About the Different Types of Listeners for Rating the Overall Listening Experience

Michael Schoeffler
International Audio Laboratories Erlangen*
michael.schoeffler@audiolabs-erlangen.de

Jürgen Herre
International Audio Laboratories Erlangen*
juergen.herre@audiolabs-erlangen.de

ABSTRACT

When the overall listening experience is rated, listeners are asked to take every aspect into account which seems important to them, including song, lyrics, mood and audio quality. The results of two previously conducted experiments revealed a significant influence of the signal bandwidth and the spatial reproduction format on the overall listening experience. In this work, a systematic analysis is applied to the results of these two experiments with the purpose to investigate listeners in more detail. Regarding rating the overall listening experience, the results show that listeners can rather be described by continuous variables which reflect their preferences than clear categorizations of different listener types. Furthermore, a regression model for predicting ratings was significantly improved by describing the listeners with such continuous variables.

1. INTRODUCTION AND RELATED WORK

The enjoyment while listening to music has been a subject of interest to many researchers from different fields. In 1972, Prince published a paradigm which describes a music listening process from the listener's point of view [1]. Prince assumes that listeners differ widely in their response to music depending on many characteristics including personality, maturation, musical training and experience, musical aptitude, and musical memory.

Several researchers have confirmed that responses to music are in fact significantly influenced by characteristics which were described by Prince. When measuring the influence of personality on responses to music, many researchers use Jungian types as basis for their work. Jungian types describe psychological types of humans and were introduced by Jung [2]. Based on Jung's theory, the Myers Briggs Type Indicator (MBTI) sorts psychological differences in four different preferences: Extraversion–Introversion, Sensing–Intuition, Thinking–Feeling, and Judging–Perception [3]. Pearson et al. investigated whether Jungian types have a correlation with music preferences [4]. They found out that the Sensing–Intuition MBTI dimension correlates with enjoyment in musical listening. Similarly, Lewis and Schmidt also used the MBTI to examine listeners' responses to music as a function of the personality variables measured by the MBTI [5]. Undergraduate non-music majors indicated characteristic responses to music on the Music Listener Response Scale (MLRS). The MLRS is a questionnaire which contains statements with responses to music (see Hedden [6]). Lewis and Schmidt found significant effects of personality types based on the MBTI on the responses given by the MLRS.

Related to the characteristic of musical training, Woody and Burns carried out a study in which they investigated how non-musicians respond to pieces of classical music that are considered highly expressive among musically sophisticated listeners [7]. The results of their study indicate that young adults who have had past emotional experience with classical music are more responsive to the expressive qualities of classical music and are more willing to listen to this style of music in their leisure time. Experiments by Stöter et al. and Schoeffler et al. indicated that musicians are able to correctly estimate the number of instruments of music excerpts more likely than non-musicians [8, 9]. However, it is not known whether the enjoyment of music listening is influenced by an awareness of the instrumentation.

Regarding the listening experience, researchers often distinguish between expert listeners and naïve listeners (or trained and untrained listeners). Expert listeners are considered to be trained in listening, familiar with audio degradation artifacts and have taken part in many listening tests. Schinkel-Bielefeld et al. investigated the differences between expert and naïve listeners while taking part in listening tests about audio quality degradations [10]. Their results showed that expert and naïve listeners have different strategies for participating in listening tests. Expert and naïve listeners differently use the rating scale and expert listeners tend to give lower ratings. Furthermore, the ratings from experts listeners are more reliable than ratings from naïve listeners. Rumsey et al. examined the relationship between expert listener ratings of multichannel audio quality and naïve listeners preferences [11]. They found out that naïve listener preferences can be predicted from expert listener ratings. However, their results also showed that naïve listeners and expert listeners are not similar groups in every aspect of audio quality evaluation. In contrast to the expert listeners, naïve listeners paid almost

*A joint institution between the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS no attention to frontal spatial audio fidelity. Also surround spatial fidelity was much more important for naïve listeners than for expert listeners. Although listening experience had an influence on the results of these studies related to audio quality, no influence on the overall listening experience was found in three studies carried out by Schoeffler et al. [12–14]. The term overall listening experience is used to describe the degree of enjoyment while listening to music [13]. Contrary to rating the audio quality, listeners take everything into account when rating the overall listening experience. Factors of influence might include the song, lyrics, audio quality, listener's mood, the listening room and the reproduction systems.

