Expressive Intention and Perception of Body Language in Popular Music Performance

Interpretation of the visual communication between the vocal performer and the audience

Line-Mathilde Espeseth

Master thesis by the Department of Musicology

UNIVERSITY OF OSLO

18/10-2018
Expressive Intention and Perception of Body Language in Popular Music Performance
Copyright Line-Mathilde Espeseth

2018

Expressive Intention and Perception of Body Language in Popular Music Performance

Line-Mathilde Espeseth

http://www.duo.uio.no

Reprosentralen, University of Oslo

IV
Abstract

This thesis explores the communication chain between the performer and audience in the physical vocal performance, and more precisely how the performer’s body language (movements, gestures and posture) is subjectively and intersubjectively perceived by the audience. This is done by analyzing 24 participants’ comments about what they believe the performer to communicate, in two videos with different emotive intention (anger and sadness). The participants were free to express themselves about every aspect of both the audible and visual performance, as in a normal performance setting, which led to a large variety of findings. This thesis will present results both from the theory applied, and from the experiment, which concerns how we perceive the performer’s body language, and how we can interpret it from a scientific perspective. First and foremost, this thesis wishes to shed light on the lack of information about expression and perception of the physical vocal performance, as the field (of music and gestures) has been concerned with experiments with instrumentalists, not vocalist. One of the findings is that the research of expressiveness in music performance needs to establish a terminology for analyzing the body language of vocalists. It also seems that illustrating and emblematic gestures, which substantiate the lyrics and mood intended, are the most effective conveyors of emotional intention. This was found especially when the gestures and movements derived from the performer’s natural gesture repertoire, understood in this thesis as emotional leakage.

Keywords: visual communication, music performance, body language, performance intention vs. audience perception
Foreword

To me, music has always been about expressing emotions. All my life I have been moved, both emotionally and physically, by music, and I have always experienced music through movement. When I was ten years old, I started writing pop songs, and I always made up a dance to go with it. I have cried and laughed to music for as long as I can remember, and in my job as a front figure and vocalist, I strive to convey these emotions to the audience on a weekly basis; Sometimes successfully, other times with challenges. My hedonic motive for this thesis is therefore to be able to understand how and when I manage to convey emotions to my audience, and to improve the communication chain between me as a performer and my audience, so to be a better performer.

There is little, or often no focus, on body language and movements in vocal education. One often says that god musicians just know how to express themselves. Still, there is a lot of common movements done by great artists for conveying certain emotions/moods. We also know that the visual presentation of a performance affects our musical experience on several levels as a spectator (Thompson, Graham, & Russo, 2005), and that the environment for where and how the music is perceived affects our experience of it. There is a lot written about how musicians can work with their anxiety before entering the stage, but not so much about what to do on stage. Woody and McPherson (2010) talks about relaxation and visualization techniques before a performance, as well as focusing on the feelings/emotions you had when writing and rehearsing the piece/song, rather that thinking about how this is perceived from the audiences’ perspective. This is an introvert perspective of performing, and as of today, I have not found an extrovert perspective for working with performance anxiety.

There has been an increased interest in the systematic study of music performance, and music as performance, since the turn of the twentieth century (Clarke, 2004, p. 77), but the research on visual performance in popular music has been neglected for a long time (Auslander, 2009).\(^1\) In addition to this, research on vocal performance in popular music is scarce (Patrik N Juslin, 2005) and most of the research material provided, both in the field and this thesis, is on instrumental music.

---

\(^1\) This is strange, given that before the gramophone, the only place to hear music was at live performances (Thompson, Russo, & Quinto, 2008).
Acknowledgements

The last two years I have had help from several professors and staff at the Department of Musicology at the University of Oslo. I would like to thank Hans T. Zeiner-Henriksen and Stan Hawkins for helping me in the early phase of this thesis. Their insight to the widespread field of popular music studies helped me track a course for the thesis and lay the foundation on which it is built. Jonna Katariina Vuoskoski, for helping me maneuver in the large field of research that music, performance and emotion is, and for the kind words of motivation. Several of the staff members have been extremely helpful, and I would like to address my gratitude to Victoria T. Berg and Mons Thyness at the administration for their patience and help during the accomplishment of the experiments. The video experiments are central to this thesis, and I would like to thank my second supervisor, Alexander Refsum Jensenius, for his help and perspective on the technical procedure of filming, in addition to help researching the software best suited for the purpose. But most of all, I would like to thank my supervisor, Hallgjerd Aksnes. The complexity of this thesis, and its accomplishment, would have been an impossible task if it wasn’t for her expertise in several musicological fields. Even though I changed the angle and theoretical framework of this thesis several times, she read up on new fields within the science so to give me the best possible foundation. I also want to thank her for the number of times I have cried during our meetings, and how she always has managed to guide me through towards the finish line.

Thank you. I would also like to thank the staff at the Department of Musicology, the University of Oslo, for allowing me to write this thesis, granting me access to the green room and its equipment, and always giving me a warm welcome when I needed assistance on issues of access and information applied in this thesis. That said, I had a lot of help creating the videos, and I would like to thank my talented colleagues Joakim Arnt Holmen, for recording and producing the guitar track, Are Næsset, for recording and producing the vocals and Kristoffer Lislegaard for helping me film and produce the videos. Last, but not least, I want to thank the 24 participants who volunteered to help out with the experiment. This thesis would have been impossible if it wasn’t for you. And to all my friends and family who have cheered me on for two years; I am forever grateful.
Table of contents

1 Introduction .......................................................................................................................... 1
   1.1 Research questions; main objectives and limitations ................................................. 1
   1.2 Overview of thesis .......................................................................................................... 5
   1.3 Literature searches .......................................................................................................... 7

2 Theoretical foundation and earlier research ............................................................... 8
   2.1 Communication .............................................................................................................. 8
      2.1.1 Non-verbal communication ................................................................................... 9
      2.1.2 Verbal communication ......................................................................................... 10
      2.1.3 Musical meaning; something to be imposed or extracted? ................................... 14
      2.1.4 Perception of musical performance: Sight over sound? ........................................ 15
   2.2 Sociological perspectives; individual preferences and subjectivity ......................... 17
   2.3 Music and emotions ..................................................................................................... 20
      2.3.1 The basic emotions sadness and anger ................................................................. 20
      2.3.2 Communicating emotions in music performance ................................................. 21
      2.3.3 Music and the Circumplex Model of Affect ......................................................... 25
      2.3.4 GEMS; emotions in response to music ................................................................. 27
   2.4 Movements and gestures to music .............................................................................. 29
   2.5 Definitions employed in the thesis ............................................................................ 32
      Performer ......................................................................................................................... 32
      Communication ............................................................................................................... 32
      Participant ....................................................................................................................... 32
      Emotion .......................................................................................................................... 33
      Body language ................................................................................................................ 33
      Gesture ............................................................................................................................ 33
      Movement ........................................................................................................................ 34

3 Method ............................................................................................................................... 35
   3.1 Phenomenological issues ............................................................................................. 35
   3.2 Empirical musicology ................................................................................................. 36
   3.3 Applied methods; limitations and usage ..................................................................... 36
      3.3.1 The song ................................................................................................................. 36
      3.3.2 The videos .............................................................................................................. 38
      3.3.3 The performance/performer intent ....................................................................... 40
      3.3.4 Movements in the videos ...................................................................................... 41
      Terms from Jensenius et al. (2010) .............................................................................. 41
Terms from Kurosawa and Davidson (2005) ................................................................. 43
Explanation of terms used in the analysis ................................................................. 44
3.3.5 The questionnaire ......................................................................................... 46
3.3.6 The software ................................................................................................. 48
3.3.7 Procedure of the experiment ......................................................................... 50
3.3.8 Interpreting the data ..................................................................................... 52

4 The video experiments ......................................................................................... 54
4.1 The song ........................................................................................................... 54
4.2 Analysis of the questionnaire ........................................................................... 56
4.3 Exploratory video analysis: performer intention and perception of body language 58
4.3.1 The angry video ............................................................................................ 58
4.3.2 The sad video ............................................................................................... 61
4.4 The experiment .................................................................................................. 64
4.4.1 Comments on lyrics ..................................................................................... 65
4.4.2 Comments on the voice ............................................................................... 66
4.4.3 Adjectives used by the participants .............................................................. 67
4.4.4 Comments on body language ....................................................................... 70
   Comments on the legs ......................................................................................... 72
   Comments on eye contact .................................................................................. 72
4.4.5 Mood and lyrics substantiating gestures ....................................................... 73

5 Results and discussion .......................................................................................... 76
5.1 A critical approach to the experiments; limitations and expectations ............ 76
5.2 Models of communication in music performance ......................................... 79
5.3 The performance analysis ............................................................................... 81
5.4 The taxonomies ................................................................................................. 84
5.4.1 Discussion of the words pain and sadness .................................................... 86
   Comments on pain in the angry video ............................................................... 87
   Comments on sadness in the sad video ............................................................. 88
5.5 Visual versus audio ......................................................................................... 89
5.6 Subjective differences ...................................................................................... 90
5.7 Significant moments ......................................................................................... 92
5.7.1 Gestures and movements as emotional leakage .......................................... 93
5.7.2 Mood induction and attunement behaviours .............................................. 97
5.7.3 Centre of moment ....................................................................................... 99
5.7.4 Eye-contact .................................................................................................. 99
5.7.5 Lyric substantiating gestures .................................................................... 101
5.7.6 Exercising and body awareness ................................................................. 102
6 Conclusion

6.1 Future work

References

Appendix 1 Video analyses (angry)

Appendix 2 Video analyses (sad)
1 Introduction

Music is often regarded as the “language of the emotions” (Patrik N Juslin, 2005; Woody & McPherson, 2010), and as we know, language is so much more than just words since “only 7 per cent of the content of a message is verbal, the remaining 93 per cent coming from vocal tone and matters of body language” (Moore, 2012, p. 109).2 But what about music? How important are the matters of body language in musical performance?

1.1 Research questions; main objectives and limitations

The intention of this thesis is to continue the research on a question asked by Dahl and Friberg (2007): “What types of movements supply the bits of information about the intent and mood of a performer?” (p. 434). This thesis is then motivated by the thought of intersubjectively detecting movements that expresses the emotions of sadness and anger and which, hopefully, can be applied to pop music performance in general. The experiment and research is inspired by the methods and findings of emotive facial expression by Ekman (1993). Even though this thesis has a different field of focus than Ekman (his focus was on facial expression and this thesis concerns the body language), the findings of Ekman and how he separated the facial expressions into emotive categories inspired my thoughts of separating different gestures and movements into emotive categories, which underlines both the theoretical and methodological foundation in this thesis. My intention is therefore to find out:

Is it possible to delimit and intersubjectively identify bodily postures, gestures and movements that can augment or attenuate the audience’s reception of the performer’s expressive intention? Which bodily postures, gestures and movements are the most effective conveyers of anger and sadness in the communication between the vocal performer and the spectator?

“...A central preoccupation in research on performance has been the nature and function of expression” (Clarke, 2004, p. 84), but, the study of expression in music has mostly concerned

2 Allan F. Moore is here quoting Mehrabian, and I can only assume that this is numbers derived from a study on verbal conversation that takes place face to face. These numbers derive from an experiment “dealing with communication of feelings and attitude” (Albert Mehrabian, n.d.) and not communication of other matters. His findings have been misinterpreted in several articles, but here it is used to establish the importance of body language, and not as a mathematical rule to follow.
instrumentalists, and there is little research on the physical expression of vocalists, and how this is perceived. This affects the theory underlining this thesis, but a lot of the theory applied can be seen in the light of vocal performance. The study of vocal performance interferes with several fields of research. The field of music psychology has been an important factor for developing the empirical studies of performance, and both psychology and sociology is interfering with, and even determining, the social act of performing (Clarke, 2004; Davidson, 2004).

As a theoretical and methodological foundation to understand the extent of the research questions, I find Tagg’s (1982) question of “why and how does who communicate what to whom and with what effect” (p. 39) underlying the experiment. The complexity of this question is too big of a task to answer in full, but for the validity of the results from this experiment, it is crucial that I try to explain how, who, what and with what effect. All of which are important aspects for understanding any communicational matter. The question of Dahl and Friberg, mentioned earlier, concerns both the performers intention (what) and how this is perceived by the participants (what effect). I have therefore applied different theories and methods to examine the communication chain in music performance, as recommended by Windsor (2004), since it is so intricate in its nature. My main experiment is a video experiment, consisting of two videos, one intended to convey the emotion sadness and one intended to convey the emotion anger. When talking about emotions, I do not mean embodied emotions (emotions that are mentally and physically experienced in that time and place) but perceived and intended emotions. This does not mean that perceived emotions, or the emotions I intend to convey as a performer, cannot be embodied. But, since I have no way of measuring the physical response of the participants, or myself when performing, I cannot know whether the participants experience some kind of emotional valence or perceives it from the performer. The knowledge of what is intended by the performer would be impossible to explain for anyone who does not possess a first-person explanation of the intension, and I am therefore using myself as the performer in the two videos. I am performing one of my own songs (that I am sure none of the participants have heard before, and therefore have no emotions attached to it), with two different body languages and emotions presented (sadness and anger). The body language was deliberately exaggerated to see if it provoked some kind of reaction from the participants. The reaction I hoped for was
that these movements would simplify the mood detection and more accurately express the performer’s intended emotion to the listener.

The shift from a highly subjective and personal description of the body language used, to a more descriptive analysis of it will be divided into first-person, I, when talking about my intentions, and third-person, the performer, when describing the observation of the movements. By using first- and third-person descriptions in the analysis, I hope to clear out some of the issues with subjectivity that underlies this whole thesis, so for the reader to understand when I am talking about my personal intention, and how I, as a researcher, observe this.3 The term body language is of particular importance in this thesis, including all aspects of the visual (physical) performance, which includes movements, gestures and postures made by the performer. (If I were to narrow down the findings in the experiment to only one of the categories mentioned, it would not have been beneficial for the thesis, as I don’t have enough participants for a valid conclusion regarding only the movements, or gestures of the performer.) I have analysed the body language in the two videos to the best of my abilities, applying terms from Jensenius, Wanderley, Godøy, and Leman (2010) and Kurosawa and Davidson (2005).

This thesis builds on the view of music as a social act, since the goal is to find out what is communicated (expressed and perceived) by the singers’ body language, which requires more than one person; both a performer and a perceiver. I have therefore gathered 24 participants who voluntarily watched and commented the two videos, which creates the foundation to answer how the performance is perceived (what effect). I chose to talk about how the performance is perceived by the participants, since this may include both the reception of a message from the performer to the participant, but also the participants subjective opinion on this. When it comes to how the performance is perceived, it is not just the audible music that is important to take in mind, but also the surrounding environment, who the performer is, and the perceiver (Hargreaves, MacDonald, & Miell, 2005). Because of the view of music as a social act, involving more than one person, it is also important to point out that there are no “one truth” in the performance, but rather a subjective, and maybe intersubjective, meaning of it. By this, I mean that what is perceived by the participants might align with my intention

3 An in-depth explanation of subjectivity will also be given in sections 2.2, and this will be the foundation for analyzing the videos in section 4.3. and further discussion in section 5.3.
as a performer, and other participants perception, which makes it an intersubjective perception, but this might not be the case for other participants. Since this experiment concerns the perceived communication via audio-visual information, the setting varies from how we normally perceive a live performance. The difference from analysing a music video, or a live music performance, is that this is an artificial setting, and in this case the video is stripped clean for all the normal “disturbing” elements in a performance for the participants to focus only on the message being conveyed or communicated, which in this case is a multimodal impression. These aspects of the experimental setting are one of the reasons why I chose to conduct the experiment with music students, since they are accustomed to analysing musical performance, both live and in videos, and often possess a technical language for describing their perception of it. For the purpose of knowing who my participants are, I have used a questionnaire for a more detailed feedback on the participants’ personal reflection and background (whom).

The use of different data collecting methods is also creating a huge problem area as the amount of data I received, from both the video experiment and the questionnaire, is too big for the thesis, and I have chosen to narrow it down. An example of the limitation process is the question “Are people who exercise on a regular basis more aware of their own and other’s body language, and therefore better at extracting the emotive message in a performance?” This was meant as a part of the research questions at the start of this process, but while conducting the experiment, I understood that I did not have enough participants for successfully answering the question. Other limitations were decided beforehand, as for the discussion of music and meaning, which is not included in this thesis but might have been relevant for a philosophical angle on the research questions. I also chose to exclude the category of facial expression, since there is a lot of research on emotive facial expressions (e.g. Ekman), and in order to focus on the body language. This is also done to narrow down the data material, since the data material collected would have been too big for a thesis of this size, in addition to simplify the experiment. The experiment is conducted on a 13-inches

---

4 The question was still used in the questionnaire (explained in section 3.3.5) and are included in both my result and discussion section (5.7.6) and as a suggestion for further research (section 6.1).  
5 Still, one can argue that expression in all forms are meaning, depending on our individual perception of it. The issue of subjective meaning will be explained in section 2.2, since it is of relevance for the validation of the experiment, but I have tried to keep a sociological perspective on this philosophical question.
screen, and the software used in the experiment is only showing the video on half the screen, which in combination makes it difficult for the participant to see the facial expressions (and especially the eyes) of the performer clearly. However, I still find that the software used is sufficient for the experiment, since the focus is on the body language of the performer, which is visible in detail.

To sum up, this thesis focuses on issues of communicating emotions through body language, and the purpose is not to find bodily expressions that are universally understandable, but to examine whether or not my intentions are intersubjectively understandable among the Norwegian students in my experiment. The motivation for this thesis is that it can help bring clarity to how visual vocal performances are perceived.

1.2 Overview of thesis

Since this thesis is divided between several fields of musicology, the theoretical foundation is widespread and interdisciplinary, applying perspectives from – most importantly – sociology, psychology and cognitive musicology. It is impossible to give a profound insight into all the disciplines in all their diversity. Thus, I have instead chosen to introduce the theoretical framework, upon which this thesis depends, in the beginning, and will further lay out some smaller, but significant, findings from earlier research that is important for the understanding of the results as they appear.

As a point of departure for this thesis, it starts with an explanation of theories on communication (section 2.1), both verbal and non-verbal communication, in addition to musical communication and how this is perceived in both the audible music and the physical performance. I will further explain the sociological perspectives in this thesis (2.2), which includes the issue of personal experiences and individual preferences, in addition to the issue of subjectivity that underlines the empirical research conducted. Section 2.3 concerns the complex topic of music and emotion. This is included for the reader to understand the underlying intention of the performer, in addition to providing a foundation for earlier research on this topic, both from a sociological and musicological perspective. As a natural extension of this topic, I continue explaining earlier research on movements and gestures to music (2.4), that will be used as a foundation for the analysis later on. I will end the chapter with the definitions of the words that are important for this thesis (2.5). The reason for
introducing the definitions after the theory is presented, is because the theory foundation abounds of differences on the definitions of central words for this thesis, and so for the clarification of my personal use of the terms I choose to present the definitions after the theoretical framework is laid out. The methods chapter (3) starts with an explanation of phenomenology (3.1) and empirical research (3.2), since this is the foundation for the conduction of the experiments. Further, I will present the applied methods (3.3) that consists of an explanation of the song used (3.3.1), the videos (3.3.2), the performer intent (3.3.3) and the terms used for analysing the movements (3.3.4). The questionnaire, and how this was construed (3.3.5), are of importance for validation of the experiment, as for an explanation of the software used (3.3.6), the procedure of the experiment (3.3.6) and how the data is interpreted (3.3.7). Chapter 4 concerns the experiments, where I will start off with a short analysis of the song used (4.1) for the reader to have a foundation when the analysis is laid out. The analysis of the questionnaire (4.2) are of importance to understand who the perceiver in the experiment is, as with an analysis of the videos (4.3) for understanding what is communicated. Section 4.4 concerns the feedback from the participants on the two videos, and I have chosen to present the comments in different sections according to what they concern (sections 4.4.1 to 4.4.5). These sections (4.4.1 to 4.4.5) are an attempt to structure the comments as to prepare a foundation for the findings and further discussion. The results and discussion chapter (5) starts with a discussion of the limitation process (5.1) before further discussion of the results from the experiment compared to the theoretical foundation applied in the thesis. The models of communication used will be discussed in section 5.2, the terms applied for the performance analysis in section 5.3, and the taxonomies in section 5.4. After this I will discuss how we perceive a multimodal expression via our senses in the light of the findings (5.5), and how we all perceive this inter/subjectively (5.6). The last section in this chapter concerns what I have choose to call “significant moments” (5.7) and are divided into different section according to what they concern (5.7.1 to 5.7.6), as with the feedback from the experiment. Last but not least, I will provide a conclusion (chapter 6) and an agenda for further research (6.1). The appendix, that includes comments that are found to be relevant to the research questions, is attached as 1 (angry) and 2 (sad) at the end of the thesis, both including a permanent link with the video it represents. The appendix consists of several columns: The lyrics and time code of the video, an “objective” description of the body

---

6 The main focus is on the perception and conduction of the physical performance, but the music in itself (melody, timbre of the voice and lyrics) is off course of significance, since if there wasn’t any music the performance would have been a speech, or a theatre play.
language of the performer, the movement categories of which the different movements are
labelled, an explanation of the performer intent and the comments by the participants. The
comments by the participants are separated by colour for the author, and reader, to be able to
distinguish between the genders. The red comments are written by females and the blue by
males. This is done to simplify the process of validation, and as a starting point for further
research, since I am not allowed to keep the participants personal information.

1.3 Literature searches

For the sake of validation, I will account for the literature searches that are part of the
theoretical and methodological foundation for this thesis. The thesis comprises many
different fields (music performance, music and movement, and music and emotion), and in
order to orient myself within this vast amount of research, I have consulted reference works
that have been introduced to me via my supervisors and in lectures as a point of departure. As
an example, I was introduced to Juslin and Sloboda’s (2010) “Handbook of Music and
Emotion: Theory, Research, Applications” by my first supervisor, Hans T. Zeiner-Henriksen,
which is a cornerstone for this thesis. I then used the book’s reference lists for further
literature searches. In addition to this, I also performed searches on specific articles and
researchers in academic databases, as well as open searches on terms such as “music +
emotion”, “music + performance” and “body language”, in various combinations in Oria
(used by the library at the University of Oslo) and Google Scholar. This led me to online
archives such as Jstore, and online journals such as “SAGE journals”, “Music Perception: an
interdisciplinary journal” and “Psychology of music”, to name the most important. Still, most
of my literature has derived from reference works where I have done searches on specific
articles.
2 Theoretical foundation and earlier research

The singer’s task is multiple, comprising both the physical techniques for producing the sound, the pronunciation of the words, the interaction with both co-performers and audience, and the conveyance of the narrative of the song (Davidson, 2005). Furthermore, the communication issues in performance concern both the performer’s intention and the audience’s perception and interpretation (reception) of it. In section 2.1, I will present a theoretical foundation for the understanding of non-verbal and verbal communication to answer Tagg’s (1982) question of how we communicate (confer page 2) that underlines the first research question (is it possible to delimit and intersubjectively identify bodily postures, gestures and movements that can augment or attenuate the audience’s reception of the performer’s expressive intention?). The attempt to throw light upon this research question will include an explanation of how we perceive musical performance both physically and individually through our senses. When it comes to musical perception, the “who communicates to whom” is understood as a sociological perspective that will be explained with both an approach to individuality and subjectivity in section 2.2. When attempting to answer the second research question (which bodily postures, gestures and movements are the most effective conveyers of anger and sadness in the communication between the vocal performer and the spectator?) I will start by discussing earlier research on music and emotion, and how we detect emotions in music (2.3). This discussion is crucial for the validation of the experiment, as it gives a theoretical foundation for the understanding of how we are able to detect different emotive states in the physical appearance of human beings (2.4).

2.1 Communication

When working with communicational issues, it is important to note that all people have individual ways of interpreting and learning processes, so it is up to the performer to convey the desired message as understandably as possible. For this to happen, several factors are at play, since you need to be aware of how you express yourself, both verbally and physically, and you need to understand who your audience is.
2.1.1 Non-verbal communication

The importance of non-verbal communication between human beings has been well established for years, especially in research on the mother-infant relation (Stern, 1985) but also in the relation of performer-audience (Kurosawa & Davidson, 2005). Through experience and human interaction, we start learning to interpret social situations via facial expression, body language and the tone of the voice before the spoken language (Wolff, 1948), but after our first years we start to use words to express ourselves in clarity, and the importance of body language is only a secondary source for information (Stern, 1985). The readings of non-verbal communication is based on affect attunement where we read the subtext of the tone and body language, to attune with other peoples’ emotions and reactions to the outer world (Stern, 1985). This type of attunement is often subconscious, and we are often unaware of the change in our physical and mental state. Affect attunement is most common in our social encounters, but it can also occur in a musical setting as it is “the performance of behaviors that express the quality of feeling of a shared affect state without imitating the exact behavioral expression of the inner state” (Stern, 1985, p. 142). An example can be if one person is depressed, he or she might, in the encounter of others, influence or affect other peoples’ body language and tone of voice, but the emotive state of the depressed person does not necessarily become embodied within the others. The experience of a “shared affect state” as mentioned by Stern (1985, p. 142) can be explained as “significant moments”, as done by Trondalen (2004), which can be applicable in several situations. Trondalen (2004) uses the term from a therapeutic point of view, but in her definition these moments are not just emotional, cognitive or physical, but describes interpersonal meetings in time and space that is hard to translate into words. Our ability to convert the perception of emotion from one sense to another is based on amodal perception (Stern, 1985), and is a prerequisite of the social reading of interpersonal communication. (Example: When you see a person with an angry face and translate this to how the person feels and how the voice would sound like.) Affect attunement might be a desired state for a music performer, where the audience attune with the performer’s intended expression, and where the reading of the performer’s emotive state has a multimodal concordance (amodal perception). This means that the body language is an important factor for understanding people’s emotive states, and ergo plays a large role in expressing emotions in social encounters.
Our gesture-language is largely international, and according to Wolff (1948) “it must correspond to the primary level of existence, comprising instincts and emotions on the one hand and an elementary knowledge of objects on the other” (p. 1). One of the most important visual aspect in verbal communication are hand gestures, which often accompany speech in a natural way, with the most energetic part of the gesture substantiating “the most prominent syllable” (Davidson, 2005, p. 222). These types of gestures, along with facial expressions and body language, helps us communicate the verbal message in clear, as to substantiate our emotional message and the subtext of the words. In our everyday life, the correct reading of body language and tone are important for our communicational process, since “when watching others speak, we can tell how committed they are to what they are saying, we can tell whether to risk trusting in them, by how they express themselves, how they accompany their words with posture and gesture” (Moore, 2012, p. 201). This process of amodal perception are linked to the findings of mirror neurons, which has been proposed as a “mechanism allowing an individual to understand the meaning and intention of a communicative signal by evoking a representation of that signal in the perceiver’s own brain” (Molnar-Szakacs & Overy, 2006, p. 235). This theory was first tested on monkeys and, during the early 2000, they found that there are some correlations between this communicative process of the monkeys and human beings. Several areas of the frontal brain are involved in the process of receiving and understanding motor actions, in addition to imitate actions (Molnar-Szakacs & Overy, 2006). This means that the mirror neuron system is a big part of our learning process when growing up and the way we communicate with each other, both verbally and physically.

2.1.2 Verbal communication

As we start using language to express ourselves the complexity of communication grows. The concept of the conduit metaphor (first identified by Reddy, 1979) gives a good indication on why this is. Our thoughts and feelings are first ejected into an “external idea space” (Reddy, 1979, p. 291) by writing or speaking them, as for music. Then they are reified in this external space, so that they exist independent of any need for living human being to think or feel them; these reified thoughts and feelings may, or may not, find their way back into the heads of living humans. (Reddy, 1979, p. 291)

This means, that the conduit metaphor is colluding our everyday language and consists of thoughts, ideas and dead metaphors that we use daily (give me an idea or grasp that thought.
These are metaphors in that way that thoughts and feelings are not something we can grasp, give or touch. In this thesis, both the audible song, the comments by the participants and my movements can be seen in the light of this external idea space. All of these elements are thoughts and feelings that are in some way manifested as verbal or gestural communication by the performer, that may or may not find its way back to the audience. According to Patrik N Juslin (2005) communication can be seen as a transmission of an *intended* message between two or more people that shares a common “code” (p. 86). This definition can be seen in the light of both verbal, and non-verbal communication, but I find the statement on how the message communicated between people should be intended incorrect. The message can be everything from an unintended body language, or tone of the voice that reveals one person’s emotional state, to a deliberate verbal message, gesture, or posture. In this thesis, the message communicated is just as much unintended as it is intended gestures, and, which I will present in my results (section 5.7.1), the unintended messages communicated from me to the participant is often the most interesting ones (e.g. the gesture of touching the heart, and the leg gestures described in section 5.7.1). When the message is received in a way that aligns with the sender’s intention it is an example of a good communication process, but when it is misinterpreted, or just interpreted in a different way than what was intended from the sender, it is still communication, but maybe miscommunication. The conduit metaphor also explains that if a sentence is misunderstood, it is the speaker to blame and not the listener (Reddy, 1979). If we were to think of thoughts as something that is to be given, like a gift, the act of unwrapping the gift is easy and can hardly be done wrong (Reddy, 1979, p. 289). In the case of this study, I am the speaker and the participants are the listeners. If we don’t communicate well (that is if the participants misunderstand the task or don’t find the performer to convey anything in the videos), according to the conduit metaphor, I have myself to blame for not communicating my intension in clear. This is a simplified view on the complexity of communication, and there is still a potential pitfall when it comes to the interpretation of the participants answers. When analysing the comments from the video experiments, the participants and myself are both speakers and receivers and the whole process will be colluded with thoughts from both participants sides (that is both me as the performer and author, and the external participants), which leaves us with several “gifts” to unwrap. Communication is always both a subjective, and intersubjective matter, and will differ both from person to person, and within different situations. What is considered verbal communication in this thesis is all aspects concerning speech, such as semantics (metaphors
and subtext) and prosody (volume, tempo, timbre, articulation and accentuation). This is first and foremost the comments made by the participants, and the lyrics of the song, since verbal communication and singing has a lot of similarities. The lyrics may contain semantics and metaphors and the melody are influenced by the same prosodic aspects as speech, both of which are explored and developed from an early age (Welch, 2005).

