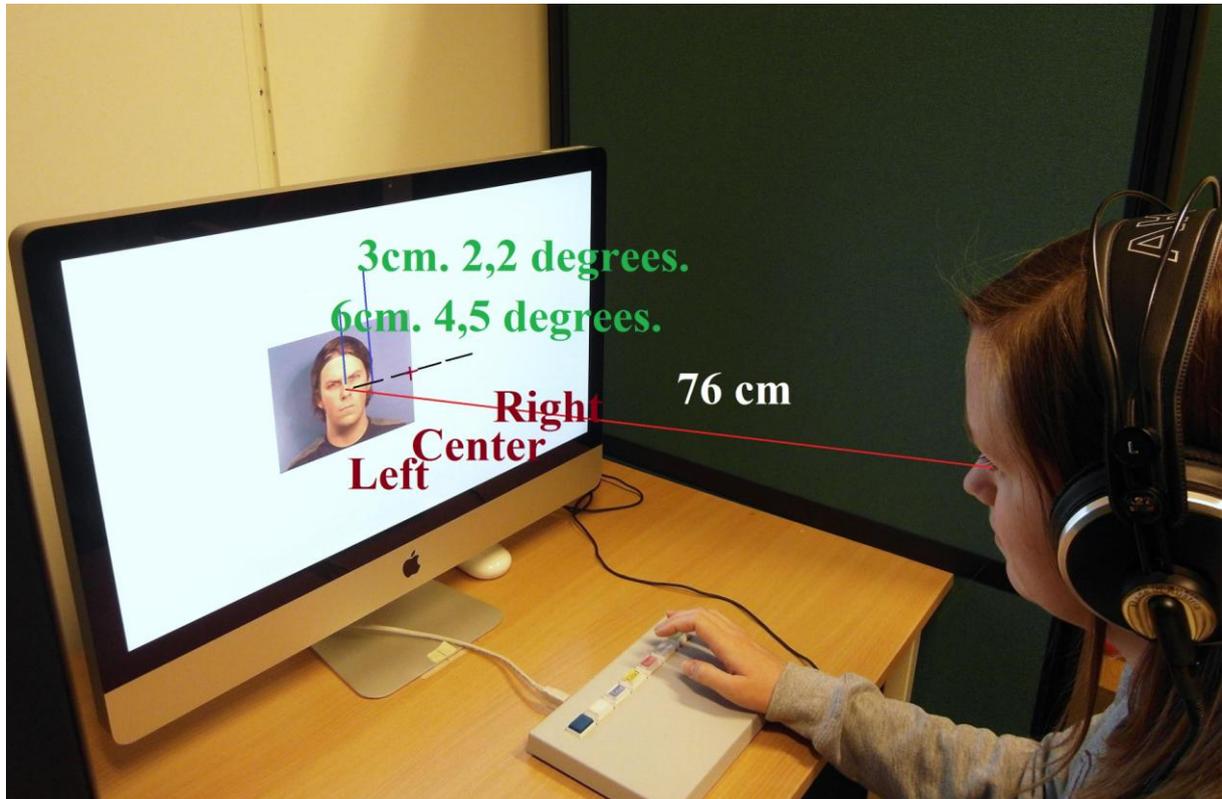


Figure 1: Picture of experiment setup with calculations and measurements for computer monitor and visual angle (Sketch: private).

The experiment was presented on an Apple iMac 7.1 with a 24 inch computer, 1920x1200 pixels. AKG K271 stereo closed dynamic circumaural studio. The sound level was set to four dots where eight dots correspond to 68 dB. The brightness levels for the screens were set at 8 dots. SuperLab version 4.5 was used to run the experiment.

Each clip was presented four times randomly resulting in 360 clips in the AV condition (five emotions x three lengths x three monitor positions x two talkers x four repetitions), 360 clips for the video-only condition (five emotions x three lengths x three monitor positions x two talkers x four repetitions) and 360 clips in the audio-only condition (five emotions x three

lengths x three ear presentations x two talkers x four repetitions), for a total of 1080 trials in the experiment.

## **Procedure**

The experiment took place at Speech lab at Dragvoll, NTNU. The study was advertised through NTNU intranet open for students and through word of mouth. The information given in the advisement asked if anyone were interested in participating in a experiment researching emotional perception. There was also listed criteria's for being able to participant; normal vision, normal hearing, native to Norway, and between the age of 18 to 30 years old. When receiving emails from possible participants more information was given and questions answered before scheduling a time for participation in the experiment. The participants were made anonymous by giving each participant a code used for all forms and results. This code consisted of the letter M or K, meaning "mann" or "kvinne" indicating gender, and a number between 1 up to 20.

Informed consent were given before pre-tests the audiometric-, Snellen test, and experiment start, containing the information required about the experiment, withdrawal, and the question if the participant would like to participate in the experiment (see appendix A). Participant who passed the tests came back to take the experiment. Before arrival the experiment leader set up the computers, response boxes and tested headphones. The placement of the screens on the tables, and the chairs were measured so that they were centered to each other. Markings on tables and floor ensured the same placement of equipment for each participant. SuperLab was opened and ready to run.

Participants entered the computer-lab and sat down in front of a computer. On the table the questionnaire - (see appendix B) and the handedness test forms where already placed and with correct participant number. The participants were instructed to fill in these forms and ask if they had any questions.

Up to four participants took the experiment at the same time. Each participant was seated in a chair 76 cm from a computer monitor. The chairs were solid four-legged chair to ensure the participants would keep the determined position to the screen and not move around during the experiment. The distance from the screen to the participant was measured by hand by the experiment leader. The participants were told to find a normal and comfortable position in the

chair that they could hold for the duration of the experiment. The experiment leader measured from the screen to middle of the participant's pupil seen from the side. The participant then moved the chair forward or backwards within the markings on the floor depending on the length measured. This procedure proceeded to the desired length was obtained. To ensure correct length the experiment leader measured one more time after the desired length was obtained to verify correct measurements. The participant was told to try to hold still and not move around in their seats and not talk to each other during the experiment. The experiment leader checked that all headphones were correctly placed on the ear and matched the left/right markings on the headphones.

The SuperLab experiment and the pre-test phase started. Information appeared on their screen regarding the different types of clips they would be presented and that the lengths of the clips might be perceived as short. They were informed to answer as fast as they could, by pressing the button that matched the emotion perceived using only their right hand. They were encouraged to answer even in they were not 100 percent sure of the emotion shown, and to keep looking at the fixation at all time. There was given information on the slides informing that if a technical situation, or if the participants at any time wanted to withdrawal they were to leave their chair and the room as quietly as possible. They were also informed that there would be 10 breaks in the experiment. In conjunction with the breaks they were reminded to always look at the fixation point and received information on which clips type would be presented next.

For each slide with information in the beginning they had to push a specific button labeled with one of the emotions. All emotions where pushed during the information phase and they were all randomized between the participants. They were given information that some example clips would be shown and if there were questions they could raise their hand and the experiment leader would come over. When the test phase was over the experiment leader left the room and closed the door.

The clips were presented in a forced-choice speeded identification task with buttons corresponding to the five different emotion alternatives. The response pads (Cedrus RB-730) with seven keys was used in the experiment (figure 2).



Figure 2: Picture of one of the response pads used in experiment.

Five of the seven buttons was marked with the five different emotions used in the experiment. The different emoticons were anger, fear, neutral, happiness and sadness. The emotions were labeled in Norwegian on the buttons. The buttons had colored backgrounds to distinguish the placement of the different emotions. Color choice was based on the wheel of emotion by Plutchik (1980). “Sint” (Angry) was written in white on a red background, “Frykt” (Fear) was written in white on a green background, “Nøytral” (Neutral) in black on a white background, “Glad” (Happiness) in black on yellow background and “Trist” (Sadness) in white on a blue background. Times New Roman font was used in size 23 for “Frykt”, “Glad”, “Trist”, 21 for “Sint” and size 17 for “Nøytral”. The different font size depended on length of emotion name and having as big letters as possible without squeezing the letters.

The placement of the five buttons on the response pad was randomized to each computer.

The experiment trials were blocked together based on audio-only, video-only and audio-visual clips. The order of blocks was randomized to each computer. Between each block a break followed, in total 10 breaks. The participants were encouraged to relax for a moment and continue when they are ready by pushing one of the buttons.

The clips were presented either to the left, center or to the right of the fixation. Participants had three seconds to respond from onset of clip. After the clip had ended a white background with a red fixation reappeared and remained on the screen until response was given or tile

three second time limit had expired. Responses after time limit was categorized as errors and ruled out of the analyses.

The experiment took 35-45 minutes depending on quickness of response and duration of breaks. After completion the experiment leader entered and started debriefing. Debriefing consisted of questions on how they had perceived the experiment and explained the experiment and pre-tests. The experiment leader also made sure that everyone left in the same state as they came. Relevant information from the debriefing was noted in a logbook. After the debriefing the participants were offered a Midtby gift card worth 200 NOK for their participation. If the gift card was accepted the participant signed for it (see appendix C).

With pre-test and the experiment the full experiment time had a mean duration of 60 minutes.

The project abided by the ethical guidelines given in the field of psychology. All information about the data collection and storage from this experiment was sent and registered at NSD and approved. All identifying data was made anonymous by giving the participants their own participant-number. Questionnaires and other non-digital information, such as email addresses or telephone numbers, were stored in a locked safe in the Speech Lab. After the project is finished, this data will be deleted or shredded.

## **Design**

The experiment had a within -and- between -subject design. There were four categorical independent variables (modality, emotion, length and direction), and one continuous dependent variable (accuracy in percent) and one categorical between subject variable; gender of participant.

## **Results**

To answer the hypothesis in this study three ANOVA's were used. The result section will start with the first ANOVA which looked at the control conditions in the experiment. Second look at support or non-support for the valence- or right-hemisphere hypothesis with the use of a second- and third ANOVA. Finally the section will examine if there is evidence for any differences between genders in the sample.

## Control conditions

Responses were tabulated based on match of emotion shown in clip and response from participant.

The SuperLab response data was exported to Excel where the responses was sorted and transformed into percentages based on the response the participants reported when presented with an emotion in a clip. This was done by as “IF..” formula in Excel. If response matched emotion shown the result was 25. If the response did not match the result was 0. The four repetitions where then added giving a max percentage of 100% for 4/4 right responses, 2/4 giving 50% and so on.