This paper is concerned with the question whether listeners can be precisely categorized according to how much their individually-perceived overall listening experience depends on the technical sound quality. In this paper, the technical sound quality representations are bandwidth-limitations of played-back music and different spatial reproduction format which are used for play back music. We measured the influence of the signal bandwidth on the overall listening experience by a previously conducted experiment [13]. The results of the experiment are used in the analysis for the listener categorization.

> *Hypothesis 1a:* Listeners can be categorized according to how much their individually-perceived overall listening experience is influenced by the signal bandwidth of the played back music excerpt.

In the first experiment, participants were asked whether audio quality had been important for their ratings. The question arises if their answers corresponds to their actual ratings.

> *Hypothesis 1b:* Listeners can reliably self-report how important the audio quality is for their individual overall listening experience.

How much the enjoyment of listeners is influenced by the single-/multi-channel system was measured by another previously conducted experiment whose results are also used in the analysis [14]. In addition to the two hypotheses of the first experiment, we define two hypotheses of the second experiment which have not been addressed so far.

> *Hypothesis 2a:* Listeners can be categorized according to how much their individually perceived overall listening experience is influenced by the spatial reproduction format.

> *Hypothesis 2b:* Listeners can reliably self-report how important the single-/multichannel system is for their individual overall listening experience.

An additional research question is whether listeners, whose overall listening experience is influenced by the signal bandwidth, are influenced by the single-/multi-channel system in the same way when the overall listening experience is rated.

> *Hypothesis 3:* If a listener's overall listening experience is influenced by the bandwidth, his or her overall listening experience is influenced by the single-/multi-channel system in the same way.

The two experiments are briefly described in Section 2, Section 3 and Section 4, including a statistical verification of the five hypotheses. The results are discussed in Section 5.

2. TYPES OF LISTENERS FOR BANDWIDTH (EXPERIMENT I)

2.1 Experiment Procedure

In the first experiment, 34 participants rated bandwidth-limited music according to the overall listening experience. The experiment was divided into one registration session and twelve listening session.

In the registration session the participants were asked by a questionnaire whether they are professionals in audio and to which age group they belong. After filling out the questionnaire, the participants rated nine songs according to how much they like them ("How much do you like this item?"). It was emphasized on the instructions that participants should take everything into account what they would do in a real world scenario (e. g. including their taste in music). The participants rated the songs by using a five-star Likert scale. The stars were labeled with "Very Bad", "Bad", "Average", "Good" and "Very Good". The ratings, retrieved from the first session, are called *basic item ratings*. A maximum of eight out of nine songs were individually selected from the registration session and bandwidth-limited by six levels (Cut-off frequencies: 1080 Hz, 2320 Hz, 4400 Hz, 7700 Hz, 12000 Hz and 15500 Hz) for the listening sessions. A maximum of eight songs had to be rated in each listening session since the same song did only occur once. As in the registration session, the participants were instructed to rate each bandwidth-limited item according to how they like it and they should take everything into account what they would do in a real world scenario. After listening to all bandwidth-limited music items, listeners had to fill out a questionnaire in each session. It was asked how important had been audio quality, the song and their mood for their ratings. The answers were given by a Likert scale with the values "Strongly Agree", "Agree", "Neutral", "Disagree" and "Strongly Disagree". The ratings retrieved from the listening sessions are called *item ratings*.

A detailed description of the experiment procedure and a basic analysis of the results has already been published [13].

2.2 Result Analysis

The Hypotheses 1a and 1b are verified on the basis of a pair of values which reflects how much a listener was influenced by the bandwidth-limitations and the song. Such a pair of values is derived by the following equations and then used for the hypotheses verification.

An item rating of the first experiment is defined as:

$$\mathrm{IR}^{\mathrm{E1}}(i, j, k), \tag{1}$$

where i denotes the participant, j denotes the song and k denotes the cut-off frequency level.

A vector with all items ratings which were given by a participant is defined as:

$$\mathbf{IR}^{\mathrm{E1}}(i) = \begin{bmatrix} \mathrm{IR}^{\mathrm{E1}}(i,1,1) \\ \dots \\ \mathrm{IR}^{\mathrm{E1}}(i,1,K) \\ \dots \\ \mathrm{IR}^{\mathrm{E1}}(i,J,K)] \end{bmatrix}, \tag{2}$$

where J is the number of songs that were selected for participant i and K is the number of all cut-off frequency levels. Vectors are indicated by bold letters.