To be able to answer Tagg’s (1982) question of *Who communicates what to whom?* (confer page 2), I have chosen to include Hargreaves et al. (2005) “Reciprocal feedback model of musical response” (p. 8). This model includes a lot of elements that can affect the listeners perception of a performance, such as the situation or state of mood. Hargreaves et al. (2005) has a different model for musical performance, but, in the performance model, they do not take in mind the listener or the spectator of the performance, which is why I find the model of musical response better suited for the purpose of this thesis. On the other hand, they include “broadcast and video” in the performance model which is important for this thesis, but I find

![Figure 1 Reciprocal feedback model of musical response (Hargreaves et al., 2005, p. 8).](image-url)
that this could also be included in the box for “music” in the reciprocal model of musical response (figure 1). Still, the model for musical response is based on the live experience with music, and in this thesis, the experiment is not coherent with a live performance. In a live performance, the communication chain would be performer-listener-performer (where the performer is affecting the audience, and the audiences’ response is affecting the performance), but when it comes to broadcast and video, as in my experiment, the communication chain is performer-listener (given that I use a video recording without any possibility to take live feedback from an audience). I find that the model for musical response (figure 1) can be seen in the light of musical communication with the minor alterations on adding performance in the box of music. A different way of explaining the communication between the performer and listener in music performance is given by Hargreaves et al. (2005) and is called the “Reciprocal feedback model of musical communication” (p. 18) (figure 2).

![Image of the Reciprocal feedback model of musical communication by Hargreaves et al. (2005, p. 18).](image)

In this model, the point of effective and good communication is given in clear (where the triangles meet in the middle), with the elements of influence from both performer and spectators point of view explained in the two gray triangles. This model can be seen in the light of the model for musical response, where the boxes for “music” and “situation and contexts” (both from figure 1) are the two grey boxes that gives rise to the performance and how it is shaped and received from both the spectator’s, and the performer’s, point of view in the model for musical communication (figure 2). This means that there are big variables from

---

7 Hargreaves et al. (2005) explain that “broadcast and video” have the same foundation for interpretation as the box for “music” in the reciprocal feedback model of musical response.
both the performer’s/composer’s side, in addition to the listener’s, that can influence whether or not the performance in itself is perceived as integral or of a personal liking for both the performer and the listener/spectator. I find that this model of musical communication is suited both for studying live performance, and broadcast and video, on the basis of its simplicity and is therefore the foundation for further methodological framework of this thesis.

2.1.3 Musical meaning; something to be imposed or extracted?

The conduit metaphor concerns speech, but when applied to this thesis, both the lyrics and the underlying text conveyed by both the voice and the body language, as for the thoughts and feelings of the performer and the participant’s answers, are put into this external idea space. This creates an even bigger communicational issue, since we are not just talking about the spoken language, but also a language that is without words. This said, both types of languages are in need of interpretation to be fully understood, if ever, which implies that both my performance in all its facets, and the answers from the participants will be interpreted in different ways. Davidson (2004) states that “in the vast majority of music-making contexts, the real or implied presence of others means that at some level social communication or interaction takes place” (p. 57). This implies that we can think of the act of creating and performing music as a social act, and not just merely as an individual process. Music only become hearable music when people interact with it, whether alone or together, which means that it always has a communicative purpose. Since the science of music is a human act, several fields of musicology will always be interpreted in different ways according to the individual applying them since “musical analysis is not an exact science and cannot be relied upon to provide an unequivocal basis for distinguishing between errors and intentions” (Clarke, 2004, p. 85). Besides the mathematical aspect of music (music as audible sound, intervals and rhythms) it is difficult, or maybe impossible, to extract findings that is common to all men when studying musicology. The theories of musical communication can be seen from different angles: One is that “a normative reading of semiotics would assert that the meaning of an artistic expression has become encoded within it, such that the role of a competent spectator, or listener, is to decode that meaning” (Moore, 2012, p. 220). Within this view, the listener, or in this case the participant, is holding the key to “decode the meaning” of the song and/or the performance, and will be “able to complete the message, or make sense of the song, through using persons, times and places from their own physical and metaphysical situation” (Tim Murphy quoted in Moore, 2012, p. 102). The opposite way to
look at it is “a model in which the meaning is not considered encoded by the listeners” (Moore, 2012, p. 221), but rather that the participant’s role is to extract the meaning that the performer imposes in the performance. (This view aligns with the conduit metaphor.) Moore (2012) states that “we cannot generalize sufficiently to explain what a particular group of listeners will understand of the meaning of a particular track, and that it is therefore obfuscatory to pretend that we can” (p. 207). Can we ever conclude with anything other than physical mathematical issues when it comes to music, or art in general, since it is so individually perceived? As will be explained later in this thesis (section 2.2), we are all shaped by our environment and upbringing, and no one can ever fully understand other peoples’ thoughts and feelings, but one can strive to take a broader outlook on these issues, as this thesis attempts. Since the foundation for this thesis relies on the view that communication is a shared social interaction between people, my view is that the participants interpretation of the performance is equally valid as my own. Musical experiences will always have as many truths as there are thinking minds and I will never be able to find out exactly what the participants believe me to convey, and how this makes them feel, and my findings are thus not universally, only contingently valid. Instead, such findings can throw light upon the polysemous, subjective and singular nature of musical meaning.

2.1.4 Perception of musical performance: Sight over sound?

How the music is presented to us, in which media, is crucial for our perception of it (Hansen, 2017; Thompson et al., 2005), and in several studies they found that visual information caused the auditor to hear musical sound differently (Auslander, 2009). In a historical perspective, our vision is more important than the other informational senses, since it helps us fit into a group by copying the body language of the others, and this can explain why we learn to copy movements before words, as explained earlier. When adapted to the understanding of music, it is necessary to state that all music derives from action (the action of drumming, playing piano or singing etc.) and that when watching, and even just hearing these actions, the frontal centre and mirror neurons are activated. That means, that we mentally imitate the action conducted to create the sound perceived (Molnar-Szakacs & Overy, 2006, p. 236).

---

8 Moore is here concerned about musical style, but this also applies to the phenomena addressed in this thesis.

15
Several studies have found that the view of a performance is equally important as the sound when it comes to grasping the expressiveness of a performance (Dahl et al., 2010). The study by Vuoskoski, Thompson, Spence, and Clarke (2016), addressing the coherence between audio and visual information in musical performance, concludes that “visual information about a performer’s movements not only has an impact on the intensity of emotional reactions evoked by the performance, but can also change how that performance sounds to an observer” (p. 469). In Davidson’s (1993) study of visual music performance, she used three different exaggerated bodily expressions (projected [as in a simulated public performance], deadpan [with reduced expression of mimic, dynamic etc.] and exaggerated [where all aspects of the performance are overstated]) and mixed the bodily expression with a more neutral sound. She had the performers wearing tight-fitted black tracksuits and used the point-light technique which is a method that “uses ribbons of glass-bead retroreflective tape attached to the major body joints, and spotlights mounted very close to the lens of a camera so that the light hitting the reflective ribbons is reflected back into the camera” (Davidson 1993, p. 105). By using this technique, one minimizes the information given and reduces the chance of the spectator judging the performer by his/her outfit and looks, since one can only see the movements of the performer. The study was done in three different ways; one with both audio and video presented, one with only the video and one with only the audible sound presented. Her conclusion was that non-musicians tend to judge a performance by the visual aspect of it, and not so much by the audible sound, and that “vision is the only mode to significantly differentiate between projected and exaggerated manner. This suggests that vision most clearly conveys the differences between the three performance manners and therefore suggest that vision is the most “effective” indicator of manner” (Davidson, 1993, p. 109). In Tsay’s article on “Sight Over Sound In the Judgement of Music Performance” (2013) they found that the participants who watched only the video (without audio) had a better chance of detecting the award-winning pianist than the ones that were given only the audio (and not the video). Their conclusion where that “both novices and experts make judgments about music performance quickly and automatically on the basis of visual information” (Tsay, 2013, p. 14583). On the contrary, Camurri, Mazzarino, Ricchetti, Timmers, and Volpe (2003), found that the sound seemed to be the primary focus of the participants, and not the visual aspects. A big difference in these studies is both the level of musical training the participants had, and the number of participants. (Tsay used both novices and experts [1164 participants], as with Camurri et al. [four out of twelve participants were
musicians], whereas Davidson only used music students who were familiar with viewing music performance [21 and 34 participants in the two experiments]). Davidson’s decision on exclusively using spectators with musical knowledge was based on findings about how observers with some musical experience have access to similar, but more detailed, information than observers without any musical experience. This means that we still don’t know whether or not the visual aspect is the predominant one when it comes to judging the expressiveness, or level of expertise, in music performance. But, seen in the light of the majority of these experiments, and how the visual sense is the key holder to our development in the early years (confer page 9), as with detecting the sound source (confer section 2.1.1), the visual aspect of the musical performance is at least important for our judgement of it, if not the predominant one.

2.2 Sociological perspectives; individual preferences and subjectivity

The readings of social behaviors (body language, facial expression and tone of voice) affects how we act and think (affect attunement), as well as our physical behaviors (Stern, 1985). The environment in our upbringing and our interpretation of it is a prerequisite for individual differences as “the same situation and the same stimulation can produce different affective states in different individuals depending on the nature of their cognitive appraisal of the situation” (Scherer, 1986, p. 146). When it comes to analyzing musical performance, there are several aspects that needs to be taken in mind since

- no analysis of musical discourse can be considered complete without consideration of social, psychological, visual, gestural, ritual, technical, historical, economic and linguistic aspects relevant to the genre, function, style, (re-)performance situation and listening attitude connected with the sound event being studied. (Tagg, 1982, p. 40)

All of these topics can be seen in the light of sociocultural analysis and Juslin’s (2005) thought on the communicated message as “a shared “code” among senders and receivers” (p. 86). This code might consist of “interpretations rules (for instance, musical systems) which are shared in a group or culture, or of inference dispositions based on personality, prior

---

9 What is important to note is that in all of these experiments, as with mine, the participants are put in an artificial situation, asking them to focus only on the video. In a live setting, there will be disturbing noises, people coming and going, and the sound can vary from where you are standing in the venue.
experiences, musical talent or musical preferences” (Zentner 2010, p. 111), but the meaning of that message might be different for the receiver than for the sender, even within the same group or culture. This said, it is not that the message can have widely different meanings from the performer to the receiver/audience, and it is still a message, but it might have a different connotation for different recipients in different cultures. Especially when it comes to musical communication, there is a high chance of the performer and the audience to have different interpretations of the message. This because, in the first place, there is a lot of factors that can influence the audience or steal the audiences’ attention away from the performer and the message that the performer intends to convey, in addition to difference in upbringing and socio-cultural references (Zentner, 2010). Clarke (2004) discusses the problem of expressiveness in performance and whether or not a performance should be regarded as an individual, personal intention, or as a part of a cultural conditioned expression. It is hard to say if anyone really can separate themselves as a person from the society they grew up in, or live in, and if not, all performances are to be regarded as both individual and cultural conventions. In my view, emotions, and to a certain degree thoughts – which both are a foundation for music creation – can only be fully understood by the one who possesses them. When we describe a feeling or thought to someone, we can never be completely sure that the other person understands precisely how we feel, since our emotive and mental self are developed in different ways throughout life and upbringing. This said, meanings or points of view might be different since a meaning often is more directly spoken, like “I don’t like this music” or “this is unfair”. In this case, the meaning itself is clear, but the grounds of why the person takes on this statement might be harder to grasp, since it is based on emotions and thoughts that derives from personal experiences. This sociological issue requires further research, but for this thesis I will say that my performance is shaped by the culture and age that has been my childhood, and me as a person. In my opinion, music cannot be studied without taking culture, production etc. in mind since the act of “musicking” (Small, 1999) is always both an individual and sociological process.

Stern (1985) explains that “Some might say that research or theory that is determined by highly personal factors should not be trusted” (p. ix), and I partially agree to this. But, my opinion is that if the researcher is aware of his/her own situation one can get as good results

---

10 It has to be stated that when Patrik N Juslin (2005) talks about musical communication he explains that a musical piece may be interpreted differently from one receiver to another, but that the “listeners usually agree about the general nature of the expression in a piece of music” (p. 88).
as any when interfering with questions of interpersonal communication. It might even lead to better results if the researcher is able to see the situation from different views than if the researcher did not have any hands-on experience with the object of observation. As human beings we always evolve, and this includes our knowledge of the inner- and outer world. With this in mind, it will be impossible for me to give a performance that I can stand for a hundred percent for the rest of my life, maybe not even to the end of this master project. Since I am ever evolving, my readings of my own body language and vocal expression might change from the time I did the recordings, until I started analysing them. My perceived emotive state in each movement and vocal cue will never be static in that sense that my perception of it, in addition to the liking of it, might have shifted only short time after the recordings, which also applies to the analysis of the experiment in total. But, I will still have the knowledge of when my gestures and movements are intended, and what emotive state I wanted to convey in an overall sense, and if I were to use a different vocalist I will have the same problem of interpreting the movements, since they might not be aware of it. Davidson conducted a study with a pianist who did several performances, with very different gestures, and according to Davidson (2001) “The pianist’s lack of knowledge of his expressive intentions made it impossible to say why his gestures had varied from one performance to another” (p. 246). By using myself in this research I can eliminate some of the problematic issues pointed out here by Davidson (2001), since this is a staged performance that I have rehearsed, and I am aware of when my movements are intended or not intended. The problem area here is what these subconscious or unintended movements are based on, if they are the result of my emotional experience through different social events in life that is not related to music, or movements that is deliberately imitated from my idols, parents or other people involved in my emotive and musical growth. Since we are all a product of our time and culture, it is hard to differ from what is the “neutral” and the “natural” me, and there will be no way of finding out exactly why I use these types of movements. But, as explained, this would be equally difficult if you use an external performer in the experiment. Even more so if you use a recording of a known performer, since the possibility of getting in touch with the performer might be difficult and asking questions about specific movements in a particular show can give the same ambiguous answer as explained in Davidson’s (2001) experiment. This type of experiment (movement analyses of live recordings of famous musicians), has been used a lot in this field of study (Davidson, 2001; Kurosawa & Davidson, 2005). In these

---

11 Davidson (2001) also used a video recording of herself performing in a preliminary study.
experiments, one or two people competent with the scientific terms and analysing tools from the field of musicology has analysed the nonverbal behaviour of recorded artists, but none of these experiments explains whether or not the scientists interpreted the artist’s expression in a way that aligns with the artist’s intention. Some movements are universal and simple to analyse – such as smiling, bowing, clapping and so on – but others may be diverse in their meaning. I wonder how it is different for one person to project his/her meaning on to other peoples’ body language than for one person to attach meaning to how his/her body language is experienced by others. If both persons strive to be objective and possess applied knowledge about the theoretical foundation of both receiving and expressing emotions, should not both approaches be equally valid? It is thus important for me to point out that this study is only analysed by one person, which means that I rely on self-reports both in my own performance and how I interpret the answers from the participants, as for interpreting the participants’ comments on the videos. “The video itself is an interpretation (normally first-order) of the track, normally suggesting an interpretation to which the listener can then respond, that is, offering (and sometimes strongly) a particular subject position” (Moore, 2012, p. 164). This means, that when relying on introspective reports from listeners, we rely on subjective opinions that is not our own for empirical findings, and is it then ever possible to say that a conclusion based on human interaction is objective?

2.3 Music and emotions

The research field for this thesis is emotional expression in music performance, and it therefore depends on theories on music and emotion. There are different theoretical approaches to this issue as “We may simply perceive emotions in the music, or we may actually feel emotions in response to the music” (Patrik N Juslin, 2005, p. 91). The focus of this experiment is how the emotive state of the vocal performer is perceived, not embodied or felt, and I will therefore lay out some groundwork on how we perceive emotions in music, as well as how emotions in music is expressed by a performer, both audibly and visually. Since I am basing my thesis on the emotive state of sadness and anger, which are both categorized as basic emotions, I find it necessary to explain some of the theories of these emotions.

2.3.1 The basic emotions sadness and anger

While the term emotion is hard to describe, researchers are more in consensus about what defines a basic emotion. Basic emotions can be said to hold neurobiological substrates that
has been developed through evolution (Kalawski, 2010), and some also assert that the basic emotions are the basis “from which all other emotions may be derived” (Patrik N Juslin & Laukka, 2003, p. 771). This is only a hypothesis and is so far not proven right, but, what is scientifically discovered is the multi-cultural recognition of facial expression of these basic emotions. In 1972, Ekman (1999b) found evidence that people from different cultures recognize six different emotions just by watching facial expressions. The emotions that became known as basic emotions were: “happiness, anger, disgust, sadness, and fear/surprise” (Ekman, 1999b, p. 316). Some of the basic emotions have some similar features when applied to other areas of research, and anger and happiness are in several studies of movement found quite similar (Dahl & Friberg, 2007). Since anger and happiness are found similar, I chose to focus on anger, because I found it easier to express in the song used. (This is a subjective point of view, and it might be that if I had chosen happiness it would have given different results, but this is only speculations and will require further studies.) Kalawski (2010) argued for implementing the emotion of tenderness in the basic emotion category which means that the debate about which emotions that can be included in the “basic emotion category” is not over. Still, both sadness and anger are confirmed as basic emotions by several different experiments conducted over time and are therefore internationally recognizable by facial expressions.

2.3.2 Communicating emotions in music performance

At some age, we learn to interpret feelings by both the sound, facial expression and body language that is stereotypical for this particular emotion, which means that our reading of emotive states is a multimodal impression (confer sections 2.1.1 and 2.1.2). When it comes to emotions in music, a lot of factors are combined for conveying expressions, and “All musical expression of emotion is conveyed in terms of pitch, intensity, duration, and extensity, which are the four elemental attributes of all sound” (Seashore, 1923, p. 323). Patrik N Juslin (2001) also explains that musical emotions are expressed by different melodies. In addition to this, the tempo of the song “has been shown to be the most potent structural factor in music to determine the type of result emotional response” (Persson, 2001, p. 283), which means that all the musical components of a song are part of creating an “emotive image”. If we look at

---

12 In the international findings there were cultural differences about the expression of fear and surprise, but there was consensus within the same culture (Ekman, 1999b).
13 Seashore talks here about instrumental music, but as I will argue later on, this is applicable for the vocal expression and performance as well.
the actions of creating the different sounds, it is quite clear that our multimodal way of interpreting emotions in facial expression and body language of human beings also is crucial for interpreting emotions in music. If the bpm/tempo increases, we naturally assume that the action of playing in this tempo increases as well (our body language and gestures are more rapid), and if the sound is low in volume, we associate it with the action of creating this sound (which can be to whisper or use smaller movements etc.). There are several studies on emotions in music and how we detect emotions in music, as well as which emotions are detected. In one particular study, explained by Juslin (2005), five composers were told to write short melodies that was conveying one emotion each: “joy, sorrow, excitement, dullness, anger, and peace” (p.94). These melodies were then played to 14 listeners who successfully recognized the intended emotive states in the melodies.\textsuperscript{14} He also explains that, according to several studies, “professional performers are able to communicate five emotions (happiness, anger, sadness, fear, tenderness) to listeners with an accuracy approximately as high as in facial and vocal expression of emotions” (Patrik N Juslin, 2005, p. 94). All of these (basic) emotions “have distinct expressive characteristics in other non-verbal communication channels; notably, the non-verbal aspects of speech and human body movement” (Patrik N Juslin, 2005, p. 100), which might be why they are easily detected and separated. Thompson et al. (2008) explains that visual ques in music performances “can augment or attenuate the perceived emotional tension of performances” (p. 2), and several studies have found that “any musical message or performer feelings will not be received by the audience unless they are manifested as acoustic properties and, in the case of a live performance, visual cues” (Woody & McPherson, 2010, p. 411). But, for the visual cues to be comprehended in a way that aligns with the performer’s intention, the performer needs to be aware of his/her use of body language. According to Dibben (2009), our body movements can reveal some of our inner emotional state in a performance and this state “depends on an interpretation of “adaptor” movements as “emotional leakage”” (p. 323). Dibben (2009) further explains physical “adaptors” as movements and gestures that people takes on in different emotional states often without being aware of it. This means, that when performers use adaptors on stage or in public appearances, people tend to “use the information from adaptors gestures to make judgements about the emotional and physical state of the speaker and therefore offer a means of investigating non-verbal communication of performer identity” (Dibben, 2009, p. 323). This concurs with the theory of amodal perception and mirror neurons (as explained in

\textsuperscript{14} In this experiment there was only audible music, and not the visual aspect to focus on.
chapter 2.1.1) and can easily be applied to music performance. An example of this might be if the performer is nervous. In this situation, his/her body language will leak this information to the audience, as the tension in their body language are perceived by the spectators’ mirror neurons and will affect the audience’s anticipation of the performance. (I will argue that this is a common socio-cultural way of adapting to any social situation. We all have a repertoire of gestures that is used in different emotive states, but this might be clearer when it comes to musicians, actors, dancers and public speakers.) Woody and McPherson (2010) explains a technique called visualization, where you mentally dwell on an image or memory of a certain mood or situation for self-induction of emotions. This technique is also used by public speakers and in meditation, and by actors it is known as method acting. My experience is that if a song is rehearsed in an exaggerated manner, possibly based on a situation in extreme emotive state, some of the movements or communicative gestures rehearsed will become a part of your natural gestural repertoire (Davidson, 2001) and will not seem as enlarged and ungainly as they might be experienced from the performer’s perspective in the first place. This exaggerated body language might be easier to execute in some emotions than others. Camurri et al. (2003) did an experiment on expressive gestures in performance where they compared the spectators’ feedback on the multimodal recording of dancers and musicians’ performances. It seems that in their experiment, the participants showed highest accuracy in detecting anger and grief, and there was a lower accuracy on detecting fear and joy. One of the reasons for this might be explained by the movements associated with the two different moods. In Dahl and Friberg (2007) study with three recorded musicians they found that “Slow and smooth movements were associated with sad performances for all three recorded musicians. Anger was primarily associated with jerky movements” (presented in Dahl et al., 2010, p. 51). In addition to moving slow, some signs of sad body language might be; Face looking down, hard to establish eye contact and high shoulders tugged in towards the chest (protective). The signs of angry body language are in complete opposite; “In all “negative” moods of anxiety, jealousy, rage and any physical discomfort the system of the extensor muscles takes the upper hand” (Wolff, 1948, p. 8). This means, that when we get angry, we tend to stand up straight exposing our chest and straighten our arms and back, in addition to looking up and move in rapid, fast movements, as explained by Dahl et al. (2010). Woody and McPherson (2010) concludes that the most
important way of expressing emotions are through eye contact (and bodily gestures), which might not be as farfetched based on how we use eye-contact in completely different ways in these two emotive states (anger and sadness). But, there are several other factors that implies our mental state and which mood we are in. In several studies they have found how emotions have different characteristics when conveyed by the voice (Scherer, 1986), and that the timbre of the voice is found as a key holder to the emotive meaning of the message (Seashore, 1923). Both intensity and duration (as mentioned by Seashore [1923] earlier) are related to the vocal quality of speech, as for sung music. Juslin & Laukka presented in 2001 a chart over emotion characteristics found in music. Their study only examined instrumental music, but some of the findings can be translated to vocal expression. Musical features of anger are among others: fast tempo, high sound level, sharp timbre, and contrast between timing and notes (Patrik N Juslin, 2005). Gabrielsson and Juslin (1996) explains “that “sharpness” and “roughness” indicate “anger”” (p. 85) in addition to the notes being accentuated or “dotted”.\textsuperscript{15} Patrik N Juslin and Laukka (2003) found that in music that was believed to be sad, there was (amongst other factors) slow tempo, low sound level and slow vibrato.\textsuperscript{16} Gabrielsson and Juslin (1996) found that “slower onsets, legato articulation and subdued timbre” (p. 71) in addition to the features already mentioned, was concurring with peoples’ subjective impression of the emotive state sadness in music. The tempo often affects the duration of notes, and intensity can be found in both the onset and volume of tones, which leads to different emotive expressions (which is coherent with the findings of Persson, 2001). According to these studies by Juslin and Laukka (2003, 2005) the vibrato is used quite different when expressing anger and sadness and can be explained by how we react physically to different emotions. Wolff (1948) explains how Charles Bell, an English physiologist, found that our organs was affected when we felt emotions, as “When the heart is affected breathing is altered too. The respiratory and circulatory systems are inseparable in their functions, and a strong emotion or feeling produces immediate repercussions on the rhythm of breathing and on the voice” (p. 21). The ability to react to emotions can be explained by both our autonomic nervous system

\textsuperscript{15} Gabrielsson & Juslin only talk about the audible performance, and there is no description of the visual performance, or whether or not the participants saw visual clips of the performances which differs from my experimental setting.

\textsuperscript{16} For further description of musical features, see Patrik N Juslin (2005).
(ANS) and our somatic nervous system (SNS) (Scherer, 1986). The ANS controls the muscles in our inner organs, heart and our glands, while the SNS controls the muscles connected to our skeleton and our sensory experiences (Jansen & Glover, n.d.). When we experience emotional responses in our self, both of these systems are at play, but in fluctuate degree. Fear is our most important survival mechanism, and in threatening situations our ANS increases the heart race and diverse the blood flow from our intestines to the heart, before our SNS distribute this to the muscles we need in the given situation (Scherer 1986). The same can be seen when in rage, where our blood pressure increases, and our respiration gets shallow and quick. This also affects our vocal production and

are most pronounced for respiration (producing changes in subglottal pressure) and for the secretion of mucus as well as salivation (resulting in changes of the resonance characteristics of the vocal tract). Because vocalizations are produced by the combined action of a great variety of muscles in the chest, throat, and head, any changes in muscle tonus and mode of functioning will have strong effects on a number of vocal characteristics. (Scherer, 1986, p. 149)

Seashore (1923) believed that the most desirable vocal quality for conveying emotions was the vibrato, and if we look at what vibrato is, he might be right. Vibrato is a pulsating change of pitch, most of the time only with the range of a halftone, created by inconsistent airflow from the lungs, through the larynx that affects the vocal chords. But, it is only some of the emotions that require the qualities of the vibrato, and that is feelings, or moods, connected to the respiratory process. Take crying for an example: When you start tearing up, you produce mucus that makes you lose control of your larynx, and when sobbing you also start breathing uneven, or even hyperventilate. Ergo you lose control of both your breath and your vocal chords, and you get a natural vibrato and will struggle with making sounds. This also affects the volume on your voice, which in most cases will be at a low point, which also happens when we become scared, nervous etc. This means that when conveying emotions in music, a lot of factors are at play, and all of them needs to be consistent and coordinated for conveying the emotional message in a way that aligns with how we experience the basic emotions in human beings.

2.3.3 Music and the Circumplex Model of Affect

17 Do not mistake crying in sadness with crying in anger. These are two very different states of mind where the ANS and SNS of the body works in different ways (as mentioned earlier).
As pointed out in the latter section, music derives from motion, and without motion there cannot be music. Motion, on the other hand, might be explained as energy, either pleasant or unpleasant. Ekman’s international study on facial expressions started with a thought about how emotions could be placed in scales of pleasant-unpleasant and active-passive (Ekman, 1999a). Ekman has since this taken back this statement, but his groundwork inspired other scientists (Russel, 1980, among others). The theories on pleasantness/unpleasantness in music and sound is an idea already proposed by Kate Hevner as early as the 1930’s. Hevner studied the elements of expression in music (Hevner, 1936) and affects in poetry and its sound (Hevner, 1937). She also arranged 66 mood related adjectives in a circular model, as Russel (1980). Hevner’s experiment with music focused on the affective value of the major and minor chords, and the importance of rhythm when inducing affects, and she found that difference in rhythms, tempo and harmonies lead to different mood inducements (Hevner, 1936).

Russel proposed in 1980 the “circumplex model of affect” (CMA) (figure 5) which was based on a similar idea that Ekman had. He lined up 28 different affects and placed them in a circle with two axes; the one going from north to south is arousal- sleepiness (energy) and from east to west, pleasure- misery. Russel explained his axes and stated “that any affect word could be defined as some combination of the pleasure and arousal components” (Russel, 1980, p. 1163). All of the emotions used by Russel can be explained as having a strong bodily connection (like tensed, distressed, exited, pleased), whereas he leaves out some of the more dualistic emotions (e.g. nostalgia, thrills, impatient, stress). These dualistic emotions may be placed in opposite ends of both the axes of arousal- sleepiness (like nostalgia and thrills) or pleasure- misery (like impatient, thrills, nostalgia, stress).

Figure 5 The circumplex model of affect (Russel, 1980, p. 1167)
2.3.4 GEMS; emotions in response to music

One study that tried to find a descriptive language for emotions felt in response to music is the GEMS study by Zentner (2010). In comparison to Russel’s circumplex model of affect which only uses 28 different affects, Zenter begins his study by finding 515 different terms for affects, moods and feelings. He then asked 252 participants to rate the terms according to their frequency of appearance when listening to their choice of music, and “this process left us with 66 terms relating to emotions which are commonly felt in response to music” (Zentner, 2010, p. 104) (figure 6). The highest ranked terms of affection evoked by music was then named “the GEMS model” (Geneva Emotional Music Scale, since the experiment was conducted at a festival in Geneva). This model found nine primary music evoked emotions and “these music emotions differ markedly from those defined by basic emotion theory” (Zentner, 2010, p. 109).
In the GEMS experiment, Zentner (2010) reports that among the 801 festival attendants that participated in the experiment fewer than a quarter “reported having felt any significant degree of emotion” (p. 110) which is an interesting finding. My thoughts are that Zentner’s findings are especially useful (and maybe confusing) in stating that music itself is not the only factor when it comes to convey and evoke emotions in the listener in a performance, since the performance is a multimodal expression.