The dataset was exported to SPSS. To handle missing responses Expectation-Maximization (EM)-analysis was used. 20 (0.2%) of 9160 (99.8%) responses were missing from the accuracy data. Prior to use of EM Little’s MCAR was completed with chi-square=,00 (df=5906;  $p=1.00$ ), which indicates that the missing responses were missing at random and therefore fulfilling the requirements for running a EM analysis.

The female- and male speaker were collapsed through the use of the compute command in SPSS with a formula of (clip male speaker + same clip female speaker)/2 to get the average percentage for further analysis.

ANOVA mixed repeated measures was performed for the control conditions 2 (gender: male and female; between subject factor) x 3(modality; audio-only, video-only, audio-visual) x 5(emotion: anger, fear, happiness, neutral, sadness) x 3 (length: 120ms, 160ms, 520ms) x 3(direction: left hemifield/ear, center/both hemifields/ears, right hemifield/ear). An alpha of .05 was used for this and all of the ANOVA’s.

As a control condition to the emotions used in the experiment neutral facial video- and/or neutral vocalization clips was used. There were found a significant difference between the emotions presented and percentage of correct responses;  $F(3,94)=26,98$ ,  $p=.001$ . With the use of Bonferroni correction the only difference was revealed to be between the neutral control and the four different emotions by having a significant higher percentage of correct responses compared to the other emotions used (see figure 3).

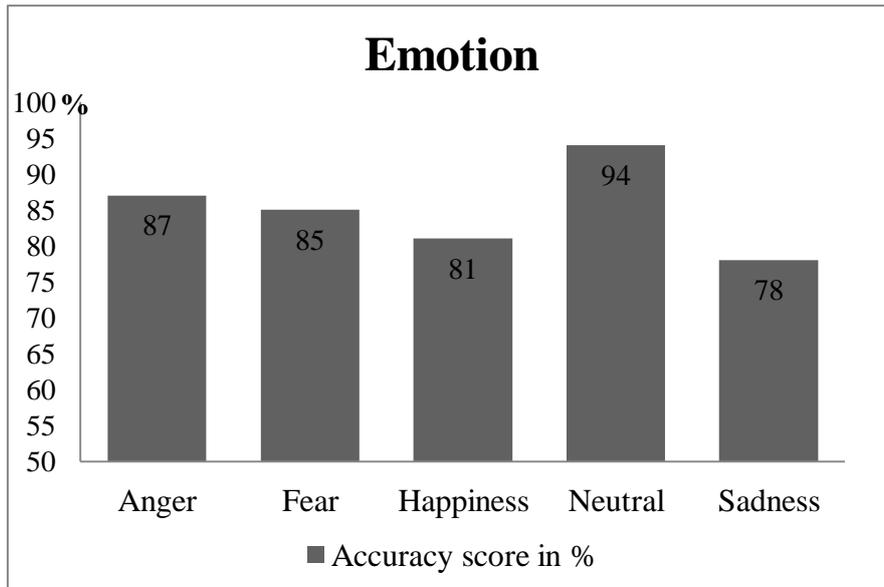


Figure 3: Mean percentages for correct response of emotion shown across length, modality and hemifield placement.

Another control condition used was the central placement on the screen corresponding to both hemifields, and sound to both ears. Presentation to hemifield was significant  $F(2,64)=5,87$ ,  $p=.005$ , and through Bonferroni correction found to be in connection to clips presented in the central hemifield/ear compared to the left- and right hemifield (see figure 4).

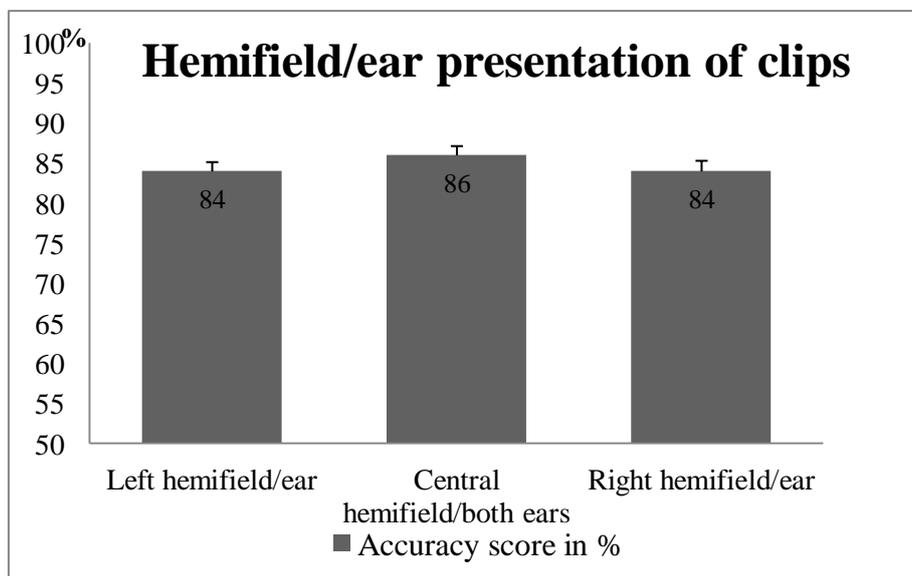


Figure 4: Mean percentages for correct responses for clips presented in the different hemifields/ears across length, emotion and modality.

There were three different clip lengths used in the experiment. The audio-only-, video-only- and audio-visual clips containing the 520ms full length of the emotion that was used as a comparison against the two shorter lengths (120ms and 160ms). Clip length was found to be significant  $F(1,41)=304,72, p=.001$ , and Bonferroni showed that the significant differences was found between all three different lengths. From the full 520ms clips having the highest percent correct responses with 93% ( $M=93\%, SD=0,8$ ) to 120ms having the lowest with 76% ( $M=76\%, SD=1,47$ ) (figure 5).

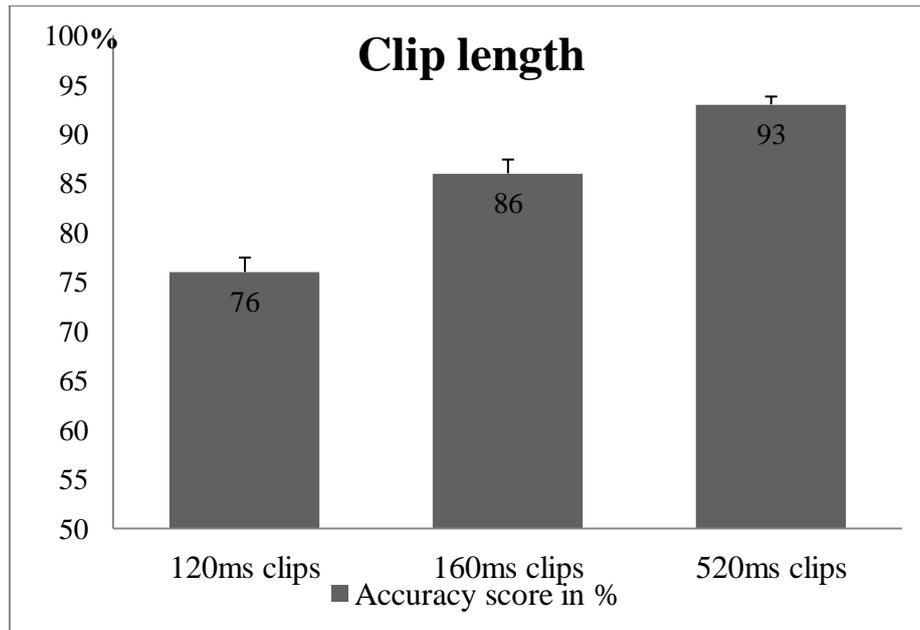


Figure 5: Means percentages of correct responses to the different clip lengths used across emotion, modality and hemifield.

The clips used in the experiment where audio-visual, only video or only audio. Modality of clip used to present the emotions where significant  $F(1,44)=152,56, p=.001$ .

The Bonferroni corrections showed that significant differences were found between all three modalities. Audio-visual clips having the highest percentage of correct responses ( $M=91\%, SD=1,3$ ) to audio-only having the lowest percentage ( $M=76\%, SD=1,34$ ) (see figure 6).

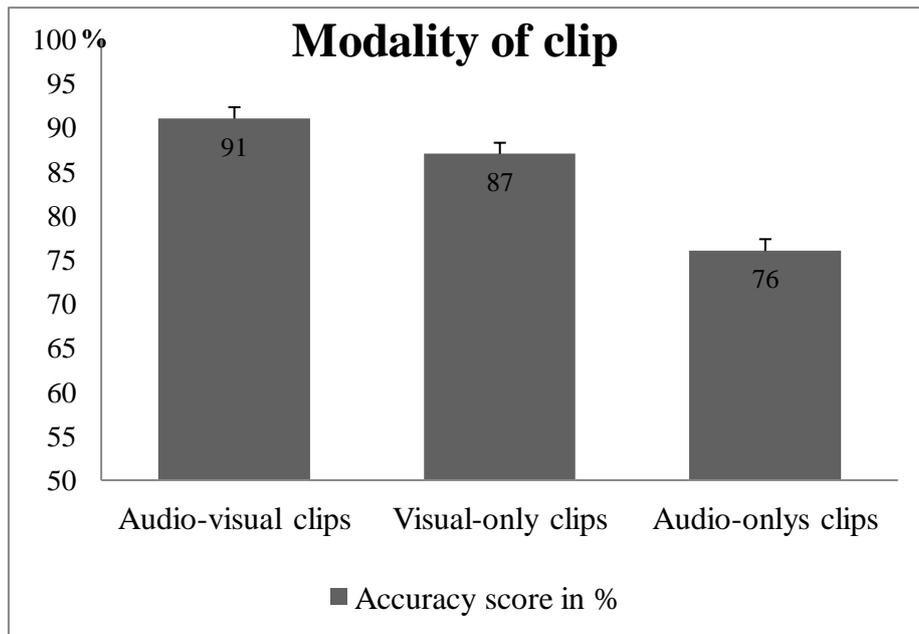


Figure 6: Means percentages of correct responses in different modalities across emotion, length and hemifield/ear.