The corresponding cut-off frequency of an item rating is defined as:

$$\mathrm{F}^{\mathrm{E1}}(i, j, k). \tag{3}$$

A vector with all cut-off frequencies that corresponds to all items ratings which were given by a participant is defined as:

$$\mathbf{F}^{\mathrm{E1}}(i) = \begin{bmatrix} \mathrm{F}^{\mathrm{E1}}(i,1,1) \\ \dots \\ \mathrm{F}^{\mathrm{E1}}(i,1,K) \\ \dots \\ \mathrm{F}^{\mathrm{E1}}(i,J,K)] \end{bmatrix}. \tag{4}$$

The corresponding basic item rating of an item rating is defined as:

$$\mathrm{BIR}^{\mathrm{E1}}(i, j, k). \tag{5}$$

A vector with basic item ratings that corresponds to all items ratings which were given by a participant is defined as:

$$\mathbf{BIR}^{\mathrm{E1}}(i) = \begin{bmatrix} \mathrm{BIR}^{\mathrm{E1}}(i,1,1) \\ \dots \\ \mathrm{BIR}^{\mathrm{E1}}(i,1,K) \\ \dots \\ \mathrm{BIR}^{\mathrm{E1}}(i,J,K)] \end{bmatrix}. \tag{6}$$

The influence of the cut-off frequency on the item ratings is measured by Kendall's tau coefficient which is a measure of rank correlation between two vectors [15]. Values of Kendall's tau range from -1 to +1. Negative values indicate negative associations and positive values indicate positive associations. A value of zero indicates the absence of any association between two variables. A pair of two Kendall's tau values are calculated for each participant. The first value is Kendall's tau of the item ratings and the cut-off frequencies:

$$\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i) = \mathrm{cor}_\tau(\mathbf{IR}^{\mathrm{E1}}(i), \mathbf{F}^{\mathrm{E1}}(i)), \tag{7}$$

where cor_τ denotes the function of Kendall's tau. How much the ratings a participant were influenced by the bandwidth-limitation of the stimulus is reflected by $\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i)$. The second value is Kendall's tau of the item ratings and the basic item ratings:

$$\tau^{\mathrm{E1}}_{\mathrm{IR,BIR}}(i) = \mathrm{cor}_\tau(\mathbf{IR}^{\mathrm{E1}}(i), \mathbf{BIR}^{\mathrm{E1}}(i)). \tag{8}$$

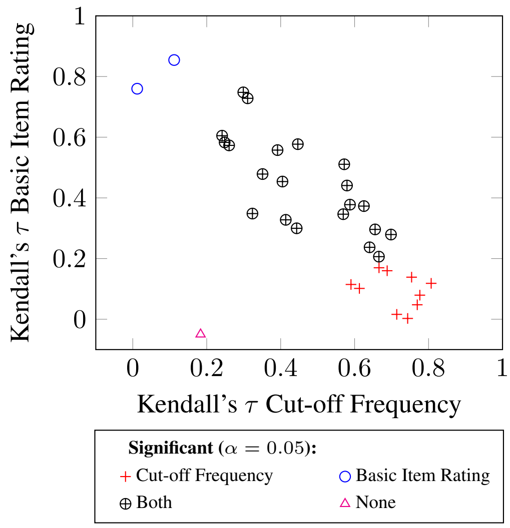

Figure 1. Kendall's rank correlation between item rating and the two variables "basic item ratings" and "cut-off frequency" for each participant. Correlations which are significant for basic item ratings, single-/multi-channel system, none or both are marked differently.

$\tau^{\mathrm{E1}}_{\mathrm{IR,BIR}}(i)$ reflects to which degree the ratings of a participant were influenced by the song. A pair of Kendall's tau values of the first experiment is defined as:

$$\tau^{\mathrm{E1}}(i) = \{\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i), \tau^{\mathrm{E1}}_{\mathrm{IR,BIR}}(i)\}. \tag{9}$$

The pairs of Kendall's tau values are depicted in Figure 1 as scatter plot.

In the lower left corner of the scatter plot is a data point which can be considered as outlier, since its corresponding item ratings have neither a significant [1] correlation with the cut-off frequencies nor the basic item ratings. The Chi-squared plot method (see Garrett for details [16]) is applied to confirm that the data point is an outlier. The Chi-squared plot in Figure 2 clearly shows one outlier which is the non-significant data point. The identified outlier is excluded from further analysis.

