

COVER PAGE

Document downloaded by @DAEL

Wed Nov 11 05:17:24 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

A Qualitative Approach to Evaluation of Perceived Qualities of Audio and Video in a Distance Education Context, Dan Nyberg and Jan Berg, Acta Acustica **vol. 100** (Number 1), 2014, pp. 154-165

DOI
https://doi.org/10.3813/AAA.918695

Acta Acustica united with Acustica
Vol. 100 (2014) 154 – 165

DOI 10.3813/AAA.918695

A Qualitative Approach to Evaluation of Perceived Qualities of Audio and Video in a Distance Education Context

Dan Nyberg, Jan Berg
Luleå University of Technology. Piteå, Norrbotten, Sweden. [dan.nyberg, jan.berg]@ltu.se

Summary
This study presents a qualitative method for collecting and analysing data to describe audio and video quality. Used in the social sciences, arts, and humanities, this approach relies on phenomenology and hermeneutics and uses interviews and questionnaires to assess the audio and video quality of master classes in classical music taught via the Internet. Although this study is only exploratory, it provides evidence that the method could successfully be used to gather descriptions of perceived qualities.

PACS no. 43.71.Gv, 43.72.Kb, 43.75.Cd, 43.75.St

1. Introduction

There are numerous contexts in which evaluation of audio quality may be performed and several evaluation methods have been developed over the years. Although each method has its own merits, specific methods may be more suitable in specific contexts. When a new context emerges, these methods need to be evaluated, refined, and considered. Existing methods, however, mainly rely on experimental data that comes from carefully manipulated situations to control experimental variables. Exclusively relying on experimental data may not actually reflect the reality that the experiment aims to represent, i.e. the experiment's ecological validity may not be sufficient. The concept of ecological validity was addressed by Brunswik [1] and later discussed by Gibson [2] who elaborated on examples relating to visual perception (e.g. drawings and how well they represent reality). For reproduced sound, Guastavino *et al.* [3] suggested that listening tests should be designed to match the aim of the study through generation of stimuli that enable subjects to "treat the test samples as potentially familiar experiences through cognitive processes elaborated in actual situations". Consequently, due to the risk of not attaining a sufficient level of ecological validity, there are situations where alternatives to experimental studies should be considered. Thus, relying only on experimental data can be an issue when both audio and video qualities are of interest and/or when other quality aspects are involved. In addition, most of these methods work by quantifying the listener's experiences, an approach that limits the type of data collected.

Received 3 September 2012,
accepted 13 October 2013.

This limitation is imminent in the context of distance education over the Internet, which involves audio, video and real-time interaction. This becomes significantly important in distance music education (MED). Here the research has increased over the last ten years [4, 5]. Typically, a MED teacher located at one school teaches a student located at another school via the Internet, often by means of a video conferencing system [4, 5].

Several findings regarding teaching possibilities and attitudes have been made in earlier research. Greenberg's white paper on video conferencing-based distance education [6] shows that teachers have to adopt their teaching to accommodate; this involves e.g. interactivity between participants and teaching strategies that have to be matched with the technology. Work by Masum *et al.* [4] in the project "MusicGrid" adds to this by noting that teaching music over distance requires both frequent demonstrations and continuous feedback during the lessons. Findings by [4, 6] also indicate that participants become quickly adapted and comfortable with the system. Both [4, 6] conclude that distance education can be cost effective and allows users to access more learning experiences without travelling. Shepard [5] takes a more technical view and explains the need of striving to have high quality audio and video, bidirectional transmission and easy-to-use system in a MED situation. Shepard also gives indications that conferencing systems using MPEG-2 encoder and decoder can work if used right.

Research efforts have also been made by Woszczyk *et al.* [7] to establish a transparent system for connecting two physical locations via a shared electronic virtual space for musical practice. However, few studies specifically focus on how MED students and teachers perceive the audio quality of a particular system. Woszczyk *et al.* [7] conclude that in addition to audio quality, other factors such as

© S. Hirzel Verlag · EAA

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

image size of video, tactile/haptic response (low frequencies transmitted via separate channels to simulate floor vibrations) and synchrony between audio and video are involved in creating a transparent MED system. They also found that when asynchrony between audio and video occurs, participants often only focus on the music.

This paper presents a selected qualitative method for data collection and data analysis that is commonly found in the social sciences, arts, and humanities and applies these methods to evaluate audio and video quality as it is perceived by teachers and students in a non- experimental distance education setting (specifically in a real-life classical music course conducted via the Internet).

The terms audio and sound quality will be used interchangeable during the course of this paper. The term *non-experimental* means that the researcher has no control over the parameters such as choice of participants (other than inviting students already enrolled in the classes), where the study took place, transmission quality between the locations, and choice of equipment. The term *real-life* means that the lessons in the classical music course are authentic lessons in which the researcher does not intervene. These qualitative methods are used because the study was conducted in non-experimental and real-life conditions. Although the results from such a study may be used to evaluate the quality of audio equipment, this issue was not the target of this study. Rather, this study intends to investigate what type of data a researcher can collect by using these qualitative methods and to show the applicability of these methods in the field of perceptual quality.

2. Earlier research on perceived audio and video quality

As stated in the introduction, MED involves more than one perceptual quality mode (e.g. auditory, visual and tactile).

Previous research on the quality of perceived audio, which is one of the modalities under investigation, shows that it has been successfully evaluated. These data collection methods often deal with rating scales [8, 9, 10, 11, 12] and descriptive analyses and use the subject's own vocabulary to evaluate the perceived audio quality [13]. What all these methods have in common is that they all are conducted in laboratory settings and follow an experimental paradigm. Toole [8] as well as Gabrielsson and his coworkers (e.g. [9]) conducted research that comprised laboratory experiment using subjects as evaluators of sound quality, in more particular, the sound quality of loudspeakers was evaluated by means of rating scales. The experimental paradigm has since then been a predominant approach for audio quality evaluation through subjective assessments. Bech and Zacharov give a comprehensive review of different approaches in their book [10].

Since then, several recommendations have been developed to evaluate subjective audio quality with the use of rating scales, for an example ITU-R BS.1116-1 [11] and ITU-R BS.1534-1 [12]. The former focuses only on assessment of small impairments of audio quality, while the latter focuses on assessment of intermediate audio quality. There has also been research employing subjects' own vocabulary in order to find an appropriate descriptive terminology that may be used for subsequent quantification of perceived audio characteristics [13, 14].

