



New Insights into Soundscape Evaluations Using the Experience Sampling Method

Jochen Steffens

Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany.

McGill University, School of Information Studies & CIRMMT, Montreal, Canada

Daniel Steele

McGill University, School of Information Studies & CIRMMT, Montreal, Canada

Catherine Guastavino

McGill University, School of Information Studies & CIRMMT, Montreal, Canada

Summary

Soundscape research draws from quantitative and qualitative methods to explore peoples' perception and understanding of their acoustic environments. One opportunity to combine the strength of both approaches is the so-called Experience Sampling Method (ESM). ESM refers to a method of data collection in which people periodically make momentary (i.e. "in-the-moment") judgments over the course of the day while naturally acting within their everyday environment. This means that reactions to environmental sounds are captured as they occur. We conducted a 7-day ESM study to investigate the relationship between momentary and retrospective soundscape judgments. In the course of the study, participants were prompted 10 times a day by a smartphone application. They were asked to evaluate their soundscape and report on further potential influencing factors (e.g. mood, activity at-hand, and the degree of attention paid to the soundscape). Additionally, they performed summary retrospective judgments at the end of each day and the whole week. Results ($n = 26$, mean age = 29.3 (SD = 5.9), 13 women) show that daily retrospective judgments of soundscape pleasantness can be predicted by the average and the linear trend of the momentary judgments, the negative peak, and the person's mood while performing the judgment. Weekly retrospective judgments, however, are governed by the positive peak and the person's mood. On theoretical grounds, the results provide new insights into the complex structure of judgment processes in soundscape perception. Especially given the biases between momentary and retrospective judgments, the ESM is a promising tool to collect naturalistic longitudinal data in the moment and situation in which the experience takes place.

PACS no. 43.66.Ba, 43.66.Lj

1. Introduction

1.1. Perception of the acoustic environment

Listening to our acoustic environment takes time. One way a sound "persists" through time is when it is perceived and remembered by listeners. Any expressed evaluation of an acoustic environment necessarily makes use of retrospection, whether we are still in the environment or remembering one from our childhood. The influence of cognitive processes, especially memory representations of a temporal experience, may lead to a weighting of certain episodes in

the course of a overall retrospective evaluation. Previous research in the field of decision-making psychology has shown that momentary (i.e. "in-the-moment") judgments of time-varying experiences do not always match retrospective judgments (e.g. [1] in the context of decision-making and [2] for pain evaluation). Studies in psychophysics indicate that initial and final portions of abstract auditory stimuli receive greater weight than the middle section when listeners are asked to judge the overall loudness of a sound (e.g. [3, 4]). Recency effects have also been found for sounds with longer durations (e.g. [5, 6]). Kahneman and co-workers could show that, beside the recency effect, the most extreme affect (peak) experienced during the episode is decisive for the retrospective evaluation of a hedonic or aversive experience [2]. Several studies revealed that an unweighted combination of

(c) European Acoustics Association

both factors (i.e. both factors are equally weighted) could well explain the variance of the retrospective judgments. The authors established their "peak-end rule" in the course of investigations on pain perception, but higher weightings of the peak and the end could also be confirmed for other kinds of affective experiences, e.g. listening to musical pieces [7].

According to Ariely and Carmon, summary assessments also partly reflect inferences about future states [1]. These future anticipations are derived from the temporal development of the experience. This assumption has gained empirical support from studies by Loewenstein & Prelec (1993) and Ariely (1998) who demonstrated that retrospective judgments are influenced by the linear trend of an experience [8, 9].

A laboratory study by the authors on the relationship between momentary and retrospective judgments of environmental sounds indeed revealed the linear trend of the experience to be a significant predictor of a retrospective pleasantness judgment [10]. The second, more important, predictor in this study, however, was the average of the momentary judgments. This factor suggests a "cognitive averaging" of all temporal elements of the experience. The authors hypothesize that the degree of "cognitive averaging" could be dependent of the properties of the sound, personality traits of the participants, and the test situation itself. In our experiment, sounds did not exceed a duration of 90 seconds. It may therefore be assumed that participants in these kinds of listening experiments are often able and willing to overlook the whole experience when making retrospective judgments and apply a kind of rational averaging, which they assume has been requested by the investigator. These methodological issues raise the need to extend the investigations to longer and more eventful sound periods, to remove the laboratory and investigator from the context, and to also take into account personal and situational variables. One method that allows the measurement of peoples' reaction over the course of hours, days, and weeks, is the so-called experience sampling method (ESM).