Hypothesis 1a is verified by applying a cluster analysis on the data set. The number of components found by the clustering algorithm can be seen as categorizations of listeners. An Expectation Maximization (EM) clustering algorithm was chosen [17] (an introduction into the EM method is given by Couvreur [18]). The EM algorithm works well with small sample sizes and does not require a training phase which are the main reasons for choosing the EM algorithm. Since the number of components is not estimated by the EM algorithm itself, an implementation of the EM algorithm by Fraley and Raftery is used [19]. Furthermore, the implementation takes several cluster models into account. The method of Fraley and Raftery uses a modified version of the Bayesian Information Criterion for choosing the number of clusters and to select the cluster model. By applying the EM algorithm, an ellipsoidal model with one component was found to be the best model (see Figure 3).

[1] The significance level α is defined as 0.05 in this paper.

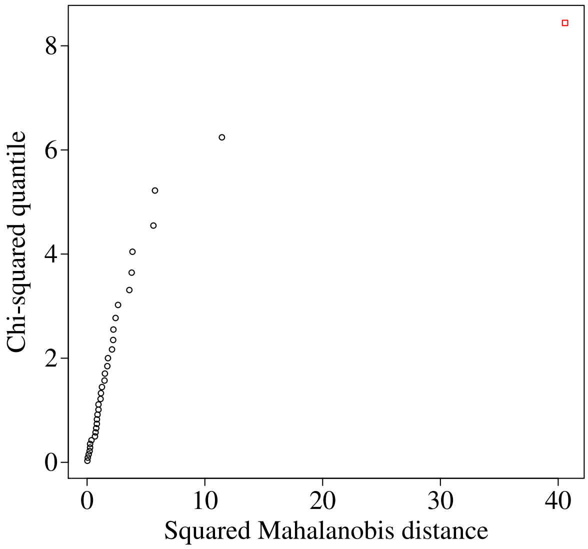

Figure 2. Chi-square plot of the Kendall's tau pairs. The detected outlier is depicted as a red square.

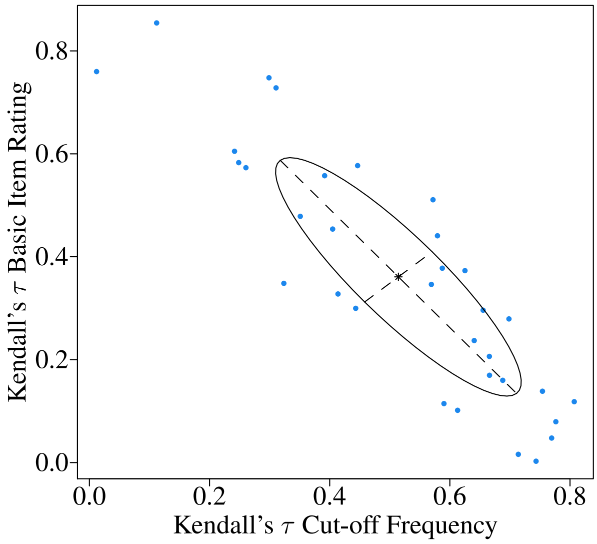

Figure 3. Clustering of the pairs of correlation values by ellipsoidal model with one components.

Based on this result, listeners can not be clearly categorized into types of listeners when rating bandwidth-limited music.

The responses of the questionnaire after each listening session are used to verify Hypothesis 1b. The correlation between the participants' $\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i)$ and their responses to the question – how important was the audio quality for participants' ratings – are investigated. Spearman's rho [2] ($r = 0.67, p = .000$) and Kendall's tau ($r = 0.56, p = .000$) show a significant correlation. Based on these results, it can be concluded that participants can reliably self-report how much they are influenced by the audio quality. The same is applied for $\tau^{\mathrm{E1}}_{\mathrm{IR,BIR}}(i)$ and the responses to the question of how important was the song for participants' ratings. Again, Spearman's rho ($r = 0.56, p = .001$) and Kendall's tau ($r = 0.46, p = .000$) show a significant correlation which indicates that participants can also reliably self-report how much they are influenced by the song.

3. TYPES OF LISTENERS FOR SINGLE-/MULTI-CHANNEL SYSTEMS (EXPERIMENT II)

3.1 Experiment Procedure

In the second experiment, 30 participants rated music excerpts according to the overall listening experience. The ratings were given by a Five-star Likert scale. The stars were labeled with "Not at all", "Not a lot", "Neutral", "Much" and "Very Much".

The experiment was divided into two sessions. In the first session, 30 participants rated fifteen music excerpts by a multi-stimulus comparison while listening through headphones. The ratings retrieved from the first session are called *basic item ratings*. In the second session, participants rated an individual selection of the music excerpts while listening through a single-/multi-channel system (mono, stereo or surround). The ratings of the second session are called *item ratings*. At the end of the second session, the participants had to fill out a questionnaire, where they were asked how important had been audio quality, the song and their mood for their ratings. The answers were given by a Likert scale with the values "Strongly Agree", "Agree", "Neutral", "Disagree" and "Strongly Disagree".