### **Valence – and the right-hemisphere hypothesis**

The second valence ANOVA; 2 (gender: male and female; between subject factor) x 3(modality; audio-only, video-only, audio-visual) x 2(emotion: positive valence, negative valence) x 2 (length: 120ms, 160ms) x 2(direction: left hemifield/ear, right hemifield/ear) with alpha of .05 used. To test the valence hypothesis the emotions were computed into negative- and positive emotion. The percentages for the clips featuring anger and sadness were added and divided by two (anger clip + sadness clip)/2 to become a weighed negative valence emotion category. The positive valence emotion category consisted of the happiness percentages only. The central placement, where clips were shown to both hemifields/ears was excluded, leaving the left hemifield/ear and right hemifield/ear. No significant effect or interaction concerning hemifield/ear were found,  $p > .05$ .

Neutral and fear was excluded from the ANOVA used to test the valence hypothesis on the basis that neutral does not contain a positive or negative valence and therefore functioned as a control. Fear which is a negative valence was excluded many of the missing cases in the experiment were found in the responses to fear.

A third mixed repeated measures ANOVA was used to test the right-hemisphere hypothesis. Here all 5 emotions were present and non-categorized based on emotions positive or negative valence. The central placement, where clips were shown to both hemifields/ears was excluded, leaving the left hemifield/ear and right hemifield/ear. ANOVA three; 2 (gender: male and female; between subject factor) x 3(modality; audio-only, video-only, audio-visual) x 5(emotion: anger, fear, happiness, neutral, sadness.) x 2 (length: 120ms, 160ms) x 2(direction: left hemifield/ear, right hemifield/ear). An alpha of .05 was used for all of the ANOVA's.

No significant effects and interactions were found in the third ANOVA concerning left hemifield/ear advantage compared to the right hemifield/ear for the five emotional clips in the experiment,  $p > .05$ . There are therefore found no support for either the valence- or right-hemisphere hypothesis in this study.

In both the RH-  $F(2,46)=178,44$ ,  $p=.001$  and valence ANOVA  $F(1,40)=235,46$ ,  $p=.001$  a significant main effect of modality was found, showing a significant difference between all three modalities in the RH ANOVA and between audio-only and the other modalities used in the experiment in the valence ANOVA. The result was similar to the modality main effect in the control ANOVA where audio-only had the lowest percentage, then video-only and audio-video with the highest percentages of correct responses to emotion shown.

Emotion as a main effect was also significant in both ANOVA's,  $F(1,32)=11,95$ ,  $p=.002$  in the valence ANOVA showing a significant difference between negative- and positive valence with negative valence having the highest percentage of correct responses of the two. And  $F(4,128)=35,25$ ,  $p=.001$  in the RH ANOVA with the difference being found comparing neutral to the four other emotions, by having the highest percentage of correct responses.

Length was also significant as a main effect in the valence ANOVA  $F(1,32)=114,09$ ,  $p=.001$  and  $F(1,32)=238,44$ ,  $p=.001$  in the RH ANOVA where the difference was found between the two lengths, where 160ms had the highest percentage of correct responses, in both ANOVA's, compared to the 120ms length.

ANOVA two and three did reveal a significant interaction between emotion and modality of the clip shown, in ANOVA two  $F(2,51)=61,72$ ,  $p=.001$  and in the third ANOVA  $F(8,256)=73,21$ ,  $p=.001$ .

Bonferroni corrections were used in both ANOVA's to find the significant differences. In the second ANOVA where emotion was in categories depending of negative or positive valence, the Bonferroni found that the percentages between the two categories were significant in each modality. Negative valence had the highest percentage of correct responses in audio-only compared to positive valence. Positive valence had the highest percentages over negative valence in both the video-only and audio-visual modality. Positive valence clips had the highest and lowest score of the two categories with a range from 39-92% compared to negative valence that ranged from 63-85%.

Positive valence had the highest percentage of correct responses when the modality of the clip was video-only, in this case a male or female smiling. Audio-visual clips were 2% less accurate and audio-only had a low percentage of 39% correct responses when presented. The same was found for negative valence, the percentage of correct responses was highest in the audio-visual modality with 85% and video-only 84% and significant lower in audio-only with 63% correct responses to valence shown(see figure 7 and table 2).

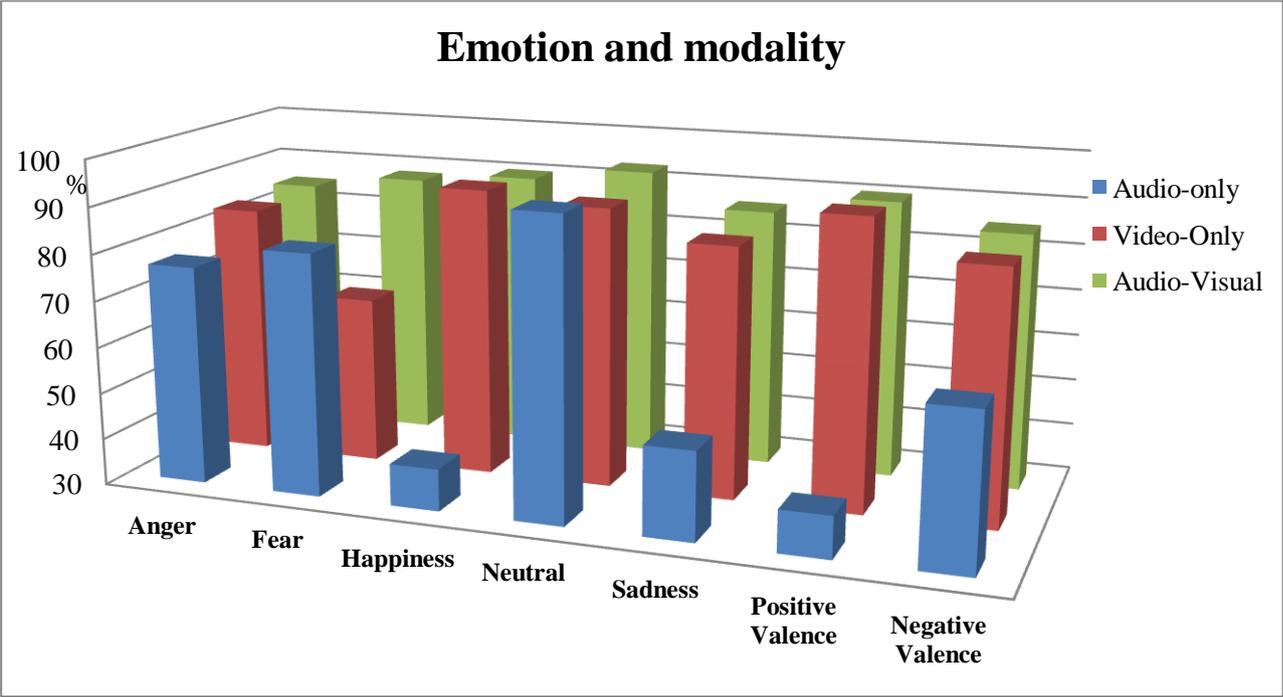


Figure 7: Means percentages of correct responses in the different emotion in different modalities across, length and hemifield/ear.

Table 2

Means percentages of correct responses and SD in the different emotion and modalities across length and hemifield/ear. Number found from ANOVA 2 and 3.

Emotion/ Valence	Audio- only	Video- Only	Audio- Visual	SD AO	SD VO	SD AV
Anger	77	84	85	3,13	2,75	2,88
Fear	82	66	88	3,13	2,09	2,27
Happiness	39	92	90	3,13	1,54	2,22
Neutral	94	90	93	0,92	1,86	1,22
Sadness	49	84	86	2,86	2,61	1,89
Positive Valence	39	92	90	3,12	1,54	2,22
Negative Valence	63	84	85	2,33	3,44	2,05

In the third ANOVA emotions were not categorized based on valence, and fear and neutral was included in the ANOVA. The interaction between emotion and modality was significant;  $F(8,256)=73,21, p=.001$ . The Bonferroni correction revealed that the percentage of correct responses to anger in the audio-only modality was significant different to the two other modalities by having the lowest percentage with 77% correct responses compared to the audio-video modality which had 85% (see figure 7 and table 2). Happiness showed the same Bonferroni difference with a low percentage of 39% in audio-only. Sadness was also significant in the audio-only modality with 49% correct responses compared to the other modalities. The Bonferroni found no significant difference between the modalities regarding the neutral emotion, having the highest percentage of 94% in audio-only, 93% in the audio-visual modality to 90% in video-only as the lowest. Fear revealed a significant difference in the video-only modality having the lowest percentage of correct responses in this modality with 66% correct responses.

When comparing emotion to emotion within each modality the audio-visual condition was not significant indicating that the percentages in the audio-visual modality were not significant different in the different emotions. In audio-only the Bonferroni found a difference in regards to neutral compared to the other emotions by having the highest percentages of correct responses in audio-only. In video-only fear was significant compared to the other emotions by having the lowest percentage of the emotions (see figure 7 and table 2).

Emotion and length was also found to be a significant interaction in the third RH-, but not the valence ANOVA,  $p > .05$ . Indicating no significant difference between the valence of emotion and length of the clip shown.

The RH ANOVA was significant regarding emotion and length  $F(4,128)=25,47$ ,  $p = .001$ . The Bonferroni found only a significant difference in the 120ms concerning neutral clips having a higher percentage of correct responses than the four other emotions. There was no difference in the 160ms duration to the different emotions.

When comparing the emotion and lengths all emotions excluding neutral had a significant higher percentage of correct responses in the 160ms condition compared to the shorter 120ms (See figure 8 and table 3).

