ASSESSING APPLAUSE DENSITY PERCEPTION USING SYNTHESIZED LAYERED APPLAUSE SIGNALS

Alexander Adami

International Audio Laboratories*,
Friedrich-Alexander Universität Erlangen
Erlangen, Germany
alexander.adami@audiolabs-erlangen.de

Sascha Disch

Fraunhofer IIS,
Erlangen, Germany
sascha.disch@iis.fraunhofer.de

Garri Steba

Ostbayerische Technische Hochschule
Amberg-Weiden, Germany
g.steba@oth-aw.de

Jürgen Herre

International Audio Laboratories*,
Friedrich-Alexander Universität Erlangen
Erlangen, Germany
juergen.herre@audiolabs-erlangen.de

ABSTRACT

Applause signals are the sound of many persons gathered in one place clapping their hands and are a prominent part of live music recordings. Usually, applause signals are recorded together or alongside with the live performance and serve to evoke the feeling of participation in a real event within the playback recipient. Applause signals can be very different in character, depending on the audience size, location, event type, and many other factors. To characterize different types of applause signals, the attribute of 'density' appears to be suitable. This paper reports first investigations whether density is an adequate perceptual attribute to describe different types of applause. We describe the design of a listening test assessing density and the synthesis of suitable, strictly controlled stimuli for the test. Finally, we provide results, both on strictly controlled and on naturally recorded stimuli, that confirm the suitability of the attribute density to describe important aspects of the perception of different applause signal characteristics.

1. INTRODUCTION

Recordings of applause signals can be very different in their sound character, depending on many factors including audience size, location, event type, recording setup, etc. Several publications in the past have shed some light on the nature of applause signals, providing a basic clap anaysis [1], attempting to synthesize applause through physical modeling of individual claps [2,3], through sound texture synthesis [4–7] or morphing of granular sounds [8]. Also the phenomenon of rhythmic applause [9, 10] was already subject of scientific curiosity. Yet, to our best knowledge, nobody has actually looked into how to perceptually characterize different types of applause signals. We propose to use the attribute 'density' to capture the predominant character of different applause signals.

To subjectively assess the suitability of the attribute density, we designed a dedicated listening test that is presented in this paper in the evaluation section.

To achieve a controlled variation of density within different applause stimuli without changing other factors like timbre, spatial properties, etc., we generated all test stimuli through layering from dry studio recordings of individual persons applauding. To mimic the geometric and physical conditions of a gathered audience, we applied a simple model of the same while layering. This is discussed in the next sections.

2. APPLAUSE DENSITY

Existing, well established perceptual attributes like loudness, pitch and timbre do not describe applause properties very well. To characterize different types of applause signals, the novel attribute of 'density' appears to be suitable. The concept of density might be rightfully attributed to all kinds of sounds and sound textures that predominantly consist of sufficiently dense transient events being distributed in a pseudo-random way like rain and fireworks.

In previous work, the idea of defining an 'impact rate' to further characterize environmental sound textures has already been roughly sketched, e.g., for describing the sound of falling raindrops [11]. However, the author concludes that such an impact rate appears to be more of a physical measure than a psychoacoustic one. Notwithstanding, the author suggests to further conduct psychoacoustic experiments to clarify this assumption. In our present paper, we are presenting the results of such experiments.

Kawahara, for example, measured a set of simple parameters of different applauses at a recording site in order to efficiently recreate similar applause sounds at a receiver site, controlled by the proposed set of transmitted parameters. These parameters include, among other things, the average time interval between two clapping events [12]. Again, the proposed parameters are related much closer to a physical measure than to a perceptual quality.