A detailed description of the experiment procedure and a basic analysis of the results has already been published in [14].

3.2 Result Analysis

Almost the same procedure as for Hypotheses 1a and 1b is applied to verify the Hypotheses 2a and 2b. A pair of values is calculated which reflects how much a listener was influenced by the single-/multi-channel system and the song.

An item rating of the second experiment is defined as:

$$\mathrm{IR}^{\mathrm{E2}}(i, j, l), \tag{10}$$

where i denotes the participant, j denotes the song and l denotes the single-/multi-channel system.

A vector with all items ratings which were given by a participant is defined as:

$$\mathbf{IR}^{\mathrm{E2}}(i) = \begin{bmatrix} \mathrm{IR}^{\mathrm{E2}}(i, 1, 1) \\ \cdots \\ \mathrm{IR}^{\mathrm{E2}}(i, 1, L) \\ \cdots \\ \mathrm{IR}^{\mathrm{E2}}(i, J, L) \end{bmatrix}, \tag{11}$$

where J is the number of songs that were selected for participant i and L is the number of all single-/multi-channel systems.

[2] Spearman's rho and Kendall's tau correlation coefficients are used for measuring the correlation between ordinal variables. Pearson's product-moment correlation is applied for measuring the correlation between continuous variables.

The corresponding single-/multi-channel system of an item rating is defined as:

$$\mathrm{S}^{\mathrm{E2}}(i, j, l). \tag{12}$$

A vector with all single-/multi-channel systems that corresponds to all items ratings which were given by a participant is defined as:

$$\mathbf{S}^{\mathrm{E2}}(i) = \begin{bmatrix} \mathrm{S}^{\mathrm{E2}}(i,1,1) \\ \cdots \\ \mathrm{S}^{\mathrm{E2}}(i,1,L) \\ \cdots \\ \mathrm{S}^{\mathrm{E2}}(i,J,L)] \end{bmatrix}. \tag{13}$$

The corresponding basic item rating of an item rating is defined as:

$$\mathrm{BIR}^{\mathrm{E2}}(i, j, l). \tag{14}$$

A vector with basic item ratings that corresponds to all items ratings which were given by a participant is defined as:

$$\mathbf{BIR}^{\mathrm{E2}}(i) = \begin{bmatrix} \mathrm{BIR}^{\mathrm{E2}}(i,1,1) \\ \cdots \\ \mathrm{BIR}^{\mathrm{E2}}(i,1,L) \\ \cdots \\ \mathrm{BIR}^{\mathrm{E2}}(i,J,L)] \end{bmatrix}. \tag{15}$$

Almost the same analysis procedure as for Experiment I is applied. Pairs of correlation values for each participant are created by calculating Kendall's tau values. The first value is Kendall's tau of the item ratings and the single-/multi-channel system:

$$\tau^{\mathrm{E2}}_{\mathrm{IR,S}}(i) = \mathrm{cor}_\tau(\mathbf{IR}^{\mathrm{E2}}(i), \mathbf{S}^{\mathrm{E2}}(i)), \tag{16}$$

where cor_τ is the function of Kendall's tau. The second value is Kendall's tau of the item ratings and the basic item ratings:

$$\tau^{\mathrm{E2}}_{\mathrm{IR,BIR}}(i) = \mathrm{cor}_\tau(\mathbf{IR}^{\mathrm{E2}}(i), \mathbf{BIR}^{\mathrm{E2}}(i)). \tag{17}$$

A pair of Kendall's tau values of the second experiment is defined as:

$$\tau^{\mathrm{E2}}(i) = \{\tau^{\mathrm{E2}}_{\mathrm{IR,S}}(i), \tau^{\mathrm{E2}}_{\mathrm{IR,BIR}}(i)\}. \tag{18}$$

The pairs of Kendall's tau values for all participants are depicted in Figure 4 as scatter plot.

As for Experiment I, the EM clustering method is applied to identify categories of listeners. An ellipsoidal model with one component was found to be the best model by the clustering method (see Figure 5). Such a results indicates that there is no clear categorization into types of listeners when the single-/multi-channel system is taken into account while rating the overall listening experience.