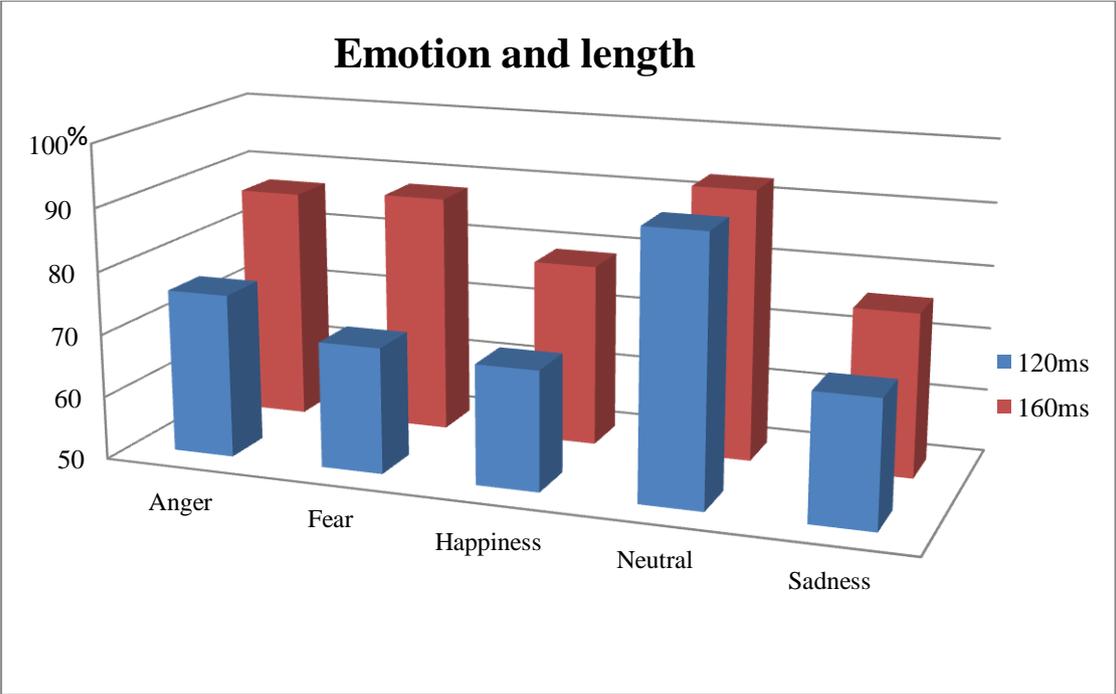


Figure 8: Interaction between emotion and length. Means percentages of correct responses and SD in the different emotion in different length, across modality and hemifield/ear.

Table 3

Interaction between emotion and length. Means percentages of correct responses and SD in different emotion and length across modality and hemifield/ear.

Emotion	120ms	160ms	SD 120ms	SD 160ms
Anger	76	87	2,85	2,47
Fear	70	88	2,34	2,28
Happiness	69	79	1,64	1,92
Neutral	92	93	1,20	1,08
Sadness	70	76	2,01	1,82

The interaction between modality and length of clip was significant in both ANOVA's. In the valence ANOVA  $F(2,52)=17,09$ ,  $p=.001$  and  $F(2,54)=10,36$ ,  $p=.001$  in the RH ANOVA. The Bonferroni correction in the valence ANOVA showed that the significant differences were found between audio-only and the other modalities in both lengths having the lowest percentages of correct responses. The same significant difference was found in the RH ANOVA in the 160ms length for audio-only. In the 120ms length there was found a difference between all modalities with the highest percentage of correct response found in the audio-visual modality, and the lowest in audio-only (see table 4)

Table 4

Interaction modality and length, in RH- and valence ANOVA. Means percentages of correct responses and SD in modalities and different length, across emotion and hemifield/ear. AV=Audio-visual, AO=Audio-only and VO=Video-only.

Modality	120ms	160ms	SD 120ms	SD 160ms
AV (RH ANOVA)	85	90	2,16	1,92
AO (RH ANOVA)	43	59	2,57	2,53
VO (RH ANOVA)	86	90	2,02	1,88
AV (Valence ANOVA)	86	90	1,94	1,59
AO (Valence ANOVA)	63	74	1,71	1,77
VO (Valence ANOVA)	78	89	1,54	1,74

Modality, emotion and length were also found to be a significant interaction in both the second and third ANOVA. In the valence ANOVA  $F(2,64)=13,97, p=.001$ .

Comparing valence of emotion to valence of emotion (positive valence compared to negative valence, negative valence compared to positive valence) in the modality, emotion and length interaction the Bonferroni revealed that in both the 120ms and 160ms length participants found it easier to recognize negative valence when presented as audio-only compared to positive valence. In video-only the opposite result is found, were positive valence had a higher percentage of correct responses compared to negative valence in both lengths. A significant difference was found in the 120ms clip between the two categorizations, negative valence having 82% correct responses and positive valence 88% in the audio-visual modality. In the audio-visual modality the 160ms long clips show no significant difference between positive and negative valence, negative valence with 90% and positive valence 91% (see figure 9 and table 5).

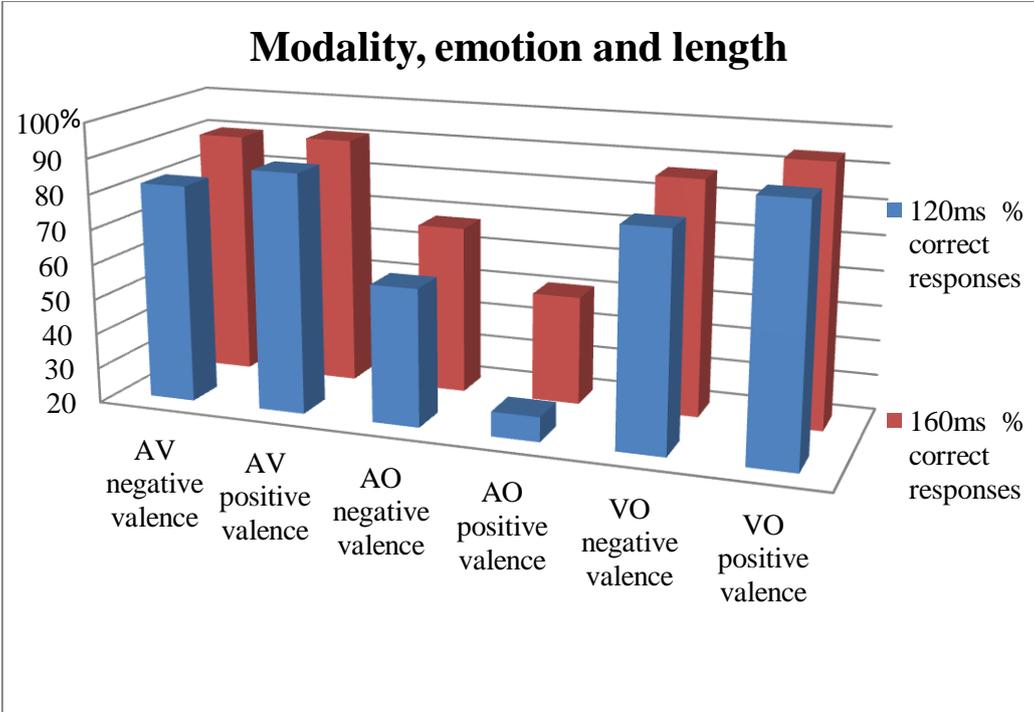


Figure 9: Interaction modality, emotion and length. Means percentages of correct responses and SD in the different modalities, emotion and lengths across hemifield/ear. AV=Audio-visual, AO=Audio-only and VO=Video-only.

Table 5

Interaction modality, emotion and length. Means percentages of correct responses and SD. AV=audio-visual, AO=Audio-only and VO=Video-only.

Valence of emotion	120ms	160ms	SD 120	SD 160
AV negative valence	82	90	2,50	1,88
AV positive valence	88	91	2,48	2,36
AO negative valence	59	68	2,80	2,21
AO positive valence	27	51	3,27	3,76
VO negative valence	81	87	2,72	2,42
VO positive valence	91	94	1,81	1,64

Comparing modality to modality in the modality, emotion and length interaction the Bonferroni revealed a significant difference in audio-only compared to the video-only and audio-visual modality, having the lowest percentages of correct answers across valence and length. Positive valence had the lowest scores of the two categories with 27% in 120ms clip and 51% in 160ms in audio-only compared to negative valence 59% in the 120ms clip and 68% in the 160ms (see figure 9 and table 5), which again is significant lower than the percentages obtain in the other modalities of clips.

When comparing length to length in the modality, emotion and length interaction the Bonferroni revealed a significant difference in between the two lengths in the audio-visual modality portraying negative valence having a higher percentage of correct answers when the clip with the longest duration (160ms) was presented compared to the shorter length (120ms). This was also evident for negative valence and positive valence in audio-only. In video-only a significant difference was found between 120ms and 160ms in negative valence. Positive valence showed the same tendency in the means, with lower percentage in the shorter length, but could not show a significant difference in Bonferroni with,  $p=0.52$  (see figure 9 and table 5).

In the RH ANOVA modality, emotion and length was also significant,  $F(5,153)=19,63, p=.001$ , Bonferroni correction revealed that when comparing emotion to emotion in the modality, emotion and length interaction 120ms happiness were significant

less recognized in the audio-only modality ( $M=27\%$ ,  $SD=3,27$ ) compared to the four other emotions (anger, fear, neutral and sadness) that had a range from 47% for sadness to neutral with the highest with 94% in the clips with a duration of 120ms.

When comparing modality to modality in the modality, emotion and length interaction in the RH ANOVA the Bonferroni revealed a significant difference in 120ms video-only clip of fear ( $M=47\%$ ,  $SD=1,84$ ) compared to the audio-visual ( $M=86\%$ ,  $SD=2,86$ ) and audio-only condition ( $M=76\%$ ,  $SD=4,46$ ) with having the lowest percentage of correct responses (see table 6).

Table 6

Interaction modality, emotion and length in RH ANOVA. Means percentages of correct responses and SD. AV=Audio-visual, AO=Audio-only and VO=Video-only.