Zentner (2010) questions the use of the CMA when it comes to define music evoked emotions. He states that Russel’s model of affect only is usable when it comes to define experienced emotion (by this I believe him to talk about embodied emotions) and is not
capable of differing emotions such as “gratitude, compassion, tenderness and contentment” (Zentner, 2010, p. 103). This might be correct, since Russel’s axe of arousal-sleepiness encourages emotions that can be experienced by physical alterations, in for example breathing and heart rate, which makes them embodied, while Zentner’s GEMS model includes emotions regardless of their bodily connection. Zentner concludes himself that the GEMS model and models of basic emotions are based on different emotions, and different ways of perceiving and explaining them. (Basic emotions are utilitarian, while music evoked emotions often are more aesthetically based.) Even though there are several studies on language used for describing music, as with the GEMS study and Hevner’s (1936) experiment on expressiveness in music, Leech-Wilkinson and Prior (2014) arise the concern about how we do not have a complete technical language when talking about music and its expressiveness. As mentioned, these studies are taking on different approaches (embodied emotions evoked by music and emotions perceived in the music), but it might indicate that music evoked emotions differ from emotions other impressions evokes, as explained by Patrik N. Juslin and Västfjäll (2008, p. 572). What we do know is that we still have not found one model that is applicable for all aspects of musical communication of emotions.

2.4 Movements and gestures to music

According to Jensenius et al. (2010) there are two main categories of musical gestures; the gestures of those who produces the sound (the musicians) and the gestures of those that perceive the sound (the “listeners”, or what Jensenius is also calling “perceivers” to account for the multimodal nature of the musical experience). They also describe musical gestures as “human body movements that goes along with sounding music” (Jensenius et al., 2010, p. 13). These categories can intertwine in a live setting, and the musicians can go from the sound-producing gestures to the sound-perceiving gestures in the same song, when communicating with the audience. This implies that when using the term gesture, there is a conscious thought, or an intended or communicated meaning, behind the movement. The term movement, on the other hand, describes the physical displacement of an object in time and space. Motor patterns are movements that we do automatically or have been trained to do for so long that we don’t need to think about how we do them, like walking, or breathing etc. Jensenius et al. (2010) talks about three different definitions/views on gestures: as control,  

---

18 There are still a lot of terms for physically perceived emotions in the GEMS model, but this is of higher importance in the CMA than in the GEMS model.
communication or metaphor. What is most relevant for my thesis, is musical gestures for communication purpose. Even though I am alone in the video, most gestures are made for communicative purposes, and my gestures are “2) individual interpretations of the narrative or expressive/emotional elements of the music, 3) the performer’s own experiences and behaviours, and 4) the aim to interact with and entertain an audience” (Dahl et al., 2010, pp. 48-49). The gestural category of motor pattern is not as relevant in my video as the one of conscious movement, but there are still a lot of gestures that is a combination of both conscious gestures and motor patterns – like singing. To sing is an action that combines both unconscious bodily movements (taking a deep breath before a long note can give an impression of drama and a high level of skills when exaggerated, but we don’t need to think about how we physically are taking a breath), conscious gestures (clapping your hand to the beat), and combined gestures (holding you hand out when hitting the long note can be a way to both communicate/interact with the audience, create a dramatic effect and a way to subconsciously help the singer intonate and create crescendo for the purpose of managing the long note). When it comes to different movements made by the sound producers (musicians) they are divided into four categories: sound-producing, communicative, sound-facilitating, and sound-accompanying (Jensenius et al., 2010).19 (Further explanation of these categories will be done in section 3.3.4.) All movements and gestures made by the sound producers can be seen in the light of these categories (that is both the conscious gestures made, and the movements that can be seen in the light of motor pattern). These terms from Jensenius et al. (2010) (also used by Dahl et al. (2010) in the same book) are based on an experiment with instrumentalists, but Kurosawa and Davidson (2005) designed their own terms based on gestures used in vocal performance: emblem, illustrators, affect displays, posture, touch, regulators, adaptors, facial expression and gaze. These terms are more detailed than the ones given from Jensenius et al. (2010) and can be used as a supplement when doing in depth analysis. (Further explanation will be given in section 3.3.4.) Jensenius et al. (2010) also uses the term “gesture space”, which is the space where most gestures are executed. The viewer will anticipate the performer not to move too far away from the gesture space, and it creates a natural focus of attention. This gesture space can help with defining the gestures made during the performance, since all sound-producing and facilitating gestures will happen in front of the microphone, and we also have a measure point for the sound-accompanying gestures.

19 It should be noted that the terms sound-producing and sound-accompanying were first coined by Godøy (2009), but I have chosen to use the classification from the overview article by Jensenius et al. since this is the most pertinent to my experiment.
Davidson (2005) found in one of her studies that when told to play expressive, or just being free to express himself, the pianist in the study moved his torso and upper body a lot more than when playing expressionless. She explains that “it could be that the pianist’s waist region functions as the central physical core for the musical expression, for that is the point from which the swaying action emanates” (Davidson, 2005, p. 219), and explains this as the “centre of moment” (p. 220). This centre of moment can be seen in a lot of instrumentalists and singers. Auslander (2009) describes a comment from a musicologist trying to analyse a guitar solo of B. B. King: His colleague meant that the way King bent his back on high notes was just something most guitarists do and did not have an additional meaning to it (Auslander, 2009, p. 310). Dahl et al. (2010) describes a study on trumpet players, where they found that the trumpet players bent their knee more often when hitting a high F than a low C (p. 57). They explained this with the need for more support when playing high notes than the low notes and described the gesture as related to the respiratory process. For a vocalist, bending your back when belting (singing high pitched notes in your chest voice) might give you some extra support from your diaphragm and abdomen, as with trumpet players, and is, therefore a physical position to help you sing. This said, it’s not a necessary movement for reaching the notes if your energy level is balanced and you have good vocal technique, but it can make the high notes look even more difficult and impressive. This means that it can be a movement for technical purposes, but when a guitarist does it, it is more likely to be of communicative and expressive purposes. I’d say that this action of “bending your body” probably derived from the action of screaming (noticed how angry people tend to bend over forward a bit when yelling?) that developed into a natural movement for both volume increases’ and showing involvement and expression of emotions. This means that the centre of movement in general can be found in the torso, especially for the instrumentalists who depends on respiration for producing sound. When it comes to dance movements they can also be seen as a “graph of gestures” (Wolff, 1948, p. 5). Davidson (2005) explains that dance movements “functions for the body to say something, with music performers using dance to draw attention to themselves as a means of “display”” (p. 228), which means that dancing are of primarily communicative and expressive purposes and not used for sound creating. This means that there are several different categories for moving to music that includes everything from barely noticeable sound-producing gestures to communicative and expressive dance movements.
2.5 Definitions employed in the thesis

**Performer**

I use the term *performer* about my role in the video, since I find that it holds information about what my role on stage (in this case on the video) is, which is to communicate with both my voice and my body. If I were to use the term *singer*, it would not have been sufficient for analysing the data from the survey in the correct way for this thesis. The term *singer* is focusing more on the audible sound than *performer*, and since my main focus is on the visual, and not audible performance, I find that *singer* gives incorrect information to the reader about the analysis. I do use the term *singer* sometimes, but this is only when referring to the participants description of me in the video, or when referring to vocal techniques. (Most participants used the term *singer*, which is quite natural in the Norwegian language.) I also use a third-party description of my own performance and find that the term *performer* is applicable for this.

**Communication**

The definition of *communication* in this thesis is a transmission of any kind of message, intended or unintended, perceived by a person who is not the sender of the message. This means that communication is always to be seen as a social act, either with the intension of social transmission or for personal use/expression. The definition of good communication is a transmission of an intended or unintended message conveyed, that is interpreted by a recipient in a way that aligns with the sender’s intention or emotional state.

**Participant**

The participant has a major role in the study, and I choose to use the term *participant* instead of the *listener* or *spectator*. Since this is a study involving both video and sound, I find it that the term *listener* and *spectator* comes to short, since we are talking about receiving a multimodal message that includes both these senses. It would also be less accurate to use the term *receiver*, since communication in this thesis is defined as something that involves two or more parts. The communication in this study will go both ways, even though the audience is being performed *to*, and are not actively participating in the performance. It consists both of what I, as the performer, want to communicate, and what the participant believes me to communicate, and in that way, the participant is an active part of the communication chain in this thesis.
Emotion
I choose to use the word *emotion*, and not feeling. This is because my understanding of emotion is that this is something universal, and more *overall* than a feeling, which can vary from person to person and be hard to describe. There is no consensus about how we should define the term *emotion* (Izard, 2010; Zentner, 2010), but a study described in Zentner (2010) came to the broad definition that emotions contained of several “psychological and behavioural manifestations: in addition to subjective feeling, they also involve action tendencies, physiological arousal, cognitive appraisals and expressive motor behavior” (p. 118). This means that the term emotion is not just an emotive thought, but something which is embodied and expressed. I use the term emotion when subjectively explaining my performer intention, and when talking about how the participant interpret my emotive intention. What is important to note is that I do not talk about the embodied emotions of the participants, what the participant *feel*, since I have no way of knowing this. I also focus on the emotion’s sadness and anger, and I use these terms as unifying terms for all the feelings it can contain (depression, anxiety, vulnerability etc. is all related to sadness).

Body language
The body language is of particular importance in this thesis and is to be seen as an umbrella category to include all movements, gestures, postures and visual expressions made by the performer. It also concern’s both conscious and unconscious movements and gestures, including all categories applied from Jensenius et al. (2010) and Kurosawa and Davidson (2005).

Gesture
The word *gesture* comes from the Latin *gestus* that means pose or posture and even attitude, and are in a historical perspective used to describe hand gestures for conveying or communicate certain feelings (Schneider, 2010). I use the term *gesture* when talking about an isolated movement, like one arm moving or the leg, and I only use the term gesture when describing a physical action, never about the music itself.\(^{20}\) In my definition of this term, I separate the communicative gestures into two categories; the conscious and the unconscious.

\(^{20}\) It is important to note that the definition of gesture in this thesis differs from Jensenius et al. (2010) usage of the term and are not to be compared.
gestures. Conscious gestures can be waving hello, lifting your hand to underline lyrics that describes a word like “up” or “heaven” or substantiate the rhythmical structure and are described as illustrators, communicative, sound-accompanying, posture, and emblem. Unconscious gestures are the ones that you do without thinking of them. In this thesis, they are described as illustrators, communicative, affect display, sound-accompanying and sound facilitating gestures, and can be seen in the light of Davidson (2001) theory on gestural repertoire.

**Movement**

The term *movement* is used if more than one body part is involved and can be both conscious and unconscious. It is also exclusively used when the torso is moving, as the torso can be seen as the centre of moment and will move or affect other limbs. Movements can be illustrators, sound-accompanying, emblem, sound facilitating, affect display, sound-producing, communicative and touch. All movements, both conscious and unconscious, can be seen in the light of these categories and will be further explained when the nuances in ulterior motive is necessary.
3 Method

There are different ways to approach music performance research, and the complexity of this thesis requires different methods. In the first two sections I will discuss phenomenological (section 3.1) and empirical (section 3.2) issues pertaining to the experiments. The next section (3.3) goes in depth in explaining the methods applied in the experiment. For validation reasons, it is necessary to explain how the music and videos used in the experiment was conducted (3.3.1 and 3.3.2) and the author´s intention with the performance (3.3.3). I will also go in depth on some of the theories explained in the latter chapter (movements and gestures to music, section 2.4), which will serve a methodological purpose (3.3.4). The remaining sections (3.3.5 to 3.3.8) are explaining the procedure of planning and conducting the experiment and collecting the data.

3.1 Phenomenological issues

Phenomenology, broadly speaking, is the study of how events are immediately experienced by our senses (Johannessen, Tufte, & Christoffersen, 2010), an infinite task without any right or wrong answer. In phenomenological analysis, one interprets the data with the aim of reaching a more profound understanding of peoples´ thoughts and opinions on different issues (Johannessen et al., 2010). Robinson, Watkins, and Harmon-Jones (2013) explains that:

> Our intrapsychic lives are dominated by two sorts of phenomena; thoughts (more formally, cognition) and feelings (more formally, emotion). Both are internal events that cannot be directly observed by others and, in this important sense, are subjective or at least particular to a person. (p. 3)

In this thesis, I have analysed both my own thoughts and feelings, in addition to the participant´s thoughts, which makes it hard, even impossible, to reach a unanimous conclusion. In addition to the complexity of the experiments, the participants in the experiment have a large variety of experience in both listening and expressing themselves about music. It is therefore not possible for me to find out why they answer the way they do, since “Selv den mest nøyaktige iakttakelse og gjengivelse kan aldri fange inn den autentiske virkeligheten, men kun et utsnitt av det som studeres” (Johannessen et al., 2010, p. 36). This means that phenomenological analysis does not arrive on one right answer, but multiple excerpts of a complex and subjective reality.

---

21 In recent years the dichotomy on these internal events are criticized; is it possible to distinguish feelings from thoughts, or vica versa?
3.2 Empirical musicology

Empirical research is a way of gaining knowledge through observation and experience and Cook and Clarke (2004) state that the line between empirical and non-empirical musicology is hard to find since there “can be no such thing as a truly non-empirical musicology” (p. 3). By this they mean that the whole field of musicology are shaped by empirical research, along with our own and the societies thoughts and previous knowledge of different subjects. Cook and Clarke (2004) further explain that “if an observation is to be regarded as trustworthy, it should be possible to make it on different occasions, and it should be possible for different people to make it” (p. 4). This does not hold true for qualitative research, that often seeks the subjective and singular opinions. In my thesis I have a limited variety of participants, both in age, nationality and background, and the results can therefore not be applied to all. My perspective is still empirical, since the thesis is built on an experiment based on observation and collecting data, but there are different empirical methods for analyzing musical performance; one is to use several recordings of a performance and compare the uniqueness and similarities in them, and the other is to look at one single performance (Clarke, 2004, pp. 84-85). Both approaches can be conducted with or without participants, but either way it is an empirical approach. I use the first approach in this thesis, since I use two videos, comparing the participant’s answers to my intensions.

3.3 Applied methods; limitations and usage

The process of creating the videos and audio tracks for the experiment started in February 2017. First the backing track was recorded, then the videos and finally the vocals. The questionnaire was in the making from fall 2016 to June 2017, before the first participants completed the experiment mid-August 2017, and the last ones in the end of September 2017. The reason for this time span in executing the experiment was that it was a lot harder to find participants then expected, and I had to adapt to the student’s schedule.

3.3.1 The song

I have chosen to use one of my own songs for several reasons, and the first is that it is unfamiliar to the participants. If I were to use a known song, the result may have differed between the people who knew the song and those who did not, and that might confound the validity of the data. The song used in the experiment, Stronger, is the only song where I feel
that my emotive expression in the performance may vary, depending on the visualisation conducted before the concert/performance, in addition to my personal intention at that time and place. This is the second reason why I chose to use this song, since my interpretation of it can change from one emotion to another which makes it easy for me to do the two different versions of it (sad and angry). The song is performed in 128 bpm, which is the original tempo. This tempo helps me manage to move in different ways, since it can be perceived as middle fast tempo when counting the triplets or fourth notes underlying the backing track, and fast when counting the sixteenths. It was therefore important for me, when choosing a song, that its tempo also could fit into different emotive expressions (given the theory of Persson [2001] about how the tempo of the song is the most important structural factor for determining the emotional response). As explained in section 2.3.2, the melody is also an important holder of emotions in music, but “if different emotions are expressed by different melodies, it is impossible to know whether the obtained effects on listener judgments or performance measures are due to the melody, the performance, or some complex interaction between the two” (Patrik N Juslin, 2001, p. 313). Since this experiment is based on the same melody, performed in two different emotive states, it excludes some of the possibilities of perceiving different emotions in the melody, but this does not mean that the participants cannot find the melody to be mood creating in one way or another. As mentioned in the introduction, this thesis focuses upon meaning as conveyed through body language, and I don’t find it necessary for the purpose of this thesis to analyse the song in depth, but I will give an overview of the structure of the song in section 4.1, so that the experiment is comprehensible to the reader. There are some places where I am explaining the phrase or melody in depth, but this is only where I find a correlation between the answers given by the participants, and the structure of the song, which will be done in chapter 4 and 5. I also chose not to include the lyrics of the song in this part of the thesis, since it is to be found in the appendix and will be explained when and where I find it to be relevant to the findings. But, as there has been some thoughts on whether or not to use lyrics in the song, I need to explain why I chose to include it. The purpose of this experiment is to become a better singer (focusing mainly on popular music), and for a singer of popular music it is unnatural to sing without words. I also wanted to find out if there was any correlation between the words sung

---

22 This said, it might be that the outcome of the participants’ comments would have been different if the tempo was different in the two videos, but then it would have been two different songs performed, which does not align with the main purpose of this thesis.
and the descriptive words used by the participants which, of course, brings in a whole new perspective of meaning.

The backing track was recorded by guitarist and producer Joakim Arnt Holmen in Whiteroom Studio, Asker, during the early spring of 2017. The song is originally four minutes long but was shortened down to an AABCADBC form, lasting 2 minutes and 18 seconds, for the purpose of keeping the experiment short and to the point. The vocals were recorded at Earport Studio at Fornebu, by the help of producer Are Næsset, after the video clips were chosen. (The reason I chose to do it in this order is for the vocals to be as authentic as possible, since it is easier to “dub” a video than to do it the other way around.) Normally, when recording pop vocals, one does several vocal takes on shorter snippets and cuts together the best phrases to create a perfect track. In this case, since I wanted a track that simulated a live performance, we recorded the whole song with the different moods a couple of times and chose the one that we believed to convey most of the given emotion. This means that there are some minor pitch differences a couple of places, but I consider this as a part of a “live” setting. Both the volume, placement of the voice and the use of effects has an effect on how the listeners receive the voice, and if they find it authentic or not, “because they specify a location and physical relationship between listener and sound source” (Dibben, 2009, p. 320). The vocals are therefore placed semi-close (in the middle of the mix) and with some light reverb on. This was done to simulate the room in the video, and the vocals were kept quite dry and unprocessed to keep it as close to an acoustic performance as possible. I have applied the mood characteristics described in section 2.3.2 to the vocal performances, which means that there are large differences in volume, onsets, articulation and timbre in the two song’s used, as to convey the two different emotive states in the performances.

3.3.2 The videos
The videos were filmed in the motion capture lab at the Department of Musicology, using a green-screen and a stationary camera. I had help from a colleague filming, Kristoffer Lislegaard, a former music student who is now working fulltime with both video and audio, so I would not have the stress of running back and forth to the camera, which might collude my performance. Since I didn’t know who my participants were beforehand, I had no way of knowing what they would focus on in this experiment and I had to take precautions. What

\[23 \text{ I chose to confer with the producer, Are Næsset, for an intersubjective opinion on the expression.}\]
surrounds us is significant for what we chose to focus on and not (Zeiner-Henriksen [2010]), and “as human beings we are, inevitably, always involved in the interpretation of our environment” (Moore 2013, p. 57). This means that our field of focus gets affected by our surrounding environment, our mood and the channel in which the music is presented, and all of these factors might change how you receive and interpret the music. If one of these factors were wrong, the chain of emotion would be broken, and the emotion would not reach the listener (Zentner 2010). Therefore, I had to exclude all elements that could disturb the participant, as done in the study by Davidson (1993) (confer page 16), both in the video and in the environment where the experiment was conducted, and I did so according to my subjective aesthetical interpretation of how I wanted to present myself. My clothes in the video are all black and contoured, selected for the purpose of creating a contrast to the white background, and for the viewer to be able to see all my gestures and movements in detail.\(^{24}\) If I had used clothes that covered up my body, it would have been difficult to see the nuances of the gestures. My hair is tied in the back for the same purpose. I also use a microphone stand for the purpose of making it a bit more similar to a concert performance and to create a gesture space (Jensenius et al., 2010).

After filming three videos (with one angry, one sad and one neutral intention) the background was made white and we created two different styles; one with the face showing, and one where you only see the body as a black shape (Figure 7). In their study of expressiveness in performance, Dahl and Friberg (2007) found that the visual image of the head was particularly important when it came to detect the emotion sadness in the performance, and after consulting with my supervisor, we decided to use the coloured videos where the head and face are showing, which is also closer to a live setting than only the contoured shape.\(^{25}\) The videos are each two minutes and eighteen seconds long, using the same guitar track. When the vocals were done, the completed track was added on top of the three different videos by Lislegaard. A decision

\(^{24}\) This appears to be a convention, and for instance Camurri et al. (2003) also showed their dancers in their experiment in black clothing against a white board.

\(^{25}\) Since this thesis is not focusing on facial expression, it might have been interesting to use the black contoured video, which might be interesting for further research.
was then made to use only the sad and the angry videos, with emotively matching vocals, since the neutral one was so called “expressionless”, and its relevance and authenticity were questioned.26

3.3.3 The performance/performer intent

For the reader to fully understand the findings of the experiments, it is important to explain the subjective intentions of the performances. As mentioned in section 2.4, most gestures and movements in the videos are made for communicative purposes. The body language in the videos are a combination of motor pattern27, unconscious and conscious gestures, but most of the visible gestures and movements are on some level meant for a communicative purpose. Most gestures are on some level rehearsed, but, the level of consciousness behind the movements vary and I have adapted the term gesture repertoire, as suggested by Davidson (2001), for the movements and gestures that is unconscious executed in the performance and have become a part of my gestural repertoire in different moods – like the act of lifting the shoulders when conveying something heartfelt – both as a person and a performer. I also took the liberty of adding natural to the term during the video analyses, as I find that the natural gesture repertoire offers a description of the gestures included in this repertoire as unconsciously developed, in comparison to gesture repertoire, that might be deliberately taken on. (An example of this type of deliberate gesture repertoire are the gesture of pointing up at top notes.) These types of deliberate communicative gestures have been adapted during years of rehearsing and experimenting within the performer/audience relation, and by using techniques of visualisation and method acting, which I often use in my everyday work as a singer. I use these techniques frequently when rehearsing new repertoire, but in a normal performance setting, I leave these thoughts behind and just focus on enjoying myself and interact with the audience. I still find that the more I rehearse a movement, the easier it is to implement in a live setting (the natural gesture repertoire is increased). According to Dibben (2014) a lot of studies “indicate that expression in human performance is often unavoidable, even when performers try to play without expression” (p. 126), so I chose to exaggerate my movements. This is done to convey the two different emotive narratives of the song (anger

---

26 This said, it could have led to different answers from the participants, but it does not align with a study of live performance.

27 The only movements that can be seen as motor pattern are the ones of breathing, and intonating, which might not be visible for the spectator since one cannot see the vocal chords. But this might still have an effect on how I unconsciously use some gestures, like sound facilitating.
and sadness) in a way that aligns with my natural gestural repertoire in different situations. This because the videos strive to be as authentic as possible, and for this reason I also move during the whole video, but in different ways adjusted for the different emotive states. (You do not see many artists standing completely still on stage, and I chose to simulate this by “digging the beat”. When performing an outgoing song on stage, I normally move more around and try to get eye contact with the audience, which is hard to do in a video setting. But I will still argue that my body language and movements are close to the movements I make naturally based on how I act in different states of mind. The character of the video is therefore still a part of both me as a person and me as a performer.

3.3.4 Movements in the videos

When it comes to analysing the body language in the two videos, I have applied the categories of Kurosawa and Davidson (2005) and Jensenius et al. (2010) (as explained in section 1.1) that aligns with my usage. I use Jensenius et al.’s (2010) four categories as principal categories, and the terms from Kurosawa and Davidson (2005) as sub categories for further explanation of the gestures and movements. When starting to analyse the videos, I found that the categories of Jensenius et al. (2010) (when used alone) where much too vide to give a profound description of the gestures and movements. I then decided to further separate the gestures, and instead of creating new categories for this thesis I decided to use the terms already established by Kurosawa and Davidson. This was done both for the simplicity of not creating new terms, but also as an experiment of its own to see if they were able to throw light upon a first-person analysis. I also wanted to find out if the categories of both Jensenius et al. (2010) and Kurosawa and Davidson (2005) could supplement each other and bring new insight to the analysis of the movements. I will present the original definitions of the terms, along with my own usage of them, before explaining how they are combined in the analysis.

Terms from Jensenius et al. (2010)

*Sound-producing gestures* are the “body movements necessary for producing sound” (Godøy, 2010, p. 110). For a vocalist, this includes the whole torso and the head since this is where the sound resonates, and not just the vocal chords that is located in the throat. However, since

---

28 The analysis of Kurosawa and Davidson are third-person analysis of gestures of famous performers, analysed through video clips without the insight of whether or not it is a deliberate gesture made.
one can open the mouth and breathe without making any sound, we cannot see the sound-producing gestures made by a singer. This category is therefore ineffective when it comes to analysing song performance, and it will therefore not be used in the analysis of the videos.

*Communicative gestures* are all movements that are meant to communicate either between the performer and co-performers, the audience or as an internal communicative dialogue. In my opinion, internal communication is often related to the intended external expression and can be based on the song character or the artist’s persona. (Example: when applying a new song technique, one needs to focus on correct physical positions and facial expression for conveying the intended audible expression, which can be a part of the performer’s song character or persona.) The communicative body language and gestures can vary from song to song, or within the song’s narrative, building on the artist’s image (how they wish to appear both for themselves and the audience which includes both the visual body image and the auditive image). The same communicative gestures can therefore be both internal and external.

*Sound-facilitating gestures* includes the preparatory gestures such as taking a deep breath before a long note, opening the mouth to the right position of the technique required in the particular phrase, and any movements made to avoid fatigue or discomfort, including the ones of facilitating the sound. In my case, this can also involve the whole torso, since the correct posture is required for accomplishing different vocal techniques. When analysing a song performance, the line between sound-producing and sound-facilitating gestures can be even harder to draw than it is for instrumentalists. When a guitarist strikes the chord and uses the vibrato arm, the whole arm moving is a sound-producing gesture, but if she/he move the arm more than necessary it is a combination of sound-producing, sound-facilitating and communicative gesture. This might be easier to detect with an instrumentalist than with a singer since the instrument are a physical object visible with all its details.

*Sound-accompanying gestures* “meaning all kinds of body movements that may be made to music, but which are strictly speaking not necessary to produce the sound, such as dancing, marching, gesticulating, nodding the head, and so on” (Godøy, 2010, p. 110).
Terms from Kurosawa and Davidson (2005)

*Emblem* is “a symbolic body movement that can be replaced with a word” (Kurosawa & Davidson, 2005, p. 115). An example of an emblematic gesture can be when the performers sings the word *up* followed by the artist raising his/her hand up in the air, or the word *heart* can be substantiated by holding a hand to your heart or putting both hands together to create a heart with your fingers.

*Illustrators* can be said to “illustrate the spoken message content, inflection and loudness, or rhythmically accent or trace ideas” (Kurosawa & Davidson, 2005, p. 115). In comparison to the emblematic gestures, the illustrating gestures can be used to substantiate the lyrics, like when the artist put his/her hand on the heart when singing about something heartfelt, but not necessary to substantiate the word *heart*. These types of gestures alone can have several meanings (in contrast to the emblematic gestures), but when put together with the lyrics we understand their purpose in the particular song, or we capture the rhythm or lyrics easier. These gestures can serve several functions as it can be used for the purpose of creating a more expressive performance, but it can also be to help the singer with his/her flow and articulation, in both rhythm and melody.

*Affect displays* are “Facial and bodily expressions which reveal emotional states” (Kurosawa & Davidson, 2005, p. 116). This could be the small gesture of smiling to your audience indicating that you are happy, or a frown when conveying something sad. Kurosawa and Davidson (2005) differentiate between *affect displays* and *facial expressions* but explain that affect displays can be both bodily and facial expressions. Since I am not focusing on facial expression, but on the totality of all the movements I will use the category of affect displays for both the facial and bodily movements that I believe to be emotional states. This is not to say that the facial expression of the performer is redundant in any performance, nor in this thesis, but given the surroundings of this experiment (the video is filmed without audience and presented to the participants on a 13-inches screen) the large facial expressions made in the videos are to be seen as gestures.

*Posture*, in this analysis, is only used when the performer stands still for a certain period of time (minimum one second). The way the artist stands, is an important holder of “information including interpersonal attitudes; liking/disliking, dominant/submissive, and tense/relaxed”
The starting posture of one song can be dominant, while the next can be submissive, and both of these can be aspects of the artist's persona, song character, or even personal self.

*Touch* are explained by Kurosawa and Davidson (2005) as physical human interaction (gestures) that increases the intimacy between the performer and the audience. I want to add to this term the touch of not only human beings but also other physical objects, like a microphone stand. The way you hold a microphone stand can also be a big part of the expression and create a feeling of intimacy or expressivity. If the singer stands close to the microphone (with the whole body), grasping the stand with both hands you would expect the performer to sing something that is of a more vulnerable and introvert expression than if the singer holds the stand with one hand, distancing his/her self from it. I therefore use this category for analysing the performer's relation to the microphone stand.

In addition to these categories, Kurosawa and Davidson (2005) use the terms *adaptors* (that are personal behaviours, or habits) and *regulators* (that “maintain and regulate the flow and content of interactions” [p. 116]). I find these terms hard to qualify from an empirical perspective, as the term of adaptors requires even a higher degree of external observation for validation reasons than other terms (such as posture and illustrators), and for regulators to be present one needs either audience or co-performers. I have therefore excluded them from the analysis. The term *gaze* is also hard to justify in this experiment since it is hard to see the performers eyes on the screen, and the experiment is conducted without audience. This is often the case in music videos, and recordings of music performances.

**Explanation of terms used in the analysis**

For the reader to fully understand my video analyses it is necessary for me to explain how I chose to use the terms (table 1). The first column to the left provides the term from Jensenius et al., the second column the terms from Kurosawa and Davidson that I have chosen to include under the principal category of Jensenius et al., and the last column consists of my explanation for how they are grouped together, which affects how the analysis is conducted.

<table>
<thead>
<tr>
<th>Gestural terms</th>
<th>Application of terms in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Jensenius et al. (2010)</em></td>
<td><em>Kurosawa and Davidson (2005)</em></td>
</tr>
<tr>
<td>Sound-producing</td>
<td>Posture, affect displays</td>
</tr>
<tr>
<td>Communicative</td>
<td>Emblem, illustrators, affect displays, posture, touch</td>
</tr>
<tr>
<td>Sound facilitating</td>
<td>Posture, illustrators, touch</td>
</tr>
<tr>
<td>Sound-accompanying</td>
<td>Illustrators, affect displays</td>
</tr>
</tbody>
</table>
difference between affect display and illustrators for sound-accompanying purpose is a grey area and needs subjective explanation and reflection on the degree of felt emotions (affect display) or rehearsed movement (illustrator).