The responses, which were given by the questionnaire of the second session, are used to verify Hypothesis 2b. The correlation between $\tau^{\mathrm{E2}}_{\mathrm{IR,S}}(i)$ and the participants' responses to the question of how important was audio quality for their ratings. Spearman's rho ($r = 0.56, p = .001$) and Kendall's tau ($r = 0.46, p = .001$) show a significant correlation. The same is applied for $\tau^{\mathrm{E2}}_{\mathrm{IR,BIR}}(i)$ and the responses to the question of how important was the song for the participants' ratings. Again, Spearman's rho ($r = 0.54, p = .001$) and Kendall's tau ($r = 0.46, p = .002$) show a significant correlation. Based on these results, participants seem to reliably self-report how much they are influenced by the audio quality and the song.

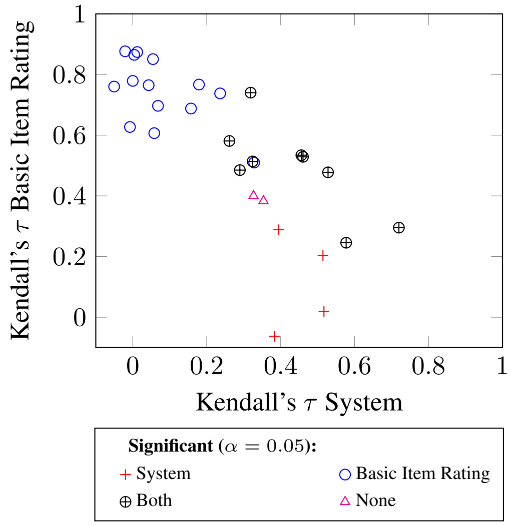

Figure 4. Kendall's rank correlation between item rating and the two variables "basic item ratings" and "single-/multi-channel system" for each participant. Correlations which are significant for basic item ratings, single-/multi-channel system, none or both are marked differently.

4. INTERACTION BETWEEN BANDWIDTH AND SINGLE-/MULTI-CHANNEL SYSTEM

Ten participants attended both experiments. Hypothesis 3 is verified by comparing their responses of both experiments. $\tau^{\mathrm{E1}}(i)$ and $\tau^{\mathrm{E2}}(i)$ are taken as a basis for further analysis. By calculating the correlation between $\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i)$ and $\tau^{\mathrm{E2}}_{\mathrm{IR,S}}(i)$ we measure whether a participant, whose ratings were dominated by the cut-off frequency in the first experiment, took also the single-/multi-channel system strongly into account in the second experiment. Pearson's product-moment correlation indicates a significant correlation between $\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i)$ and $\tau^{\mathrm{E2}}_{\mathrm{IR,S}}(i)$ ($r = 0.69, p = .023$). In addition, a significant correlation between $\tau^{\mathrm{E1}}_{\mathrm{IR,BIR}}(i)$ and $\tau^{\mathrm{E2}}_{\mathrm{IR,BIR}}(i)$ is indicated by Pearson's product-moment correlation ($r = 0.68, p = .032$). The significant correlation between $\tau^{\mathrm{E1}}(i)$ and $\tau^{\mathrm{E2}}(i)$ indicates that a listeners, who strongly takes the bandwidth-limitation into account, is also strongly influenced by the single-/multi-channel system while rating the overall experience.

To confirm the relationship between $\tau^{\mathrm{E1}}(i)$ and $\tau^{\mathrm{E2}}(i)$, the values of $\tau^{\mathrm{E2}}(i)$ are used to predict the responses of the first experiment. The responses of the first experiment are predicted by a cumulative link model without interactions (an introduction into regression models with ordinal data is given by McCullagh [20]). The cumulative link model predicts the item ratings by the predictor variables: "basic item rating" and "cut-off frequency". How

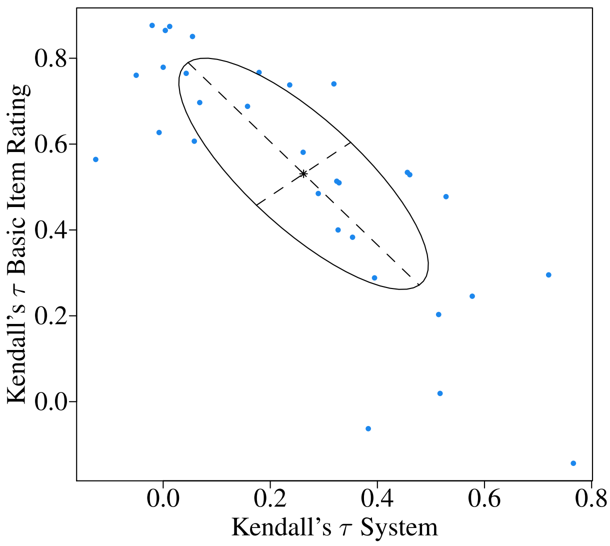

Figure 5. Clustering of the pairs of correlation values by ellipsoidal model with one component.