Emotion	120ms	160ms	SD 120ms	SD 160ms
AV anger	79	90	3,61	2,46
AV fear	86	90	2,86	1,94
AV happiness	88	91	2,48	2,26
AV neutral	94	93	1,47	1,40
AV sadness	84	88	2,20	2,18
AO anger	71	84	3,70	2,88
AO fear	76	89	4,46	3,10
AO happiness	27	51	3,27	3,76
AO neutral	94	95	1,40	0,92
AO sadness	47	52	3,29	3,11
VO anger	82	87	2,98	2,82
VO fear	47	85	1,85	3,03
VO happiness	91	94	1,81	1,64
VO neutral	90	91	1,98	2,02
VO sadness	80	88	3,17	2,42

Happiness presented in the 120ms and 160ms lengths a significant lower percentage of correct answer when presented as audio-only clip than in audio-visual or video-only clip. Happiness having 88% correct responses in the audio-visual modality when presented with a clip of 120ms duration, and 91% when the clip had a duration of 160ms a significant difference to the 27% and 51% in audio-only modality. In sadness audio-only had a significant lower percentage of correct responses in the two lengths 47% in 120ms and 52%

in clips with 160ms duration compared to video-only 120 ms 80% and 160ms clips 88% and in the audio-visual modality with 84% in 120ms clips and 88% in the 160ms (see table 5).

When comparing length to length in the interaction between modality, emotion and length only anger and sadness revealed a significant difference between the 120ms and 160ms duration of clip in the audio-visual modality, both with significant higher percentage of correct responses when the clips with the longest duration was presented. In the audio-only modality the same increase of correct responses the longer length of clip was found for fear, happiness and for anger, and fear and sadness in video-only.

## **Gender**

Gender as a between subject factor were non significant in all three ANOVA's ,  $p > .05$ .

The interaction between gender of the participant and modality of the clips (audio-visual, audio-only or video-only) revealed a significant difference in the third RH ANOVA,  $F(2,64)=6,71, p=.002$ , and in the second ANOVA  $F(2,64)=4,90, p=.010$ .

Bonferroni corrections in the valence and right-hemisphere ANOVA revealed that the significant differences were between in the modality itself within each gender and not between the genders. The valence ANOVA Bonferroni revealed an audio-only difference with a significant lower percentage of correct responses in audio-only for both genders. In the RH ANOVA a significant difference was found in all modalities for each gender showing that the percentages in all modalities were significant different compared to each other.

The interaction between gender and modality was significant in the RH ANOVA, ,  $p=.010$ , and the means between the two genders in the audio-only modality were notable different, where the females had a lower percentage of correct responses ( $M=66\%$  ,  $SD=2,32$ ) than the male sample ( $M=71\%$  ,  $SD=2,23$ ), the Bonferroni corrections revealed that the difference was not significant,  $p=.088$ .

## **Discussion**

The present study investigated gender differences in cerebral lateralization in perception of emotion by presenting video- and audio-visual clips of emotional facial expressions to the left, center and right hemifields, and by presenting prosody audio clips dichotic to only one or both ears as behavioral tests of laterality.

The accuracy responses to the emotion shown did neither give support to the valence hypothesis as in (e.g. Erhan, et al., 1998; Silberman & Weingartner, 1986) or the right hemisphere advantage hypothesis (e.g. Lavadas, Umiltà & Ricci-Britt, 1980; Heilman, & Bowers, 1990; Tucker, 1981; Rose, 2011) since hemifield/ear was non significant as main effect and in all interactions.

This is in line with previous studies that have not found evidence to support the right-hemisphere hypothesis in lateralization of emotions (Caltagirone, Ekman, Friesen, Gainotti, Mammucari, Pizzamiglio, & Zoccolotti, 1989; Mammucari, Caltagirone, Ekman, Friesen, Gainotti, Pizzamiglio, & Zoccolotti, 1998) or support for the valence hypothesis as in Borod (1992) and studies that found no support for either approach when comparing both as in the meta-analysis consisting of 65 neuroimaging studies in Wager et al., (2003).

## **Gender**

The lack of gender as a significant between subject factor is in contrast to Collignin et al. (2010) findings where a gender difference was found with the use of the same MAV audio clips (Belin, Fillion-Bilodeau, & Gosselin, 2008) and MPAFC video clips (Simon, Craig, Gosselin, Belin, & Rainville, 2008). The non significant gender differences is also in contrary to Wager et al. (2003)'s meta-analysis showing that men exhibit more lateralized activation in the brain than females, since this study found no effect of interaction concerning a hemispheric lateralization advantage. This behavioral experiment can therefore not show support for a gender differences as previous studies have revealed through anatomical and neurobiological studies (Good, Johnsrude, Ashburner, Henson, Friston, & Frackowiak, 2001; Schneider, Habel, Kessler, Salloum, & Posse, 2000; Hines, Chiu, McAdams, & Bentler, 1992; Russo, Persegani, Papersch, Nicoloni, & Trimarchi, 2000; Wager et al., 2003; Good et al., 2003).

Modality of clip and gender resulted in a significant interaction, but with the use of Bonferroni this interaction revealed only a ( $p=.088$ ) value showing in the means for the female sample had a lower percentage of correct responses in the audio-only condition compared to the male sample, but not significantly.

## **Audio-visual emotion perception**

The MAV audio clips and MPAFC video clips have been used by other researchers and have a solid empirical background (Collignin et al., 2010; Simon, Craig, Gosselin, Belin, & Rainville, 2008; Belin, Fillion-Bilodeau, & Gosselin, 2008). Where Collignin et al. (2010) used congruent and incongruent clips of disgust and fear, presented as audio, video or as audio-visual clips, and Simon, Craig, Gosselin, Belin, & Rainville (2008) centered on the emotion of pain, the current study used 5 emotions: happiness, anger, sadness, neutral and fear, and categories of positive- (happiness) or negative valence (anger, sadness) shown congruently with audio-only, video-only or audio-visual clips. Contrary to Collignin et al. (2010) that found that the females outperformed the male sample in all modalities this study did not find support for this claim.

In regards to the use of audio-visual also called bimodal clips the experiment gives support to previous research as (Collignin et al.,2008;2010; Kreifelts et al., 2007 ) with audio-visual clips resulting in the highest percentage of correct responses of the three modalities in all the ANOVA's as a main effect. In the RH and control ANOVA the three modalities resulted in significant different percentages with the highest percentages of correct responses in the audio-visual modality then in video-only and lowest in audio-only. In the valence ANOVA audio-only revealed the only significant difference in modalities with having the lowest percentages of modalities similar to the other ANOVA's.

Clips in this study that contained both audio and video emotion content were recognized more than when containing only emotion through prosody, or a video of a facial expression in support of Collignin et al. (2010). In this study prosody differentiated itself from the other modalities by having a significantly lower percentage of correct responses indicating that it was easier for the sample in this study to recognize emotion and valence when presented as a facial expression or facial expression and prosody. Both negative- and positive valence emotion showed this effect and all emotion with the exception of fear and neutral emotion.

Neutral did not have a significant lower percentage in audio-only compared to the two other modalities. In neutral audio-only and the audio-visual modality was only separated by 1%, and only a 4% difference from audio-only to the video-only modality. Fear had a significant lower percentages of correct responses with 66% in the video-only conditions compared to 82% in audio-only and 88% in the audio-visual modality in contrast to findings in Collignin et al.(2010). Findings like these show that neutral expression and prosody is recognized in a

different way from clips containing emotion in this study. Fear on the other hand does not show the same pattern in the percentages within the modalities as the other emotions by having the lowest percentage of correct responses in the video-only modality. Still modality as a main effect and as an interaction show support for the use of audio-visual modality in emotion perception and a high ecological validity (Collignin et al., 2008;2010; Kreifelts, Ethofer, Grodd, Erb & Wildgruber, 2007;de Gelder & Vroomen, 2000a;2000b; Simon, Craig, Gosselin, Belin, & Rainville, 2008). Still there is need for more research and the use of different research tools as in Pourtois et al., (2005) with PET- or even back to analyzing WADA data (Goldstein, 1939 found in Tan, 2013) or the use of other approaches (Alves, Fukusima, & Aznar-Casanova, 2008; Demaree et al., 2005; Gadea et al., 2011) to get a deeper understanding and looking beyond the main effects to understand better how perception and recognition of emotion is related to different modalities and for explanation for results that do not fit, like fear in studies like these.

An advantage in this study was the use of different lengths of clips and the use of different emotions not only fear and disgust as in Collignin et al. (2010). The interaction between the clips modality and length was significant in both the valence- and RH ANOVA. The valence ANOVA found that audio-only had the lowest percentage of correct responses compared to the other modalities in both lengths. In the RH ANOVA all modalities when presented as a 120ms clip had significant different percentages, but the highest in the audio-visual modality and the lowest in audio-only. When the clips were 160ms the significant difference between video-only and audio-visual disappeared leaving only the difference in audio-only having the lowest percentage of correct responses of the modalities.

This relates to the interaction between emotion, modality and length of clip presented that was significant, in both ANOVA's. In the valence ANOVA a significant difference was found between the 120ms clips in the audio-visual modality, where positive valence had a significant higher percentage of correct responses compared to negative valence. Interestingly in the 160ms clip in the same modality there is no difference and the accuracy score in positive- and negative valence, going from 82% and 88% in 120ms to 90% and 91% in the 160ms length. This result has similarities with Behne et al.(2008) indicating a change within the audio-visual modality in connection to recognition, while in Behne et al. (2008) the change was seen in McGurk clips changing from visual- to audio recognition when clip length reached 160-200ms, in this study the change was in favor for positive valence when the presentation was short(120ms), and when presented as 160ms clips both positive and negative valence result in

the same level of correct percentages with negative valence catching up, indicating that positive valence is earlier recognized in the shorter length, but when the length of a clip reaches between 120-160ms the negative valence emotion is recognized as accurate as positive valence. This was only seen in the audio-visual modality.

The RH ANOVA in the two lengths in audio-visual modality showed in Bonferroni that anger and sadness had significantly higher percentage of correct responses in the longest length compared to in 120ms. Happiness, fear and neutral showed no significant difference between the percentages in the two lengths supporting the same pattern as in the valence ANOVA, showing a different pattern compared to anger and sadness.

Although this does not tell us anything about which hemisphere processed what emotion or valence category, it does indicate a difference in how emotion is most accurately processed when being presented as a brief clip. Positive valence/happiness, fear and neutral in the audio-visual condition being recognized with similar accuracy in the two lengths compared to anger and sadness that had an increase of correct recognitions the longer the length of the audio-visual clip with all emotion ending up with near identical accuracy rates.