There are also a variety of methods designed to evaluate perceived video quality. These data collection methods range from scales [15, 16] to interpretation-based estimations of image quality [17] or both [18]. These approaches, however, only look at one modality.

In multimodal research, the focus shifts since both audio and video are involved in a MED situation. In the multimodal field, several approaches are used to evaluate multimodal quality. Some of the approaches use a mixed-methods approach where both numeric and descriptive data are analysed. For example, Strohmeier *et al.* [19] used a mixed-methods approach to evaluate multimodal quality perception of audio and video, inspired by sensory evaluation methods; they used the method Open Profiling of Quality to understand quality perception. The Open Profiling of Quality contained three parts: a method they call psychoperceptual evaluation (the subject uses rating scales to evaluate perceived audio quality); sensory profiling (collecting individual quality attributes); and external preference mapping (constructing links between the psychoperceptual result and the sensory profiling). Using analysis of variance (ANOVA) and Principal Component Analysis (PCA), they analysed the psychoperceptual evaluation and sensory profiling, respectively.

Jumisko-Pyykkö *et al.* [20] used a qualitative approach to evaluate audio-visual quality by combining a single-stimulus method, for quality evaluation, with semi-structured interviews, to establish quality evaluation criteria. The data analysis combined a qualitative analysis based on grounded theory and a quantitative statistical analysis, Bayesian modelling. Jumisko-Pyykkö *et al.* [21] also applied an entirely qualitative approach (semi-structured interviews) and compared the qualitative results with quantitative results in order to understand the relationship between produced and experienced quality in the context of interactive audio-visual systems. The analysis used here also followed the grounded theory approach. This qualitative approach showed that it was possible to collect 13 different categories of quality criteria from the subjects and that these categories were related to the quantitative results. This is an indication that a qualitative method can complement quantitative measurements in order to understand multimodal quality.

Other researchers have employed single-method approaches [22, 23, 24, 25] in multimodal quality research. Beerends and De Caluwe [22] tested the influence of video quality on perceived audio quality (and vice versa) using a nine-point absolute category rating scale. The data analyses were made using ANOVA. Bech *et al.* [24] analysed the interaction between audio and visual factors in the context of a home theatre system. The experiment collected numeric data from rating scales using defined attributes with anchors and the data were analysed using ANOVA.

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

All methods presented above have shown to be well suited for the intended purposes and delivered valuable results; however, these methods are designed for experimental or laboratory contexts and some of the methods could be hard to adapt to a real-life situation (non-experimental situation) such as the MED situation. One cannot control all parameters involved in an MED situation, contrary to the experimental research presented above. Clearly, laboratory testing cannot completely account for the real life situations found in distant learning MED, a situation that is relatively unstudied and may require new methods to develop a complete picture of its usefulness.

3. A qualitative approach

When considering a suitable method in a relatively new, non-experimental, and multi-modal context, such as master classes in classical music over the internet, another approach may be required that uses qualitative methods that focus more on describing and understanding the experience than on quantifying it.

An example of viewing sound quality from a somewhat different angle has been suggested by Blauert and Jekosch [26]. They present a perceptionism view of the world, a view that relies on the belief that all knowledge is based on sensory perception. This view allowed them to create a layer model of audio quality based on the amount of abstraction involved for each judgment of audio quality. The model enabled them to show the situation dependence of different audio qualities based on the degree of abstraction. One of their findings, the quality called "Aural-communication Quality", which involves the highest amount of abstraction, can be investigated with psychological and cognitive tests in real cases such as in tests of usability, comprehension and dialogue quality. The situation dependence in their model relates well to the previously discussed concept of ecological validity, which further strengthens the argument for considering non-experimental studies outside the laboratory when assessing audio quality in certain contexts. This example indicates that it is possible to adapt and employ methods used in other fields, in this case psychology, but also in the humanities and social sciences. Since MED also focus on usability, and dialogue, methods related to those discussed by Blauert and Jekosch are of interest.

The following sections initially present the implications of using a qualitative approach and some qualitative points of departure, followed by how data collection and analysis methods may be applied in the MED context, given the qualitative approach.

3.1. Implications of using a qualitative approach

When researchers use qualitative methods, they need to consider the possible limitations of such an approach. In qualitative research the question of validity focuses on the employment of the "reduction" process (analysis) that leads to a result and not whether the data can be replicated, a common view issue in the field of audio quality research. The qualitative research focuses on how clearly the analysis is conducted and described by the researcher and how well the descriptions of the data relates to the collected data. In hermeneutics for an example the hermeneutic circle comes into play, see [27, 28] and in phenomenology, bracketing, phenomenological reduction and horizontalisation are the major methods employed, see [28, 29]. All these methods are designed to understand the data and creating a description of the collected data.

One implication of qualitative research is that two researchers do not produce exactly the same results when faced with the same task. This poses no problem, however. That is, competent and skilful researchers produce results easily recognized by other researchers. Each exploration/study should bring a different perspective on the phenomenon under study and each perspective creates a more comprehensive picture of the phenomenon [27].

3.2. Qualitative points of departure

Empirical phenomenology and hermeneutics, approaches frequently used in psychology [27, 28], are suitable points of departure for evaluating audio and video quality in a distance education context such as MED conducted via the Internet. Empirical phenomenology has the following goal:

> To understand the general psychological meaning of some particular human way of being-in-a-situation [...] through a number of descriptions of this way of being-in-a-situation from people who have lived through and experience themselves as so involved [27, p. 40].

Moustakas, agreeing with Tesch [27], states that one can determine the underlying structures of an experience by interpreting the description of a situation where the experience occurs [28]. As for hermeneutics, Moustakas summarizes it this way:

> [R]eading a text[1] so that the intention and meaning behind appearances are fully understood [28, p. 9].

This view agrees with Tesch's explanation of hermeneutics [27]. Hermeneutics looks at human experience by studying the lifeworld[2] in order to form a whole, including a description of experience. Interpretation of an experience shows what is hidden behind a phenomenon [28]. This understanding means that an empirical phenomenological and hermeneutic departure may help evaluate audio quality, video quality, and their interaction in order to extract, interpret, and describe the experience of the persons involved in a MED situation.