well the cumulative link model predicts the item ratings is measured by Cragg and Uhler's Pseudo-R^2 (see Cragg and Uhler [21]). If the cumulative link model correctly predicts all item ratings, Cragg and Uhler's Pseudo-R^2 of the cumulative link model is 1.0. The basic cumulative link model without any data from $\tau^{\mathrm{E2}}(i)$ has a Pseudo-R^2 of 0.60. In the next step, the model is modified by data from the second experiment. The predictor variable "basic item ratings" is interacted with $\tau^{\mathrm{E2}}_{\mathrm{IR,BIR}}(i)$ and the predictor variable "cut-off frequency" is interacted with $\tau^{\mathrm{E2}}_{\mathrm{IR,S}}(i)$. By modifying the model, the Pseudo-R^2 is increased to 0.65. All interactions effects of the modified model are significant at a significance level of $\alpha = 0.05$ except for $[\text{basic item rating} = \text{“Neutral”}]$ with $\tau^{\mathrm{E2}}_{\mathrm{IR,BIR}}(i)$ $(\beta = 0.431, p = .090)$. By having R^2 increased, a significant relationship between $\tau^{\mathrm{E1}}(i)$ and $\tau^{\mathrm{E2}}(i)$ is confirmed.

5. DISCUSSION

The verification of Hypothesis 1a by a systematic cluster analysis indicates that there exists no clearly distinguishable types of listeners when the overall listening experience of bandwidth-limited music excerpts is rated. These results imply that there exists a continuum which describes to which degree the listener is influenced by bandwidth-limitations. Furthermore, the results confirm our previously made assumption based on a different experiment that there exists a continuum rather than clearly distinguishable types of listeners [12]. In this work, the influence of bandwidth-limitation on the overall listening experience for each participant is only measured by Kendall's tau correlation coefficient. Further research could confirm these results by applying other approaches to measure the influence of bandwidth-limitation on the overall listening experience.

Hypothesis 1b was positively verified based on the significant correlation between $\tau^{\mathrm{E1}}_{\mathrm{IR,F}}(i)$ and participants' responses to the question of how important audio quality was for their ratings. That participants can reliably self-report how important the audio quality is for their individual overall listening experience is also confirmed by the significant correlation between $\tau^{\mathrm{E1}}_{\mathrm{IR,BIR}}(i)$ and the responses to the question of how important was the song for participants' ratings.

Verifying Hypothesis 2a resulted in almost the same outcome as the verification of Hypothesis 1a. No clearly distinguishable types of listeners in Experiment II were found by the cluster analysis. This indicates that there exists also a continuum which describes to which degree the listener is influenced by the single-/multi-channel system.

Hypothesis 2b was also positively verified. Based on the significant correlation, listeners can reliably self-report how important the single-/multichannel system is for their individually perceived overall listening experience.

The positive verification of Hypothesis 1b and 2b shows that the prediction of the overall listening experience can be improved by responses of a post-questionnaire. This confirms the results of previous research in the field of psychology, where the perception of music was investigated according to responses of questionnaires [4, 5].

Hypothesis 3 turned out to be true since a positive correlation between the influence of the bandwidth-limitation and the influence of the single-/multi-channel system was found. Such a correlation was not fully expected since limiting the bandwidth of a music excerpt and reproducing the music excerpt over a different single-/multi-channel system are two different types of "audio degradation". The question arises whether an individual listener could be roughly described with a single attribute that reflects his or her influence of all types of audio degradation. By improving the cumulative link model of Experiment I with the values of $\tau^{\mathrm{E2}}(i)$ which were only retrieved from Experiment II, we showed that such a single attribute might exist. Further research could deal with finding approaches to retrieve such a single attribute by questioning the listener or conduct a short listening test.

6. CONCLUSION

A systematic analysis on the results of two experiments was applied to identify types of listeners based on ratings of the overall listening experience. The results show that no clearly distinguishable types of listeners exist for the investigated factors of bandwidth-limitation and single-/multi-channel system. Instead, the influence of audio quality degradation on the perceived overall listening experience of a listener can be described by two continuous correlation values. These correlation values retrieved from the second experiment have been successfully used to improve the prediction by of the first experiment's result. Furthermore, the results indicate that listeners whose overall listening experience ratings are influenced by the signal bandwidth, also take the single-/multi-channel system much into account when rating. In both experiments, listeners could reliably self-report how much an audio quality degradation influenced their individually perceived overall listening experience.