Even with MAV and MPAFC clips having a solid background, and showing a high percentage of correct responses in this study, both the audio- and video clips have the potential for improvement. In this experiment and in (Collignin et al., 2010) the MAV audio was matched with the MPAFC video to create an audio-visual modality. In both cases this combination of prosody and facial expression resulted in a higher amount of correct recognitions when compared to clips with only video or audio (Collignin et al., 2008;2010). Despite the audio-visual advantage in this study the audio and video did not match, the sound of laughter did not match the smile, the participants heard an angry scream, but the video showed an angry face with the teeth clenched. It is therefore natural to believe that having both audio-visual clips from one speaker would yield even higher accuracy score and give better ecological validity than the clips used in this experiment.

### **Binaural hearing and emotion**

The use of dichotic listening in this experiment did not result in a predicted main effect or interaction in audio-only or audio-visual clips in connection to direction/to which ear the prosody was presented to, contrary to other studies (Erhan, et al., 1998; Gadea et al., 2011; Haggard & Parkinson, 1971; Hatta & Ayetani, 1985; Henry, 1983; Safer & Leventhal,1977).

Modality of clip and gender resulted in a significant interaction, but with the use of Bonferroni this interaction revealed only a ( $p=.088$ ) tendency between the two genders. The possible tendency showed that the females had a lower percentage of correct responses in the audio-only condition compared to the male sample. Even if this is not significant the difference in means between the genders is interesting, possibly indicating that females are less accurate in recognizing emotion prosody when presented through dichotic listening to only one ear at the time compared to the males when presented with short prosody clips. It is hard to speculate the reason since the interaction between genders, modality and ear presented with the emotion (direction) was not significant.

### **Vision and emotion**

Video-only and the audio-visual clips were shown to both hemifields/ears by presenting them in the middle of the screen/both ear revealing a higher percentage of correct responses than when presented to in the left- or right hemifields/ear, supporting the assumption that people answer more accurately when information is received by both hemispheres (e.g. Collignon et al., 2010). The use of the 520ms length of clips as a control condition was on the same basis, that the more information given easier the recognition of emotion would become for the participant. An effect the experiment gave support to, but even with this advantage for the center location on the screen the use of divided visual field technique failed to produce any significant differences when emotion was present to the left or right of the fixation point on the screen. Is this because a hemispheric lateralization difference for emotion does not exist or is hard to uncover with the use of divided visual field technique or could it be a possible other reason for this result?

The high accuracy scores in the left and right hemifield (see figure 4) compared to Thayer & Johnson (2000) scores collected from both hemifield may be an indication that the divided visual technique failed and both eyes had the possibility to perceive the clip even with the use of short clip length and taking in account saccades as mentioned in (Jordan & Thomas, 2007; Rayner, 1998; Bourne, 2006).

The use of the right hand to respond is also a factor (Rose, 2011). Information in the left hemifield are received and first processed by the right hemisphere before the information is shared through the corpus callosum to the left hemisphere that again is connected to control of the right hand (Bourne, 2006; Jordan & Thomas, 2007; Rayner, 1998). Önal-Hartmann, Paulio, Ockleburg, & Güntütkün (2012) used divided visual technique combined with the use

of flexion and extensor response, and changed which hand participants responded with during the experiment. And Collignin et al. (2010) had participants use both index fingers to respond.

Early in the project this “right-hand factor” was discussed, at that time the plan was to use the reaction times from the experiment along side with accuracy percentages. With that in mind the use of the left hand when the participants were predominantly right-handed would affect the reaction times and the decision was made to use only the right hand. It would therefore be interesting for future research in similar behavioral experiment to use both hands, use both index fingers or changing the hand used during the experiment.

The production of the different clips can also be a limitation to this study and others. As told was there no possibility to create in perfectly synchronized audio-visual clips from the MAV and MPAFC material. This experiment used clips with lengths down to 120ms. When using so short audio- and video clips the onset of emotion cues are of utter importance. The clips were edited with precision in mind, removing frames before onset that was without emotional cues (see table 1), but this process was not a simple one since some of the clips started with an open mouth making it hard to determine the start of onset for the emotions. The MAV clips for the female speaker portraying fear had also its limitations. The MAV audio clip was only 440 ms long. Since the 520ms long clips was the full clips it seemed right to keep fear even with 440ms since this was the full vocalization of fear. This clip as a whole had a duration of 520ms as the other clips, with video for 520ms and audio for 440ms. The probability that this being a important factor for the results are low, especially since the 520ms of this clips was used as a control against the two other clip lengths, but at the same time a factor worth noting.

The length of the clip presented was significant resulting in a higher percentage of correct answers the longer the clip was revealing a increase from 76 % when presented as 120ms clips, 86% in 160ms and 93% in the longest 520ms full clips (see figure 5), indicating that more information give more accurate recognition to which emotion shown across all modalities. This was also found in the RH- and valence ANOVA with 120 and 160ms, in length as main effect, and in most notably in emotion with positive valence/happiness in audio-visual modality in the interaction between modality, emotion and length.

Collignin et al. (2010) used only one clip length with a duration of 500ms and Rose (2010) 250ms , 150ms in Lavadas, Umiltà & Ricci-Britt ( 1980) and 180ms in Önal-Hartmann, Pauli,

Ocklenburg & Güntürkün, (2012). The use of only one length makes it difficult to say anything about a possible increase or decrease in accuracy in connection to clip length. Behne et al. (2008) used multiple lengths of clips and revealed how responses to McGurk audio-visual clips changed depending on the length of the clip, which is similar to what this study found support for. There was found a higher accuracy for positive- than negative valence in the 120ms length in the audio-visual modality. When presented as 160ms the difference disappeared indicating that positive valence was earlier recognized, but when the length reached 160ms negative valence was recognized as accurately as positive valence. An indication that positive valence may have an early accuracy advantage compared to negative valence that disappears in between the 120ms-160ms range. The same early advantage can be found in anger and sadness with a significant increase in the longest clip length, while happiness, fear and neutral have no significant differences in accuracy score in the different lengths in the audio-visual modality.

It would therefore be interesting to see more than one clip length being used in future studies when exploring emotion perception to see if this early advantage that dissipates before 160ms in positive valence/happiness is replicable in other conditions, modalities and with other stimuli set.

A technical limitation of this study concerns missing values and the difficulty of determining if the missing responses were missing at random or not. The video-only clip of fear with a duration of 120ms had 16 of the 20 missing responses found in the whole dataset. Which may help to explain some of the findings in regards to fear as having the lowest accuracy when presented as video-only compared to the two other modalities (table 2 & 6). This stands out since no other emotion or valence in this study show this effect except neutral. Little's MCAR was used to check if the response was missing not at random. Little's MCAR approved for EM analysis and the missing values was calculated based on the other responses to the clip from the data of other participants. The missing values could come from a technical glitch in SuperLab that was used to run the experiment. They could be from the editing of the clips, taking in account the difficulty of determining onset of emotion cues and the short presentation.

A limitation of the experiment that has to be considered especially because of the non significant finding regarding the hemifield/ear presentation is that the divided visual field technique as flawed in this study. The visual field technique is vulnerable; a participant's

movement of body or head would change the premise that the clip was received by only one hemisphere. When the clips were shown in the center location of the screen, to both hemispheres, center had the highest accuracy rate, and the lengths of the clips should be too short for both eyes to receive the clip, one can still not eliminate that the divided visual field technique did not work optimally. At the same time a technical flaw in the setup for the divided visual technique does not explain why dichotic listening did not yield results as in (Haggard & Parkinson, 1971; Hatta & Ayetani, 1985; Henry, 1983; Safer & Leventhal, 1977; Erhan et al., 1998). The audio was presented directly to one ear while the other ear was exposed to pink noise and would not be affected by head or body movements. A possible reason could be the brief duration of the clips. Length was significant as a main effect and in all emotions except neutral having a higher percentage of correct recognitions in the longer the clip. The use of longer audio clips may have resulted in a higher percentage of correct responses in this experiment. At the same time the use of longer audio, but not video would make the two modalities less comparable and in the control ANOVA the 520ms audio and presentation to both ear did not yield a significant difference one then would expect in interaction between the modality (audio-only) and direction (audio to both ears), and would decrease the ecological validity even more.

### **Emotion perception**

The use of neutral clips showed a significant higher percentage of correct responses compared to the four emotions indicating that a video clip of a neutral face or a non-prosodic sound is processed in another way than a face or voice expressing an emotion (Simon, Craig, Gosselin, Belin, & Rainville, 2008). When comparing valence of emotion, negative valence have the highest mean percentages of correct response in emotion as main effect, and with the highest percentage of correct recognition in audio-only compared to positive valence in the interaction between emotion and modality. Positive valence had higher accuracy scores in both video-only and audio-visual modality. Positive valence clips had the highest and lowest score of the two categorization with a score range from 39-92%, compared to negative valence that ranged from 63-85% (see figure 7 and table 2). In the interaction between modality, emotion and length a difference was uncovered between positive- and negative valence of emotion in the length in audio-visual clips. Positive valence/happiness showed no significant difference in the two lengths with the percentage of negative valence increased with the increase of length. This was also evident in the RH ANOVA in the audio-visual modality with happiness, fear and neutral having a similar percentage of accuracy score in

both lengths while anger and sadness had an increased percentage in the longest length compared to the short length.

In regards to the percent correct responses to the clips shown, the experiment yielded 94% correct recognitions when shown neutral, and 87% when anger was presented and impressive increase compared to Thayer & Johnsen (2000) where the female sample recognized the emotions 45,7% of the time and the male sample 39,7%, compared to this study which had a range from 78%-94% (see figure 3) giving support for the use of audio-visual clips or the use of video and not a still picture as Thayer & Johnsen (2000).

Future research on gender and cerebral lateralization on perception of emotion would benefit from the use of audio-visual clips, especially even more ecological synchronized audio and video portraying emotion. If emotions are categorized into positive and negative valence as in this experiment, the categorization would benefit from having more emotions in each category. Especially in positive valence that in this experiment only contained happiness. The use of more emotions in each category would increase the reliability and validity of positive valence as a category. For similar behavioral experiment setup like in this study with the use of divided visual technique and dichotic listening I would recommend the use of a chin rest for eliminating the possibility for both eyes/hemisphere to perceive the clip resulting from a participants movements, or the use of an eye-tracker to ensure the experiment setup is functioning as planned.