Other researchers have already applied grounded theory in the form of semi-structured interviews to evaluate the perception of audio-visual quality [20, 21] (as discussed

[1] The term "text" in a hermeneutic sense is very wide and incorporates human action, interview transcripts, and texts [27].

[2] Lifeworld is a translation from German, *Lebenswelt*: "All the immediate experiences, activities, and contacts that make up the world of an individual or corporate life" (*New Oxford American Dictionary*).

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

Table I. Top: Questionnaire 1. Bottom: Questionnaire 2: Questions in bold are "yes" and "no" questions and the subject could only give a "yes" or "no" answer. The remaining questions are presented as open-ended questions in the questionnaire. The X indicates to which category of persons the questions are directed.

1	What are the positive aspect/aspects of distant musical learning, based on this occasion?
2	What are the negative aspect/aspects of distant musical learning, based on this occasion?
3	Explain the sound quality in this carried out lesson.

Nr	Teacher	Student	Both	Question
1	X			Did you change your teaching methods on this occasion?
2	X			In what way did it change?
3			X	In what way were you able to interact, by speaking and singing/playing, with the teacher/student?
4		X		**Can you perceive the teacher's instructions clearly?**
5		X		If "Yes", describe in your own words the perceived sound.
6		X		If "No", describe in your own words the perceived sound.
7		X		**Can you perceive the music examples from the teacher clearly?**
8		X		If "Yes", describe in your own words the perceived sound.
9		X		If "No", describe in your own words the perceived sound.
10			X	In what way did the communication between you and the teacher/student work?
11			X	What could further enhance the lesson?
12			X	Are there any positive aspects with the use of distance learning?
13			X	Are there any negative aspects with the use of distance learning?
14			X	What could be improved with the technology based on this lesson, according to your opinion?
15			X	What could be improved in the teaching, based on this lesson, according to your opinion?

in section 2). The aim in these cases was to collect and understand the unknown underlying characteristics of multimodal quality when evaluated quantitatively,

Grounded theory, an approach that is often found in sociology, generates a theory from data [30, 31]. When conducting grounded theory research, a researcher should collect data on a particular subject/topic before developing theories and hypotheses about it rather than developing theories and hypotheses before collecting data [27].

Grounded theory has similarities to the proposed approach above, as we want the data to give information to us about a phenomenon (in this case audio and video quality and interaction between teacher and students), but our purpose is not to generate a theory or a hypothesis about the phenomenon; rather our focus is to describe the perception of what is experienced in this specific context.

3.3. Collection of data

The two data collection methods used, *questionnaires* and *interviews*, can be applied post the MED situation, thus they do not influence student-teacher interaction. Both these methods are often found in qualitative research. Interviews are used in phenomenology research and to some extent in hermeneutics [27, 28, 32]. Compared to numeric-based data collection methods, both questionnaires and interviews can provide detailed descriptions of how audio and video quality are perceived, separately and together. Similarly, questionnaires have been used in earlier studies [33, 34, 35, 36] with good results. Open-ended questionnaires can gather many responses from the subjects and give an overview of the field under study. Of course, questionnaires can contain questions where the responses are quantitative, e.g., by using scales, lines, checkboxes, etc., but such approaches are not considered here as they represent a methodology outside the scope of this paper. Interviews provide insight into how subjects experience the learning environment–audio, video, and student-teacher interaction. This approach also allows more flexibility in data collection because follow-up questions can be asked [32], which is not possible with questionnaires. Collecting descriptive data from interviews could be beneficial when the research field is relatively unexplored [20].

The data collection is designed in the following way: The design can be seen as a funnel, with broad questions at first (Questionnaire 1), secondly more narrow questions (Questionnaire 2) still remaining open with but with certain focus points and thirdly interviews with participants in order to investigating what particular users perceive. In the following sections it is shown how these data collection methods are applied in a MED context.

3.3.1. Using questionnaires

Two sets of questionnaires were distributed during several MED sessions. Their design was partially influenced by previous studies on distance learning [4, 5]. As parts of the questionnaires included broad and/or open-ended questions, no pre-tests were considered. Questionnaire 1 is a brief questionnaire containing a set of open-ended questions regarding positive and negative aspects of the MED sessions as well as audio and video quality aspects. Based on the answers collected from the first questionnaire, see Table I, a second more elaborated questionnaire was designed (Questionnaire 2), see Table I. In this respect, Questionnaire 1 partially serves the function of a pilot test. Questionnaire 2 had more specific open-ended

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

questions and two yes and no questions that led to follow-up questions ("if Yes, please describe ...", "if No, please describe ..."). Questions 1 and 2 in the questionnaire were designed to capture the teachers' perspective and questions 4–9 were specifically designed for capturing the students' perspective. A total of 15 questions were used in the questionnaire. This new set of questions was designed to encourage the participants to reflect on whether they perceived the audio quality and video quality as good or bad and why they perceived the qualities in these particular ways. The questions were designed to capture a comprehensible view of the participant's perception of the overall quality of the MED system and the participant's experience. Data from Questionnaire 2 were systematically analysed (described below).

3.3.2. Using interviews

As a next step, personal semi-structured interviews with the participants in the MED situation (students and teachers) were used to further shed light on the results from the questionnaire and obtain an even more individual description of the experience of perceived quality. An interview guide combining both open-ended and circumscribed questions was used to guide both the interviewer and interviewee to obtain the descriptions of the users' experience [28]; see Table II for the questions used in the interview guide. The question asked from the interview guide varied depending on whether the interviewee were a teacher or a student.

Table II. Questions used in the interview guide. Translated from Swedish.