7. REFERENCES

[1] W. F. Prince, “A Paradigm for Research on Music Listening,” *Journal of Research in Music Education*, vol. 20, no. 4, pp. 445–455, 1972.

[2] C. G. Jung, *Psychological types: or the psychology of individuation.* Oxford, England: Harcourt, Brace, 1923.

[3] I. Briggs Myers, M. H. McCaulley, N. L. Quenk, and A. L. Hammer, *MBTI Handbook: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, 3rd ed. Consulting Psychologists Press, 1998.

[4] J. L. Pearson and S. J. Dollinger, “Music preference correlates of Jungian types,” *Personality and Individual Differences*, vol. 36, no. 5, pp. 1005–1008, Apr. 2004.

[5] B. E. Lewis and C. P. Schmidt, “Listeners’ Response to Music as a Function of Personality Type,” *Journal of Research in Music Education*, vol. 39, no. 4, pp. 311–321, Jan. 1991.

[6] S. K. Hedden, “Listeners’ Responses to Music in Relation to Autochthonous and Experiential Factors,” *Journal of Research in Music Education*, vol. 21, no. 3, pp. 225–238, Jan. 1973.

[7] R. H. Woody and K. J. Burns, “Predicting Music Appreciation with past Emotional Responses to Music,” *Journal of Research in Music Education*, vol. 49, no. 1, pp. 57–70, Jan. 2001.

[8] F.-R. Stöter, M. Schoeffler, B. Edler, and J. Herre, “Human ability of counting the number of instruments in polyphonic music,” in *Proceedings of Meetings on Acoustics Vol. 19*. Montreal, Canada: Acoustical Society of America, 2013.

[9] M. Schoeffler, F.-R. Stöter, H. Bayerlein, B. Edler, and J. Herre, “An Experiment about Estimating the Number of Instruments in Polyphonic Music: A Comparison Between Internet and Laboratory Results,” in *Proceedings of 14th International Society for Music Information Retrieval Conference*. Curitiba, Brazil: International Society for Music Information Retrieval, 2013.

[10] N. Schinkel-Bielefeld, N. Lotze, and F. Nagel, “Audio Quality Evaluation by Experienced and Inexperienced Listeners,” in *Proceedings of Meetings on Acoustics*. Montreal, Canada: Acoustical Society of America, 2013.

[11] F. Rumsey, S. Zielinski, R. Kassier, and S. Bech, “Relationships between Experienced Listener Ratings of Multichannel Audio Quality and Naïve Listener Preferences,” *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3832–3840, 2005.

[12] M. Schoeffler and J. Herre, “About the Impact of Audio Quality on Overall Listening Experience,” in *Proceedings of Sound and Music Computing Conference 2013*, Stockholm, Sweden, 2013, pp. 48–53.

[13] M. Schoeffler, B. Edler, and J. Herre, “How Much Does Audio Quality Influence Ratings of Overall Listening Experience?” in *Proc. of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, Marseille, France, 2013, pp. 678–693.

[14] M. Schoeffler, S. Conrad, and J. Herre, “The Influence of the Single/Multi-Channel-System on the Overall Listening Experience,” in *Proc. of AES 55th International Conference on Spatial Audio*, Helsinki, Finland, 2014.

[15] M. G. Kendall, “A New Measure of Rank Correlation,” *Biometrika*, vol. 30, no. 1/2, pp. 81–93, Jun. 1938.

[16] R. G. Garrett, “The chi-square plot: a tool for multivariate outlier recognition,” *Journal of Geochemical Exploration*, vol. 32, no. 1-3, pp. 319–341, Apr. 1989.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[18] C. Couvreur, “The EM Algorithm : A Guided Tour,” in *Proc. 2nd IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing*, Pragues, Czech Republik, 1996.

[19] C. Fraley and A. E. Raftery, “Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering,” *Journal of Classification*, vol. 24, no. 2, pp. 155–181, Sep. 2007.

[20] P. McCullagh, “Regression Models for Ordinal Data,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 42, no. 2, pp. 109–142, 1980.

[21] J. G. Cragg and R. S. Uhler, “The Demand for Automobiles,” *The Canadian Journal of Economics*, vol. 3, no. 3, pp. 386–406, Aug. 1970.