The ability to recognize the emotions of people around us is important. This study used brief (120ms, 160ms, 520ms) audio prosody, visual - and audio-visual clips of congruent emotion (anger, fear, happiness, neutral, sadness, positive- and negative valence), divided field technique and dichotic listening to investigate a possible gender difference in cerebral lateralization of emotion. No support for the right-hemisphere - or the valence hypothesis, or gender as a between subject factor was found. The female sample revealed a difference in mean percentages with lower accuracy score in audio-only modality compared to the male sample, but this difference was non significant in after Bonferroni review.

Clips containing both audio and video had the highest accuracy score of all modalities as in (Collignin et al.2008; 2010). Audio-only prosody clip had significant lower accuracy score compared to video-only and audio-visual clips in modality main effect, and in the interaction between modality and length. There was found a higher accuracy for positive- than negative valence in the 120ms length in the audio-visual modality. When presented as 160ms the

difference disappeared, indicating that positive valence was earlier recognized, but when the length reached 160ms negative valence was recognized as accurately as positive valence. An indication that positive valence may have an early accuracy advantage compared to negative valence, and that this dissipates in between the 120ms-160ms range. The same early advantage can be found in anger and sadness, with a significant increase in the longest clip length, while happiness, fear and neutral have no significant differences in accuracy score in the different lengths in the audio-visual modality.

The use of more than one clip length in future studies when exploring emotion perception would be interesting in regards to the early advantage that dissipates before 160ms in positive valence/happiness in audio-visual clips is replicable in other conditions, modalities and with other stimuli set.

Differences in perception of emotion can result in new and more information about the gender differences in mood anxiety disorders and depressive disorders where females are more at risk. Or help explain why there is a higher frequency of males that suffers from disorders as autism. This study cannot explain previously found gender differences in emotion perception or the cerebral lateralization of emotion, but from this study it is evident that people recognize emotion most accurately by the sight and prosody of an emotion, when the audio and the visual are combined than when only one modality is present.

## References

- Alves, N.T., Fukusima, S.S. & Aznar-Casanova, A.J.(2008).Models of brain asymmetry in emotional processing.*Psychology & Neuroscience, 1 (1)*, 63-66.  
DOI:10.3922/j.psns.2008.1.010
- Behne, D., Wang, Y., Belsby, S., Kaasa, S., Simonsen, L., & Back, K. (2008). Visual field advantage in the perception of audiovisual speech segments. International Conference on Auditory-Visual Speech Processing, Tangalooma, Australia.
- Belin, P., Fillion-Bilodeau, S. & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect burst for research on auditory affective processing, *Behavior Research Methods, 40 (2)*, 531-539. DOI: 10.3758/BRM.40.2.531
- Boersma, P. & Weenink, D.(1995). *Praat*. Downloaded 05.06.2012,  
<http://www.fon.hum.uva.nl/praat/>
- Borod, J.C.(1992). Interhemispheric and Intrahemispheric Control of Emotion: A Focus on Unilateral Brain Damage. *Journal of Consulting and Clinical Psychology, 60 (3)*, 339-348. doi: [10.1037/0022-006X.60.3.339](https://doi.org/10.1037/0022-006X.60.3.339)
- Bourne, V.J. (2006). The divided visual field paradigm: Methodological considerations. *Laterality 11 (4)*, 373-393.
- Caltagirone, C., Ekman, P., Friesen, W., Gainotti, G., Mammucari, A., Pizzamiglio, L., & Zoccolotti, P.(1989). Posed emotional expression in unilateral brain damaged patients,*Cortex, 25(4)*,653-663.

Carver, C.S. & Harrison-Jones, E.(2009). Anger is an approach-related affect: Evidence and implications.*Psychological Bulletin, 135*(2),183-204.

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F. & Lassonde, M.(2010). Women process multisensory emotion expressions more efficiently than men, *Neuropsychologia, 48*, 220-225.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, S., Lassode, M. & Lepore, F.(2008). Audio-visual integration of emotion expression.*Brain Research 1242*,126-135.

Darwin, C.(1872). *The expression of the emotion in man and animals*. London: Albemarle.

de Gelder, B. & Vroomen, J.(2000a). The perception of emotions by the ear and by the eye.*Cognition and emotion, 14*(3),289-311.

de Gelder, B. & Vroomen, J.(2000b). Bimodal emotion perception: integration across separate modalities, cross-modal perceptual grouping or perception of multimodal events?*Cognition and emotion, 14*(3),321-324.

Demaree, H.A., Everhart, D.E., Yongstrom, E.A., & Harrison, D.W. (2005). Brain lateralization of emotional processing: Historical roots and a future incorporating “dominance”. *Behavioral and Cognitive Neuroscience Reviews, 4* (1), 3-20. DOI: 10.1177/1534582305276837

Ekman, P. (1992). An argument for basic emotions. *Cognition and emotion, 6*, 169-200.

- Erhan, H., Borod, J.C., Tenke, C.E., & Bruder, G.E. (1998). Identification of Emotion in a Dichotic Listening Task: Event Related Brain Potential and Behavioral Findings. *Brain and Cognition*, 37, 286-307.
- Frijda, N. (1989). *The emotions*. Cambridge:CUP
- Gadea, M., Espert, R., Salvador, A. & Martí-Bonmatí, L. (2011). The sad, the angry, and the asymmetrical brain: Dichotic listening studies of negative affect and depression. *Brain and cognition*, 76, 294-299.
- Gohier, B., Senior, C., Brittain, P.J., Lounes, N., Law, V., Phillips, M.L. & Surguladze, S.A. (in press). Gender differences in the sensitivity to negative clips: Cross-modal affective priming study. *European Psychiatry*, DOI:10.1016/j.eurpsy.2011.06.007
- Good, C.D., Johnsrude, I., Ashburner, J., Henson, R.N., Friston, K.J., & Frackowiak, R.S. (2001). Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage*, 14(3), 685-700.
- Haggard, M.P. & Parkinson, A.M. (1971). Stimulus and task factors as determinants of ear advantages, *Quarterly Journal of Experimental Psychology*, 23 (2), 168-177. DOI: <http://dx.doi.org/10.1080/14640747108400237>
- Hatta, T., & Ayetani, N. (1985). Ear differences in evaluating emotional tones of unknown speech, *Acta Psychologica* 60, 73-82.
- Heilman, K.M., & Bowers, D. (1990). Neuropsychological studies of emotional changes induced by right and left hemispheric lesions. In Stein, N., Leventhal, B., & Trabasso,

- T. (Eds.), *Psychological and Biological Approaches to Emotion* (pp. 97 -113). Hillsdale: N.J.
- Henry, R.G.I.(1983). Monaural studies eliciting an hemispheric asymmetry. *Perceptual and Motor Skills, 56*, 915-918.
- Hines, M., Chiu, L., McAdams, L.A., Bentler, P.M., & Lipcamon, J. (1992). Cognition and the corpus callosum: verbal fluency, visuospatial ability, and language lateralization related to midsagittal surface areas of callosal subregions. *Behavioral Neuroscientific, 106 (1)*, 3-14.
- Hugdahl, K.(2000). Lateralization of cognitive processes in the brain. *Acta psychological 105*, 211-235.
- Jordan, T.R., & Thomas, S.M.(2007). Hemiface contributions to hemispheric dominance in visual speech perception. *Neuropsychology, 21(6)*, 721-731.
- Kimura, D.(1987). Are men's and women's brains really different?. *Canadian Psychology/Psychologic Canadienne, 28(2)*, 133-147.
- Kimura, D.(1961). Cerebral dominance and the perception of verbal clips. *Canadian Psychology/Psychologic Canadienne, 15(3)*, 166-171.
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage, 37*, 1445-1456.
- Kring, A.M. & Gordon, A.H.(1998). Sex differences in emotion: expression, experience, and physiology. *J. Personal. Soc. Psychol, 74(3)*, 686-703.

- Lavadas, E., Umiltà, C. & Ricci-Britt, P.E.(1980). Evidence for sex differences in right-hemisphere dominance for emotions.*Neuropsychologia*, 18,361-366.
- Levine, S.C., & Levy, J.(1986), Perceptual asymmetry for chimeric faces across life span.*Brian Cogn.*, 5(3), 291-306.
- Mammucari, A., Caltagirone, C., Ekman, P., Friesen, W., Gainotti, G., Pizzamiglio, L., & Zoccolotti, P.(1998). Spontaneous facial expression of emotions in brain-damaged patients.*Cortex*, 24(4),521-533.
- Önal-Hartmann, C., Pauli, P., Ocklenburg, S. & Güntürkün, O.(2012). The motor side of emotions: investigating the relationship between hemispheres, motor reactions and emotional clips, *Psychological Research*, 76,311-316.DOI: 10.1007/s00426-011-0337-4
- Pourtois, G., de Gelder, B., Bol, A. & Crommelinck, M.(2005). Perception of facial expressions and voices and of their combination in the human brain.*Cortex*, 41, 49-59.
- Plutchik, R.(1980). *Wheel of emotion*.Downloaded 08.08.2012,  
<http://www.fractal.org/Bewustzijns-Besturings-Model/Nature-of-emotions.htm>
- Rayner, K.(1998). Eye movements in reading and information processing: 20 years of research.*Psychological bulletin*, 124(3), 372-422.
- Rose, K.E.(2011).*Lateralization of emotion, reaction time, and skin conductance responsiveness*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI Number: 3465547)