- How was the lesson/master class?
- On the topic of user freedom
  - Did you feel limited in your practice?
    - Is it the technology?
    - Is it the distance between you and the teacher/ student?
    - Is it the lack of presence by the teacher/student?
    - Is it the communication that poses problems?
  - If you feel free in your practice
    - What is it that makes you feel free in your practice?
    - Can you do the things that you want to do?
    - To what extent does the feeling of freedom exist?
- Could you complete the master class without being affected by the technology/system used?
- Did you perceive the technology as a hindrance or as a tool?
- What worked and what didn't work during the master class?
  - Technology
    - Was it good?
    - Was it bad?
    - What could be improved? *Describe them*
  - Pedagogics
    - Was it good?
    - Was it bad?
    - What could be improved? *Describe them*
- How did you perceive the sound quality?
  - Can you compare it to a known format/media?
- How did you perceive the sound and video quality?
- If you exclude the video, how did you perceive the sound quality?
- If you exclude the sound, how did you perceive the video quality?
- Did you perceive any "delay/latency" between the sound and video?
  - If yes, which came first according to you?
  - Did this delay pose any problem for you?

3.4. Analysis of the collected data

A phenomenological and hermeneutical approach is used as the point of departure for the analysis. The analysis focuses on describing the user's perception of audio and video quality as well as the interaction between teachers and students. This section presents the data analysis methods.

3.4.1. Using questionnaire analysis

An ad-hoc analysis is used on the data and includes theme analysis and meaning categorization analysis [32, 37]. In the theme analysis, themes are searched for in the collected data [37]. The theme analysis is accomplished by first listing all the collected answers under each question and reading all the answers for similarities and common answers [23]. This part of the analysis is analysed hermeneutically.

The meaning categorization analysis is an analysis that is inspired by Verbal Protocol Analysis [38]. Each answer is coded and counted according to its identified properties. The verbal protocol analysis requires an algorithm or description on how to handle each verbal unit [38, 39]. Berg [14, 39] categorised the data into descriptive features and attitudinal features. Each category was then divided into two subsets. For descriptive features, the units were divided into unimodal (only audio modality) and polymodal (other sensory modalities). The attitudinal features were divided in the same way but into emotional/evaluative attitudes or attitudes related to naturalness. The current analysis, however, differs from Berg's approach: each answer to the open-ended questions for both its descriptive and its attitudinal features was analysed as each answer may contain both. Both the descriptive and attitudinal features of each answer were interpreted based on the context (each question) from which they were taken. The sorting processes were conducted in the following way: Sorting 1– each answer was sorted after its interpreted descriptive feature using the labels below; e.g., if the answer were related to audio quality, it was given a sound quality-related label (Sqr); and Sorting 2–each answer was sorted after its interpreted attitudes, i.e., positive, negative, both, or blank. The following labels were used when sorting each answer into its respective topic and attitude:

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

Sorting 1. Descriptive features: Sound quality related (Sqr), video quality related (Vqr), sound and video quality related (Sqr&Vqr), communication related (Com) ("communication" here refers to speaking and interaction among the participants), teaching related (Tch), technology related (Tec), teaching and communication related (Tch&Com), and diverse statements (Div).
Sorting 2. Attitudinal features: positive (+), negative (-), positive and negative (+ & -), and attitudes that contained no positive or negative statements (blank).

Consequently, sorting 1 in the meaning categorization analysis provides an overview of the number of answers on each perceived quality. The second sorting (Sorting 2) provides an overview of the answers' attitudes (sorting 2): positive, negative, both, or blank (general statements).

3.4.2. Using interview analysis

The interview data were analysed using an approach commonly used in phenomenology studies [27, 32], meaning condensation [32]. One or several persons can do this type of analysis. When looking at the data from a phenomenological point of view, the interviewer and/or researcher must set aside his/her pre-understanding of the phenomenon to obtain objectively rich and clear descriptions of the phenomenon under study [32]. This is referred to as "bracketing". The bracketing started during the design of the interview questions by formulating open-ended questions to facilitate the subjects' own descriptions of their experience and continued throughout the analysis process.

Meaning condensation concentrates the uttered meaning into more essential meanings and contains five general steps: Step 1–reading the interview to obtain an overview of its content, establishing a sense of the whole; Step 2–creating units of meaning (the answers) as the interview subject expresses them; Step 3–creating themes that dominate the units of meaning; Step 4–asking questions to the units of meaning based on the research purpose; and Step 5–creating a summary of the interviews' central themes and presenting them in one descriptive statement per interview [27, 32]. The steps of the interview analysis used the original method as a baseline for the analysis. The steps used in our analysis are listed below:

1. Transcribing the interview (transcription methods used were structured in colloquial language [32]);
2. Creating units of meaning of each transcribed answer;
3. Creating themes;
4. Sorting the units' answers into corresponding themes; and
5. Summarizing each theme in the text for each interview.

All steps of the evaluation method are presented in a block diagram in Figure 1. Each step (block) is followed by its result.

4. Evaluation of perceived qualities of a master class

The qualitative approach presented in section 3 was developed to evaluate perceived audio quality during real MED situations while considering the perceived video quality and interaction between the teacher and student. The MED situations were real master classes in classical music taught over distance and connected via video conferencing systems and an IP network (Public Internet). That is, the teacher was at one location and the student at another. This section presents the results from the qualitative approach.

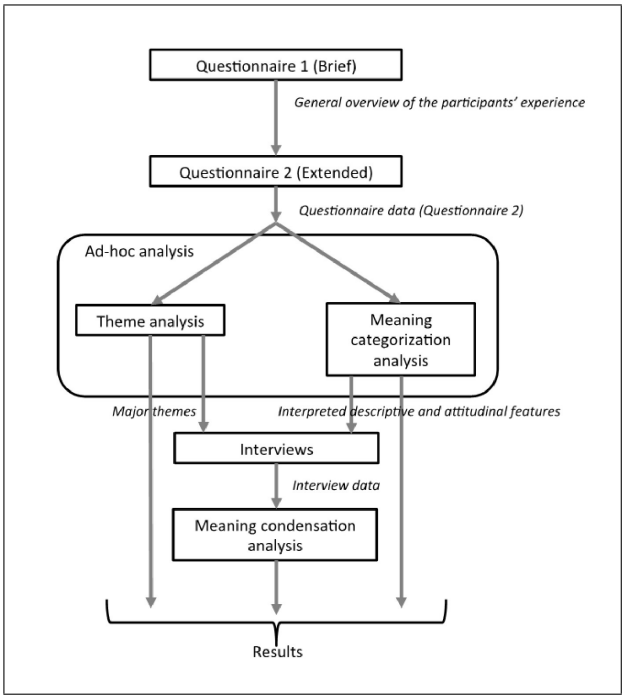


Figure 1. The steps and the results for each step used in the evaluation method.