Table 1 Explanation of terms used in the analysis

<table>
<thead>
<tr>
<th>3.3.5 The questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>When doing empirical experiments, it is important to “know your respondents” (Fink, 1995, p. 10), and this is why I chose to use a questionnaire in addition to the video experiments. If I where only to use the video experiments I would not have any information about who the participants are, and this might collude the data and my interpretation of the answers (eg. Tagg’s (1982) question of whom, section 1.1). The work with the questionnaire went on for eight months, with help from my supervisors, before sending a formal application to the NSD (Norsk Senter for Forskningsdata AS). Some of the original questions were removed (like “how are you feeling today?”, since it is hard to get an honest answer on this, and a question of ethnicity for ethical reasons), and some were added. The result where then seven questions of personal matter, related to the participant’s experience with music, and two for the purpose of improving the experiment, all approved by NSD before starting the experiment in August 2017. The questions were as following:</td>
</tr>
<tr>
<td>1. Gender (Male/Female/Do not want to state)</td>
</tr>
<tr>
<td>2. How many hours do you spend on a weekly basis consuming music? (Five options ranging from 1-5 hours, till 21 hours or more)</td>
</tr>
<tr>
<td>3. How many hours do you spend on a weekly basis producing/writing/rehearsing music? (Five options ranging from 1-5 hours, till 21 hours or more)</td>
</tr>
<tr>
<td>4. Do you conduct any form of physical exercise? (Five options ranging from “No, none” till “daily”)</td>
</tr>
<tr>
<td>5. What are your preferred musical genres? (11 options available)</td>
</tr>
</tbody>
</table>
6. Have you experienced that music was a contributing factor for putting you in a certain mood? (Yes/No/Not sure)

7. Have you experienced that a live concert moved you emotionally? (Yes/No/Not sure)

8. Do you find that the questions in the experiment and the information given beforehand correlate? (Yes/No/Not sure)

9. Do you have anything to add? (Open)

Questions two and three is included because there is a “strong empirical evidence for a direct relationship between the quantity of practice and resulting musical expertise” (Windsor, 2004, p. 200). Since all of my participants for the experiment are music students who weekly rehearse or create music, I have to assume that they are above average qualified to give feedback on the experiment. But this would only have been speculations if I did not ask them how much time they spent rehearsing and consuming music. Question number 4 might seem irrelevant at first, but I decided to include this question because of the hypothesis that people who exercise on a regular basis is more aware of their own and others body language, and if this was right it would have been important for the results of the thesis. Another reason why I found the question of physical exercise important for the thesis, is because the foundation of this thesis is not the audible sound, but the movements. (If I were to conduct this experiment with a group of dancers I believe the comments given on the videos would have been quite different, since their focus is mainly on movement, and secondary on the sound.) The question of the participants preferred musical taste worked as a control list when analyzing the comments, since “it is important to note that a listener’s competense with the style is relevant” (Moore, 2012, p. 166). Friberg and Bisesi (2014) explains that “there are specific performance conventions for different musical genres” (p. 255) and if not familiar with these, the message from the performer to the audience might not get through and can “even vitiate their emotional response” (Zentner 2010, p. 111). If, let’s say a participant where to comment only on the sound recording, or comment the movements in a generally negative way, it will be easier for me to read the comments accurate if I know whether or not the participant is competent with the musical style in the experiment. Questions number 6 and 7 might seem redundant in this experiment (since the participants (music students) obviously is more than average interested in music) but I decided to include them in the questionnaire as an insurance that the experiment was conducted in the best way possible. If all participants said
that music has been a factor for establishing a mood, but none was able to detect any communication from the singer concerning emotional state or emotional message, I would have to assume that my body language, or vocals, in the video were not expressing the intended emotions as preferred. This would have forced me to rehearse and record the whole experiment afresh and would have delayed the whole process. Luckily, this was not the case. The last two questions are included for own observation and learning processes. In addition to the questionnaire, I wrote down if the participant was a bachelor or master student on the registration form with their contact info, for the purpose of easily getting in contact later on (which was approved by all participants).

3.3.6 The software

The software used in this experiment is called VideoAnt and is a tool for video annotation, developed by the University of Minnesota\(^{29}\) where you create a private user profile and upload videos from YouTube. After trying out several annotation programs, recommended by my second supervisor, Alexander Refsum Jensenius, we (Jensenius and I) found that VideoAnt was the easiest to use for the purpose of my experiment, with both me and the participants in mind. After creating a private link for the video on your own profile on YouTube, you simply log onto VideoAnt and paste in the link for your private video. You can add as many videos you like and skip back and forth to them in the video annotation tool. (Figure 8 showing the home page in VideoAnt.) You can even create duplicates of your videos which came in hand for my experiment.

![Figure 8 The main page of your profile in VideoAnt. Here you can skip from one video to the other, and you can see how many duplicates of the video is created.](https://ant.umn.edu/welcome) (Private account registered august 2017)
When clicking on a video, an annotation box shows up and you can stop the video whenever you like, with one button, to comment. The video is only shown in one half of the screen, and the second half is for your comments (figure 9).

![VideoAnt](image)

Figure 9 The screen when playing a video in VideoAnt. The participants comments showed to the left.

Under the video is a timeline showing where you are in the video, and you can click the line to jump back and forth. There is also a play/pause button to the left, and a comment button to the right of the time shown. If you click the comment button, the video automatically stops, and a comment field is shown on your right. Here you can type a title for your comment and/or just use the larger field to describe it in depth. The comments are shown underneath the video as a grey vertical line, and also with a time code. When the video is done you can choose to export the comments as a text spread, or just save it in the software. I choose to export the text and wipe clean the comments after each participant just to be sure that no clues of what the others had replied were to be shown by mistake. Since I was going to analyse the comments in text I found this the easiest way to keep a system, both during the experiment and for the data analysing process. A downside with this annotation tool, is when using a 13-inches screen, the size of the video is small, and it might be difficult to see the performers facial expression clearly. Groves et al. (2009) found that the visual syntax affects the answers given by participants when using web surveys (p. 157), and this means that the layout of the video and how it is presented might affect how the participants respond, and what they focus on. Despite this, I still find that this software is the most adequate for the purpose of my thesis, and only one of the participants expressed dissatisfaction with the small screen, while the rest managed to accomplish the task using the software. (That said, it might
be that it effected the participants without them knowing, and that the results would have been different if they were able to see the performers facial expression.)

3.3.7 Procedure of the experiment

The video experiments were executed from mid-August 2017 till late September 2017 and was conducted with students from the Musicology department at the University of Oslo (the participants will be presented in section 4.2). I chose to observe the participants while they conducted the experiment both for validation and practical reasons. If I were to leave the room while they conducted the experiment they would have to interrupt the experiment for a longer period of time if they had questions concerning the procedure or the software than if I was present. By supervising the experiment, I exclude all other disturbing elements (like if the telephone rings or one needs to go to the bathroom) which might also be both beneficial or negative for the experiment. If the participant were in the need of using the bathroom, he/she might not be able to focus on the task given, but since the experiment were to take approximately fifteen to twenty minutes I found this not a major issue for colluding the experiment. According to an experiment explained by Wolff (1948), two groups of students conducting the same experiment with and without supervision were found to have almost the same handicap which was that “the subject was over-conscious of an inner process and was therefore on the lookout for certain results, while he should completely relax and leave his subconscious free to wander in any direction” (p. xi). This means, that when supervising the experiment, I might push the participant into an unnatural state of over consciousness, but it might also be the case if not supervising it. The first 7 participants did get pizza as a reward for participating in the experiment, but the rest of the 17 people participated without any compensation for their time. Davidson (2004) explains that to give the participants of the study a material reward can put them “under a certain obligation to respond” (p. 63), but I have no way of knowing whether or not this affected their answers. The instructions given the participants beforehand was minimal. Before they entered the room for the experiment they were told that I was writing my master thesis on music performance, and that the test was conducted anonymously and were to run for approximately fifteen minutes. When entering the room, they were told that the test consisted of watching and commenting on two videos in addition to answering a small questionnaire. The participant was then handed the questionnaire and I gave them a number for registration purposes, so that the answers given was anonymized. (As mentioned, I kept the participants name and contact info in a shielded
document for private use, in case I had to contact the participant after the experiment, and this was approved by all participants.) A pre-test was run before starting the experiment for the participants to be comfortable with the software used. They were asked to comment on a music video by Daley feat. Jessie J (randomly chosen from YouTube) after I had given them instructions on how the software worked. Only a couple of the participants asked a technically related question about the software during the experiment, but as far as I can see, this had no effect on the answer given or the completing of the task. The information the participants’ got after they had filled out the questionnaire and completed the test run of the software was to comment on “what you believe the singer to convey and why. That means everything you perceive as communication and that can be everything from the tone of the voice to facial expression and movement. I want you to see this as a simulated concert experience, and as on a concert, I want you to write down the immediate thoughts that pops up. If nothing comes up, that is ok too.” (This has been translated from Norwegian.) I stressed the bit about what they believed the singer to communicate, and that they were to comment their every thought. From a pedagogic point of view, we all learn in different ways (confer section 2.2) which might create a challenge when explaining the participants so little, but in this case, I believe this decision to be right. If I were to tell the participants what I was looking for with the experiment, they would have a different field of focus than what is natural for them, and the answers given would not have been from their point of view but might be coloured by my expectations of their answers (as in the experiment mentioned earlier by Wolff [1948]). They were then handed a headset for excluding any disturbing noises before shown the videos. Every second participant was shown the angry video first, and the others the sad. This decision was made because “perceptual learning processes allow us to improve our ability to differentiate between subtle variations” (Zeiner-Henriksen, 2010, p. 125). Even though the song is completely new to all the participants, I am using the same song in both videos (though performed with different vocal techniques, intensity, volume and body languages) and it will be impossible for the participants to not compare the two performances. When hearing the song for the first time they are probably considering the more “overall” components like melody, instrumentation and maybe lyrics, while the second time they are more familiar with the song and can listen in more detail on the vocal qualities and possibly even focus more on the video. When the participant had conducted the experiment, they answered the last two questions in the questionnaire (about information and completion of the experiment) and the whole experiment took approximately twenty minutes.
3.3.8 Interpreting the data

Before starting the data analysis, several decisions were made to narrow down the data material. The comments that had no significance for this thesis and only involved the technical procedures of the video, or said nothing about the song, performance, mood or the participants own interpretation of any of these subjects where not taken in the preliminary schema (comments like “the black clothing and the white background create contrasts” or “the screen should have been bigger”). Some of these comments could have been interesting for a different kind of study, but since they only describe the participants meaning and interpretation of the execution of the experiment it is not relevant for the research questions in this thesis (they are still stored for future work). After the first limitation process, 96 comments on the sad video and 113 comments on the angry video were collected in the preliminary schema made. Most of the comments were written in Norwegian and was translated by the author to the best of my ability. This to say, that all comments have been interpreted and might differ slightly from its original meaning. (There are a couple of the comments in the appendixes where the Norwegian word is written in parenthesis for explanatory reasons.) After collecting all the comments in one excel sheet, a second limitary process was made before gathering all comments relevant for the thesis in a word schema for each video (appendix 1 and 2). Some comments only described the musical sound (like “the sound should have more reverb” or “off pitch”) and where ruled out of the appendix. Since the information given the participants beforehand was minimal and very open (they were not told to only focus on the visual communication), comments on the audible sound was expected. The analysis then consists of the comments that gives; 1) a clear indication of the participants’ interpretation of the movements, 2) a description of the movements or large facial expression made, 3) a combination of the sound and movement, 4) focus on the vocals and how the participants interpret it´s timbre, and 5) an indication of a mood detected by the participant. This makes it 84 comments on the sad video and 96 on the angry video, all of which are included in the appendix schemas. The difference in the amount of comments seems logical to me, since it’s a lot more happening in the angry video than in the sad video, and the sad video is subtler in its expression and movements. For the purpose of clarifying the findings (which will be presented in chapter 4 and 5), I have separated the comments into different categories, but some comments are included in several categories. An example might be “Body sigh; the body fell down “I can`t take this any longer”- continuation of the lyrics. Her body can actually not take it any longer” (01:03-01:04 in appendix 2). This
comment is both concerning the lyrics, the movement made, and how the movement helps substantiate the lyrics. This means that it is included in three different categories in section 4.4: “Comments on lyrics”, “Comments on movements” and “Mood and lyric substantiating gestures”. Since the participants were free to express themselves in their own way it is impossible for me to separate the comments into strictly delimited categories. I also had to read some of the comments in context of the video. If a participant commented “falls down here as well, but with attitude” (01:02-01:03) five seconds after the action of “falling down” happened I have taken the freedom to place this comment where the movement happened in the appendix schema. A lot of comments are not related to a specific movement, and I have then placed them according to where the participant wrote it. Davidson (2004) points out that in these types of research on music as social behavior, the researcher has to be open-minded and creative in their approach to the audience, even if you use a precise questionnaire.

For the data analysis, I used Microsoft Excel both when collecting the comments from the videos and the answers from the questionnaires. The editing work started in November 2017, with formatting the data collected, by hand, since the most significant data (the comments on the video) is hard to standardize. A software program would not be able to separate the relevant comments from those who are irrelevant for this thesis, and it would take a lot of time to learn a new type of software, so to be time efficient I decided to do it by hand, using Microsoft excel. This editing process took four weeks, including translating the comments from Norwegian to English, together with the procedure of ruling out the relevant comments from the irrelevant ones and dividing them into several categories. The final schemas were created in Microsoft Word (appendix 1 and 2).
4 The video experiments

This section is concerning the experiments and I would like to repeat the underlying concern of the research questions which is: “why and how does who communicate what to whom and with what effect” (Tagg, 1982, p. 39). To answer the question of what is communicated, I have analysed the song used in the video experiments in section 4.1. The reason why I have chosen to do the song analysis in this chapter is for the reader to have the structure and tonality of it in mind while reading the findings from the experiment. In addition to understand what is communicated from the performer’s perspective (as presented in section 3.3.3) it is important for the reader to know who (whom) the receivers in the communicational chain is. This will be presented in section 4.2 where I analyse the participants answers in the questionnaire. To find out if it’s possible to delimit and intersubjectively identify bodily postures, gestures and movements that can augment or attenuate the audience’s reception of the performer’s expressive intention, we first need to analyse the movements used in the performance, which will be done in section 4.3. Since I am the subject in the analysis (as problematized in section 2.2), I have chosen to do both a general description of the physical movements conducted, and a subjective explanation of some of these. This to say that the upcoming video analysis is both analytical and explanatory, and I will shift from a first-person (I) when describing my subjective intentions of the gestures/movements, and third-person (the performer) when describing the more “objective” observations. The part of the experiment that throw light upon the second research question (which bodily postures, gestures and movements are the most effective conveyers of anger and sadness in the communication between the vocal performer and the spectator?) are presented in section 4.4.

4.1 The song

Since this thesis is focusing on the visual communicative aspect of the performance, I found that a complete score analysis of the song is beside the issue of this assignment and would acquire a lot of time and resources in comparison to the beneficial information it would bring to this experiment. For further research, it would be interesting to do a complete analysis of the music and compare the timbre and vocal structure with answers from the participants. But in that case, the audible track should be even shorter (given the issue of time it takes to do a complete analysis of two minutes of sound) and it might benefit from a third-person analysis.
for validity reasons. That said, it is important for the reader to have some insight to the audible aspect of this experiment, and I have chosen to give a small overview of the tempi, structure, tonality and vocal placement of the audible track used. Still, the videos in its whole are to be found online (permanent links in the appendix). I will refer to time codes in both videos throughout the analysis of both audio and video. (It has to be noted that there is a difference of less than a second between the videos. This has to do with the onset of the track in the video and has been counted for in both appendixes. I have chosen to use the angry video, appendix 1, as a reference in the song analysis.)

The song is 128 bpm and the meter is 4/4, and it starts with an intro played by one acoustic guitar lasting two bars (00:00-00:15), in the same pattern as the verse, Bbm- Bbm- Gb- Ab x2 (two chords per bar). The vocal starts on the Bb note (00:15), using only Bb, C and Eb for the whole verse. This establish the keynote and tonality of the song, which is Bb minor (aeolian mode), and help create some tension when the chord hits Gb, before ending on the third note in the Ab chord, establishing the change from minor to major chord. One round of the verse lasts two bars, that is one round of the chords mentioned, and the first verse is sung on two rounds, that is four bars (00:15-00:30). After the verse (00:30-00:47), the chord sequence changes to Bbm- Ab- Fm- Gb. It is played in the same rhythmical pattern as the verse, with a basis on the eight notes, accentuating the syncopated sixteenth on 1st, 4th and 7th, but the Gb chord is played on the sixteenths to build up the tension (00:38-00:39 and 00:45-00:47). This chord sequence is played throughout the pre-chorus and chorus, but with an additional guitar playing a more melodic pattern on four notes (C- Db- Bb- Ab) in the chorus (00:47-01:01). The vocal in the pre-chorus is increasing its range and are expanding in a descending pattern ranging from a high C to a low C. The chorus is centred on the higher octave, using only four notes in a rhythmical pattern similar to the verse. This is a classical way of building up a pop-song, with the verse in a lower part of the singing register than the chorus, and the pre-chorus in between the two octaves to build up for the chorus. After the first chorus, there is an instrumental break lasting two bars, that is one round of the first chord sequence (01:02-01:09). After this, the structure is repeated, this time with a longer verse lasting six bars, with the first round sung in the same pattern as the first verse (01:09-01:15) and a small pause in the guitar (two beats at 01:15-01:16) before the melodic structure changes and increase in tonality range (01:17-01:32). This is followed by a second pre-chorus (01:32-01:38) and a second chorus (01:48-02:03) before an instrumental ending of two bars (one round, 02:03-
02:06). The vocals in the two videos are audible different for reasons explained in section 3.3.1. The sad vocals are sung with more air on the voice, and a change from chest to head voice in the chorus, and ergo in less volume. When singing in less volume, you stand closer to the microphone than singing in high volume, of sound facilitating reasons. This makes it sound a bit “closer” in the mix, and you can hear the breath between the notes in some places. The angry vocals are sung consistently in chest voice which makes it higher in volume and sung with a speech-approach. (This to say that the notes are shorter, and the onset are more pronounced.) The sound is also placed further out in the mouth and nose, which gives the vocals a sharper, more “edgy” sound.

### 4.2 Analysis of the questionnaire

There were 24 participants in total, 17 men and 7 women. This means that less than one third is females, which is close to the gender balance in average on the department (M. Thyness, personal communication, February 28th, 2018). All of the participants are musicology students whereas 6 are attending the master programme and 18 from the bachelor programme.\(^{30}\) I did not find age relevant for the experiment but most participants where in their twenties, with an exception of a couple of the master students. Of the 24 participants only one replied in English, one in Swedish and the rest in Norwegian. The participant who responded in English is well known with the Norwegian language, and the experiment was conducted in Norwegian without any need for translation.

The biggest group of participants (eleven people) replied that they spend 11-15 hours a week consuming music (streaming/listening), seven participants spend 6-10 hours and the rest varied from 1-5 hours till 21 hours or more. Only three participants replied that they spend 21 hours or more per week listening to music, all of which are male. Only one participant (male) spent 1-5 hours listening to music. The number of hours spent on consuming music were wider spread for the males then for the females. Eight participants spend 11-15 hours a week rehearsing/writing/producing music, six participants spend 6-10 hours and the rest varied from 1-5 till 21 hours or more. Only two participants spend 16-20 hours (both male), and 21 hours or more (both male) rehearsing/writing/producing music. Six people spent 1-5 hours

---

\(^{30}\) It has to be stated that I went to school with one of the participants for two years, but that is ten years ago, and he have not seen me perform since. We are acquaintances and meet occasionally on concerts and random social events. I have also taken some lessons with a couple of the master students who participated, but they don’t know of my experiment, and have never seen me perform. The rest of the participants are unfamiliar to me.
rehearsing or writing music all of which spent more than 6 hours a week listening to music. Here as well, the number of hours spent on creating or rehearsing music are wider spread for the males than for the females. Nine people replied that they exercised 1-4 times a month, whereas there was 5 people on each of the other categories named “no, I don’t exercise”, “1-2 times a week” and “3-5 times week”. All the participants who replied that they did not exercise are male.

![Preferred musical taste](image)

*Figure 10 The participants preferred musical taste.*

The text in question number 5 were “which genres do you prefer?”, but they were allowed to give multiple answers (figure 10). Almost all of the participants checked off the “rock” box. The box for pop and jazz were the second largest, with instrumental music on a fourth place. The downside with giving the participants the opportunity of multiple answers in question number 5 (“which genres do you prefer?”) is that some of the participants checked off all the boxes, or close to all, while some only checked off the ones they listen to the most.

All of the participants replied “yes” to the questions if they have ever experienced that music was a part of setting an emotional mood and as to whether they have experienced a live performance that moved them emotionally. Most participants were pleased with the information given and the general achievement of the experiment, but there was some constructive feedback given. Some of the feedback were on the size of the screen, and how it was difficult to see the performers eyes, and that the lack of eye contact was unsatisfying (further discussion in section 5.7.4), and one responded that he was not sure what to look for. One of the participants explained that it was confusing that it was the same person who performed and conducted the experiment (which was problematized in section 2.2). But other
than these, in total four constructive comments, there was five comments on how the participant found the experiment interesting and satisfying.

4.3 Exploratory video analysis: performer intention and perception of body language

Even though I have strived to keep an objective view in the analysing process by applying theories for the execution of this analysis, it is important for me to stress the fact that this is not entirely possible (since I function as both the subject with first-person opinion of the analysis, and the subject to construe the third-person opinion of the analysis), nor is this the intent of this thesis, or even beneficial for the results (confer section 2.2). If I were to analyse the two videos in full, it would require a complete thesis of its own, and so I have chosen to give an overall summary of the analysis according to the theories on movements and gestures explained in section 3.3.4. The terms from Jensenius et al. (2010) are used as a the main categories (that is communicative, sound-accompanying and sound facilitating gestures) and the terms derived from Kurosawa and Davidson (2005) as descriptive sub categories of these (that is illustrators, affect display, emblem, touch and posture) (this applies for both video analysis). The analysis’ with corresponding videos are attached in full length as appendix 1 (angry) and 2 (sad).

4.3.1 The angry video

In the angry video, my intention was to convey as much energy as possible. I move my body on the eight notes and the movements are rapid and sharp (rehearsed as to imitate the action of yelling and fighting with someone [method acting], and to bring out my natural gesture repertoire in these situations). Since I move throughout the entire video, I have not analysed every movement, but the ones that stands out from the “standard movement” of swaying to the beat on the eight notes. Through repeated observation of the video, and applying categories described in the theory chapter (section 3.3.4) I arrived at a total of 48 movements in the angry video. (In the beginning of the video I have included these types of “standard” movements, but this is of significance to the results. A lot of people commented early in the video on their interpretation of the body language, and it is therefore necessary to analyse my movements here.) Of the 48 analyses of movements in the angry video, 44 are described as
communicative, 31 as illustrators, 21 as affect display, 13 as sound-accompanying, 5 as emblem, 7 as sound facilitating, 16 as touch, and only one as posture.

The communicative category is the biggest, with 38 out of 42 movements analysed in the angry video described as communicative, but it always includes other subcategories, like illustrators and affect display. That means that close to every movement I execute in the video are to be seen as communicative. This was expected, since the category of communicative movements does not say anything about what is communicated, and the sub categories were needed for further explanation of the performer’s intentions and a more objective description of the movements. This means, that categories like affect display and illustrators always are subject to the main label of communicative movements. The only time communicative stands alone is at the very end, where the performer’s movements are merely restless swaying back and forth, and I did not find any other suitable descriptions given from the theories applied. The sound-accompanying gestures are mostly paired with the category of communicative movements, or one of the subcategories. Only two times it is used alone, and these are places in the video where the movements are simple and for the sole purpose of finding the beat (after the first chorus at 01:02-01:04 and 01:04-01:07). Illustrators is also used often, always as a sub category of communicative, but it is also combined with all other categories at one or several points in the analysis. The illustrating movements have a substantiating purpose to underline the lyrics and its textual content, or the rhythm of the melody. It is also used for emotive communicative movements, like lifting the shoulders (when the performer intent was to communicate an emotion in a rehearsed manner), or gestures that illustrate and substantiate certain words, but only when the gestures are rhythmical, and not held for longer than the word it is underlining. Some of these more emotive movements are also labelled as affect display, but this is only the movements where I intend to convey an emotive state, or an actual revealing of emotions, or emotional “leakage” (Dibben 2009), and not when it is merely to substantiate words or rhythm. The category of affect display is always used when the performers legs are lifted from the ground (which happens in the first pre-chorus and chorus, up till 01:04), since this is a none-deliberate, natural gesture from my perspective and shows an emotional leakage. It is also used when describing clear facial expressions that has been of significance for the participants’ interpretation of the performance, or a couple of places where I know that I did not intend the movement or gesture made (like 01:48-01:53 described in appendix 1). This
means that affect display is used for both gestures that shows emotional leakage and illustrating gestures that is deliberately rehearsed with the technique of method acting. The movements explained as *emblem* are only three, but in addition to this I describe movements that's on “the verge of emblem” twice. These movements are found at 00:20-00:21 and 01:09-01:10. The first time the lyrics read “swimming around in your mind” and I move my hands in circles at the height of my ears to underline the word *around*. The second time, at 01:09-01:10, the lyrics read “shake it off” and the performer does a shaking movement with her right hand. I decided that the movements in itself not is compelling as emblematic, but in addition to the lyrics they seem to have an emblematic, or substantiating, effect (which is emphasised by one of the participants: “Says *shake it off* and does a shaking movement with her hand”). I deliberately use this movement to substantiate the lyrics and to give an impression of the performer as a calm person in control. This means that it is an illustrating gesture, given that it illustrates the word sung in the rhythm of the melody, but it is also a gesture that can be understood without any lyrical explanation, though it might be interpreted with a slightly different meaning (as to say, “get out”, “get it off” or “get rid of it”). This makes it not evident as emblem, but “on the verge of emblem”. The other three times where *emblem* is used as a description of the movements are when the performer points at herself while the lyrics clearly states that she is singing about *her* or *her* thoughts and feelings. Pointing at one self is hard to misunderstand as anything other than a first-person narrative, and therefore described as a communicative emblem. *Posture* is only used once, at the transition from second pre-chorus to second chorus (01:46-01:48), where the performer stands still holding the head high and pointing the right hand up to the word “start”. This is the only time where the performer holds a pose for a short period of time, the rest of the song she tends to move or gesticulate with either the hands or the right leg. The category of *touch* is used 16 times, but it is only described as the gesture of touch the first time hands are held at the mic or the stand. (Sometimes it is held for a long time, and this is clarified in the appendix, but if I were to break the analysis down to describe every single movement or gesture it would be too complex for this thesis.) What is interesting is that the performer tends to hold the microphone stand when she is singing in low volume in the verse, and the mic at the pre-chorus and chorus, when singing in high volume. Subjectively, I will explain this with different needs for sound facilitating; 1) when singing in low volume, the need for working with microphone techniques are less than when singing in high volume, since the standard is set from regulating the input of sound at the lowest point, so to prevent the
microphone input to peak, one pulls away as the volume increases, and 2) My movements are also not as enlarged in the verses as in the pre-chorus and chorus, and therefore I have more control of where the microphone is, and can only hold the stand and not the mic (If you move a lot it is hard to use a stand, since you need the microphone to follow the sound). The more I move in a performance, the more I tend to hold the mic and not the stand for sound facilitating, and physical reasons. The action of holding the microphone, or stand, is then often combined with the category of sound facilitating, for the reasons mentioned. Sound facilitating is used 7 times, only one time alone which is at 00:47-00:49 to prevent the vocals from peaking in volume (the performer pulls her head away from the microphone). Even though it stands alone in the descriptive column, it is still combined with the action of holding the microphone stand. At this point in the video, we’re at the beginning of the first pre-chorus and the vocals increase in volume, which makes the need for sound facilitating more urgent, and the head is held a bit further from the microphone. Three times this category is used to describe the action of distancing the head from the microphone. 1) At the transition from first pre-chorus to chorus (00:45-00:47), 2) at the beginning of the first chorus (00:47-00:49), and 3) in the middle of second pre-chorus (01:39-01:42). This action is done several times, but only at these points in the video in a large and effective manner, which is why I chose to include them. It is used three times to describe the gesture of a hand rising to build tension and keep intonation on longer notes (00:36-00:38, 00:45-00:47 and 01:37-01:39), which means that at 00:45-00:47 it is used to both describe the action of distancing the head from the microphone, and the rising of the hand for intonation reasons. These gestures of pulling the hand and head in opposite directions helps me create a crescendo in the voice to keep the intonation, as it helps visualise the air flow. Only once it is used for describing the cover up of a preparatory breath (00:42). All of the six times this category is used it is combined with the gesture of holding either the mic or the stand.

4.3.2 The sad video

In the sad video, the performers shoulders are generally high, and her head hanging down or looking to the sides, avoiding the camera. My intent was to create an introvert body language with slow, small and calm movements that is kept close to the body. I always keep one hand on the microphone or the stand, as to express a wish to hide, and for the microphone stand to create a sense of comfort or the need for something to lean on. In this video, the performer tends to stand more still, and there are therefore only 30 analyses of movements or gestures.
according to the limitations applied for the analysis of the angry video (explained in section 4.3.1). Here, the standard movement is slightly swaying the upper body to the fourth notes, and movements that stands out from this are included. (In the beginning of the video I have included these types of swaying movements, and also at the instrumental part in the middle of the video, after the first chorus, where there is no other gesture made.) Some places the movements are static and held throughout the phrases, and a decision was made to only analyse the first movement, and not repeat the analysis. Out of the 30 analyses of movement, 28 is described as communicative, 16 as affect display, 12 as illustrators, 3 as sound-accompanying, 3 as sound facilitating, 3 as posture, 5 as touch and only one as (on the verge of) emblem.