1.2. The experience sampling method (ESM)

ESM refers to a method of data collection in which people periodically make momentary judgments over the course of the day while going about their everyday activities [11]. To our knowledge, the ESM has not been used in soundscape research (except for a related diary study by Schulte-Fortkamp & Genuit [12]); yet, the method appears highly suitable for investigating "the acoustic environment as perceived or experienced and/or understood by a person or people, in context" (definition of soundscape proposed by the ISO working group TC43 SC1 WG 54 [13]). In contrast to laboratory experiments, the ESM can provide quantitative and qualitative "in-situ" data -

valuable perceptual data of people actually experiencing real acoustic environments. Due to the development of modern, multifunctional smartphones, the ESM is currently experiencing a renaissance. While developers of the method, such as Csikszentmihajli (1977) had to provide their participants with pagers and stacks of paper on which they recorded their responses when signalled [11], nowadays smartphone applications serve as all-in-one alarm clocks, electronic questionnaires, and measurement instruments.

1.3. Soundscape, the person, and the situation

The ESM is in line with Brunswik's concept of a representative design of experiments that allow the sampling of stimuli from the person's natural environment which are representative of the population of stimuli to which the person has adapted [14, 15]. Brunswik further highlighted the importance of understanding how various psychological factors function and interact in different real-life situations. Psychological factors and their role in sound evaluations have already been investigated in numerous studies (e.g. [16, 17]). Västfjäll (2002), for example, observed an effect of mood on annoyance reactions to a sound of a chainsaw [18]. In general, the effect of mood or affect on judgment processes is a well-known phenomenon in cognitive psychology. In his Affect Infusion Model (AIM), Forgas (1995) assumes that the impact of affect on judgment processes varies along a continuum of four different judgmental strategies [19]. According to the model, judgments requiring heuristic or substantial processing are more likely to be infused by affect than direct access or motivated judgments.

Beside situational factors, it has been shown that different (stable) thinking styles and personality traits can have an influence on judgment processes. One well-known trait related to auditory perception is noise sensitivity. This trait covers attitudes toward a wide range of environmental sounds [20]. In many studies, it has been shown to be a major factor explaining noise annoyance reactions [21, 22].

1.4. Aims of the study

By taking into account potential moderating variables like mood and personality traits and by sampling the experience "in-situ", our ESM study has a high level of ecological validity and should provide findings that are representative of everyday life listening. The study can also be considered a validation of the own aforementioned laboratory study on the relationship between momentary and retrospective judgments [10]. It extends the findings to real-life settings, longer sound durations and a wider range of soundscapes.

To sum up, we hypothesize that daily retrospective judgments of soundscape pleasantness are governed by the average of the momentary judgments obtained

over the course of the day, the linear trend, the positive and negative peaks (maximum and minimum values), and the end (last measuring point). Last but not least, we assume an influence of the personality traits mentioned earlier and the person's mood while making the retrospective judgment.

2. Method

2.1. Participants

26 participants, 13 women and 13 men with a mean age of 29.3 (SD = 5.9) participated in the study. They were recruited via the mailing list for current students, postdocs, faculty and staff of the School of Information Studies (McGill) and via Craigslist. They received \$30 CAD for their participation and were naive with regard to the hypotheses under test. 65.4% of the participants were students; the rest (34.6%) had a job outside the university. One participant reported hearing problems, and none of them dropped out from the study.