The study was conducted under non-experimental conditions. Measurements of latency between the location and audio and video were not possible due to inaccessibility to measurement equipment. Table III provides information about what equipment was used at each location.

4.1. Participants

The master classes were conducted in Oulu (Finland), Helsinki (Finland), Piteå (Sweden), Olos (Finland), and Rovaniemi (Finland) on several occasions during the autumn and winter of 2009 and spring and autumn of 2010. The instruments/ensembles used in these master classes were violin, French horn, cello, and string quartet. Singing was also a part of some classes.

The participants in the study were college/ conservatory/ university music students and teachers who were recruited by means of invitations trough their home institutions. Hence, they may be regarded as experienced performers compared to the general public as well as to music students at preceding levels of the education system. Their participation was voluntary. In the beginning of the study (Questionnaire 1 and 2, see below), teachers and students directly involved in actively singing or playing their instruments as well as students observing the players par-

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

Table III. Table with the equipment for each location.

Location	Equipment
All locations	Tandberg MXP Edge 95 video conferencing systems, 50-52 inch LCD television screens
Piteå	2x Neumann KM184 microphones (occasionally a Microtech Gefell UMT 70s microphone) 2x Genelec 1030A speakers
Helsinki	2x Neumann TLM-103 microphones 2x Genelec 8030A speakers
Olos	2x Neumann TLM-103 microphones 2x Genelec 8030A speakers
Oulu	2x Neumann TLM-103 microphones 2x Genelec 8030A speakers
Rovaniemi	TV speakers 1x Clockaudio limited C600 microphone

ticipated. In a later step (Interviews), only active teachers and students partook.

4.2. Procedure

The steps used for evaluation of the master classes were as follows:
1. Questionnaire 1 (Brief);
2. Questionnaire 2 (Extended);
3. Ad-hoc analysis of the questionnaires;
4. Interviews; and
5. Analysis of the interviews (meaning condensation).

Both questionnaires were completed and the interviews were conducted after each lesson so as not to interrupt the performance of the teachers and students. The time between the closing of the master class and the distribution of questionnaires and/or conducting the interviews was no longer than one hour.

During the first master class, Questionnaire 1 (three open-ended questions) was distributed to obtain a general overview of the participants' experiences: their evaluation of the overall sound quality and their evaluation of their overall distance learning experience. A total of six questionnaires were collected. From this information, a second questionnaire was designed and distributed to collect more detailed descriptions from the participants. The questionnaires were distributed to all participants (teachers, students, and observers participating in the master classes). A total of 22 questionnaires were collected over a period of three master classes. The participants were allowed to complete the questionnaires at home for later submission. Hence, the time used by the participants for completion of the forms differed from less than one hour up to several days. The use of questionnaires is outlined in section 3.3.1.

After the distribution of the questionnaires, a total of six interviews were conducted by the first author: two interviews with students participating in the master classes and this master class study and four interviews with two master class teachers. Both students had previous experience of master classes and some experience with virtual distance communication but had no pervious experience with MED situations before this project. Both teachers had some experience of MED situations previous to this project. The interview followed the principles outlined in section 3.3.2.

One of the teachers was interviewed three of the four times in order to collect additional information from each session. The responses were recorded on an audio recorder and later transcribed. The second teacher was interviewed via e-mail and several follow-up e-mails were sent to encourage further reflection on topics deemed interesting by the interviewer. This e-mail correspondence can be comparable to a regular interview because of the nature of asking follow-up questions to inquire interesting leads as well as asking for clarification on certain answers.

4.3. Analysis

The questionnaire data and interview data was analysed by using the traditional data analysis methods that goes along with phenomenology and hermeneutics, presented in section 3.4. For simplicity, one person did the coding and meaning condensation, as the hermeneutical point of origin does not preclude the use of a single interpreter (coder) provided that this person is aware of his/her possible preconceptions and prejudices. Although, later in the process an investigator triangulation has also been applied, i.e. different evaluators to review the findings in order to reduce potential bias [40].

4.4. Results from the analysis

4.4.1. Questionnaire analysis (ad-hoc analysis)

The first part of the analysis, finding major trends, gave four major themes/trends. Quotations are translated to English and originate from Swedish, English and Finnish. The sound quality-related trend contained statements on the perception of the sound. Several answers stated that the instruction and music examples played from the teacher/student were perceived as good and clear:

"Good natural sound".

"Sound was very clear; there were no problems understanding what the teacher said".

The sound and video quality-related trend included the perception of an asynchrony between audio and video:

"The delay could be shorter".

"The sound and the picture should be in the same time".

The teaching-related trend showed that the participants thought that the distance master classes offered more opportunities for participating in master classes and that they offered the opportunity to play in front of different teachers. This trend also meant travel was not required:

"The teaching was excellent".

"You don't have to travel far away to get lessons".

The communication-related trend contained the perception of having a hard time communicating with the student or the teacher and playing together with the teacher:

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

"It is more difficult to communicate with the teacher".
"There should be a clear signal to the students so that they would know when to stop playing".

The second part of the ad-hoc analysis – meaning categorisation analysis (Table IV) – shows the total number of responses of the categories. The purpose of the categorisation is to outline the major trends/distribution of collected questionnaire answers. Table IV also shows the distribution of the received answers: the results show that the most frequently occurring responses are positive and are related to teaching. The second largest quantity of responses is positive and related to sound quality. The third largest quantity of responses does not contain any interpreted attitudes of positive, negative, both, or general statements (blank). These qualities are related to sound and video quality. The data also show that positive attitudes are most common. The second most common attitude is general statements without any positive/negative or both attitudes in the answers. The third most common attitude is negative and the fourth most common contains both positive and negative attitudes. This analysis does not allow for more detailed analysis more then showing rough chunks of data and major trends. As shown by Raimbault [41] one runs the risk to overlook certain patterns by looking at the data only from one perspective. Since this research is exploratory, the choice of only looking at the major trends is a conscious decision made by the authors in order to discover major tendencies in the teachers and students responses.

4.4.2. Interview analysis (meaning condensation)

The descriptive texts collected from each interview (step 5, Interview analysis) are summarized in this section into corresponding themes, i.e., making one descriptive text/description for each theme based on all six interviews, one person three times and three other persons, one time each. This strategy is done to maintain the richness of the information in the collected data. A total of three themes were created.