The same conditions for the communicative category as in the angry analysis applies here. It is never used alone, always in addition to one or two categories, mostly affect display or illustrators, but twice in the beginning also with posture and sound-accompanying. (This is before the vocals come in, and the performer only sways slightly from side to side, that is the “standard movement”.) Affect display occur 16 times, all of which are a sub category of communicative, but it is also combined with posture two times, and illustrators two times. The first time it is combined with posture is at 00:06-00:15, before the vocals come in and the performer stands calm for a short period of time with her head behind the microphone stand. The second time is at the very last phrase, 02:01-02:04, where her body freezes in a pose at the word “longer”. Both times affect display is combined with illustrators are when left hand is dropped in the rhythm of the words sung, at 00:31-00:36 on the word meant, and 01:39-01:40 on the word falling. These are the only times where the performer uses her hands in a slightly rapid movement, and they align with the vocal accentuation of the words. An example of the use of affect display can be at the end of the first chorus (01:03-01:04), when the performer drops her whole upper body as a “body sigh”. This gesture is the same as in the angry video, but much smaller. It started out as an unintended gesture, but once I was aware of it, I started use it deliberately, and in this video recording it is done to emphasise the emotive state of being tired and giving up. Illustrators are used a lot less in this video than the angry one, and this because there are less lyric and rhythm substantiating gestures in the sad performance. It is also exclusively used as a sub category to communicative, for the same reasons as with affect display, but the gestures labelled as affect display is often slower than the ones labelled as illustrators. In the angry video, the category of illustrator is used a lot
with sound-accompanying, but in the sad video it stands mostly alone with communicative. *Sound-accompanying* is only used three times. Two times in addition to communicative and illustrators when describing the motion of swaying to the music (the beginning of the video at 00:02-00:05) and at the second verse when the left arm is used to substantiate the rise in the melody (at 01:23-01:29). Sound-accompanying is used alone at 01:04-01:09, at the instrumental part after the first chorus, when the performer sways to the half notes. *Sound facilitating* is used three times, once alone (at 01:31-01:32, the end of the second verse) and two times together with communicative and affect display (01:18, middle of second verse, and at 01:46-01:48, held from the end of the second pre-chorus on the word “start” and into the second chorus). The first time it is used, I pull away from the microphone shortly to cover up a preparatory breath, and the last two times (at 01:31-01:32 and 01:46-01:48) the gestures made are to help create tension and intonation of the tone sung (as explained in the analysis of the angry video, 4.3.1). Since the performer stands more still in this video, the category of *posture* is used three times, slightly more than in the angry video. Two of these times occur at the beginning before the vocals come in, and the last time at the very end, when not singing. *Emblem* is only used once (00:28-00:31), and not as a strong emblematic gesture, but “on the verge of emblem”. At this point in the video, the lyrics read “when I slammed the door to my own heart” and the performer does a small hand gesture with left hand that imitates the action of “slamming” something, before the hand is moved close to the heart. These two gestures (slamming and touching the heart) are both described as communicative and illustrating, but there are still some differences between those two. If looking at them objectively it seems like they both are made for the same purpose, that is to underline the lyrics and accentuate the meaning of the phrase. But, from my subjective point of view, the first gesture (slamming) are made deliberately and is an illustrating gesture, whereas the second (touch the heart) derives from my natural gesture repertoire and can be seen as an affect display that is “on the verge of emblem”. In the angry video, the performer does not use this movement (holding her hand to her heart) but the slamming gesture with the right hand are deliberately exaggerated and therefore labelled as illustrator. Even though the performer uses the microphone, and microphone stand, a lot more in this video than the angry, the category of *touch* is only used 5 times. The mic and/or stand is constantly in use, but only five times the performer changes hands, or the position of the hands. (As explained in the beginning, I have only described the gesture when it’s conducted the first time and have not repeated the analysis when the posture is held for a longer period of time. I have commented in the appendix on how long the hands
are held at the same position.) The action of holding the mic or stand is opposite in this video compared to the angry video. In the angry video, the performer tends to hold the microphone in the chorus (for sound facilitating reasons as mentioned in section 4.3.1) but in the sad video she holds the stand during the chorus, and the microphone during the verse. My explanation of this is the same as for the angry video, that it is of sound facilitating reasons, since the verse is sung in lower volume and ergo I have to be closer to the mic. If holding the stand and not the mic when singing in low volume it is easy to get a feeling of losing control, and also to hit the mic with your teeth if one should “get caught up in the moment” (the moment when you are completely in the state of conveying your own emotions). By holding the mic, you prevent this lack of control, and create a closeness between you as the producer of the sound and the receiver of the sound (in this case the microphone). When I reach the chorus, I am singing in normal volume, and do not have to be as close to the mic as in the verse and stand freer to move away from the sound receiver (microphone). This action of holding the mic or the stand for sound facilitating reasons are so thoroughly rehearsed that I do not need to think of it, which makes it a part of my natural gesture repertoire. The action of changing hands from the stand to the mic, and vice versa, are of these reasons mainly taking place in the beginning (00:02-00:05), at the transition from pre-chorus to chorus (00:44-00:46, 01:41-01:45), and after the first chorus (01:10-01:17). The last time it is used is when the melody changes at the variation of the second verse (01:19-01:21).

### 4.4 The experiment

For the presentation of the findings it is important that the reader keeps the limitary process of the comments in mind (as explained in section 3.3.8). The comments in the analysis is those who gives: 1) a clear indication of the participants interpretation of the movements, 2) a description of the movements or large facial expression made, 3) a combination of the sound and movement, 4) focus on the vocals and how the participant interpret it’s timbre, and 5) an indication of a mood detected by the participant. This makes it 84 comments on the sad video and 96 on the angry video, all of which are included in the appendix schemas. I will analyse them according to the categories they are placed under, each with its own sub-section. (Further discussion of the comments and the presentation of results will be given in chapter 5.)
4.4.1 Comments on lyrics

Of the 96 comments on the angry video (appendix 1), there are only 22 that includes direct comments on lyrics. Some of the comments, like “accentuate death” (00:15-00:17), are only concerned about the lyrics sung but there were no comments that only repeated the lyrics (only wrote “death”, “heart” or “pain” etc.). Four comments expressed a link between the vocal style which the text is performed in, and the lyrics, like “Serious way of singing, reflects the lyrics” (00:35). These four comments are the only ones who did not include any significance of the gesture or movement made at that time in the video. The rest of the 18 comments on lyrics either mentioned a gesture made at the same time of the word described, like “Head a bit up when she sang the lyrics head up” (00:47-00:49), or a combination of the body language and facial expression used- “she uses her hand a lot to substantiate the meaning of the lyrics. Uses the body in a kind of dance. Grimaces that makes you think of anger/evil” (00:21). These multimodal impressions are combining the participants observation of both lyrics and gesture/movement/body language that happens in synchrony, the audible vocals, and/or the participants interpretation of the combination.

In the sad video (appendix 2), there are 19 comments on the lyrics, or that mentions the lyrics, of a total of 84 comments. Only one of the nineteen comments in the appendix is not making a direct remark on the lyrics in combination with a physical gesture or movement, which is “Compassion; because of the lyrics; the pain makes me stronger” (00:48-00:54). This comment in particular is interesting in that sense that it describes a feeling that the participant recognises as compassion, but as far as one can read from the comment, the participant perceives this feeling from the lyrics alone, and not the circumstances around the lyrics (at least that is what he explains in writing). It might be that this participant is only focusing on the lyrics alone or the audible performance, or it might also be that he is unaware that the impression from the physical performance impact his judgement of the audible performance, and therefore reads the whole impression into the lyrics. (One might ask if anyone with the audible and visual senses in god shape is capable of distinguishing between these two.) The remaining 18 comments on lyrics in the sad video are describing a combination of the lyrics and gesture/movement, the audible vocals, and/or the participants interpretation of the combination of these, but I find that my interpretation of some of the comments in this category needs to be explained further. One comment that is hard to interpret is: “a bit more helpless now. I can’t take this any longer. Earlier on it seemed
angrier, more controlled, now it seems like she has given up” (01:03-01:04). The way this participant phrases himself, *it seems*, is not a word to describe the audible sound, but the visual impression given. Even though he does not comment directly on the movements, the way he uses the words gives an indication that this also is a multimodal impression that is perceived. If he believed that the emotive state of giving up, or being helpless, came from the lyrics or vocals alone, I believe he would have used *sounds* instead of *seems*, but this is merely an assumption. Another comment that is focusing on the lyrics, and not in direct relation to a movement made, is this: “Depressing lyrics that substantiate that it makes you stronger, but a bit passive communication with the listener” (01:46). This participant is telling us his disliking of the expression in the performance, but also that he has picked up on some of the lyrics, and that he has perceived a mood of depression *from* the lyrics. He uses the word *communication*, which may indicate that there is something beside his comprehension of the lyrics and mood indication that he finds *passive*, but he does not say what he finds passive. It might be the vocals, the body language or gestures, or the lack/use of eye-contact. The rest of the comments on lyrics are including a remark of a physical gesture made, or the body language in total, as in the angry video.

### 4.4.2 Comments on the voice

When it comes to comments that focuses on the sound of the voice, and not exclusively the words, there are 21 in the angry video. Some of them are a combination of both the sound of the voice and the words, like “More typical “soul” timbre, deeper and more sensual. Sings about getting “hurt”, so it might fit the song better than the first” (00:27). Other comments, though truly outnumbered, commented only on the voice: “the vocal feels somewhat closed” (00:36-00:38) or “Much more powerful singing, gets more of a power ballad-feel than the tender emotion from the other song” (01:02-01:04). Since the participants did not get any instructions on commenting solely on the body language, these types of comments were expected. They still tell something about what the participant mean is conveyed by the singer, but in a way that is hard to interpret. The comment “More typical “soul” timbre, deeper and more sensual. Sings about getting “hurt”, so it might fit the song better than the first” (00:27), is an explanation of the participants interpretation of how the voice influences his opinion on the performance. In this case, he says that this vocal style *might* fit the song better, but he has not reached a conclusion on whether or not he believes it to be so. The comment on the closed vocals (“the vocal feels somewhat closed” [00:36-00:38]) can be interpreted in several
ways. It might be that the participant found the performer to be more closed in general (both in body language and timbre of the voice) which makes it a multimodal impression. Or it can be that he was only focusing on the vocal timbre or found that the vocal expression and the bodily expression did not correspond and therefore got the impression of a *closed-up* performance. Still, most comments were a combination of the audible and visual performance, like “Narrow timbre: sings in a pleasant register but puts a lot of acting into it. Theatrical expression in her voice that is increased by her hand gestures” (00:20-00:21).

In the sad video, there are 25 comments on the voice or the audible sound. There are several comments on the low volume of the voice and these are not of particular interest in this thesis. But, several of the comments shows how the participants are interpreting the voice and explains how it helps create an impression of the overall mood in the video, or simply explains that the participant picks up on an emotive state, like “Fake “emotion” in the voice, “pressing” on the voice” (00:24-00:25), “Sings lower in volume. More air on the voice. “Closer” expression” (00:31-00:36), “Vocals; starts very careful and calm, as to underline the emotions established” (00:31-00:36), “Style of singing- toned down. More intimate, less confronting” (00:31-00:36), “Falsetto makes the experience more intimate” (00:48-00:54), “Sings in head voice to keep the emotion in the song calm/sad, which also gives a good contrast/climax” (00:48-00:54). These comments mentioned above are not just talking about the audible sound, but also the participants interpretation of the sound and how this helps communicate an expression of a mood or emotion. The comment “Sings lower in volume. More air on the voice. “Closer” expression” (00:31-00:36) is combining the participants description of what happens technically in the voice, and his personal interpretation of the style of singing. If he were just to write “sings lower in volume” it would have been a purely descriptive comment of what happens in the music, but he is also saying that he finds that this helps to create a “closer expression”. (This technique, where the singer stands close to the microphone and almost whisper the vocals, was intended when recording the vocals for the sad video.)

### 4.4.3 Adjectives used by the participants

Even though the participants were not given any instructions on what to comment on before starting the task, there was several comments on emotions detected, and the participants different interpretation of the performance. The adjectives and nouns used to describe the
performer’s appearance in the angry video is: determination (00:02-00:03, 00:27-00:30), attitude (00:02-00:03, 01:02-01:03), openness (00:02-00:07), lively (00:07-00:12), roughness (00:07-00:12), energetic (00:07-00:12, 01:48-01:53), groovy (00:12-00:15), theatrical (00:20-00:21, 01:27-01:28, 01:04-01:09), evil (00:20-00:21), anger (00:20-00:21, 00:57-01:01), fierce (00:27-00:30), aggressive (00:36-00:38), gustily (00:42), intensity (00:42), strength (00:49-00:52, 01:07-01:08, 01:48-01:53), sad (00:57-01:01, 01:07-01:08, 02:04- 02:05, 02:04- 02:05), power (01:27-01:28, 00:27-00:30), pain (00:30-00:35, 00:55-00:57, 00:57-01:01, 01:46-01:48, 01:48-01:53), spontaneity (01:32-01:34) and self-confidence (01:07-01:08, 01:15-01:16). Some of the words are very open for interpretation and hard to connect to a certain emotive state or a movement (as attitude), but others are clearly meant for the physical performance (as gustily, openness, determination, lively and energetic). Off course the audible song can be lively and energetic, but all of the comments that concern these terms (as with determination and openness) are describing how the movements creates this particular impression, and it is also clear that they are concerning the physical appearance since they are used before the vocals come in (the first fifteen seconds). All of the three comments on strength only states that the singer is “showing strength”, or seem strong, but none of the comments explains why the participants finds this to be conveyed. It might be that the strength comes from the body language that was deliberately taken on in this video (that is a combination of holding the head up, pushing the chest out and straightening the back, which in total can give the impression of self-confidence, strength and power, but if not in relation to each other it can give a totally different expression of the persons mental self) or the vocal techniques used, or a combination. (All of the comments that uses the words strength, power and self-confidence are uttered after the vocals come in.) Anger is used twice, but both times it is hard to link it to either the voice, the body language, or the movements. The comments that includes the word sad are divert in their opinion and hard to interpret. One of the comments on sadness explains that the participant finds the singer to be not sad, but three of the comments only says “sad” (02:04- 02:05), “sadness” (02:04- 02:05), and “gets a bit angry-sad” (00:57- 01:01). The two comments that only says “sad/sadness” are written at the very end of the video, where the performer is swaying restlessly, holding the head down (but it has to be noted that there are two points of eye-contact at this time where the performer looks into the camera for a short while). Some of the descriptive words acquires a deeper explanation from the participant of several reasons; 1) Phrases as “gets a bit angry-sad” (00:57- 01:01) says nothing about whether it is the vocals, the track or the body
language the participant is talking about, and 2) some of the adjectives used (as with cool and intensity) is impossible for me to read a particular meaning into. Intensity can be both an intense communicational experience, where the participant experiences direct contact with the performer, or to describe a vocal or physical performance that has a lot of energy in both a positive and negative manner. (Since I am on a schedule with this thesis I don’t have the time contact the participants for further explanation, which might not even be possible, given the amount of time since some of them answered.)

In the sad video, the descriptive words used by the participants are different: vulnerability (00:06-00:15, 00:24-00:25, 00:48-00:54, 02:09), sad (00:06-00:15, 00:56-01:03, 02:09, 01:41-01:45), calm (00:06-00:15, 00:16-00:18, 00:48-00:54), serious (00:06-00:15, 00:16-00:18, 02:09), introvert (00:06-00:15, 02:09), fervency (01:16-00:18), closed (00:21-00:23), hurt (00:24-00:25), intimate (00:28-00:31, 00:40-00:43), sensitivity (00:37-00:40), tenderness (00:37-00:40), sore (00:40-00:43, 02:09), intensity (00:44-00:46, 01:23-01:29), enclosed (00:48-00:54) small (00:46-00:48, 01:46-01:48), nice (00:46-00:48), slow (00:48-00:54, 01:03-01:04, 02:01-02:04), helpless (01:03-01:04), exaggerated (01:03-01:04), mystical (01:33-01:37), anonymous (01:41-01:45), longing (01:41-01:45), emotional (01:49-01:57), frustration (02:01-02:04), painful (02:09). All of these words are used about the performers body language, gesture or a combination of the gestures and the sound of the voice. Some of them, like slow and intimacy, are almost exclusively used to describe the movements, and in this case, mostly the hand gestures (example: “Uses the arms to drag us closer. Build intimacy” [00:28-00:31]). Calm is used both about the appearance of the singer, movements and the voice (example: “Fervency; soft, calm movements, also in the voice” [00:16-00:18]). Intense, or intensity, is used both about a gesture made and the vocals. The adjective sad, or sadness, is used four times in very different places in the sad video, and none of the times it is related to a specific gesture or movement. Introvert is only used twice, once at the beginning and once at the end, both times when there are no vocals, which means that the participants find the body language/posture, or the gestures made, introvert. The adjective vulnerability is used several times throughout the video. The first two times at the beginning (both times at 00:06-00:15) they describe the posture of the singer, the third time (00:24-00:25) it describes the hand gesture, and the last time it is used (02:09) it is a description of the participants general impression of the performers body language. It is only one time (at 00:48-00:54) where the adjective vulnerability is used in a way that is hard to interpret- “Vulnerability,
enclosed, get a feeling of curiosity of what the singer will convey” (00:48-00:54). At this point in the video, the first chorus have just started, and the right arm is slowly lifted to the head and held there throughout the first phrase of the chorus. The body language is in general calm and the arm gestures are kept close to the body. Since the adjective vulnerability is used four out of five times to describe the physical movements or body language of the singer, one can interpret this comment (00:48-00:51) as concerning the physical appearance of the singer, and not the audible sound, but this is merely speculations.

4.4.4 Comments on body language

One of the most interesting part for this thesis is the comments on the body language in general. Out of the 96 comments on the angry video in appendix 1, 64 is saying something about the body language used, a movement or gesture in particular or the movements in general. In the sad video, there are 50 comments in appendix 2 concerning any type of movement, gesture or body language. Still, some of these comments include some audible remarks, but given the subject of this experiment (music performance), and its participants (music students), anything other would have been surprising. For the purpose of this thesis I need to explain the comments that is included in this commentary group (all examples are from the angry video, appendix 1). Most of the comments on movements are using words as “movement”, “body language” or “posture”, but a lot of them are also describing a particular movement, as “Looks away, avoid eye contact but she is not as sad/devastated and owns it more” (01:07-01:08). (This comment describes the action of looking away, which is a movement that the participant interpreted as emotional loaded, but are also concerning eye-contact, and I have therefore included it in both the body language category, and the eye-contact category.) Some comments are not concerning a specific movement/gesture, but it is still included in the commentary statistics on movement because of the words used to describe the impression, like “The singer is showing strength” (00:52). In this case, the word showing is closely related to the visual sense and I have therefore interpreted the comment as based on the visual perception. This also applies for some of the adjectives or nouns used in comments, like “conveys good self-confidence” (01:15-01:16). Self-confidence has, too me, a lot to do with posture, body language and the person as a whole. It is not often you hear someone differ between a confident voice and a confident appearance, and I find this to be inseparable. In addition to that, posture is used as a movement category in the video analysis and it would be confusing to not include it as a physical movement in both sections of this
thesis. I have also chosen to include one comment that is focusing on the action of making a facial expression in this category but excluded a couple that only concerned eye contact (example: “first clear moment of eye-contact” 00:43-00:45) since the category of eye-contact is excluded from my movement category. The comment on the facial expression reads “It looks like she screams with a sad/painful expression on her face. This takes a lot, emotionally, from her” (01:48). Since the action of screaming is a large grimace, and the participant chose to use the visual verb to look, I have decided to include this in the gestural statistic. The comment “Seems more aggressive” (00:36-00:38) is excluded in the movements category since the original word used by the participant is “virker”, which is not necessarily dominated by the visual sense, which has to be seen as my subjective decision.

Out of the 64 comments on some sort of visual cues in the angry video, there are only four who makes a direct comment on how the movements or gesture was of their personal liking or made a positive impact in the video. With this, I mean comments like: “Cool hand gesture” (01:09-01:10), “The movements substantiate what happens in the song in a good way” (00:30-00:35), and “Exiting and spontaneous movement. Movements makes it easier to follow, and you don’t get bored as easily” (01:32-01:34). Six comments expressed the participants disliking of the amount of movements, like “Too much movements in hands and dance in comparison to the music” (00:27-00:30), “A lot of movements, and it does not fit so good with the song” (00:49-00:52), amongst others, but none made a remark on a particular gesture made.

In the sad video, there are only two comments, out of 50, that expresses the participants liking of the visual appearance of the singer, “Much more calm movements, more comfortable to look at” (00:06-00:15), “Look: cool how she looks down before she starts to sing” (00:06-00:15), and four concerning the opposite, like: “I start to empathize, getting used to the style of singing and moving and believing more in the singer, although the movements seems to be a bit exaggerated” (01:03-01:04), and “Movement, stands still with the legs but a lot of good movements in the upper body that shows that the artist enjoys the music, but gets stiff with so little movements in the legs.” 01:41-01:45). The same variations on comments on audible and visual remarks, or the combination of them, as in the angry video are found here as well.
Comments on the legs
In the sad video, there is only one comment on movements in the legs: “Movement, stands still with the legs but a lot of good movements in the upper body that shows that the artist enjoys the music, but gets stiff with so little movements in the legs.” (01:41-01:45). Compared to the angry video, where there are four comments on the legs: “Foot is lifted up, gives me the impression that she feels the song” (00:38-00:40), “Grimaces in her face can remind of pain. A lot of substantiating the syllables in the lyrics with arms and legs” (00:55-00:57), “She uses her legs to express emotions. Feels like she can’t hold it in any more” (00:55-00:57), and “Not as much movement in the legs as in the upper body” (02:06). The gesture of lifting the leg is used seven times in total in the angry video (all between 00:15-01:03), in comparison to none in the sad video, all of which are described as communicative affect displays.

Comments on eye contact
In the sad video, there are 11 comments in total on eye-contact, or the lack of it, and the action of looking. Most of the comments on eye-contact in the sad video are focusing on how the performer avoided it, and how this helped to create an intimate and vulnerable feeling of the song: “Sad opening; the singer looks to the ground, seems like she is going to convey something sad” (00:06-00:15) and “Introvert vulnerability; the singer rounds of as in the beginning; no eye contact- looks down- introvert body language- gets a sad/dejected feeling. Its sore and painful” (02:09-02:18). There are also some comments that is not concerning eye-contact in direct but explains the action of avoiding it, like “Eyes are closed. Matches the sore and intimate expression of the song.” (00:40-00:43), “Stands very still, often with her eyes closed to show a more calm/melancholic sensation/emotion.” (00:48-00:54). Most of the comments on the lack of eye-contact are described in the beginning and the end of the song in the sad video, whilst the ones about looking away, or avoiding eye-contact, are to be found throughout the video.
In the angry video, there are only 5 that concern the same gestures, and they are more divert in their opinions than in the sad video: “Determent; Looks up, more open” (00:02-00:03), “Eyes open, shakes her head and body and look around. Expresses more openness and attention. Makes the audience wake up” (00:04-00:06), “First clear moment of eye contact” (00:43-00:45), “Looks away, avoid eye contact but she is not as sad/devastated and owns it more.” (01:07-01:08), “I get engrossed but start to miss eye contact” (01:35-01:37). They are also not clustered in a particular point of the video but are spread throughout.

### 4.4.5 Mood and lyrics substantiating gestures

After reading the comments from the experiment, I chose to separate the comments that give some indication of mood or lyric substantiating gestures (all of which are also included in the category of “comments on body language”, 4.4.4). All the comments included in this category explain that the participants found 1) the gesture or movement at one point, or in general, helped substantiate the lyrics, or created an overall impression of cohesiveness between the lyrics and the performance, or 2) helped them perceive a mood, or emotive state, in the performance, in comparison to a more general description of body language in the latter section (4.4.4).

There are in total 18 comments on how my gestures substantiated the lyrics, or a perceived mood, in the angry video. Some of them are explaining that the participant found a movement/gesture in direct relation to the lyrics sung: “Head a bit up when she sang the lyrics head up” (00:47-00:49), “Says shake it off and does a shaking movement with her hand” (01:09-01:10), and “On clear she does a gesture toward her head” (01:30-01:32). Even though these three comments clearly explain that the participants have seen the communicative and illustrating gesture made at this point in the videos in consensus of the lyrics at that time (like holding the head up), we don’t have any information about what emotive state they believe the performer to be in when she conducts the movement, or what they believe her to communicate with the words. A lot of the comments that gives an overall description of the participants perception of the performance are difficult to interpret in one way, like “Many and large hand gestures and determined voice gives the impression of power and determination” (00:27-00:30). This comment talks about the movements that has been used the first thirty seconds in plural and does not describe one gesture in particular. If I read this comment in an “objective” way, it is both the quantity, and the size of the movements
used so far in the video, in addition to the “deterrent voice” that the participant find conveys this expression (power and determination). Other comments are vaguer in its explanation of a mood, or the perceived meaning of the gesture/s: “Hand gestures; gives more power to the message sung” (00:38-00:40), “Hand gestures amplify” (00:38-00:40), “She uses her legs to express emotions. Feels like she can’t hold it in anymore” (00:55-00:57), “The artist substantiates the lyrics by giving the impression of extra effort” (00:57-01:01), “Point at herself again; gives more power and direct the attention toward herself and the message” (01:27-01:28), and “She hold her head up to match the emotional state of the lyrics” (01:48-01:53). Some of these comments are describing a particular gesture (like the ones at 00:55-00:57 and 01:48-01:53), but none of them explains which emotions they find to be expressed or substantiated.

In the sad video, there are only 10 comments that express a link between the gesture or movement made and the lyrics. In this video, there were no comments on how the overall gestures substantiated the lyrics, like in the angry video, only on specific gestures and places. The comments of this category are also more clustered in certain places in the video, compared to the angry video where they are spread throughout. The first comments of lyric substantiating gestures are found at 00:28-00:31: “Substantiate the lyric my heart by touching her heart”, and “Touches her heart at the same time as the lyrics read my heart. The hand kind of helps the lyrics”. Both these comments are concerning the gesture made at that time (touching the heart) in direct relation to the lyrics sung (“when I slammed the door to my own heart”). These types of comments are also found at the beginning of the first chorus (00:48-00:54): “Sings something about the head and lifts her hand slowly to her head”. At 01:03-01:04, five participants commented on a specific movement made: “The body rushes to the side; when she sings can’t take this any longer she moves slowly to the side until her body “collapses”. Looks heavy”, “A kind of climax in the chorus is given weight by the one hand “spelling” the words before the vocalist is pulling away dramatically” (figure 12).
In both videos, the vocal starts after fifteen seconds, but before the vocals come in, some of the participants commented about their expectations of the song based on the performer’s movements. In the angry video, six participants commented on the movements or body language in total, like “The motions of the singer makes me anticipate an energetic song” (00:07-00:11), and “A lot of movements, gives the impression of a lively song” (00:07-00:11). In the sad video, there are five comments about the performers body language before the vocals start. These comments are much more specific and explanatory than the ones at the start of the angry video: “The posture; a kind of vulnerability- no eye contact- looks down- “tight” body language”, “Vulnerability; how the singer presents herself with how she stand/poses signals vulnerability”, “Sad opening; the singer looks to the ground, seems like she is going to convey something sad”.
5 Results and discussion

When working with such complex phenomena as music perception and expression, the discussion of the limitation processes requires further discussion, which will be done in section 5.1. The next two sections will be devoted to discussing the theories applied in the light of the findings. I will first explain the importance of a model of musical communication that is applicable for all aspects of communication in music performance (5.2), before I lay out the findings concerning the terms (5.3) and taxonomies (5.4) applied in this thesis. Further on, I will discuss the comments on visual versus audible performance in section 5.5, and how the empirical findings are highly subjective in section 5.6. The discussion of the experiment is potentially infinite, and so to present the findings in a comprehensible way, I have chosen to do so under the section called “significant moments” (5.7), as adapted by Trondalen (2004).

5.1 A critical approach to the experiments; limitations and expectations

As explained in section 2.2, some of these analyses would not have been possible to do without the interpretation and explanation of the performance from the performer’s perspective, but I would like to continue this discussion for validation reasons. Moore (2012) explains that “the video itself is an interpretation (normally first-order) of the track, normally suggesting an interpretation to which the listener can then respond, that is, offering (and sometimes strongly) a particular subject position” (p. 164). This first-order interpretation of the track can be seen from both my perspective and the participants. Since I am the performer in the track, the video is my first-person interpretation, meaning that I have a subjective opinion about the performance as the performer, and a third-person interpretation as the executor of the experiment. The participants watching the performance holds a second-person interpretation of it, whilst I, as the prosecutor of the experiment, hold a third-person interpretation of their second-person interpretation (puh!). The complexity of such an experiment is multifaceted, and, as explained in section 2.2, this thesis is heavily influenced by subjective thoughts and opinions, both from my perspective as performer and executor of the experiment, and the participants.
The situation in which the experiment was conducted, deviates from normal music performance, as explained in section 1.1, and the experiment procedure might have had an effect on the participants answers and my interpretation of them. Since I am asking my participants to write down their comments, I am pushing them into a cognitive state of reflection that might have been affected by both attunement behavior between the executor of the experiment and the participant, as well as their own observation of their social readings. (When talking about attunement behavior in this scenario, I am including both the encounter between me as a fellow student and researcher [which is two very different positions] as well as the participant’s encounter with the performance in the videos.) In retrospect, I see that there is a possibility for the participants to attune with my mood (both body language and prosody) when entering the room and conferring about the upcoming experiment. This could have had a different effect on different participants, and my mood might have changed within the time frame it took for the participant to complete the task (15-20 minutes), and during the time frame it took to complete the experiment (from mid-August 2017 till late September).

As mentioned earlier (section 3.3.7), there are pros and cons related to supervising the experiment, and the varying degree of mood attunement from the researcher’s perspective might be one of them. That said, I did not find a correlation between the negative or positive comments made by some of the participants and the way we communicated during the experiment, but this is a highly subjective opinion, and the participants might see this from a different point of view. (It will never be possible for me to measure the degree of mood attunement, in this setting or elsewhere, since how we react to people’s appearance is a highly subjective matter.) When pushing them into a cognitive state of reflection, I might also send the participants in a different direction than what is natural for them (as in the experiment mentioned in section 3.3.7 by Wolff). This might have interrupted their natural chain of thoughts, which in combination with the setting might vary from how they think about a live performance and push them into a state of self-observation. I have no way of knowing if this social encounter between me and the participant, and the unnatural cognitive situation the participant is placed in, benefits or debilitates the experiment. One of the questions that was removed from the questionnaire (“How are you feeling today”) could also have shed some light on how to interpret the comments, though the honesty of the answer given might have varied depending on the participant’s openness (which is the reason why it was removed). It might be that an experiment like this would have benefitted from an in-depth qualitative study, interviewing the participants over a longer period of time. Still, the
same problem with pushing them into a different cognitive state might be applied in this scenario as well.