2.2. Procedure

Participants were recruited on the basis of having an Android smartphone that they regularly carried with them. Those who did not have an Android phone but who wished to take the study were provided with an alternate device. Participants were invited to the Multimodal Interaction Laboratory (MIL) at McGill for an entrance interview and training session. At the interview, they were guided through the installation of the smartphone app that would run the study over the following week. Participants were instructed that the app would cause their phone to ring at 11 random points throughout the day (10 momentary judgments and one daily retrospective one (plus one extra weekly retrospective judgment on the last night of the study)), at which point they would conduct a 1-minute questionnaire. Participants were asked to provide a 12-hour time frame of their typical day where these alarms would catch them while they were "awake and active". Participants were then given a practice alarm and questionnaire that they filled out while seated next to the researcher.

The training defined soundscape for them as "the word we are using to describe your acoustic environment as you perceive it. The soundscape includes the people in it and those you are talking to as well as the background sounds." The training also guided them through samples of each of the characteristics described in the following section. Lastly, to capture the "peak" moments, participants were asked to start the app manually to make extra momentary judgments in the case that they experienced soundscapes that were exceptionally positive or negative. At the end of the week-long study, participants were also asked to

return to the MIL to fill out questionnaires and have the exit interview with one of the investigators.

2.3. Measures and Design

The questionnaires were designed by means of a web-based administration interface in combination with an Android application specifically designed for experience sampling (Movisens XS, Germany). In total, three different questionnaires were presented in this study. One questionnaire was designed to report momentary judgments of the soundscape over the course of the day and occurred 10 times per day. Amongst others, this "momentary form" contained questions on the soundscape pleasantness, ("How do you rate the pleasantness of the soundscape?"), eventfulness, and familiarity. This question had to be answered on a 7-step Likert scale ranging from unpleasant (1) to pleasant (7), not eventful (1) to very eventful (7), and unfamiliar (1) to familiar (7). Additionally, the participants were requested to report on the predominant sound source in the soundscapes, the pleasantness of the visual environment and situational factors such as the location, the degree of attention paid to the soundscape, current mood and activity-at-hand. However, since these questions addressed other research questions than the relationship between momentary and retrospective pleasantness judgments, the results will not be presented in the context of this paper.

The second form used in the study is the "daily summary". This form also contained questions on soundscape pleasantness, eventfulness, and familiarity. For this questionnaire, participants were requested to perform retrospective judgments of the whole day (e.g. "How do you rate the pleasantness of the soundscape(s) over the whole day?") using the same seven-step Likert scales as for the momentary judgments. Participants were also requested to report on their current mood (from bad (1) to good (7)) and activity-at-hand (free-format responses). The participants were prompted to fill out the daily summary at a fixed time - five minutes after the end of the selected time-frame (e.g. at 10.05 pm for a time-frame from 10 am-10 pm). In case the participants should miss this prompt, they were also given the chance to manually start the daily summary from the time of the daily prompt until midnight (e.g. from 10.05 pm-12.00 am for a standard time-frame from 10am-10pm).

The third and last form employed is the "weekly summary" which provided overall retrospective judgments of soundscape pleasantness, eventfulness, and familiarity over the course of the whole 7-day period (e.g. "How do you rate the pleasantness of the soundscape(s) over the whole week?"). The weekly summary appeared on the participants' phones on the seventh day ten minutes after the end of the selected time-frame (e.g. at 10.10 pm for a time-frame from 10 am-10 pm). In the exit interview, participants

were also asked to report on basic demographic information (birth year, gender, highest education, current job, and reported hearing problems) and to fill out four standardized personality questionnaires. The personality questionnaires were: a short version of the Big Five Inventory (BFI-10; [23]), Weinstein's noise sensitivity scale (WNS-6B; [24]), the "Rational-Experiential Inventory for Adolescents" (REI-A; [25]), and the Barratt Impulsiveness Scale (BIS-11; [26]).

2.4. Data analysis

Since the declared aim of the study was to investigate which features of the momentary judgments contribute to the retrospective judgments of a day and the whole week, two different regression analyses were conducted (SPSS 20 for [Windows/Macintosh]). Thereby, the average, the maximum value ("Peak (max)"), minimum ("Peak(min)"), and the last one ("End") of the momentary judgments were considered potential independent variables of the daily summaries. Additionally, the linear trend over the course of the day was calculated. It is the standardized regression coefficient of a linear regression analysis with the independent variable, time (time regression analysis, method of least squares [27]). Furthermore, the reported mood while performing the daily summary was hypothesized to directly influence pleasantness judgments and therefore considered another independent variable. For the regression analysis on the weekly summary, the single daily summaries were regarded as further potential independent variables. Thus, the same calculations as for the momentary judgments were performed: Average, "Peak (max)", "Peak (min)", End (= daily summary of the last day), trend, and mood while performing the weekly summary.