- Russo, P., Persegani, C., Papersch, L.L., Nicolini, M., & Trimarchi, M. (2000). Sex differences in hemisphere preference as assessed by paper-and-pencil test. *Intern. J. Neuroscience* 100 (1-4), 29-37.
- Safer, M.A., & Leventhal, H.(1977). Ear differences in Evaluation Emotional Tones of Voice and Verbal Content, *Journal of Experimental Psychology: Human Perception and Performance*, 3 (1), 75-82.
- Schneider, F., Habel, U., Kessler, C., Salloum, J.B. & Posse, S.(2000). Gender differences in regional cerebral activity during sadness.*Human Brain Mapping*, 9,226-238.
- Silberman, E.K., & Weingartner, H.(1986). Hemispheric lateralization of functions related to emotion. *Brain and Cognition*. 5 (3), 322-353. DOI: [http://dx.doi.org/10.1016/0278-2626\(86\)90035-7](http://dx.doi.org/10.1016/0278-2626(86)90035-7)
- Simon, D., Craig, K.D., Gosselin, F., Belin, P. & Rainville, P.(2008a). *Montréal Pain and Affective Face Clips*. Downloaded 09 October 2012, from <http://amor.cms.hu-berlin.de/~simondan/>
- Simon, D., Craig, K.D., Gosselin, F., Belin, P. & Rainville, P. (2008). Recognition and Discrimination of Prototypical Dynamic Expressions of Pain and Emotions. *Pain* 135 (1-2), 55-64 .
- Studdert-Kennedy, M. & Shankweiler, D.(1969). Hemispheric specialization for speech perception. *The journal of the acoustical society of America*, 9(5), 579-594.
- Tan, K.A.M.(2013). Effects of handedness on affective lateralization: a divided visual-field study. (Master's thesis)Downloaded from;

[http://www.researchgate.net/publication/236736732\\_Effects\\_of\\_handedness\\_on\\_affective\\_lateralization\\_a\\_divided\\_visual-field\\_study/file/3deec5192bf67d1e3d.pdf](http://www.researchgate.net/publication/236736732_Effects_of_handedness_on_affective_lateralization_a_divided_visual-field_study/file/3deec5192bf67d1e3d.pdf)

- Thayer, J.F. & Johnsen, B.H.(2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology*, 41,243-246.
- Tucker, D.M. (1981). Lateral Brain Function, Emotion, and Conceptualization. *Psychological Bulletin* 89 (1), 19-46.
- Vytal, K. & Hamann, S.(2010). Neuroimaging Support for Discrete Neural Correlates of Basic Emotions: A Voxel-based Meta-analyse. *Journal of Cognitive Neuroscience*,22(12),2864-2885.
- Wager, T.D., Phan, K.L., Liberzon, I., & Taylor, S.F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *NeuroImage*, 19, 513-531. doi: 10.1016/S1053-8119(03)00078-8
- Williams, S.M.(2008). *A major revision of the Edinburgh Handedness Inventory*.  
Downloaded 1.03.2010,  
<http://homepage.ntlworld.com/steve.williams7/A%20major%20revision%20of%20the%20Edinburgh%20Handedness%20Inventory.pdf>

**APPENDIX A:**  
**INFORMED CONSENT FORM**

*Emosjonell Persepsjon eksperiment*

*Høst 2012/Vår 2013*

Norges teknisk-naturvitenskapelige universitet  
*Psykologisk institutt, NTNU*  
7491 Trondheim

Forespørsel om deltakelse i forskningsprosjekt:

**Audiovisuell persepsjon av emosjoner**

Studien vil undersøke hvordan vi oppfatter forskjellige emosjoner gjennom lyd- og videoklipp plassert på forskjellige steder på en skjerm. Du vil derfor bli presentert for videoklipp av forskjellige personer med forskjellige ansiktsuttrykk og ved noen tilfeller lyd som uttrykker en emosjon. Som for eksempel et videoklipp av et sint ansiktsuttrykk eller et lydklipp som uttrykker sinne.

Noen av klippene er med video og lyd, mens noen er bare video eller lyd. Lyden som blir brukt i dette eksperimentet blir ikke presentert i stereo, altså til begge ører, men til bare ett av ørene.

Deltakerens vil gjennom å trykke på en knapp på en responsboks gi til kjenne hvilket emosjonelt uttrykk som oppfattes, og gjør dette så raskt som mulig. Lengden på video- og lydklippene varierer og noen kan kanskje oppleves som korte og derfor vanskelig å oppfatte. Selv om du ikke er 100% sikker på hva rett svar er, trykk likevel på en av responsknappene.

Studiens utvalg vil bestå av opptil 80 personer: 40 kvinner og 40 menn. Deltakerne vil ha normal hørsel, normal eller korrigert til normalt syn med linser, og god oppmerksomhetsevne på dagen ved hjelp av spørreskjema og før-tester.

Undersøkelsen vil finne sted ved Talelaben, Psykologisk institutt, Dragvoll.

Total varighet av forsøket er beregnet til ca 60 minutter, eventuelt delt over 2 tidspunkter. Dette inkluderer en synstest og en hørselstest på 15 min før forsøket starter. Synstest og hørselstest vil bare undersøke aspekter ved deltakerens syn og hørsel som er direkte relevante for forsøket. Ettersom eksperimentatorene ikke har optometrist- og audiografutdanning kan de ikke diagnostisere eller anbefale behandling. Andre tester deltakerne vil gjennomgå er fargeblindhetstest, øyedominanstest og hendthetstest.

Prosjektet er basert på frivillig deltakelse, og man kan når som helst trekke seg underveis og be om å få data slettet uten begrunnelse. Man er ikke forpliktet til å gjennomføre, og en eventuell avbrytning vil ikke få noen konsekvenser.

Dataene fra eksperimentet vil benyttes i en masteroppgave og en eller flere faglig presentasjoner eller forskningsartikler som vil bli vurdert for publisering. Alle data som samles inn vil bli behandlet konfidensielt. Når prosjektet avsluttes i utgangen av 2013 vil all informasjon som kan knyttes til forsøksperson bli makulert (kontaktinformasjon som e-post adresser etc). Annen informasjon som vil bli beholdt i form av data fra eksperiment, vil ikke kunne føres tilbake til forsøksperson ved å være helt anonymisert.

Eksperimentet er godkjent av Psykologisk Institutt, Dragvoll NTNU og er meldt inn til Personvernombudet for forskning (NSD). Eventuelle spørsmål og henvendelser kan rettes til Gunn Kristin Halvorsen (gunnkrha@stud.ntnu.no).

## SAMTYKKEERKLÆRING

**Prosjekttittel:** Emosjonell persepsjon

Jeg har lest informasjonsskrivet og jeg har hatt mulighet til å stille spørsmål angående min deltakelse i eksperimentet. Jeg sier meg villig til å delta i prosjektet.

.....

Sted

Dato

Underskrift

*Prosjekt ansvarlig:*

*Dawn Behne, Førsteamanuensis, Psykologi, NTNU*

*Tlf: 73591978 epost: [dawn.behne@svt.ntnu.no](mailto:dawn.behne@svt.ntnu.no)*

**APPENDIX B**  
**QUESTIONNAIRE**

*Høst 2012/Vår 2013*

*Prosjekt om audiovisuell persepsjon av emosjon*

**Spørreskjema**

Deltager \_\_\_\_\_

Dato \_\_\_\_\_

Tester \_\_\_\_\_

Informasjonen som samles i dette spørreskjema vil bli behandlet konfidensielt. Svarene her vil bli koblet til svarene dine i eksperimentet, men ikke til ditt navn eller annen identifiserende informasjon. Informasjon skal kun brukes for å forstå resultatene fra eksperimentet bedre. Når prosjektet avsluttes i utgangen av 2013 vil informasjonen fra spørreskjemaene bli makulert.

For å svare på spørsmålene nedenfor, vennligst skriv tydelig, eller sett kryss der det passer/hvis det gjelder deg.

1) Alder \_\_\_\_\_

2) Kjønn.

Kvinne \_\_\_\_\_

Mann \_\_\_\_\_

3) Har du bodd over ett år i et annet land enn Norge?

Nei\_\_\_

Om ja, hvilket land?\_\_\_\_\_

- 4)** Hvis kvinne, bruker du p-piller, p-ring eller annen hormonell prevensjon? (Dette informasjon er ønskelig siden hormoner kan påvirke emosjon.)

Ja\_\_\_ Nei\_\_\_ Ønsker ikke svare\_\_\_

Om du ikke bruker hormonell prevensjon når opplevde du siste menstruasjonsdag? (om du ikke husker nøyaktig skriv ca når du tror det var/kommer til å være).

Dato\_\_\_\_\_ Ønsker ikke svare\_\_\_\_\_

- 5)** Studerer du, om ja, hvor ? ( NTNU, HIST osv)

\_\_\_\_\_

Hvilket fakultet tilhører du?\_\_\_\_\_

- 6)** Bruker du noen form for syns-korreksjon?

Ja\_\_\_ Nei\_\_\_

Hvis ja, hva (Briller, linser osv.)?\_\_\_\_\_

- 7)** Vil du si du har normal hørsel?

Ja\_\_\_\_\_ Nei\_\_\_\_\_ Usikker\_\_\_\_\_

- 8)** Føler du at du har hatt tilstrekkelig med søvn i natt, normal for deg?

Ja\_\_\_ Nei\_\_\_

- 9)** Har du drukket større mengder alkohol i løpet av de siste 24 timene (Det vil si en slik mengde at det påvirker din oppmerksomhetsevne i dag)?

Ja\_\_\_ Nei\_\_\_

**10)** Har du tatt medikamenter i dag som kan påvirke oppmerksomhet, syn eller hørsel?

Ja\_\_\_\_ Nei\_\_\_\_

**11)** Har du noen helsehistorikk som kan påvirke oppmerksomhet, syn eller hørsel eller evne til å oppfatte ansiktsuttrykk? (hjernerystelse siste 6mnd, epilepsi, ADHD , autisme, depresjon etc)?

Ja\_\_\_\_ Nei\_\_\_\_

*Prosjekt ansvarlig:*

*Dawn Behne, Førsteamanuensis, Psykologi, NTNU*

*Tlf: 73591978 epost: dawn.behne@svt.ntnu.no*

**APPENDIX C**  
**PARTICIPANT RECEIPT**

Norges teknisk-

Naturvitenskapelige universitet

NTNU – Trondheim

**Deltakelse i et persepsjonseksperiment**

Jeg kvitterer herved for å ha mottatt en 200 kroner Midtbysjekk i forbindelse med å ha deltatt i et eksperiment om emosjonell persepsjon i regi av Talelaben, Psykologisk institutt, NTNU.

.....

Sted

Dato

Underskrift

*Prosjekt ansvarlig:*

*Dawn Behne, Førstemanuensis, Psykologi, NTNU*

*Tlf: 73591978    epost: dawn.behne@svt.ntnu.no*