Since I decided to exaggerate my emotional state and body language for the recorded video, I was prepared for negative comments regarding the usage of large gestures. After collecting the comments, only a small percentage (approximately 5% in both videos) were of a negative character, which is smaller than expected. My subjective explanation of this is that all of the movements and gestures made in the video are coming from my natural gestural repertoire, (confer section 3.3.3), and therefore look natural enough to be interpreted as expressive and not just disturbing. Another reason why I find the amount of negative comments surprisingly small, is that the experiment is based on music performance, which is a very subjective field, and I expected more diversity in the participants personal taste. This both on the basis that the art of music is highly subjective, but also because my participants are studying music, which gives a deeper insight to a broader variety of genres and techniques. Still, maybe because I am dealing with music students, it might be that they are more omnivorous when it comes to listening and analysing music and musical performance than most other people. But, these negative comments on the body language and movements are largely outnumbered in comparison to the ones who expressed their liking of the movements. (It is still possible that some of the participants who wrote positive comments, or was just participating in the experiment at all, had some critique about the performance but did not write it down in order to please me.)

With the explanation of Davidson (2004) about giving the participants a compensation for their time in mind, one would believe that the participants who got pizza as a small compensation for participating would use longer time or feel more obligated to answer than the others, but I cannot find any indication for this in the questionnaire or in the comments on the videos. Some were quick, and some took their time responding, but it does not seem that the time spent on the video has anything to do with the respondent’s ability to analyse the video. (Some of the participants who spent less time than others on the videos wrote down descriptive comments that are included in the appendix.) I would guess that this has more to do with an interest in the music used in the experiment, the field of musicology, or even a personality trait. But these are only speculations and would require further inquiries in the field of music sociology. One of my expectations was that the master students’ comments
would correlate more with my intended expression in the videos than the bachelor students’, since they have more knowledge of the different fields of musicology, and that this insight might give a broader ability to reflect on the experiment. This was not the case, since some of the master students commented less (and without any usable information for this study) than some of the bachelor students, and vica verca. The comments I had hoped for was of this character: “Points at herself again; gives more power and directs the attention toward herself and the message” (01:27-01:28). Here, the male participant makes a direct comment on a gesture made by the performer and what he perceived it as. This gesture made at the particular time in the angry video both substantiates the lyrics and message in the song and gave him the impression of power. A similar comment made by one of the female participants are: “Pointed upward on the top note; now we have reached the climax!” (01:46-01:48). This also describes a particular gesture and the participants interpretation of how she found that it substantiated a climax in the song or performance. Still, the participant does not explain what her expectation of the climax is (if this is positive or negative), or even if there is any emotive state involved from the participants point of view, but this gesture of pointing up on the top note clearly creates some kind of climax. Another expectation was that the women who participated in this study would understand what I wanted to communicate to a higher degree than the men. This is only based on my subjective experience of my own music performances and the feedback I have gotten from colleges and random audience, especially in the last two years. I found that the women who replied (although they replied a little less in average than the men) had a higher rate of comments that I found relevant to the purpose of this thesis (to be found in the appendix which, for practical reasons, only includes comments found to be relevant to the research questions) than the males. That said, it is not possible to say that one of the genders’ responses correlated with my intensions better than the other, since my expectations of the participants response were too high.

5.2 Models of communication in music performance

As a foundation for the communication chain in this thesis I used the Reciprocal feedback model of musical communication (Hargreaves et al., 2005) (confer section 2.1.2, figure 2). When reading the comments in the light of the feedback model of musical communication, I found that the model is a bit unclear in its presentation of how the performer communicates with its audience. It is clear that some of the participants, although the minority, did not find the performance of their personal liking. Even though some of the participants who were
critical to the performance checked off the box for pop music in the questionnaire, and ergo understands the genre, there was something in the performance, either the physical or auditive performance, or the song used in the experiment, that did not appeal to them. (Comments like “Theatrical: gives good contact and presence, but too much body language” [01:27-01:28] and “Way too much movements” [02:04-02:05]. Both participants checked off the box for pop music, and both comments are from the angry video.) In the model (figure 2) the point of interaction between the performer and the listener is marked with a star and is narrowed down from the starting point that is broad and include all possible information (situation, context etc.). I believe this model is a good example of a performance that was a “musical fit” (Hargreaves et al., 2005, p. 10) (when the performer and the listener connect through the music in a way that aligns with the performers intended expression and the listeners expectation) and can be seen in the light of the conduit metaphor. As explained in section 2.1.2, the conduit metaphor explains a simplified image of verbal communication, where the speaker is to be seen as the holder of the message, and for the receiver to understand that message in clear, the speaker needs to deliver it in a way that is easily comprehensible to the listener (“musical fit”). However, this is not always the case, and I would suggest some changes to the Reciprocal feedback model of musical communication for it to be applicable for all musical performance (figure 13). Instead of the figure being two triangles, I suggest a rectangle containing the model of Hargreaves et al. (2005) and drawing in some horizontal lines to underline the different levels of communication.

![Diagram of the reciprocal feedback model for musical communication by Hargreaves et al. (2005) with the authors alterations on levels of communication](image)

**Figure 13** The reciprocal feedback model for musical communication by Hargreaves et al. (2005) with the authors alterations on levels of communication
By placing the model inside a rectangle, one illustrates all the possible grounds for communication. The lines indicating a person’s point of view in the given situation, taking both the situation, personal history and preferences in context, and how it might be difficult to hit the spot of musical fit when the performer is taking on a different view. (If a listener’s viewpoint is at the far top of the right triangle, it is a long way to the point of coalition if the performer’s intention is at the lower left, and vica versa). In the case of these experiments, the Reciprocal feedback model of musical communication is not applicable for all the participant’s experience of the performance, maybe even none, since this point of communication is highly subjective and relies on both the performer and participant to find the message conveyed expressive of some matter. If I were to ask the participants after watching the videos in what degree they felt that the performance was of their personal liking, we could have investigated if this point of musical fit was attained with some of the participants, but this is a different aspect of musical communication than what is examined in this thesis. I therefore question the use of the model in any situation that relies on subjective interpretation of a complex communicative matter, such as musical performance.

I will also raise a question about the using the term listener in the Reciprocal feedback model of musical communication. When talking about music performance, and not just the audible music, I believe a better term would have been audience, which is not connected to one sensual impression. This because of the surrounding environments (the venue, audience, lights, stage and musicians) are all a part of creating the expression that the audience is picking up during the whole concert (confer section 1.1). (This is also the case when only listening to the audible music, since we are always affected by our surroundings, and our personal emotional life.) If looking at the comments by the participants, it is clear that some of them also was influenced by the surroundings in this situation (like “the black clothing and the white background create contrasts” and “the screen should have been bigger” that was excluded in the limitary process). This means that being a part of a concert is more than just taking in the audible information, as suggested when using the term listener, and I mean that the term audience might be more suitable.

5.3 The performance analysis

After repeatedly watching the videos and analysing the gestures and movements according to the categories applied, I found that the movement categories adapted in the thesis is not
satisfying for analysing the movements of a vocal performer. Some of the categories applied in the video analysis, like communicative, are much too vide when it comes to the needs of this thesis (44 out of 48 gestures in the angry video, and 28 out of 30 in the sad video is labelled as communicative). These terms from Jensenius et al. (2010) are derived from studies of instrumentalists, and it might be that they are not suited for describing a vocalist’s movements and gestures. The vocalist is often freer (in that sense that they do not have an external instrument to focus on) and are concerned with conveying and communicating lyrical content, in addition to the melodic and rhythmical structure. The category of communicative is combining gestures for both internal and external communication (both with the audience and co-performers) so for a satisfying analysis of vocal performances, one has to use sub-categories for further explanation of the gestures, as done in this thesis. For an instrumentalist, one might be able to distinguish between the gestures made for sound facilitating and communicational purpose clearer because of the physical instrument. As mentioned in section 3.3.4, the sound-producing gestures of a vocalist is not possible to see, since one can open the moth without making a sound or even sing with your mouth shut (hummin), which also affects the sound facilitating. This is also a reason why, in this analysis, the label of sound facilitating is used fewer times than others (7 times in the angry video and 3 times in the sad) and almost always combined with other categories.

After doing the video analysis, I found that the use of the illustrating label aligns with how we use our hands in verbal communication. According to the information given by Davidson (2005) in section 2.1.1, we use our hands in verbal communication to substantiate the words of importance (the message) to the recipient. In this analysis, the label of illustrator is used when the hands, or legs, accompany the rhythmical pattern of the melody, substantiating the desired message, and especially when my performer intentions were to substantiate a certain word, or help with a vocal technique. Even though this is a subjective analysis of my own body language, it might be that a lot of the research on verbal communication and its gestures might be applied for further research on vocal performances in music that includes lyrics. (It might be that in music which do not include lyrics, the gestural repertoire serves a different purpose, as mentioned with the trumpet player in section 2.4.) In this analysis, the illustrating label serves a purpose of rhythm/melody/audible substantiating gestures, but it might also be accurate to label rhythmical gestures with emotional meaning as illustrators. What I found when analysing the movements in the angry video, are that the movements of the legs (which
is not a conscious movement from my perspective and can be seen as emotional “leakage”)
serves both an illustrating and affect displaying purpose (00:38-00:40). This is only
noticeable from a first-person (subjective) perspective, since I know that this is not a
rehearsed movement, and the analysis of the movements might have been different from a
third-person (objective) perspective. In this case, I believe that the analysis would have
benefitted from a label that included both of these categories. If we had a category that
included both the rhythmical nature of the gesture and the emotional leakage that it derives
from, it would have been easier for the reader of the analysis to understand the nature of the
gesture without watching the video. This gesture of lifting the foot (at 00:38-00:40, figure 11)
is now labelled both illustrating and affect display, whilst some of the other gestures of lifting
the foot is only labelled affect display. The difference in these gestures are the onset and
rhythm of the leg. When the leg is lifted it is held throughout the word it substantiates, and
most of the times this means that it is held above ground for a longer period of time (at least a
second). Whilst the gesture of lifting the leg on a short word (like at 00:38-00:40) is seen as a
rhythmical gesture (illustrator), but also as an emotional leakage (affect display). (This
applies for other gestures of affect display, mostly large facial expressions, but also hand
gestures.) In retrospect, I believe that the term adaptor could have been used to describe the
rhythmical movements that shows an emotional leakage, in a way that aligns with Dibben’s
(2009) usage of the term. As explained in section 2.3.2, Dibben (2009) explain that we all
have adaptor movements that reveals some of our inner emotional states, and which leaks this
information in encounter with others. Kurosawa and Davidson (2005) also use the term
adaptor, but explain them as personal habits, and as far as I could read they are not linked to
emotional leakage. Still, one can argue that emotional leakage is personal habits, as emotions
are individually expressed, still but personal habits does not necessarily need to be emotional
leakage. This means, that the term adaptors could have been useful in the analysis when
applied in a way that aligns with Dibben’s (2009) usage of the term. In addition to this, I
found that there are grey eras between the categories emblem and illustrator. As explained by
Kurosawa and Davidson (2005), the category of emblem is used when the gesture made can
replace a specific word and still be understood as that word by the recipient. As explained
earlier, there are several gestures in the video analysis that can be seen as “on the verge of
emblem”, which is a combination of several categories like illustrating, communicative,

31 Still, as explained in section 2.3.2, the basic emotions have a lot of internationally recognizable features both
in vocal ques, and facial and bodily expression.
sound-accompanying and emblem. These gestures might differ slightly from the person interpreting it (like the gesture of “slamming” at 00:27-00:30 in the angry video) and do not exclusively mean one word in particular. But, to only label it as illustrating would be incorrect since it is executed with the purpose of substantiating and explaining the word, not the rhythm of the word. (You would not interpret the gesture of “slamming” as “heart” or “me”.) It might be that when analysing a vocal performance using lyrics, one could benefit from differentiating the gestures that are lyrics substantiating of a narrative character (illustrator and emblem like the gesture of “slamming”) and the rhythm substantiating gestures (illustrator). Some of these terms used in the video analysis are therefore not sufficient for giving a profound explanation of the body language in vocal performance that is analysed from both a first-person and third-person perspective. The usage of these terms might also differ according to the person applying them, but it might benefit future research to continue the experimentation and elaboration of terms for analysing the movements of a vocalist.

5.4 The taxonomies

Two of the theoretical foundations on the issue of music and emotions in this thesis was the GEMS model (confer section 2.3.4) and the circumplex model of affect (CMA) (section 2.3.3). Both the GEMS experiment and the CMA was executed with only voluntary participants, which is the same foundation as in my own study. All my participants are voluntarily studying at the Department of Musicology and participating in this experiment. (With this in mind, we can conclude that all of the participants in this study where somehow intrigued about joining the experiment, or else they would not have participated.) Already here, we have a deviation. These 24 people have already shown interest above average in the field of study, since they have chosen to study musicology, in addition to that they all planned on spending time on the experiment (as opposite to just stopping people at a festival asking them questions without any further notice as with the GEMS experiment). However, the main difference in the experiment in this thesis and the GEMS model, is that the GEMS model concerns the emotions that are commonly evoked in response to music (subjectively
embodied emotions), whilst this analysis is concerning the emotions that the participants are perceiving from the performer (subjectively perceived emotions). This means that the field of study is actually quite the opposite (which is substantiated by the participants replies on the performances). The CMA, on the other hand, concerns bodily emotions and was not conducted with any artforms in mind, but as a sociological experiment. The difference in field of study is even clearer when looking at the emotive states that the performer intended to convey in the videos (anger and sadness). In the GEMS model (figure 6, page 28), anger is not used when describing emotions felt in response to music, but as explained by both Patrik N Juslin (2005) and Camurri et al. (2003), anger is used for describing an emotive expression in music and music performance. The difference is then that the music itself does not evoke the emotive state of anger in the listener (if so, it would have been included in the GEMS model), but the emotive state of anger might be perceived in the performance by the audience. Anger is described in the CMA as an affective state of high arousal and misery, which makes it an embodied emotion with high energy level. Sadness, on the other hand, is found in both taxonomies. In the CMA, it is on the scale of low arousal and misery, and in the GEMS model it is placed under unease, which indicates that the emotional state of sadness is both perceived in the music and felt in response to music, but in a low energy level.

If analysing the adjectives from the sad video compared to the GEMS model, there are some similarities (words found in the GEMS model in parenthesis). Sad (sadness), calm (peacefulness) and tenderness are words used by the participants in this experiment to describe the performance. Besides from the adjective sad, none of the adjectives and nouns used by the participants matches the ones from the GEMS model accurate, and they are in need for interpretation to be placed in the GEMS model categories. “Calm” (as used in this experiment) and “peacefulness” (used in the GEMS experiment) might be two very different states of mind, but there is still a description of something that has a low energy level. When I tried to find similarities between the descriptive words in the angry video and the GEMS model there was only strength and energy (in the GEMS model they use the word strong which is found under the category energy). This means that there is very little consensus between the emotive adjectives and nouns in the GEMS model and the emotive adjectives and nouns found in this experiment. If looking at the participants comments in the light of the circumplex model of affect (CMA), it might be a slightly higher correspondence. Some of the
words in this experiment that corresponds with the CMA are calm, sad and angry (both the emotional states that I intended to convey). In addition to these (that is found in the CMA without any interpretation), a lot of the words used by the participants confers with the idea of arousal/sleepiness and pleasure/misery, and have a strong bodily connection (like calm, energy, pain, strength etc.). This means that the participants have perceived my body language as some kind of valence, concurring with the scale of arousal and sleepiness, in addition to perceiving these movements or gestures as pleasurable or painful. I believe that the reason for this is the different fields of research, as mentioned earlier, where the GEMS model concerns the participants interpretation of the audible sound, whilst this thesis concerns the participants interpretation of the physical performance. This said, a lot of the words from the GEMS experiment concerns adjectives and nouns that might be used for describing a physical state of the body (like relaxed, nervous and irritated among others), but the layout of the CMA is easier to use as a foundation for the descriptive analysis.

When looking at the descriptions the participants made of the two videos it is clear that they convey different emotive states, and the only word that is found in the description of the performance in both videos is pain (five participants used the word pain to describe what they found the performer to convey in the angry video, and one in the sad video). Pain is not found in the GEMS model, but it can be seen in the light of misery in both a high or low level of arousal in the CMA. Depression (which is known to be painful), as used in the CMA, is a state of misery, as with being angry. But, the amount of energy in these two emotive states are often opposite, which makes it two extremities of the misery scale. In the sad video, there is only one comment at the very end that uses the noun, and in the angry video it is used more overall in the video.

5.4.1 Discussion of the words pain and sadness

There is a marginally higher frequency of the same adjectives used in the sad video than in the angry video. In the sad video, ten of the words are repeated by different participants versus nine in the angry video. (I believe this can be seen in the light of the amount of movements that can be interpreted differently in the angry video, versus the sad video that has less movements.) But, only two descriptive words are used in both videos, sadness and pain, and I find it important, and interesting, to further explain how the participants used these nouns.
Comments on pain in the angry video

The comments on pain in the angry video is particularly interesting. The first comment (at 00:30-00:35) is where I deliberately changed my mood and body language from anger and fury to express frustration and vulnerability, which I, subjectively, associate with the emotional state of pain. The comment reads: “She turns away from the microphone. Can be interpreted as pain, for example” (00:30-00:35). In this particular place in the video the performer pulls her head away from the microphone just for a second on the word “meant”, and I have called this particular gesture a communicative, illustrating, sound facilitating and affect displaying gesture, but it can be seen as emotional leakage (Dibben, 2009 p. 323). This because the movement in itself is not intended as an expression of pain, or anger, (it is merely just a technique for volume control that is so rehearsed I no longer think about it), but the emotive state I try to express at this point in the performance is the one of frustration, and anger. It is then a possibility that this technique is more expressive than I believed it to be. Another participant wrote a more diffuse comment using the word pain, in addition to quote the lyrics at that point in the performance: “Pain; I can’t take this any longer” (00:57-01:01). Once again, it might be that he finds that the lyrics expresses pain, or it can be a merge of impressions. A lot of the participants have recited some parts of the lyrics, but very few have written down a whole sentence. I therefore thought that he probably was shown the sad video first, and that he perceived the lyrics because he had heard it once before. This is not the case, since he was shown the angry video first, which means that this time in the video is the first time he ever hears the lyrics. It might then be that he is particularly interested, or just able to, catch the lyrics only based on audio, but it also might be that something in the performance substantiated the lyrics at this point and made the lyrics clearer. Since he is not just reciting the lyrics, but also writes pain, it is natural to conclude that he perceived something else at this point, in addition to the lyrics. The word “pain” is used in the second phrase in the chorus, and he might have noticed the lyrics (unconsciously or consciously) and found that it fits his interpretation of this part of the performance. One of the other comments on pain is also interesting: “Strong and energetic even though the lyrics is about pain” (01:48-01:53). At this time in the angry video, the lyrics read “I hold my head up high cause the pain just makes me stronger”. Since both the words “stronger” and “pain” is used in the lyrics it is probably not a coincident that the participant uses these words to describe what she finds is communicated, as with the comment at 00:57-01:01. At this point in the angry video we just reached the second chorus and the gestures are rapid and large. The chorus starts with right
hand doing a short gesture upwards on the word *head* and the left hand gesticulates rapidly throughout the phrase (described as communicative, illustrators and affect display). On “stronger” the whole upper body is doing a small crunch while the face makes a grimace (affect display). But, the same argument of what is described, the audio or the physical movements, applies here. The participant gives no clue to what she describes as “strong and energetic” and this could be both the body language in total, one set of movements/gestures, the posture or the voice. At this point in the audio, the vocals are increasing in volume to make a clear message of fury, anger and frustration in the chorus, to separate the verse from the chorus, and to be able to hit the high C in chest voice. This volume increase also comes naturally with the amount of energy being used physically and creates a natural symbiosis of energy in movements and vocals. To say that the comment is affected by both the video and the audio is therefore a natural conclusion. The other comments that includes the word pain is more focused on the facial expression than movements and makes it clear that the facial expression/grimace of the performer is giving them the impression of pain. This means that most of the comments that includes the word pain are concerned about the body language in some way, either a large grimace or the movements used.

**Comments on sadness in the sad video**

The first comment on sadness reads “Sad opening; The singer looks to the ground, seems like she is going to convey something sad” (00:06-00:15), the second “Differentness; The lyrics and the expression seems to knowingly be mismatching and creates an effect of sadness” (00:56-01:03), third “The melody (the register) is giving me a feeling of sadness and longing” (01:41-01:45) and the last “Introvert vulnerability; The singer rounds of as in the beginning; no eye-contact- looks down- introvert body language- gets a sad/dejected feeling. It’s sore and painful” (02:09). The first and last comment are both explaining that the way the singer avoids eye-contact and looks to the ground creates an impression of sadness (which one of the participants found to be painful), and the third (01:41-01:45) is explaining that it is the register of the melody that creates this expression. The second comment on sadness (00:56-01:03) is more focused on the lyrics, but he does not explain if he means the lyrics in an overall view or at that exact point in the video (the lyrics read “pretend I never cry I can’t take this any longer” at this point). He is neither saying whether it is the body language that conveys something else than the lyrics, or the audible sound and vocals. What is interesting with this comment is that he obviously read the lyrics and the general performance
expression as two different emotive states that, when put together, creates a sad expression. The diversity in the emotive state of the performance is not intended from my perspective, and just underlines how we all perceive the performance from different perspectives. From my point of view, the lyrics expresses a suppression of emotions which is something I consider to be sore and painful, but I can see how this also can be interpreted as something else. I am reading the comment at 00:56-01:03 as a more general description of the video so far that gives an impression of sadness. This means, that the introvert body language, and the action of looking down, was perceived as sad by at least three participants, when combined with the lyrics and melody. This concurs with the findings in both the GEMS model and the CMA, that sadness can be used for describing both music, and a physical state.

5.5 Visual versus audio

The findings in this thesis can be seen in the light of “sight over sound” (section 2.1.4). Out of the 96 comments on the angry video, 64 concerns the body language used, a movement in particular or the movements in general. In the sad video, 50 out of 84 comments concern any type of movement, gesture or body language. (As mentioned in section 4.4.4, some of these comments include some audible remarks and these also included in the categories of lyrics or sound of voice.) Of the 96 comments on the angry video, only 22 includes a direct comment on the lyrics. Only four of these comments did not include any of the gestures or movements made at that time in the video, the rest of the 18 comments on lyrics all did mention a gesture made at the same time of the word described. In the sad video there are 19 comments on the lyrics, of a total of 84 comments. Only one of the 19 comments in appendix 2 is not commenting in direct on the lyrics in combination with a physical gesture or movement. When it comes to comments that focuses on the sound of the voice, and not exclusively the words, there are 21 in the angry video. Most of them are a combination of both the sound of the voice and the words or gestures that went to it, others, though truly outnumbered, commented only on the voice. In the sad video, there are 25 comments on the voice or the audible sound, and most of them concerns the voice, and audible expression, and not the gestures that went along with it. This means that there is a marginally higher amount of comments on the voice in the sad video (25 comments) than in the angry video (21 comments), in addition to the comments in the sad video being more concerned about the voice, and not as much the body language. This indicates that the participants were more concerned about the sound (or focusing on it) in the sad video than the angry. One
explanation for this can be the increase of movements and physical details happening in the angry video compared to the sad. The performers body language is moving a lot more in the angry video, which might have distracted the participants from paying attention to the sound (like the comment “Still a lot of movements. It’s very disturbing and I cannot focus on the song!” [01:04-01:07]) or just focus on the visual information rather than the audible. Still, only three of the twenty-four participants focused mainly on the vocals or audible sound and did not comment directly on what they believed was communicated, or the visual aspects of the performance.

After the discovery of the visual versus auditive comments, I went back to the preliminary schema to see just how many comments who focuses on the voice, included those who were left out from the appendix. In the angry video, 24 comments are focusing solely on the audible performance when including the ones left out, and in the sad video there are 29. This means, that the comments on movements is still almost three times as high as the ones that includes remarks on lyrics, or the audible vocals (in total). It might be that the information by Davidson (2001) about how our vision is the dominant perceptual sense (section 2.1.4) can hold some truth. As mentioned in section 2.1.1, we first start learning via our visual sense, adapting to other peoples’ body language before start learning words. Applied to the findings in this thesis it might be that the visual sense is superior to the audible in a performance setting, but we cannot know if anyone is ever capable of distinguishing between audible and visual input in a situation like this experiment.

5.6 Subjective differences

One of the reasons why I wanted to do this experiment is to be able to grow as a performer. I find that it is easier to convey song characters that are not so closely related to my personal self than to reveal me as a person, and this means that I often take on a more expressive personality on stage than what I believe my personal self to be. In this experiment, I tried to be more of my real self in the sad video, and this might be just what the participants picked up on when interpreting my performance as “painful”, or “theatrical”, since I was a bit uncomfortable with revealing my personal self in this way. This might also be a good explanation for why some participants found my performance in the angry video “theatrical”, because it partially is theatrical in the sense of me using a song character based on me as a
person, and not as much myself as a person, as in the sad video. Davidson (2001) explains that

there can be a “surface” level of movement – a kind of rhetoric – which the performer ads to the performance. On the negative side, if these gestures are not consistent with the intention of the performer, they can create physical tensions which may inhibit technical fluency and disturb observers with the incongruity between the gesture adapted and the performance intention. (p. 240)

Most of my songs are complex and differs in the emotive expression from one part to another and this way of separate the emotional themes is not common ground for me. Davidson (2001) explains that if a performer is using too many movements it might seem exaggerated, and when standing still it might be boring. In the experiment, there are eight comments in both videos that are expressing their disliking of the performance, or parts of it.32 Approximately half of these are concerning the participants dislike of the audible sound, and the remaining half are concerning the physical performance. In the angry video, there is one man and two women who expresses their dislike of the amount of movements: “The movements in the instrumental parts destroys the impression of calmness and the impression given earlier. Could stand more still” (01:08), “Too much movements in hands and dance in comparison to the music” (00:27-00:30) and “Theatrical; gives good contact and presence, but too much body language” (01:27-01:28). All of these comments explain that the performance given in the angry video is not of their personal liking. Both females who criticized the performance checked of the box for pop music, among a couple of other alternatives, and are spending between 16-20 hours a week consuming music. This means that they are both in the average of the females who participated in the experiment, and there are no indications in the questionnaire that they should have any other experience with the performances than the others. Both of the female participants, who uttered that the movements were a bit “too much” in the angry video, also commented on the same subject in the sad video: “Movement, stands still with the legs but a lot of good movements in the upper body that shows that the artist enjoys the music, but gets stiff with so little movements in the legs.” (01:41-01:45), “The singer seems to be very focused, she is getting into the “feeling” of the song, but it seems a bit fake to me (generic way of holding a microphone and swaying to the music.)” (00:06-00:15) and “I start to empathize, getting used to the style of singing

---

32 When included all comments that I, subjectively, interpreted as negative, like “The vocals feel somewhat closed” (00-36-00:38), which also concerns the audible sound. Still, it might be that the participant did not mean this in a negative way, and that if I myself was not the subject of investigation in the experiment, the interpretation of the comment would have been different.
and moving and believing more in the singer, although the movements seems to be a bit exaggerated” (01:03-01:04). What would have been interesting is to hear what the participant who wrote the latter comment means by “believe in the singer” and if it is the specific movements (holding the mic and swaying) that she finds to be exaggerated, or the overall impression of the videos. As the comment stand now, I can only assume that she is explaining her overall impression of the performance, and that she found the movements in total to be too much. That said, she is the only one who has pointed out that the movements in the beginning of the sad video seem fake, which is something I agree on myself. The performance of this song has (when performed in concerts) always been explosive and energized, and I was struggling with converting the performance from the normal way I do it, to a “sad” – or introvert - version. This process was maybe rushed a bit and created a mental situation that was not normal for me during the recording of the video. Therefore, I find the first comment about the performance being perceived as “fake” (00:06-00:15) interesting, although I don’t agree on the action of holding the microphone as generic and fake. (The performer holds the microphone with her right hand, and left hand are held at the middle of the stand, which is a normal pose that most singers do from time to time, especially in slow songs.) Her observation of how “the singer getting into the “feeling” of the song” is quite right (since what I intended with these movements is to get into the feeling of the song), and I can agree that the body swaying is a bit fake for the reasons mentioned. Besides these two comments on how the movements are a bit fake for the reasons mentioned. Besides these two comments on how the movements are a bit fake for the reasons mentioned, two males write that it “Is perceived as very anonymous” (01:41-01:45) and “The expression is more homogeneous. Music, lyrics and appearance says the same. It’s a little less interesting” (02:01-02:04). The participant who wrote the comment at 01:41-01:45 explained after the experiments was conducted that he did not know what to look for, and the other participant (02:01-02:04) commented earlier in the video that he found the expression of the performer to be more introvert (00:06-00:15).

Introvert was a part of what I tried to convey, which means that the participant who found the expression introvert made a reading of my body language that aligns with my intention, though it was not of his personal liking (02:01-02:04).