Perceived audio and video quality

The teachers perceive that the video quality is good and works for distance master classes. The teachers could also imagine how the sound of each instrument sounded live, based on experience, even though the instruments did not sound natural during the master class. One teacher perceived the sound quality as "metallic" and "boring", but with significant direct sound, no room sound, and good dynamics. The same teacher compared the sound to a high fidelity MP3. One teacher perceived a delay between audio and video, with the audio leading. With respect to sound quality, one teacher found it hard to distinguish between the system's limitations and the student's limitations although the teacher could distinguish this difference in a later master class. One teacher could easily see the student's playing technique, but could not evaluate it because the sound was difficult to hear.

Table IV. The number of responses in each category from the meaning categorisation analysis from the extensive questionnaire survey. The bold numbers show the categories attaining the largest number of statements in total and for each attitude.

Descriptive features	Attitudinal features				
	+	-	+&-	Blank	Total
Tch	**35**	4	3	1	**43**
Sqr	**17**	4	8	7	**36**
Sqr&Vqr	3	4	5	**14**	**26**
Diverse	6	3	1	8	18
Com	4	5	1	7	17
Vqr	2	1	0	6	9
Tec	0	5	0	1	6
Tch&Com	2	0	0	1	3
Total	**69**	**26**	18	**45**	

Both students compared the video quality as equal to or better than a YouTube clip. One student could only see the main features but not the contours or the proportions of the image. One student perceived the audio quality as "far away, distant, and a little muddy" and compared the audio to a YouTube clip. In addition, the same student found it hard to understand the teacher's voice. One student compared the sound quality to a MP3 coded sound, but worse than a movie although the student could distinguish between a normal spoken voice and a softly spoken voice. One student perceived the delay between audio and video as strange.

Perceived problems and possibilities

The delay between audio and video and the locations are perceived as a problem that makes it hard to communicate with the teacher/student. Small details in the music disappeared. Not knowing what is sent to the other locations, both related to video and sound quality, is perceived as a problem. During a distance master class, it is also hard to perceive the playing technique used by the students and how the students control their muscles. Distance learning is also problematic for music classes because it is difficult for teachers to evaluate their students' playing techniques and muscle control.

Several topics –controlling tempo, intonation, articulation, and phrasing– were adequately dealt with during the classes. According to one teacher, if the delays between audio and video were short and the teachers were aware of it, the delay could be less problematic. Both teachers perceived the technology in an overall positive light. One student and one teacher state that there are indirect benefits of distance master classes since they provide means to connect with teachers/students without traveling.

Perceived differences and similarities between regular and distance master classes

The perceived similarities were meeting with the student personally and discussions with the student. In addition, from a pedagogic point of view, the distant learning master class was similar to a regular master class. Two differences

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

were identified: correcting playing technique was difficult and verbal explanations rather than hands-on demonstrations were required to explain new positions to the student, a situation that required more time. In addition, creating a relationship with a teacher/student during distance master classes was more difficult than during a regular master class. One teacher also stated that a physical/personal contact established before starting long periods of distance lessons would be helpful.

5. Discussion

5.1. The results

Seen together, the ad-hoc analysis and the meaning condensation analysis provided an overview of what type of trends/themes exist in the data and how the perceived audio quality relates to the perceived video quality and to the interaction between the users.

The results from this study can be seen in the following points:

- The subjects could complete the master classes without major difficulties.
- The audio and video quality was not optimal but sufficient.
- There was a lack of synchrony between audio and video.
- There was a perceived delay between locations.
- Teachers had to adapt their teaching to accommodate for the system used.
- MED is perceived as cost-affective.
- Participants can access more teaching and learning experiences with less travelling.

These results confirm previous research findings on several counts. Masum *et al.* found that teachers and students found the tools (system used) comfortable [4]. This finding coincides with the authors' results; the subjects could complete the master classes without any major difficulties.

The authors' results also indicate that the perceived audio and video quality in the system is not optimal but is sufficient for this type of music education. There are also problems with lack of synchrony between video and audio and latency between the locations, but the subjects' statements indicate that they can work around these problems and manage a MED situation successfully. Woszczyk *et al.* also report similar results [7].

The results also align with previous research when it comes to teaching. To conduct successful teaching with the use of video conferencing systems, the teacher needs to adapt the content to handle the pedagogical situation [6]. In addition, it is cost-effective to bring teachers and teaching experiences to a large population of students [6]. Clearly, videoconference systems also allow teachers to train and teach in places other than their home location [4]. These previous results all coincide with the results from the current study. The implications of these findings will be discussed the section 5.2.

The subjects' initial attitude towards using such a system could be a bias that affected the results. That is, a student may have entered the study with a preconceived idea about distance learning. Such preconceived attitudes need to be considered when dealing with subjective responses. Before the study, based on their experience of sound quality, the authors expected the subjects to be negative about the sound quality; however, the results did not indicate this. Another possible bias, which may be connected to the positive attitudes collected, is that all the participants volunteered to participate in the study, showing they had an initial interest in MED; that is, the participants were self-selected on some level.

5.2. The methodology

Using qualitative methods including analysis shows a potential for arriving at a set of data that can be usable when evaluating several perceived qualities in one system. This method, with post questionnaires and post interviews, allows the participants to complete their task without interruptions and encourages the subjects to use their own words to describe their experience. This approach helps create a broad picture of what is happening in the study. This method also allows for a simultaneous description of what is perceived and what is affected even if there are more parameters affecting the perceived total quality than just the audio quality. Comparing this method to other subjective assessment methods of audio quality proposed by other studies [10], one could not say a particular method is better than another method because they collect different sets of data; however, some subjective assessment methods [8, 9, 10, 11, 12, 15] have predetermined verbal descriptors and factors that aim at a defined part of the perceived quality. The approach in this paper enables the subjects to reflect on what they have perceived with few restrictions. Thus, the information can shed light on broader aspects of the perceived quality. This broad approach makes it possible to discover unexpected and possibly important factors, related either to audio quality or to other qualities that affect how a specific situation or implementation is perceived. Hence, factors outside the audio domain may also be considered.