Since mood and the measured personality traits Extraversion, Agreeableness, Neuroticism, Openness to experience, Conscientiousness, Noise sensitivity, rational and experiential thinking style, and impulsiveness were hypothesized to have a moderating effect, standardized (z-transformed) values of these variables were computed. The interaction terms are products of the potential standardized moderator variables and the features of the momentary judgments.

Every participant had to perform 7 daily summaries and 1 weekly summary on soundscape pleasantness. These overall judgments were considered as dependent variables in the different regression calculations. The judgments performed over the seven days are repeated measurements and can be correlated in a way that violates independence assumptions required for common linear models. Thus, a mixed linear regression model including time (day 1-7) as a potential random effect were calculated [28]. A further advantage of mixed models compared to common linear models (e.g. ANOVA) is that datasets of participants are included in the calculation even if they are not complete, e.g. when participants missed single prompts (ibid.).

However, the authors decided to only include datasets with at least six momentary judgments per day for the regression on the daily summary, and at least 6 daily summaries for the regression on the weekly summary.

3. Results

The results show that the momentary pleasantness averaged across participants and situations is 4.7 (SD=1.4) and is therefore higher than the center of the scale (4). Both the retrospective pleasantness judgments of the day and the week are even higher (Retro(day): 4.9, SD = 1.1; Retro(week): 5.0, SD = 0.7). At the same time, a decrease of the standard deviation can be observed. In the next step, regression calculations were performed with the retrospective judgment considered dependent variables. The calculation of the linear mixed model (covariance matrix: compound symmetry) reveals four fixed effects predicting the daily retrospective judgments: average, the linear trend, the negative peak (peak(min)), and the mood of the person while making the judgment. The regression statistics are reported in Table 1.

Table I. Mixed Models Analysis of the Relationship of Retrospective Judgments and Features of the Momentary Judgment - Estimates of Fixed Effects

Variable	Estimate	Std. Error	t	Sig.
Intercept	.66	.51	1.28	.20
Average	1.00	.13	7.74	.00
Trend	.53	.16	3.25	.00
Peak(min)	-.18	.07	-2.50	.01
Mood	.32	.07	4.37	.00

The results are partly in accordance with our hypotheses and confirm the validity of the model obtained in the aforementioned laboratory experiment. In contrast, the positive peak, the end, and direct or moderating effects of personality traits did not significantly influence the daily retrospective judgments. The personality trait correlating best with the retrospective judgments is the Weinstein noise sensitivity (Pearson product moment correlation $r = -.223$, $p < .01$). Furthermore, time (day 1-7) did not have a fixed or random effect on the retrospective judgments or the relationship between retrospective and momentary judgments. This means that the retrospective judgments of each day can be considered independent from each other.

For the prediction of the weekly retrospective pleasantness judgments, a standard linear regression model could be calculated (no within-subject-effects because of only one judgment made by a participant). Within both step-forward and step-backward regression analyses two significant predictors could be obtained: peak

(max) and mood. The standard regression coefficients β are .63 for peak(max) ($t = 4.8$, $p < .01$) and .52 for mood ($t = 3.9$, $p < .01$). The model explains 72.3% of the variance of the weekly retrospective judgments ($R^2 = .72$; $F = 20.9$, $df_1 = 2$, $df_2 = 16$) which can be considered high, especially for studies on experience sampling.