5.7 Significant moments

I have chosen to call this section significant moments, adapted by Trondalen (2004), to explain and discuss certain moments in the video experiments I find particularly interesting.
5.7.1 Gestures and movements as emotional leakage

The first interesting place in the sad video where the combination of lyrics and gesture seems to have been particularly effective for communicative purposes is at 00:28-00:31 where two female participants have commented “Substantiate the lyric my heart by touching her heart” and “Touches her heart at the same time as the lyrics read my heart. The hand kind of helps the lyrics”. The lyrics read “when I slammed the door to my own heart” (00:28-00:31), and the performer does a small hand gesture with left hand that imitates the action of “slamming” something, before the hand is moved close to the heart (figure 14). The two gestures of slamming and touching the heart (at 00:28-00:31 in the sad video) are both described as communicative and illustrating, but there are still some differences between those two. If looking at them objectively it seems like they both are made for the same purpose, that is to underline the lyrics and accentuate the meaning of the phrase. But from my subjective point of view, the first gesture (slamming) are made deliberately and the second (touch the heart) is not. In the angry video, the performer does not use this movement (holding her hand to her heart) but the slamming gesture with the right hand are deliberately exaggerated. It was not until I read the two comments that I saw my own gesture of touching the heart. (I was actually sure before I looked at the video that the two participants had misinterpreted it.) But, there is a clear view of the left hand being held in front of the heart, and this means that I cannot have done this gesture deliberately, and it therefore derives from my natural gesture repertoire, which can be seen as emotional leakage. Both participants who picked up on this combination of lyric and gesture are women who spend on an average of 15 hours a week listening to music, and they both checked the boxes for pop and rock music. (One of them checked off the box for blues and others as well, but since they both only checked off two and four boxes I assume that this is their preferred musical taste.) Since they both checked the boxes for pop, I have to assume that they are familiar with the genre performed, in the light of the information given in section 3.3.5. This might give them a more nuanced view of the performance since they were able to pick up this small gesture and put it into the lyrical context. Both these participants were shown the sad video before the angry, but they only commented this word, heart, in the sad video. The reason for this might be that the performer does not use this movement (holding her hand to her heart) in the angry video and ergo there is no illustrating gesture
made at this point to the lyrics. It might also be that this gesture, as emotional leakage, were more of their own personal liking than other rehearsed communicative gestures. At this particular place in the video (at heart), is the only place where they commented at the same time, but, they both make other significant comments that states that the gestures made are substantiating the lyrics, like: “Body sigh; the body fell down I can’t take this any longer-continuation of the lyrics. Her body can actually not take it any longer.” (01:03-01:04), and “Hand movement; shows frustration clearly with the hand gesture towards the end where she reaches her head. Substantiate the lyrics.” (02:01-02:04).

The second interesting place in the sad video is the beginning of the first chorus at 00:48-00:54, where the performer sings “I hold my head up high” and lifts her right arm towards her head whilst her shoulders moves upward (figure 15). At this point in the video, two participants commented on how the gesture made by the hand help substantiate the word head: “Touches her head when the lyrics read down my head” (the participant wrote wrong lyrics, but the word head is still correct), and “Sings something about the head and lifts her hand slowly to her head”. There are also five other comments made at this point in the sad video whereas two more (in addition to the two mentioned) perceives the word head but does not make a clear connection of the gesture made and the word. Since this is the beginning of the chorus it is a natural increasing in the soundscape with an extra guitar track coming in, and a transition from chest voice to head voice in the vocals. There is no noticeable change in the body language or movements of the performer, and I believe that the dynamic in the audio has a lot to do with this cluster of comments at this particular time in the video. Even though none of the participants mention this change in the soundscape at this chorus, it is staring that so many commented in this particular place.

The last place where several other people have commented on the substantiating gesture in the sad video is at 01:03-01:04. This is the end of the first chorus and when the last phrase ends, the performer drops her whole upper body as a communicative and affect displaying gesture (figure 12, page 75). This gesture is the same as in the angry video, but much smaller. It started out as an unintended gesture, or emotional leakage, but once I was aware of it I
started to use it as an “affect displaying illustrator”. In this video recording, it is done deliberately to emphasise the emotive state of being tired and giving up, but it still derives from my natural gesture repertoire. At this point, seven participants commented in total, whereas five of the comments include the movement made. Three of them are combining the words sung with the movement: “Body sigh; The body fell down I can’t take this any longer-continuation of the lyrics. Her body can actually not take it any longer”, “Dropped her body after the phrase. Matched the lyrics.”, and “A kind of climax in the chorus is given weight by the one hand “spelling” the words before the vocalist is pulling away dramatically”. Several of the participants have described this movement as a body sigh, or a collapse, which I find very telling. Other comments explain that the emotive state of giving up, or just a general expression of heaviness, is shown by this particular movement; “The body rushes to the side; when she sings can’t take this any longer she moves slowly to the side until her body “collapses”. Looks heavy”, “Dejected movement: almost like a sigh”, and “A bit more helpless now. I can’t take this any longer. Earlier on it seemed angrier, more controlled, now it seems like she has given up”. This particular movement is clearly communicating the intended mood of “giving up” to some of the participants, in addition to substantiate the lyrics.

Another significant moment of substantiating gesture is concerning the leg gestures in the angry video. I lift my right leg seven times in total in the angry video (all during the first verse and chorus, 00:15-01:03), in comparison to none in the sad video, all of which is described as an affect display (appendix 1). The leg gestures are not something I have rehearsed or put in for the purpose of conveying or communicate anything, and they are described as emotional leakage that derives from my natural gesture repertoire (section 3.3.3). Some of these movements are small and barely noticeable, but there are several places where these leg gestures are interesting. The first large leg gesture of importance happens in the first pre-chorus, at 00:38-00:40 in the angry video (figure 11, page 72). After this, several large leg gestures are made, one in each lyrical phrase to be exact (00:40-00:42, 00:43-00:45). The first one, at 00:38-00:40, happens on the word what, which is the accentuated top note in the phrase. This movement is big and are combined with a downward going hand gesture. These gestures are not intended, but the emotive state that it derives from is. The communicative purpose of the bodily expression in total are of conveying the emotional state of anger and frustration, and the body movements that happens is an unconscious reaction to
this emotive state. If we apply the theory on why instrumentalists bend their back while hitting the high notes (as implied earlier by Auslander, 2009; Dahl et al., 2010) the leg gestures made in the angry video at 00:38-00:40 can be of sound facilitating reasons, in that sense that it helps create more energy that is needed for accomplish singing in this volume in this area of the vocal register, in addition to communicative, affect display and illustrating purposes. What is interesting is some of the comments made by two participants at the point of the first pre-chorus in the angry video: “Foot is lifted up, gives me the impression that she feels the song” (00:38-00:40), and “A lot of movement to express that the song means something to her” (00:38-00:40). Both of these comments are pointing out that the movements made in this part of the video are giving them the impression of authenticity of the message sung. At this particular point in the angry video, it is clear that my movements (both my right arm and leg) helps me convey the inner emotional state to the participants. What is interesting is that the participant who made the comment “Foot is lifted up, gives me the impression that she feels the song” left ten comments in this video, but only this one concerns some indication of an emotion being expressed. Even though it is not a direct comment on what emotion he is perceiving, he is getting the impression that the performer “feels” something. All the other comments made by this participant are merely explaining what happens in the video, like “arms move in a natural flow” (00:27-00:30), “head a bit up when she sang the lyrics head up” (00:47-00:49), and “falls down here as well, but with attitude” (01:02-01:03), which means that the emotional leakage of lifting the right leg is the only place where he describes an emotion being communicated. The latter comment contains some information about how he perceives the singer, but the adjective attitude is not compelling with one single emotional state. This participant uses the adjective attitude two out of ten times to describe the performance, once at the beginning (00:02-00:03) and this other time after the first chorus, whilst in the sad video he is not using it. This means that he at least reads the body language of the performer in the two videos differently (even though I cannot say if he uses the word attitude as a synonym for engagement, adaption, posture or the all the above). At 00:55-00:57, another participant commented “She uses her legs to express emotions. Feels like she can’t hold it in any more”. At this point in the video we are at the end of the first chorus, where I use my legs the most. Like

Figure 16 the performer's leg movement at 00:55-00:57 in the angry video

96
mentioned, I do not use my legs in a conscious way, so the readings that these participants have done of my leg movements are based on emotional leakage. That is, I literally couldn’t hold it in, and my legs became the visible cues where my emotional state was manifested (figure 16).

It is clear that these movements (touching the heart, lifting the leg and “body sigh”) helped substantiate and communicate a mood to the participants, even though it was more or less unintended from my perspective. But, it can be described as emotional leakage just because it was unintended, which might be the reason why it was received by so many participants – I, as a performer, was in a state where I did not control my movements and just enjoyed performing, which might be found realistic. As mentioned, I have used the technique of visualization and method acting when rehearsing and recording the videos, and as explained by Woody and McPherson (2010), these techniques might help performers communicate emotions and moods to the audience. From the findings in this experiment, it seems that the gestures and movements that derives from the natural gesture repertoire can serve as emotional leakage, that is easily detected, and found expressive in some manner, by the audience. This means, that if performers use techniques (as visualization) to access the natural gesture repertoire in different moods, it might benefit the communication chain with the audience.

5.7.2 Mood induction and attunement behaviours

The movements made by the singer in the beginning of the videos (00:02-00:15) has a sound-accompanying purpose of mobilizing energy and increase the tension to prepare the intended volume of the vocals, by using simple dance moves (swaying the body from side to side in both videos, and lifting arms upwards in the angry video, and holding the microphone stand in the sad video [00:02-00:15]). In addition to this, the movements can be seen as communicative and illustrating, for the purpose of establishing and substantiate the rhythm and groove in the guitar in both songs, along with communicating the energy and vibe in the performance to the audience. In the sad video, the swaying is on the fourth and half notes, but in the angry video, it’s on the eight notes. This means, that in both videos the communicative, sound-accompanying and illustrating gesture of swaying are done for the same reasons, which is to build up under the different moods in the two performances. In the angry video, a total of six people made a direct comment on the body language used and the mood that was
established the first fifteen seconds. Some of the comments where more general in its description of the mood established, like “The motions of the singer make me anticipate an energetic song” (00:07-00:11), “A lot of movements, gives the impression of a lively song” (00:07-00:11) and “Attitude from the start” (00:02-00:03). Here, the movements actually reach its purpose of communicative and sound-accompanying, since they are intended for conveying energy. Lively is here used to describe a briskly mood (in Norwegian “livlig”), which I read as a positive word used for describing a happy, energized mood. This means that the movements in total, the first twelve seconds, gives the participants this impression of a positive, energized mood. In addition to the comments mentioned, other comments were “determent; looks up, more open” (00:02-00:03), “Much more groovy movements of the singer bids to dance” (00:12-00:15) and “Eyes open, shakes her head and body and look around. Expresses more openness and attention. Makes the audience wake up” (00:04-00:06).

In the sad video, five participants uttered that the posture and body language of the singer conveyed something of a more melancholic and introvert vibe (00:06-00:15). These comments are much more specific than the ones at the start of the angry video: “The posture; a kind of vulnerability- no eye contact- looks down- “tight” body language”, “vulnerability; how the singer presents herself with how she stand/poses signals vulnerability”, “sad opening; the singer looks to the ground, seems like she is going to convey something sad”, “posture; it seems to be a personal, serious song”, and “more introvert; looks to the ground” (all from 00:06-00:015). When starting to analyse this, I thought that the order that the videos were presented to the participants could have something to do with how they interpreted the beginning, or that it affected the first impression given in the videos. This does not seem to be the case, since there is an equal number of participants (who commented on the posture and mood in the beginning of the song) who were shown the angry video first, as for the sad. These findings (though only a small amount of the participants commented on the beginning) indicates that the posture and movement of the performer is equally important before he/she start singing, as during the vocal accomplishment, for establishing the intended mood. This might indicate a state of attunement behaviour (Stern, 1985), which could be the ultimate state of communication for a performer (you don’t want your audience to be sad, but you want them to understand what is communicated and maybe...
get emotionally affected in some way). If the performers body language is slow and introvert, the audience might perceive a mood of sadness – or vulnerability – which they expect the song to follow. They might even non-deliberately imitate the body language of the performer, but they will not necessarily become sad (although they might).

5.7.3 Centre of moment

In the light of my own findings, the centre of moment (explained by Davidson (2005) in section 2.4) might be the waist, as explained with the study of the pianist, or the core, that holds the function of expression. This is clear in the findings of the leg movements, where I constantly crunch my stomach while lifting the legs and pulling one elbow toward my core on the notes that I want to substantiate, are higher pitched, or simply needs more power (figure 11, page 72). This increase of movements from the core in the angry video can be explain with the need for more support since I am singing in a higher volume and register, as explained in section 2.4 by Dahl et al. (2010). In the sad video, the centre of moment is nearer the waistline, since the need for respiration support is smaller (the volume is lower). This means that the theory of Seashore (1923) on how the vibrato, or the breath (as explained in section 2.3.2), is the holder of emotion in vocalising, might not be an aberration. The ability to express emotions is closely connected to the ability to use the support techniques required for conveying different emotions, which comes from the abdomen, diaphragm and core muscles. This is somewhat simplified, since we know from the studies of Ekman (1999b) that facial expressions play a huge part in conveying different emotions – and ergo is essential in musical performance – but this is not mutually exclusive. Even though facial expressions of basic emotions are international recognisable and can give an indication of both unintended and intended emotive states, it is the combination of support technique, vocal placement and volume that in combination creates an overall impression of emotive state in a vocal performance, which will affect the singers body language, and to a certain extent, the movements.

5.7.4 Eye-contact

Woody and McPherson (2010) states that eye-contact is the most important way of expressing emotions (confer page 24). One of the participants uttered at the very end of the angry video that she “Did not comment since I did not see the face clear, and neither the eyes of the performer”. Because of the equipment and media used in this experiment (a 13-inches
screen and the software showing the video on half the screen), it is hard to create a feeling of eye contact, in addition to the video being filmed without audience. If this were a recorded concert, there would probably have been several points of eye contact, or at least “imagined” points of this type of communication. But, what is interesting is that in both videos there is equally hard to see the performers eyes, since when they are not closed they tend to focus on the microphone stand or at the sides of the room. There are just as few moments in the sad video, as for the angry, where the performer looks into the camera (which can give the participant a feeling of eye-contact). The lack of eye-contact in the sad video are intended, and I deliberately kept my head down for the purpose of creating an introvert body language and sad expression. But, when recording the angry video, I was consciously avoiding looking into the camera, since I find this uncomfortable. This might make the experiment deviate from a live performance, but, since the point about eye-contact was neither the focus of this thesis, I chose to rather focus on the differences of these findings in the videos. For some of the participants it is quite clear that eye contact is important, but this seems more so in the sad video than in the angry one. Most of the comments on the lack of eye-contact in the sad video are uttered in the beginning and the end of the song, where there are no vocals, but the performer keeps her eyes closed almost throughout the entire performance. At the end, two participants utter that they “Don’t feel like being sung to” (01:49-01:57), and that it is “Depressive lyrics that substantiate that it makes you stronger, but a bit passive communication with the listener” (01:49-01:57) which is not a direct comment on the lack of eye-contact but might be seen in the light of it. The first participant makes three comments during this video, all of them concerning the audible track or his personal disliking of the appearance of the performer, and none on the body language or the direct visual performance. At the time in the video where he feels like he’s not being sung to, the performer stands with her eyes closed most of the time while singing and looks to the ground when not singing – as in the whole video – but there are less movements than in the middle of the video. The last forty seconds she stands with her left hand on the microphone stand and the right hand slowly substantiating the rhythms while held next to the microphone and let both hands down at the very end. The other participant, that explains that he finds the performer to be a bit passive in her communication with the listener, comments five times in the sad video, but this is the only comment where he does not concern any bodily gesture or movement. In the angry

33 On stage, or just performing with an audience, I have no problem with making direct contact, but in an artificial setting, like this is, I deliberately avoid it.
video, he makes two comments about eye-contact, “First clear moment of eye contact” (00:43-00:45) and “I get engrossed but start to miss eye contact” (01:35-01:37). Both these comments only explain the participants experience, or lack of experience of eye-contact, and does not explain what he believes the singer to convey when it happens, or why he misses it. His last comment in the sad video, “Looks down in the ground, does not feel like a part of the performance.” (02:06-02:08) concerns the gesture of looking to the ground. But, if we read this comment in the light of the ones written in the angry video, it seems that to him, the experience of eye-contact is important for his impression of the performance. (Maybe even more so than for the other participants.) Others has also commented on this, like “Introvert vulnerability; the singer rounds of as in the beginning; no eye contact- looks down- introvert body language- gets a sad/dejected feeling. Its sore and painful.” (02:09-02:18). It is clear, since there are a lot more comments on the lack of eye-contact at the beginning of the sad video (00:00-00:15) and how the singer looks away and her “passive way of communication” closer to the end of the sad video (01:49-02:18), that something is different at these points. Both at the end and the beginning, where the performer does not sing, she stands very still with only some small, sound-accompanying movements in the upper body and head, while holding the microphone stand. Her head is often held behind the microphone as an affect display intended as an expression of wanting to hide (as explained in appendix 2). Since the movements in general in this video is conducted with arm gestures for the purpose of lyric substantiating (both communicative, illustrating and sound-accompanying), it makes sense that there is a lack of gestures when the vocals stop. This means, that in addition to one less element to focus on (the vocals), there are also less gestures to focus on, and this gives the participants the opportunity to focus on other aspects of the video, like eye-contact. It also makes sense that eye-contact is more urgent in the sad video than the angry one, since there is a lot less movements here than in the angry video. The video is intentionally more static, and as several participants points out (00:21-00:23, 00:48-00:54) this creates a more introvert and melancholic vibe, which was the performer intention.

5.7.5 Lyric substantiating gestures

Since almost none of the comments on lyric or mood substantiating gestures are uttered at the same time in the videos, it is impossible for me to find out which movements or gestures the participants found mood creating. But, from the comments that reads “Point at herself again; gives more power and direct the attention toward herself and the message” (01:27-01:28),
and “She hold her head up to match the emotional state of the lyrics” (01:48-01:53) (both from the angry video), we can see that some of the gestures made (like pointing at oneself [figure 18] and holding the head up) were, according to the participants, in consensus of the lyrics at that point, which created an overall impression of cohesiveness between the lyrics, posture or the gesture made. In addition to the positive comments on movements already mentioned in section 4.4.4, one of the participants wrote a longer, more descriptive comment that can be seen in the light of lyric substantiating gestures: “It seems easier to comprehend the lyrics since so much of the syllables and emotions are substantiated by gestures and facial expressions. I think these aspects are more important than the fact that I´ve heard the song once before” (02:01-02:03). Even though he does not write down the lyrics in words (the lyrics in its precise formulation), I have to assume that he at least heard parts of it, or the whole lot. It is also possible that the participants who wrote these types of comments have their own interpretation of the lyrics and the meaning of the song, this I cannot control. Some of the comments on mood or lyric substantiating gestures are difficult to interpret, like “Many and large hand gestures and determined voice gives the impression of power and determination” (00:27-00:30). The mood this participant is describing, the one of power and determination, is not coherent with one particular emotive state (power can be both angry and happy) and is neither coherent with one gesture made. Still, a lot of the participants explained that the communicative and illustrating movements and gestures made substantiated certain words, whole sentences or a mood.

5.7.6 Exercising and body awareness

---

34 It has to be stated that I went to school with this participant for three or four years, and we are acquaintances. That said, he does not know me very well and I don’t find this colluding his participation. If he were the only participant who interpreted the video in a way that aligns with my intended expression, I would have to rethink our personal relationship and read his comments differently, but this is not the case since several others also commented in a way that I find to be interesting for the purpose of this thesis. What is interesting is that he wrote several comments on lyric substantiating gestures that I find align with my intentions in the angry video: “She uses her hand a lot to substantiate the lyrics. Uses the body in a kind of dance. Grimaces that makes you think of anger/evil” (0:20-0:21), and “says shake it off and does a shaking movement with her hand” (1:09-1:10).
There was nine people who replied that they exercised 1-4 times a month, and 5 people on each of the other categories named “no, I don’t exercise”, “1-2 times a week” and “3-5 times week”. All the participants who replied that they did not exercise are male. Four of the people who checked the box for exercising 3-5 times a week are male, and one female. My hypothesis, before running this experiment, was that the people who exercised on a regular basis – and certainly as often as 3-5 times a week – were more aware of the usage of body language and the complexity of movements in a performance than people who did not exercise. What is interesting is that the participant who checked off the box for exercising 3-5 times a week almost exclusively commented on the audible sound. Only five people, all males, checked the box for “No, I don’t exercise at all”, and almost all of them seemed to focus on the audible music. There are still some comments on movements, but they are more diffuse, like “small, but nice movements” (00:46-00:48, sad video), “Posture; it seems to be a personal/serious song” (00:06-00:15, sad video), “Solidarity; because of the comfort that the movements show” (00:30-00:35, angry video) or “cool hand gesture” (01:09-01:10, angry video). The rest of the comments were more focused on the voice or sound, like “substantiate death”, “cool use of the voice” or “more introvert singing”. Still, one of the participants commented a lot on movements, and this small group of participants aren’t enough to conclude with a correspondence between physical exercise and awareness on how we use our body language.
6 Conclusion

Now returning to the research questions presented in the introduction: Is it possible to delimit and intersubjectively identify bodily postures, gestures and movements that can augment or attenuate the audiences’ reception of the performer’s expressive intention? Which bodily postures, gestures and movements are the most effective conveyers of anger and sadness in the communication between the vocal performer and the spectator? Due to the experiment design, it was of course not possible to perfectly recreate how we normally perceive, and experience, a performance. However, there are a number of interesting findings from both the applied theory and the experiments. The results are tentative, as the sample of participants is too small to give a statistically significant answer to the research questions, but the research design (with free feedback from the participants) afforded a phenomenological validity – and depth – that might easily be lost in a quantitative multiple-choice survey. Thus, this study might also serve well as a pilot for further investigations.

To conclude this thesis, I will start with the most important findings from the literature studies. First, (a) the field of music performance studies would benefit greatly from a more developed taxonomy that focuses on intended – and perceived – embodied emotions in the physical performance of music, and not only on perceived emotions in the audible music. Based on the many findings expressing intersubjective commonalities regarding the participants’ perception of the valence and activation contours of the performances, I believe that such a taxonomy could benefit from applying the curve of high/low valence as a point of departure, as in Russel’s (1980) Circumplex Model of Affect. This model could also account for similarities between the many seemingly disparate adjectives and nouns with an embodied subtext found in this experiment (e.g. the high activation contours of painful sadness and anger). This leads to the findings of (b) how we need a terminology for describing the physical vocal performance that is applicable from both first- and third-person perspective. (c) The field of musical communication is also in need of a model for the communication chain between the performer and audience that more fully takes into account its embodied multimodality, including both the musical fit and the misfit, as suggested in figure 13 (page 81).
As presented in section 2.1.4, several experiments have shown that judgements of musical performances are based more on visual than audible aspects of the performance. This thesis supports the earlier findings, as (d) there are almost three times as many comments on visual aspects of the body language than on the audible sound of the performance. The performer’s body language gave the participants an indication of what to expect in the song (mood), even though we cannot know to which extent they attuned to it through their comments alone. (f) It seems that the illustrating and emblematic gestures, which substantiated the lyrics and mood, were the most effective conveyers of the intended message, especially if the gestures and movements derived from the performers’ natural gesture repertoire (emotional leakage). The found indeterminacy of body language implies that we cannot arrive at a final answer to the research questions of this thesis. However, the found pertinence of illustrating and emblematic gestures, when derived from a natural gesture repertoire, can provide us with a starting point for further research.

6.1 Future work

In retrospect, it could have been interesting to combine the emotive videos with neutral sound, to see if the results would have been different, as well with having people with different backgrounds participating in the experiment. As explained in section 3.3.8, several comments were ruled out before starting the in-depth analysis. Some of these comments (like “the screen should have been bigger”) actually concerns the methods used in the experiment. What could have been interesting, was to show the participants the same video on two different screens (sizes), or divide the group into two, where one was shown the videos on a large screen and one on a small screen, to see if they focused on different communicational aspects. This might have shed some light on the issue of eye-contact and could be an interesting approach for further research on the subject of expression and perception of music performance. It would also be interesting to see if there is any consensus about the amount of physical exercise and awareness of others body language. Even though the findings in this small group of people indicates that the people who exercised less was more concerned about the audible sound, it is not enough evidence to support this claim, and requires further research.

35 It is thus important to point out that a performance is a multimodal impression, and it is difficult, if even possible, to distinguish between the sensual impressions and how they affect each other, since this is highly individual and subjective.
The field of music performance could benefit from new techniques to work with anxiety in vocal performances, and the findings of (f) movements as emotional leakage can be used as a foundation for an extrovert rehearsing technique for future work with stage anxiety. Still, further empirical research will be needed to test the relationship between music, lyrics and gestures, and how this affects our perception and judgement of performance.

References


**Appendix 1 Video analyses (angry)**

Permanent link to video in Video Ant; [https://ant.umn.edu/crbuudjsnx](https://ant.umn.edu/crbuudjsnx)

<table>
<thead>
<tr>
<th>Time</th>
<th>Lyrics</th>
<th>“Objective” description of movement</th>
<th>Movement category (derived from Jensenius, Wanderley, Godøy, &amp; Leman [2010], and Kurosawa &amp; Davidson [2005])</th>
<th>The performer’s movements intention</th>
<th>Comments by the participants (male in blue font, female in red)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:02-00:03</td>
<td>Grasping the microphone with right hand, lifting</td>
<td>communicative; affect display (leg),</td>
<td>Mobilizing energy and</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>“Determent; Looks up, more open”</em></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Movement Descriptions</td>
<td>Illustrators</td>
<td>Actions &amp; Comments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:04-00:06</td>
<td>Swaying body movement to the rhythm restlessly</td>
<td>Communicative, sound accompanying; illustrators (StR)</td>
<td>Mobilizing energy and getting into the groove.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*“The movements made me think more <em>rocked</em> than what happens in the song”</td>
<td></td>
<td>*“Eyes open, shakes her head and body and look around. Expresses more openness and attention. Makes the audience wake up”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:07-00:12</td>
<td>Continuous swaying, raising right lower arm upward. Grasps microphone stand for a second with left hand.</td>
<td>Communicative, sound accompanying; illustrators (StR)</td>
<td>Arm movement to increase tension in song. I tend to touch the microphone stand inattentively.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*“A lot of movements, gives the impression of a lively song”</td>
<td></td>
<td>*“She owns it. I sense more roughness (tøffhet)/courage.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*“The motions of the singer make me anticipate an energetic song”</td>
<td></td>
<td>*“Creates an impression that she is <em>in the</em>”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Movement Description</td>
<td>Communicative; illustrators (StR)</td>
<td>Movement Description</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>00:12-00:15</td>
<td>Swaying, right arm raised upward and hand hold beside head.</td>
<td>Arm movement to create tension and build up before vocal starts.</td>
<td><em>“Much more groovy, movements from the singer bids to dance”&lt;br&gt;</em>“The person moves more and expresses her likening of the song”&lt;br&gt;<em>“Looks like she’s dancing to a drum beat, but the guitar is all there is”&lt;br&gt;</em>“Moves her body, and “er mer med”/gets more involved”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:15-00:17</td>
<td>If this is death&lt;br&gt;Right hand beside the head, restless movements in whole body. Left hand on microphone stand.</td>
<td>Conveying anger/frustration</td>
<td>*“accentuate death”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:18-00:19</td>
<td>I’ve been dead the whole time</td>
<td>Right hand gesticulates with melody, left hand is held at microphone stand.</td>
<td>Communicative, sound accompanying; illustrators (melody)</td>
<td>“The way she performs the lyrics is rougher-substantiate that is seems like she owns the situation more. A bit sassy/too cool, ha ha.”</td>
<td></td>
</tr>
<tr>
<td>00:20-00:21</td>
<td>Swimming around in your mind</td>
<td>On “around” singer adds a gesture with both arms that makes half a circle from her upper body.</td>
<td>Communicative; illustrators, on the verge of emblem</td>
<td>The arm gesture acts as a natural upbeat to the accentuated “around”.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Narrow timbre: sings in a pleasant register but puts a lot of acting into it. Theatrical expression in her voice that is increased by her hand gestures”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“She uses her hand a lot to substantiate the meaning of the lyrics. Sings with a narrow tone in her voice. Uses the body in a kind of dance. Grimaces that makes you think of anger/evil”</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Description</td>
<td>Gesture/Touch</td>
<td>Hand Movement/Touch</td>
<td>Comment</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>---------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>00:21-00:22</td>
<td>Screaming for love</td>
<td>On “screaming” hands go out fast to each side at ear level.</td>
<td>Communicative</td>
<td>Illustrators</td>
<td>Hand gesture is meant to underline the word “scream”.</td>
</tr>
<tr>
<td>00:23-00:24</td>
<td>If this should hurt</td>
<td></td>
<td></td>
<td></td>
<td>“The song is a statement”</td>
</tr>
<tr>
<td>00:24-00:27</td>
<td>I’ve been hurt worse before</td>
<td>Grabs microphone stand with left hand, gesticulating with right hand.</td>
<td>Communicative, concurring sound, illustrators (melody), touch</td>
<td>Hand movements are used to create tension. I still grab the stand inattentively but believe this is to feel comfort in the performance and gain some confidence.</td>
<td>“The singer uses hand gestures to communicate a mood. She drags me into the experience.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Gesticulating substantiation of lyrics.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“The singer communicates drama through intensity and the voice.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“This version seems less vulnerable, both with the way she almost snarls the words and her body language.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“More typical “soul” timbre, deeper and more sensual. Sings about getting “hurt” so it might fit the song better than the first.”</td>
</tr>
<tr>
<td>Time</td>
<td>Event Description</td>
<td>Hand Gestures</td>
<td>Additional Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00:27-00:30</td>
<td>When I slammed the door to my own heart</td>
<td>On “slammed” the right hand does a slapping movement.</td>
<td>Communicative, sound accompanying; illustrators (melody) Hand gesture to underline the word “slammed” and to accentuate “slammed”. “Too much movements in hands and dance in comparison to the music” “Body language; fierce” “Arms that move in a natural flow” “Many and large hand gestures and determined voice gives the impression of power (Kraftfull) and determination”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 00:30-00:35 | It isn’t meant to be the time I see with                                              | Shoulders shrunk higher, left hand to the microphone (held for 5 seconds) and right hand held beside the microphone, eyes closed. On “meant” and “time” the head goes backward as the volume in the voice increases on these notes. The left hand goes | Communicative; illustrators, affect display, sound facilitating, touch High shoulders and a tighter body language to express frustration and vulnerability. The head movements on the top notes is a microphone technique that derives from my natural gestural repertoire. The hand gestures are not intended but comes natural to “Solidarity; I experience it as if she has accepted the pain with help from the security that the movements show” “High note; turns away from the microphone. Can be interpreted as pain.” *“The movements substantiate what happens in the
<table>
<thead>
<tr>
<th>Time</th>
<th>Action</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:35</td>
<td>You</td>
<td>Opens up the body after “you”, both hands fall down to the sides and upper body sways swiftly to the left side</td>
<td>Communicative; affect display</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Give the impression of giving up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Serious way of singing. Reflects the lyrics.”</td>
</tr>
<tr>
<td>00:36-</td>
<td>Falling apart</td>
<td>Rising right hand on “apart”, left hand held at the stand (less than 4 seconds)</td>
<td>Communicative, sound facilitating; illustrators, touch</td>
</tr>
<tr>
<td>00:38</td>
<td></td>
<td></td>
<td>Right hand movement to help with the crescendo on “apart”.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Arm gesture shows that what is conveyed has a certain significance, a bigger meaning”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Seems (virker) more aggressive”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“The vocals feels somewhat closed”</td>
</tr>
<tr>
<td>Time</td>
<td>Text</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>00:38-</td>
<td>it isn’t what it seems</td>
<td>On “What” right hand falls down and right foot lifts from the ground. Left hand is held on the microphone stand throughout the next phrase (3 seconds).</td>
<td></td>
</tr>
<tr>
<td>00:40</td>
<td></td>
<td>Communicative, sound accompanying; illustrators (arms), affect display (leg), touch.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The hand foot combination is not intended but happens often when I am feeling the frustration in the song.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“Foot is lifted up, gives me the impression that she feels the song”</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“Hand gestures; gives more power to the message sung”</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“A lot of movement to express that the song means something to her”</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“Hand gestures amplify”</em></td>
<td></td>
</tr>
<tr>
<td>00:40-</td>
<td>these horror dreams</td>
<td>Right leg slightly lifted from the ground on “dreams”.</td>
<td></td>
</tr>
<tr>
<td>00:42</td>
<td></td>
<td>Communicative; affect display (leg)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The leg movement is not deliberately executed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“Movements are kind of gustily. Expresses intensity”</em></td>
<td></td>
</tr>
<tr>
<td>00:42</td>
<td></td>
<td>Turns away from the microphone.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sound accompanying, Sound facilitating</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Covering up the preparatory breath. This is done automatically.</td>
<td></td>
</tr>
<tr>
<td>00:43-</td>
<td>I’ll make them my</td>
<td>hand movements and right foot slightly off the ground. Left hand to the mic (just a second)</td>
<td></td>
</tr>
<tr>
<td>00:45</td>
<td></td>
<td>Communicative; illustrators (arms), affect display (leg), touch</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“First clear moment of eye contact”</em></td>
<td></td>
</tr>
<tr>
<td>00:45-</td>
<td>start</td>
<td>Right hand lifted up when the tone changes interval.</td>
<td></td>
</tr>
<tr>
<td>00:47</td>
<td></td>
<td>Communicative, sound facilitating; illustrators,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>To help with intonation and crescendo. The</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>“The verse has a loose, partially big body language”</em></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Action Description</td>
<td>Movement Details</td>
<td>Comments</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>00:47-</td>
<td>The head goes backward as the volume in the voice increases on “start”. Left hand</td>
<td>touch</td>
<td>and towards the chorus she takes it up a notch (drar på); here comes the message!</td>
</tr>
<tr>
<td>00:49</td>
<td>to the stand (held for two seconds).</td>
<td>head movements on the top notes is just a microphone technique.</td>
<td></td>
</tr>
<tr>
<td>00:49-</td>
<td>I hold my head up high</td>
<td>The head is held a bit further from the microphone during the whole chorus.</td>
<td></td>
</tr>
<tr>
<td>00:52</td>
<td>Cause the pain just makes me</td>
<td>Shifts from holding left hand on stand, to right hand on the microphone.</td>
<td></td>
</tr>
<tr>
<td>00:54</td>
<td>Stronger</td>
<td>Right arm emphasise the rhythm of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicative, sound</td>
<td></td>
</tr>
</tbody>
</table>

*“Emotional”
*“Head a bit up when she sang the lyrics “head up””

*“A lot of movements, and it does not fit so good with the song”
*“Higher notes in this section. More intense. Volume also up.”
*“The singer is showing strength”
*“A lot more pressure on the vocals now, sounds more confident.”