On the topic of analysis, the data presented in Table 1 do not say anything in detail about the content of the responses; the data only show the distribution of answers, which demonstrates that simple quantitative observations can also be made using this method. A difficulty with this categorisation/numeric summary as with all categorization methods is that one can choose other categories and thus get a different distribution of the answers. In this study, the blank category refers to that no attitudinal response could be discovered. Consequently, blank responses may indicate that no strong emotions were evoked in those particular cases and possibly that the responses were more of a descriptive character. The lack of further detailed information, e.g. in subcategories, is a result of the chosen approach [41].

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

By using open-ended questions in a questionnaire, a variety of answers can be elicited. This strategy can be an advantage if a participant gives an answer that sheds light on a new area that the researcher has overlooked, or it can be fruitless if the participants "don't know" or do not even answer the question. Open-ended questions used during an interview can encourage the interviewee to answer freely while still allowing the interviewer to guide the interviewee into interesting topics if they arise. In the latter situation, the interviewer needs to be very responsive to follow-up possibilities and needs to understand what to ignore. Such a strategy, of course, is a source of bias that has to be considered carefully. Researchers using questionnaires and conducting several interviews during an evaluation process must be aware that the data size quickly becomes large and is hard to manage and time-consuming to analyse.

As stated in section 3.3 the study design can be seen as a funnel moving from at set of broad answers collected from the questionnaires 1 and 2 to more detailed answers in the interviews. This design was used since the questionnaires worked similarly to a pilot study to facilitate narrower questions and topics in the interviews.

In the meaning condensation of the interviews, each unit of meaning was categorized into themes. This division made the answers more clear and provided a better overview and a better understanding of the interviewed subject's perception. The study includes repeated interviews with one teacher. This will unavoidably create learning effects that affect the response. Learning effects was not studied separately as the major trends were in focus this time. However, for future studies of the method's characteristics, this may be of interest.

As all data collection was done after the completion of the event, this collection method relies on the subject's memory. When answering questions after an event rather than during it, some shift in the recollection of perceived sensations may occur. This shift can be a disadvantage compared to other more direct methods used for rating and assessing. Small differences between sessions may not be captured, as they would be harder to detect when one experience is compared to another with a time gap in between. Further studies with these methods might be enhanced by a stimulated recall approach by means of video recording the master classes allowing the subjects to watch and comment and thus facilitating the subject to recall what happened during the master class.

Because the study was conducted during a real MED situation, some other limitations did arise. One limitation of the methodology is that in this particular case only one person conducted the coding. As pointed out, the hermeneutical approach does not disqualify a single interpreter from doing the coding, provided that this person is aware of his/her possible preconceptions. Though, in several applications more than one person performs coding of the data. In the current study, the findings were discussed among the authors in order to reduce the potential bias and when using a phenomenological or hermeneutical approach for the analysis, the methods used requires the researcher to be aware of his/her own preconceptions and prejudices and exclude them during the analysis. The author performing the analysis was aware of this possible bias (examples in section 5.1) during the design of questions as well as under the analysis of the results.

As can be seen in section 4, Table III, the equipment was changed occasionally between different locations and sessions. This can, of course, result in a bias, but the primary focus was not to link a set of experiences to particular equipment, but to evaluate the perceived audio and video quality in an ecological valid situation during live distance master classes. Another restriction was the unavailability of measuring the delay between the locations and between audio and video. This could have shed some light on when delay was present and when the delay was not present.

As indicated, the approach used in the current study yields different information compared to most of the previously used methods that evaluate audio quality. Hence, this method cannot be used interchangeably with existing methods to obtain the same type of data. However, by adding information that is not available from other methods, this approach will increase the knowledge of the subject's experience. The results may also be used in an exploratory way as a means of observing what subjects perceive as noticeable, which in its turn can be used to develop evaluation scales used in existing methods. The rich verbal data resulting from a qualitative approach may provide a more holistic representation of an audio event, improving our understanding of how the event is experienced. For further research on qualitative research applied on perceived sound and video quality the method and framework of introspection (examining ones own conscious thought and feelings [42]) might be fruitfully applied. See [42].

The high correspondence between the results from this study and the studies quoted in section 5.1 can be seen as a successful triangulation and a verification of that the method employed in this study enables the extraction of results that have previously been found in similar contexts. In addition, the method allows for unexpected features to be discovered e.g. the interaction influence on perception of audio and video quality, making the audio quality for an example good enough. The method is shown also to be potent for extracting similar data from real-life situations as data extracted from experimental conditions. Altogether this strengthens and encourages future development and use of this type of methods.

Acknowledgement

This work is part of the Vi R Music project and was partially financed by the European Union program Interreg IV A Nord. The authors would like to thank all the participants, students, and teachers who participated in the surveys and interviews. A thanks also goes to colleagues, technicians, and staff members who made this work possible. The authors also want to express a special thanks to

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

Dr. Anders Persson, Jonas Ekeroot, and Jon Allan for their valuable comments on the drafts.

References

[1] E. Brunswik: Organismic achievement and environmental probability. The Psychological Review **50** (1943) 255–272.

[2] J. J. Gibson: The ecological approach to the visual perception of pictures. Leonardo **11** (1978) 227–235.

[3] C. Guastavino, B. F. G. Katz, J.-D. Polack, D. J. Levitin, D. Dubois: Ecological validity of soundscape reproduction. Acta Acustica united with Acustica **91** (2005) 333–341.

[4] H. Masum, M. Brooks, J. Spence: MusicGrid: A case study in broadband video collaboration. http://firstmonday.org/ojs/index.php/fm/article/view/1238/1158.

[5] B. K. Shepard, G. Howe, T. Snook: Internet2 and musical applications. Proceedings, National Association of schools of music, The 84th annual meeting, Seattle, USA, November 24, 2008. http://www.briankshepard.com/pdf/2008Proceedings.pdf.

[6] A. Greenberg, R. Colbert: Navigating the sea of research on videoconferencing-based distance education. Wainhouse Research, Polycom, Inc., February 2004. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.8045&rep=rep1&type=pdf.

[7] W. Woszczyk, J. Cooperstock, J. Roston, W. Martens: Shake, rattle, and roll: Getting immersed in multisensory, interactive music via broadband networks. Journal of Audio Engineering Society **53** (2005) 336–344.