4. Discussion

The results of our study confirm our assumption that retrospective judgments of soundscape pleasantness are not only governed by "cognitive averaging" processes but also in particular by specifically unpleasant and pleasant peak moments. Moreover, the results provide further empirical evidence that judgment processes in auditory perception are influenced by a person's mood and anticipation how the soundscape experience might go on (derived from the linear trend). This is, to a large extent, in line with the literature mentioned in the introductory section and, last but not least, with our own laboratory experiment [10]. Comparing the three different time scales of both our studies (minutes, hours, days), it is shown that the judgmental strategy of "cognitive averaging" disappears with an increased time span that has to be recalled by a person in the course of an overall judgment. This is probably due to limitations of the long-term memory since it is much easier for a person to recall the last minute compared to the whole last week. If not all pieces of information can be recalled from memory, a heuristic judgment based on the most salient information (e.g. the positive and negative peak moments) is highly likely [29].

It must be stated, that the findings regarding the daily and weekly retrospective judgment are partly contradictory. While mood turns out to be a significant predictor in both the daily and weekly summaries, obviously different peak experiences are recalled and "used" to establish the retrospective judgments, namely the negative peak for the daily and the positive peak for the weekly summary. One explanation could be the special evaluation context of the weekly summary. It is the last judgment process in the course of the study week, and participants could desire to perform a "benevolent" judgment and emphasize more on the positive aspects experienced during the week. The effect of the positive peak on the retrospective judgment could also be related to a mood-congruent memory bias which describes the improved recall of information congruent with one's current mood [30]. The average mood of the participants while performing the judgment can be considered high ($\bar{x} = 5.4$; $SD = 1.1$; not reported in the results section). The high mood scores could lead to a better recall of positive moments experienced during the week. A second explanation addresses one limitation of the study. The sample size, specifically regard-

ing the estimation of the predictors of the weekly summary, must be considered relatively low. The model for the weekly summary only contains 19 datasets (26 participants, 7 cases excluded because of missing values). In contrast, the calculation of the mixed model for the daily summaries contains 162 datasets. The small sample size may also be a reason why no significant effect of personality on soundscape judgments could be observed. However, the correlation between Weinstein noise sensitivity and the daily retrospective judgments is in line with literature mentioned earlier.

A further limitation addresses the method itself. Since no variables are experimentally varied within the study and relationships are only observed on a correlational basis, the ESM does not allow the researcher to draw definite conclusions about causal relationships [31]. Furthermore, the ESM is time-consuming, expensive, and demanding for both the participant and the researcher (*ibid.*).

However, the ESM proves to be a promising method to investigate "the acoustical environment as perceived or experienced and/or understood" in the ecologically valid context as suggested by the ISO definition of soundscape. It provides a meaningful insight on contextual factors governing soundscape evaluations. Last but not least, the research topic itself addressed in the study, the relationship between momentary and retrospective judgments, provides empirical evidence that it is necessary to sample "in-situ" while the experience takes place and to obtain on-line judgments which are not affected by "cognitive distortions" such as memory biases.

Acknowledgement

This project is funded by the German Research Foundation (GZ: STE 2331/1-1), a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN 327392-13), and a William Dawson award from McGill University. The study was reviewed and approved by the McGill Research Ethics Board.

References

- [1] D. Ariely, Z. Carmon: Gestalt characteristics of experiences: the defining features of summarized events. *Journal of Behavioral Decision Making* **13** (2000) 191-201.
- [2] D. Kahneman, B. L. Fredrickson, C. A. Schreiber, D. A. Redelmeier: When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science* **4** (1993) 401-405.
- [3] D. Oberfeld: Temporal weighting in loudness judgments of timevarying sounds containing a gradual change in level. *Journal of the Acoustical Society of America* **123** (2008) 3307.
- [4] B. Pedersen, W. Ellermeier: Temporal weights in the level discrimination of time-varying sounds. *The Journal of the Acoustical Society of America* **123** (2008) 963-972.