*“A large movement”
<table>
<thead>
<tr>
<th>Time</th>
<th>Action</th>
<th>Illustration</th>
<th>Movement Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:54</td>
<td>Word and right foot lifted from the ground.</td>
<td>accompanying; illustrators (arms) affect display (leg)</td>
<td>A movement I often do as a preparation before using a lot of energy. It is a conscious movement, but the scale of how the movement is conducted is not conscious.</td>
</tr>
<tr>
<td>00:55- 00:57</td>
<td>Whole upper body leans down to the left side, left hand holding the stand.</td>
<td>Communicative; illustrators, affect display</td>
<td>My mood shifts from angry to frustrated.</td>
</tr>
</tbody>
</table>
| Pretend I never cry | Right foot barely touches the ground.                                 | Communicative; affect display | *“Grimaces in her face can remind of pain. A lot of substantiating the syllables in the lyrics with arms and legs.”*  
*“She uses her legs to express emotions. Feels like she can’t hold it in any more”* |
| 00:57- 01:01 | I can’t take this any longer Arm gesticulating, right foot lifted before “longer” | Communicative, sound accompanying; illustrators, affect display | I am really feeling the vibe here and the movements are not intended.                |
|                                                   |                                                                                             |               | *”Gets a bit angry-sad (sinnaleiseg)”*  
*“The artist substantiate the lyrics by giving”* |
<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
<th>Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:02-01:03</td>
<td>Whole upper body leans down to the left side, right hand holding the stand. Left foot barely touches the ground.</td>
<td>Communicative; illustrators. This movement is a typical movement that feels right after I have pushed the volume in the voice. The movements are intended but the scale of how the movement is conducted is not conscious.</td>
</tr>
<tr>
<td>01:02-01:04</td>
<td>Establishing balance and contact with the ground again.</td>
<td>Sound accompanying. I am “finding the beat” again.</td>
</tr>
<tr>
<td>01:04-01:07</td>
<td>starts swaying the whole body.</td>
<td>Sound accompanying.</td>
</tr>
</tbody>
</table>

*“Pain; I can’t take this any longer”*

*“Falls down here as well, but with attitude”*

*“STRONG expression, good lift in the chorus.”*

*“Ends “longer” with an “eh” sound. Rhythmical expression. Powerful and less vulnerable.”*

*“Much more powerful singing, gets more of a power ballad-feel than the tender, vulnerable emotion from the other song.”*

*“Still a lot of movements. It’s very disturbing and I can’t focus on the song”*

the impression of extra effort.”
<table>
<thead>
<tr>
<th>Time</th>
<th>Action/Description</th>
<th>Communicative; touch</th>
<th>Purpose</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:07-01:08</td>
<td>Grasps microphone with left hand (held throughout the next four bars).</td>
<td></td>
<td>To prepare for the vocals.</td>
<td><strong>“Looks away, avoid eye contact but she is not as sad/devastated and owns it more.”</strong></td>
</tr>
<tr>
<td>01:09-01:10</td>
<td><strong>Just shake it off</strong>&lt;br&gt;the singer does a shaking gesture with the right hand</td>
<td>Communicative; illustrators, on the verge of emblem</td>
<td>This is to emphasize the lyrics and give the impression of the singer being in control and calm.</td>
<td><strong>“Cool hand gesture”</strong>&lt;br&gt;<strong>“I get sucked into the mood (stemming)”</strong>&lt;br&gt;<strong>“Says shake it off and does a shaking movement with her hand”</strong></td>
</tr>
<tr>
<td>01:11-01:15</td>
<td>This terrible fear I’ve been mistaking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Description</td>
<td>Behaviors/Techniques</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>01:15-01:16</td>
<td>my own</td>
<td>Pointing at herself with right hand.</td>
<td>*“Conveys good confidence” <em>“Point at herself to substantiate the lyrics”</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>At the end of the phrase, the upper body fall shortly to the right side</td>
<td>This is to emphasize the lyrics. The “falling” movement is natural and not rehearsed.</td>
<td></td>
</tr>
<tr>
<td>01:17-01:19</td>
<td>It’s not like they told me</td>
<td>Left hand leaves the mic but touches the stand less than a second a few times.</td>
<td>*“Difference in her movements when she don’t sing and sings, where she often look at the microphone to express focus and involvement” *“Changes between a playful, teasing voice and a more “real/felt” soul voice. Gives an impression of dialogue, even though it is not the case as far as I understand.”</td>
<td></td>
</tr>
<tr>
<td>01:20</td>
<td></td>
<td>At the upbeat before the lyrics both arms are raised up to the head on each side</td>
<td>I tend to do this movement before difficult phrases, as a preparation, but this is an</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Gesture Type</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>01:21- 01:27</td>
<td>It’s not like they told me the pain wasn’t bad at all</td>
<td>Arms move up at the beginning of the phrase and helps create a “crunch” with both arms and upper body at the word “told” (which is the top note at the phrase). Both legs are slightly bent at “told”.</td>
<td>This movement is not deliberated. It is to help me substantiate the word “told” and give more power to the top note.</td>
<td></td>
</tr>
<tr>
<td>01:27- 01:28</td>
<td>It made me see</td>
<td>Pointing at herself again</td>
<td>Communicative; emblem, illustrator</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:30- 01:32</td>
<td>Clear</td>
<td>Both hands go up and follows out from the head as the note is held</td>
<td>Communicative, sound accompanying; illustrator</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>This is meant to be a gesture of desperation and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Bigger arm gestures”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Theatrical; gives good contact and presence, but too much body language”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Points at herself again; gives more power and directs the attention toward herself and the message”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*“Goes from doing short, sharp movements to bigger, flowing
<table>
<thead>
<tr>
<th>Time</th>
<th>Action and Gesture Description</th>
<th>Communication Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:32-01:34</td>
<td>It isn’t meant to be the</td>
<td>Communicative; Affect display</td>
<td>Meant to convey the feeling of giving up.</td>
</tr>
<tr>
<td></td>
<td>On “meant” both arms fall down, and search for the mic stand.</td>
<td></td>
<td>&quot;Exiting and spontaneous movement. Movements make it easier to follow, and you don’t get bored as easily”</td>
</tr>
<tr>
<td>01:35-01:37</td>
<td>Time I see with</td>
<td>Communicative; Touch, Illustrator</td>
<td>&quot;I get engrossed but start to miss eye contact”</td>
</tr>
<tr>
<td></td>
<td>Right hand holds the mic and are held there until the last phrase of the pre-chorus (1:45). Left hand substantiating the melodic rhythm.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Movement and Affect Display</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------</td>
<td>------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>01:37</td>
<td>You</td>
<td>Turn quickly to the left side after you</td>
<td>Communicative; affect display</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Like a sigh movement, giving up and tired.</td>
</tr>
<tr>
<td>01:37</td>
<td>Falling apart</td>
<td>Left hand raised up on the long note “apart” again</td>
<td>Communicative, sound facilitating</td>
</tr>
<tr>
<td>01:39</td>
<td></td>
<td></td>
<td>To help with intonation and crescendo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Moves her head in addition to the arms that has been in use the whole song. Becomes more of a show.”</td>
</tr>
<tr>
<td>01:39</td>
<td>It isn’t what it seems</td>
<td>On “what”, left hand goes down and grasps the stand shortly, while head goes back.</td>
<td>Communicative, sound facilitating; illustrators, touch</td>
</tr>
<tr>
<td>01:42</td>
<td></td>
<td></td>
<td>To help substantiate the textual content and to maintain the volume of the voice.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Moves her head in addition to the arms that has been in use the whole song. Becomes more of a show.”</td>
</tr>
<tr>
<td>01:42</td>
<td>these horror dreams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:45</td>
<td>I’ll make them my</td>
<td>Pointing at herself with both arms. Making a grimace of wrinkling her nose.</td>
<td>Communicative; emblem (arms), affect display (face)</td>
</tr>
<tr>
<td>01:46</td>
<td></td>
<td></td>
<td>To emphasize the lyrics</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Moves her head in addition to the arms that has been in use the whole song. Becomes more of a show.”</td>
</tr>
<tr>
<td>01:46</td>
<td>Start</td>
<td>Right hand pointing up to the ceiling, left hand at the mic stand. Head is held further from the microphone than in the verse.</td>
<td>Communicative; illustrator, posture, touch</td>
</tr>
<tr>
<td>01:48</td>
<td></td>
<td></td>
<td>I tend to put my hand in the air before an upward interval, this is a mental preparation before making a vocal leap. My head is held further from the</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Moves her head in addition to the arms that has been in use the whole song. Becomes more of a show.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Moves her head in addition to the arms that has been in use the whole song. Becomes more of a show.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Moves her head in addition to the arms that has been in use the whole song. Becomes more of a show.”</td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>01:48-01:53</td>
<td>I hold my head up high cause the pain just makes me</td>
<td>Right arm down in a sharp movement at the beginning of the phrase. Quick gesture with right arm towards the head when the word “head” is sung. A lot of grimaces in the face. Shifts from holding left hand at the stand to right hand holding the mic on “cause”. (Right hand is held at the mic throughout the song.) Left arm gesticulates in the rhythm of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicative; illustrator, touch, affect display</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>None of the arm gestures are intended.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Hand gesture into the chorus vocal.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“It looks like she screams with a sad/painful expression on her face. This takes a lot, emotionally from her”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Strong and energetic even though the lyrics is about pain”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“She is holding her head up to match the emotional state of the lyrics”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Something”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Increasing volume, more power. Gets the impression of strong emotions with power.”</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Action Details</td>
<td>Movement Description</td>
<td>Affect Display</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>01:53-01:55</td>
<td>stronger</td>
<td>Tensing up the body, arms before the chest and shoulders up high, right foot almost off the ground.</td>
<td>Communicative; affect display</td>
</tr>
<tr>
<td>01:56</td>
<td></td>
<td>Upper body falls down towards left side again.</td>
<td>Communicative; affect display</td>
</tr>
<tr>
<td>01:56-02:01</td>
<td>Pretend I never cry I can’t take this any</td>
<td>Raising right arm above head while dropping upper body, holding microphone with left hand.</td>
<td>Communicative, sound accompanying; affect display</td>
</tr>
<tr>
<td>02:01-02:03</td>
<td>Longer</td>
<td></td>
<td>Giving up</td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Comment</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| 02:04-02:05 | Turning away from the microphone for a short while | aspects are more important than the fact that I’ve heard the song once before.”
|        | Communicative; illustrators                                           | “The words “stronger” and “longer” feels more meant, and more intense.” |
| 02:06  | Letting go of microphone stand. Restless swaying throughout the song  | “This version of the song was an explosion of emotions, whilst the other was a journey through the singer emotional life”
|        | Communicative                                                         | “Way too much movements” |
|        |                                                                       | “Sad” |
|        |                                                                       | “Sadness” |
|        | This is just to keep the rhythm and focus throughout the song.       | “I did not comment because I could not see the singer face” |
|        |                                                                       | “Her body language seems to close up at the end. Maybe she didn’t own it as much as she thought.” |
“Not as much movement in the legs as in the upper body”

## Appendix 2 Video analyses (sad)

Permanent link to video in Video Ant: [https://ant.umn.edu/xuqgzozkds](https://ant.umn.edu/xuqgzozkds)

<table>
<thead>
<tr>
<th>Time</th>
<th>Lyrics</th>
<th>“Objective” description of movement</th>
<th>Movement category (derived from Jensenius, Wanderley, Godøy, &amp; Leman [2010], and Kurosawa &amp; Davidson [2005])</th>
<th>The performer’s movements intention</th>
<th>Comments by the participants (male in blue font, female in red)</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:02-00:05</td>
<td>Lifts right hand up towards the microphone (held until 00:45) and left hand to the microphone stand (held until 00:16) slowly, starts swaying body to the half notes.</td>
<td>Sound accompanying + communicative; illustrators, posture, touch</td>
<td>The posture taken on in the video is generally an introvert body language with high shoulders and small arm gestures.</td>
<td>“Grabs the microphone; I am ready, I am going to convey something”</td>
<td><strong>“The posture; a kind of vulnerability- no eye contact-looks down- “tight” body language.”</strong></td>
</tr>
<tr>
<td>00:06-00:15</td>
<td>Head turns away from the camera and are held at the side, or behind, the microphone.</td>
<td>Communicative; affect display, posture</td>
<td>To avoid eye contact for the purpose of communicating sadness. Head is held mostly</td>
<td><strong>“The posture; a kind of vulnerability- no eye contact-looks down- “tight” body language.”</strong></td>
<td></td>
</tr>
</tbody>
</table>
behind the microphone for the purpose of hiding and expressing the emotion of feeling vulnerable.

*“Vulnerability; how the singer presents herself with how she stands/poses signals vulnerability.”

*“Sad opening; the singer looks to the ground, seems like she is going to convey something sad.”

*“Looks down; “wait for it”

*“Much more calm movements, more comfortable to look at.”

*“Posture; It seems to be a personal, serious song.”

*“More introvert; looks to the ground.”

*“The singer seems to be very focused, she is getting into the “feeling” of the song, but it seems a bit fake to me
<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
<th>Body Language</th>
<th>Notes</th>
</tr>
</thead>
</table>
| 00:16-00:18 | If this is death                                                             | Lifts shoulders and left arm, eyes are shut, and eyebrow is frowning.         | “Listens to the music. Small movements. Has ‘respect’ for the expression, and not just when one has something to actively do oneself.”
|            |                                                                             | Communicative; affect display                                                 | “Look: cool how she looks down before she starts to sing.”           |
| 00:18-00:21 | I’ve been dead the whole time, swimming                                    | Right arm still holding the microphone, left arm slowly lifted toward it.     | “Stands calm during the intro, communicates that this is something that should be taken seriously.”
|            |                                                                             | Communicative; affect display                                                 | “Fervency; soft, calm movements, also in the voice.”                |

(generic way of holding a microphone and swaying to the music.)

*“Listens to the music. Small movements. Has “respect” for the expression, and not just when one has something to actively do oneself.”

*“Look: cool how she looks down before she starts to sing.”

*“Stands calm during the intro, communicates that this is something that should be taken seriously.”

*“Fervency; soft, calm movements, also in the voice.”
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:21-</td>
<td>Screaming for love</td>
<td>Small gesture with left arm substantiates the rhythm of the melody</td>
<td><em>“Not very related to the audience; In the beginning (she) shows a closed personality to the audience, sticks to this further into the song.”</em></td>
</tr>
<tr>
<td>00:24-</td>
<td>If this should hurt</td>
<td>Left arm slowly down again</td>
<td><em>“Procurement (formidling); gets a feeling that the singer is hurt from the sound of her voice.”</em></td>
</tr>
<tr>
<td>00:25-</td>
<td></td>
<td></td>
<td><em>“Fake “emotion” in the voice. Pressing on the voice.”</em></td>
</tr>
<tr>
<td>00:27</td>
<td>I’ve been hurt worse before</td>
<td>At “worse”, left arm makes a short lifting gesture to</td>
<td><em>“Moves the arm in a “come in”-movement.”</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Gesture Description</td>
<td>Movement Description</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 00:28-00:31 | When I slammed the door to my own heart | Small slamming gesture with left hand before left hand is raised up and held before the heart. | First movement is to substantiate the word and action of *slamming* something and are intended. The second movement of touching the heart derives from my gestural repertoire and are not intended, but too small to be called an emblem. | *“Uses the arms to drag us closer. Build intimacy.”*  
*“The hand takes focus now.”*  
*“Substantiate the lyric “my heart” by touching her heart.”*  
*“Touches her heart at the same time as the lyrics read “my heart”. The hand kind of helps the lyrics.”* |
| 00:31-00:36 | It isn’t meant to be the time I see with you | Left hand gesticulates and drops on “meant”. | To give the impression of disappointment and giving up. | *“The sound and aesthetics are perceived as more integral (helhetlig) and feels more like a “live-setting” than video 1.”*  
*“Movement with the arm that does not rest on the microphone stand. The gesture of the arm matches the*
<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:37-00:40</td>
<td>Falling apart</td>
</tr>
</tbody>
</table>

**“Conveys sensitivity, tenderness.”**

**“Closed, more fragile.”**

The phrasing of the melody.”

“Sings lower in volume. More air on the voice. “Closer” expression”

“Vocals; starts very careful and calm, as to underline the emotions established”

“Aims towards a person; seems like she tries to reach a certain person.”

“Hand gesture; as if she was to stroke the microphone. Maybe to “give comfort”.”

“Style of singing- toned down. More intimate, less confronting.”
<table>
<thead>
<tr>
<th>Time</th>
<th>Line</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:40-00:43</td>
<td>it isn’t what it seems these horror dreams</td>
<td>*“Eyes are closed. Matches the sore and intimate expression of the song.”</td>
<td>*“Sways to the music”</td>
</tr>
<tr>
<td>00:44-00:46</td>
<td>I’ll make them my</td>
<td><strong>Right hand is lifted from the mic and left hand is held at the microphone stand (until 1:09). Right hand is lifted toward the microphone and head</strong></td>
<td>This shift of hands is not deliberate.</td>
</tr>
<tr>
<td></td>
<td>start</td>
<td><strong>Communicative; illustrators, touch</strong></td>
<td>*“The other hand is joining; the song builds up, and now she raises the other hand from the microphone to help build it- more intensity.”</td>
</tr>
<tr>
<td>00:46-00:48</td>
<td>I hold my head up high cause the pain just</td>
<td><strong>Communicative; illustrators</strong></td>
<td>*“Small, but nice movements.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*“The movements are still “swaying along” and hand gestures, but they are smaller.”</td>
<td>*“Smooth transition to the chorus where both the vocal and the pitch goes up.”</td>
</tr>
<tr>
<td>00:48-00:54</td>
<td>I hold my head up high cause the pain just</td>
<td><strong>Communicative; illustrators</strong></td>
<td>*“Vulnerability, enclosed, gets a feeling of curiosity of what”</td>
</tr>
<tr>
<td></td>
<td>makes me stronger phrase. Shoulders is ascending throughout the phrase.</td>
<td>the singer will convey.” *“Touches her head when the lyrics writes “down my head”. (yes, they write the wrong lyrics) *“Dropped the mick before a higher note- get ready.” *“Hand to the head- head ache.” *“Sings something about the head and lifts her hand slowly to her head.” *“Hand up to the head, communicates that she’s thinking about something.” *“Stands very still, often with her eyes closed to show a more calm/melancholic sensation/emotion.”</td>
<td></td>
</tr>
</tbody>
</table>
**Falsetto makes the experience more intimate.**

**Sings in her head voice to keep the emotion in the song calm/sad which also gives a good contrast/climax.**

**Compassion; because of the lyrics; “the pain makes me stronger”.”

<table>
<thead>
<tr>
<th>Time</th>
<th>Action</th>
<th>Communicative; affect display</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:55</td>
<td>Right arm dropped heavily down, deep breath.</td>
<td></td>
<td>Give the impression of giving up.</td>
</tr>
<tr>
<td>00:56-</td>
<td>Pretend I never cry I can’t take this any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:03</td>
<td>longer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both arms held at waist height, left hand</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>holds the microphone stand. Small</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gesticulating movements with right hand</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>substantiate the rhythm of the melody</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:03-</td>
<td>Whole upper body drops down to the left</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:04</td>
<td>side,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>**Differentness: the lyrics and the expression seems to knowingly be</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mismatching and creates an effect of sadness.”</td>
</tr>
</tbody>
</table>

**“The body rushes to the side; when she sings...”**
right hand holding the stand.  
being tired, giving up.  

can’t take this any longer she moves slowly to the side until her body “collapses”. 
Looks heavy.”  
*“Body sigh; the body fell down I can’t take this any longer-
continuation of the lyrics. Her body can actually not take it any longer.”  
*“Dejected movement: almost like a sigh.”  
*“Dropped her body after the phrase. Matched the lyrics.”  
*“A bit more helpless now. I can’t take this any longer. Earlier on it seemed angrier, more controlled, now it seems like she has given up.”  
*“I start to empathize, getting used to the style
of singing and moving and believing more in the singer, although the movements seems to be a bit exaggerated.”

*“A kind of climax in the chorus is given weight by the one hand “spelling” the words before the vocalist is pulling away dramatically.”

<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
<th>Sound accompanying</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:04-01:09</td>
<td>Small swaying movement with the whole body on half notes.</td>
<td></td>
</tr>
</tbody>
</table>

*“Intensity; She don’t sing with as much intensity at this point in the song as in the last video. Can be good for building up the song.”

*“The volume goes les up in the chorus. Less theatrical.”

*“Technique; sings in the throat in the chorus, together with the
<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
<th>Gesture Description</th>
<th>Gesture Purpose</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:10-01:17</td>
<td><strong>Just shake it off this terrible fear</strong>&lt;br&gt;I’ve been mistaking for years it was my own</td>
<td>Both hands are raised slowly toward microphone, left hand is holding the mic (until 1:24) and right hand are held beside the mic, grasping it lightly. Shoulders held high and face frowned.</td>
<td>Communicative; affect display, touch</td>
<td>To emphasise the emotive state of being scared.</td>
</tr>
<tr>
<td>01:18</td>
<td><strong>Deep breath, head away from the microphone</strong></td>
<td></td>
<td>Communicative; sound facilitating, affect display</td>
<td>To emphasise the emotive state of being tired.</td>
</tr>
<tr>
<td>01:19-01:21</td>
<td><strong>It’s not like they told me</strong>&lt;br&gt;Facial grimace; eyes shut, frowning. Right hand is held at the mic (until 1:43).</td>
<td></td>
<td>Communicative; affect display, touch</td>
<td>To simulate crying and pain.</td>
</tr>
<tr>
<td>01:23-01:29</td>
<td><strong>It’s not like they told me the pain wasn’t bad at all</strong>&lt;br&gt;Left arm raised up toward the face as the melody goes up and falls as the melody goes down</td>
<td>Sound accompanying and communicative; illustrators</td>
<td></td>
<td><strong>“The melody is higher, both in pitch and volume and the vocalist makes a grimace. Maybe to show the increasing intensity.”</strong></td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Description</td>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>01:29-01:30</td>
<td>It made me see</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:31-01:32</td>
<td>Clear Arm reaches out</td>
<td>Sound facilitating to help create tension and intonation in the tone.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:33-01:37</td>
<td>It isn't meant to be the time I see with</td>
<td>Small gesticulating movements with left hand.</td>
<td>Communicative; illustrators; authentic and not deliberately done.</td>
<td></td>
</tr>
<tr>
<td>01:38</td>
<td>You Left arm lifted horizontally</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01:39-01:40</td>
<td>Falling apart Left arm drops</td>
<td>Communicative; illustrators, affect display</td>
<td>To substantiate the words and give the impression of giving up and that everything is heavy.</td>
<td></td>
</tr>
<tr>
<td>01:41-01:45</td>
<td>It isn't what it seems Left arm is grasping the microphone stand</td>
<td>Communicative; Touch</td>
<td><em>“Clouds; Right here, with the arm movements and “Clouds; Right here, with the arm movements and”</em></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Gesture Description</td>
<td>Comment</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>01:46</td>
<td><strong>I’ll make them my Right hand lifted up towards microphone</strong></td>
<td>Communicative</td>
<td>*“Increasing volume, more intense, stronger emotions.”</td>
<td></td>
</tr>
<tr>
<td>01:46-01:48</td>
<td><strong>Start Right hand and head are slightly pulling opposite directions</strong></td>
<td>Communicative; affect display, sound facilitating gesture</td>
<td>This movement is rehearsed as a sound facilitating gesture, but at</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*“Smaller hand gestures, gets more contact.”</td>
<td></td>
</tr>
</tbody>
</table>

*“These horror dreams (held there throughout the song)."

*“The use of falsetto and acoustic guitar I get a feeling of clouds, like a timeless dream world.”

*“Is perceived as very anonymous”

*“Movement, stands still with the legs but a lot of good movements in the upper body that shows that the artist enjoys the music, but gets stiff with so little movements in the legs.”

*“The melody (the register) is giving me a feeling of sadness and longing.”
<table>
<thead>
<tr>
<th>Time</th>
<th>Description</th>
<th>Description</th>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:49-01:57</td>
<td><em>I hold my head up high cause the pain just makes me stronger</em></td>
<td>Small gesticulating gestures with right hand in the rhythm of the melody.</td>
<td>Communicative; illustrators</td>
<td>To substantiate the words and message sung.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*“Hand gestures; makes a stronger impression.”</td>
</tr>
<tr>
<td>01:57-02:01</td>
<td><em>Pretend I never cry I can’t take this any</em></td>
<td>Facial grimace; eyes shut and frown. Shoulders high.</td>
<td>Communicative; affect display</td>
<td>*“The use of this mixed vocal timbre and moderate use of air on the voice, together with the lyrics “stronger” gives me a feeling of someone who tries to stand up straight in the wind. It’s something tall”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*“Body language/facial expression; emotional”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*“Depressive lyrics that substantiate that it makes you stronger, but a bit passive communication with the listener.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*“I don’t feel like being sung to.”</td>
</tr>
<tr>
<td>Time</td>
<td>Action</td>
<td>Posture/Display</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------</td>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>02:01-02:04</td>
<td>Longer Body freeze’s, shoulders high.</td>
<td>Communicative; affect display, posture</td>
<td>The same as for the start of the video; I avoid eye contact and have a closed body language to represent the emotive state of sadness and giving up.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Hand movement; shows frustration clearly with the hand gesture towards the end where she reaches her head. Substantiate the lyrics.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Just dropped the hand down slowly this time.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“The expression is more homogeneous. Music, lyrics and appearance says the same. It’s a little less interesting”</td>
<td></td>
</tr>
<tr>
<td>02:06-02:08</td>
<td>Stares down at the ground.</td>
<td>Communicative; affect display</td>
<td>“Looks down in the ground, does not feel like a part of the performance.”</td>
<td></td>
</tr>
</tbody>
</table>
| 02:09-02:18| Turn away from the camera, stares at the ground. | Communicative; affect display | “Introvert vulnerability; the singer rounds of as in the beginning; no eye
contact- looks down- introvert
body language- gets a
sad/dejected feeling. Its sore
and painful.”
*“Turns away from the
microphone to indicate that the
song is over.
Focus is shown
until the guitar is
over by holding
the eyes shut.”
*“Stands more
still and keeps her
head down. This
makes the lyrics
seem more serious
and sad in this
performance.”
*“More
introverted, more
vulnerable.”