[8] F. E. Toole: Subjective measurements of loudspeaker sound quality and listener performance. Journal of Audio Engineering Society **33** (1985) 2–32.

[9] A. Gabrielsson, B. Lindström: Perceived sound quality of high-fidelity loudspeakers. Journal of Audio Engineering Society **33** (1985) 33–53.

[10] S. Bech, N. Zacharov: Perceptual audio evaluation – Theory, method and application. John Wiley & Sons Ltd., Chichester, England, 2006.

[11] BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. ITU-R Recommendations, 1994–1997.

[12] BS.1534-1: Methods for the subjective assessment of intermediate quality level of coding system. ITU-R Recommendations, 2001–2003.

[13] G. Lorho: Individual vocabulary profiling of spatial enhancement systems for stereo headphone reproduction. Presented at the 119th AES convention, New York, New York USA, October 7-10, 2005.

[14] J. Berg, F. Rumsey: Identification of quality attributes of spatial audio by repertory grid technique. Journal of Audio Engineering Society **54** (2006) 365–379.

[15] BT.500-12: Methodology for the subjective assessment of the quality of television pictures. ITU-R Recommendations, 09/2009.

[16] K. Seshadrianthan, R. Soundararajan, A. C. Bovik, L. K. Cormack: Study of subjective and objective quality assessment of video. IEEE transactions on image processing **19** (2010).

[17] J. Radun, T. Leisti, J. Häkkinen, H. Ojanen, J.-L. Olives, T. Vuori, G. Nyman: Content and quality: Interpretation-based estimation of image quality. ACM Transactions on Applied Perception **4** (2008) Article 21.

[18] G. Nyman et al.: What do users really perceive – probing the subjective image quality. Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol 6059, 2006, 605902.

[19] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze: Open profiling of quality: A mixed method approach to understanding multimodal quality perception. – In: Advances in Multimedia, Vol 2010. Hindawi Publishing Corporation, 2010, Article ID 658980.

[20] S. Jumisko-Pyykkö, J. Häkkinen, G. Nyman: Experienced quality factors – qualitative evaluation approach to audiovisual quality. Proceedings of SPIE - The International Society of Optical Engineering, 6507, 2007, art. No. 65070M.

[21] S. Jumisko-Pyykkö, U. Reiter, C. Weigel: Produced quality is not perceived quality – A qualitative approach to overall audiovisual quality. 3DTV Conference, May 7-9, 2007.

[22] J. G. Beerends, F. E. De Caluwe: The influence of video quality on perceived audio quality and vice versa. Journal of Audio Engineering Society **47** (1999).

[23] M. P. Hollier, A. N. Rimell: An experimental investigation into multi-modal synchronisation sensitivity for perceptual model development. Presented at the 105th AES Convention, San Francisco, CA, September 26-29, 1998.

[24] S. Bech, V. Hansen, W. Woszczyk: Interaction between audio-visual factors in a home theatre system: Experimental results. Presented at the 99th AES Convention, New York, October 6-9, 1995.

[25] J. Häkkinen, V. Alatonen, M. Schrader, G. Nyman, M. Lehtonen, J. Takatalo: Qualitative analysis of mediated communication experience. Quality of Multimedia Experience (QoMEX), Second International workshop, June 2010.

[26] J. Blauert, U. Jekosch: A layer model of sound quality. Journal of Audio Engineering **60** (2012).

[27] R. Tesch: Qualitative research: analysis types and software tools. RoutledgeFalmer, 2 park Square, Milton Park, Abingdon, Oxon, Ox14 4RN, 1990.

[28] C. E. Moustakas: Phenomenological research methods. SAGE publications, Inc., Thousand Oaks, California, 1994.

[29] S. B. Marriam: Qualitative research, a guide to design and implementation. Jossey–Bass, San Francisco, CA, US, 2009.

[30] B. G. Glaser, A. L. Strauss: The discovery of grounded theory: Strategies for qualitative research. Aldine Publishing Company, 200 Saw Mill River Road, Hawthorne, N.Y. 10532, 1967.

[31] A. L. Strauss, J. Corbin: Basics of qualitative research – Techniques and procedures for developing grounded theory, second edition. SAGE Publications, 2455 Teller Road, Thousands Oaks California 91320, 1998.

[32] S. Kvale, S. Brinkmann: InterViews – Learning the craft of qualitative research interviewing. Second edition. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks, California 91320, 2009.

[33] X. Fang, S. Chan, J. Brzezinski, C. Nair: Development of an instrument to measure enjoyment of computer game play. International Journal of Human-Computer Interaction **26** (2010) 868–886.

[34] S. Wolfson, G. Case: The effect of sound and colour on responses to a computer game. Interacting with computers **13** (2000) 183–192.

[35] L. E. Nacke, M. N. Grimshaw, C. A. Lindley: More then a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. Interacting with computers **22** (2010) 336–343.

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only

[36] S. H. Hsu, M.-H. Wen, M.-C. Wu: Exploring user experiences as predictors of mmorpg addiction. Computers & Education **53** (2009) 990–999.

[37] G. Ejlertsson: Enkät i praktiken - en handbook i enkätmetodik (andra upplagan). Studentlitteratur, 2005. ISBN 91-44-03164-5.

[38] E. Samoyleko, S. McAdams, V. Nosulenko: Systematic analysis of verbalisation produced in comparing musical timbres. International Journal of Psychology **31** (1996) 255–278.

[39] J. Berg: Systematic evaluation of perceived spatial quality in surround sound systems. Publication 2 3-4; Doctoral dissertation, Department of Music and Media, Piteå, Sweden, 2002.

[40] N. K. Denzin, Y. S. Lincoln: Introduction: Entering the field of qualitative research. – In: Handbook of qualitative research. N. K. Denzin, Y. S. Lincoln (eds.). SAGE, Thousand Oaks, CA, 1994.

[41] M. Raimbault: Qualitative judgements of urban soundscapes: Questionning questionnaries and semantic scales. Acta Acustica united with Acustica **92** (2006) 929–937.

[42] P. Vermersch: Introspection as practice. Journal of Consciousness Studies **6** (1999) 17–42.

Document downloaded by @DAEL, 2026-11-11 05:17:26 - personal use only