- [5] D. Västfjäll: The “end effect” in retrospective sound quality evaluation. *Acoust. Sci. & Tech.* **25** (2004) 170–172.
- [6] C. A. Schreiber, D. Kahneman: Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology* **129** (2000) 27–42.
- [7] A. Rozin, P. Rozin, E. Goldberg: The Feeling of Music Past: How Listeners Remember Musical Affect. *Music Perception: An Interdisciplinary Journal* **22** (2004) 15–39.
- [8] G. F. Loewenstein: Anticipation and the Valuation of Delayed Consumption. *The Economic Journal* **97** (1987) 666–684.
- [9] D. Ariely: Combining Experiences Over Time: The Effects of Duration, Intensity Changes & On-Line Measurements on Retrospective Pain Evaluations. *Journal of Behavioral Decision Making* **11** (1998) 19–45.
- [10] J. Steffens, C. Guastavino: (Tr-)end Effects of Momentary and Retrospective Soundscape Evaluations. *Acta Acustica united with Acustica* (in review).
- [11] M. Csikszentmihalyi, R. Larsson, S. Prescott: The ecology of adolescent activity and experience. *Journal of Youth and Adolescence* **6** (1977) 281–294.
- [12] B. Schulte-Fortkamp, K. Genuit: The acoustical diary as an innovative tool in soundscape evaluation. *The Journal of the Acoustical Society of America* **115** (2001) 2496.
- [13] ISO: ISO 12913-1:2014 Acoustics – Soundscape – Part 1: Definition and conceptual framework. Geneva: International Organization for Standardization (ISO). 2014.
- [14] E. Brunswik: Systematic and Representative Design of Psychological Experiments: With Results in Physical and Social Perception. *Proceedings of the Berkeley symposium on mathematical statistics and probability* (1947).
- [15] M. K. Dhami, R. Hertwig, U. Hoffrage: The role of representative design in an ecological approach to cognition. *Psychological Bulletin* **130** (2004) 959–988.
- [16] J. Blauert, U. Jekosch: Sound-Quality Evaluation - A Multi-Layered Problem. *Acta Acustica united with Acustica* **83** (1997) 747–753.
- [17] R. Guski: Personal and social variables as co-determinants of noise annoyance. *Noise & Health* **1** (1999) 45–56.
- [18] D. Västfjäll: Influences of current mood and noise sensitivity on judgments of noise annoyance. *Journal of Psychology* **136** (2002) 357–370.
- [19] J. P. Forgas: Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin* **117** (1995) 39–66.
- [20] W. Ellermeier, M. Eigenstetter, K. Zimmer: Psychoacoustic correlates of individual noise sensitivity. *The Journal of the Acoustical Society of America* **109** (2001) 1464–1473.
- [21] Job, R. F. S.: Community response to noise: A review of factors influencing the relationship between noise exposure and reaction. *The Journal of the Acoustical Society of America* **83** (1988) 991.
- [22] S. M. Taylor: A path model of aircraft noise annoyance. *Journal of Sound and Vibration* **96** (1984) 243–260.
- [23] B. Rammstedt, O. P. John: Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* **41** (2007) 203–212.
- [24] H. Kishikawa, T. Matsui, I. Uchiyama, M. Miyakawa, K. Hiramatsu, S. A. Stansfeld: The development of Weinstein's noise sensitivity scale. *Noise and Health* **8** (2006) 154.
- [25] R. Pacini, S. Epstein: The Relation of Rational and Experiential Information Processing Styles to Personality, Basic Beliefs, and the Ratio-Bias Phenomenon. *Journal of Personality and Social Psychology* **76** (1999) 972–987.
- [26] J. H. Patton, M. S. Stanford, Barratt Ernest S.: Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology* **51** (1995) 768–774.
- [27] N. R. Draper, H. Smith: Applied regression analysis. 3rd ed. J. Wiley, New York (N.Y.), 1998.
- [28] N. Bolger, J.-P. Laurenceau: Intensive longitudinal methods: An introduction to diary and experience sampling research. Guilford Press, New York, NY, 2013, (Methodology in the social sciences).
- [29] A. Tversky, D. Kahneman: Judgment under uncertainty: Heuristics and biases. Springer, New York, 1975.
- [30] P. C. Watkins, K. Vache, S. P. Verney, A. Mathews: Unconscious mood-congruent memory bias in depression. *Journal of Abnormal Psychology* **105** (1996) 34–41.
- [31] P. N. Juslin, S. Liljeström, D. Västfjäll, G. Barradas, A. Silva: An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion* **8** (2008) 668–683.