For sure, the perception of applause signals is influenced by several factors, whereas the most obvious one is certainly the number of people clapping simultaneously. Still, we suggest that the sensation of density is an abstract psychoacoustic quantity in its own right albeit closely coupled to the actual physics behind the generation of such sound textures. Additionally, parameters like spatial properties, room acoustics, and distance of a listener to the applauding crowd have an impact on the resulting applause impression: the amount of reverb determines how much individual claps get smeared and possibly blended, whereas the distance to the crowd influences the perceived near-by to far-off clap ratio, i.e., the ratio of individually distinguishable foreground claps and

* A joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany

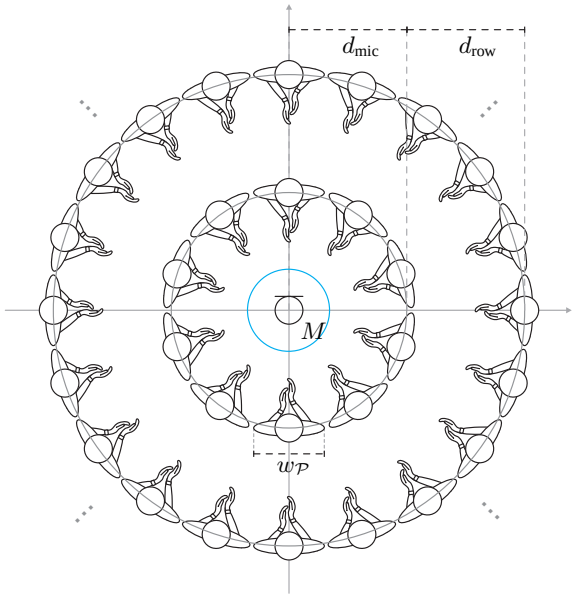


Figure 1: *Illustration of the basic applause model: virtual clapping people with width $w_{\mathcal{P}}$ at rows with the inter-row distance d_{row} and d_{mic} denoting the distance between the microphone M and the first row.*

the more homogenous background noise floor resulting from many claps being superimposed. Applause density appears to be a suitable choice for an attribute which accommodates all mentioned properties additional to the impact rate. Consequently, we investigate in this paper if there is a perceptual attribute like applause density and if it is consistently perceived.

3. SYNTHESIZING APPLAUSE SIGNALS

To assess applause density perception, it is necessary to have control over as many parameters of the applause signals as possible. For that reason, we came up with a simple model which allows to synthesize applause signals using layering of individual recordings of single people clapping. Note that the layering approach is not aiming at synthesizing the most realistic applause signals but providing a tool producing sufficiently natural and plausible applause to investigate density as a suitable perceptual attribute. In this section, the basic physically motivated applause model and some assumptions which are used to synthesize the applause signals are described. Furthermore, the algorithm will be introduced.

3.1. Basic Applause Model

Figure 1 depicts the assumed physical model for the applause synthesis. In the model, the virtual applauding people $\mathcal{P}_{r,p}$ are arranged on concentric circles, which will be called rows, around a virtual microphone M in the center, where r and p denote the one-based row and person indices. Every virtual person produces a corresponding clap signal $C_{r,p}$. The distance between the microphone and the first row of virtual people is given by d_{mic}, and the inter-row distance, i.e., the distance between consecutive rows is given by d_{row}. Varying d_{mic} will result in different near-by to far-off clap ratios, i.e., a small d_{mic} will result in many intensive and individually distinguishable near-by claps whereas a large d_{mic} will result in a more homogenous and noise-like signal.

The maximum number of virtual people allowed on a specific row is determined by the space a virtual person consumes and the radius of the circle the row is placed on. In the r-th row, the maximum number of virtual persons is given by

$$P_r = \text{floor}\left(\frac{2\pi(d_{\text{mic}} + (r-1)\cdot d_{\text{row}})}{w_{\mathcal{P}}}\right), \tag{1}$$

with $r = 1...R$ and R denoting the total number of rows. This also determines the angular spacing of the virtual persons on that row, i.e., a person in row r is placed every $\Delta_{\varphi,r} = \frac{360^\circ}{P_r}$ degree. Assuming all spots in a row are occupied, the overall number of persons P_Σ is then determined by

$$P_\Sigma = \sum_{r=1}^{R} P_r. \tag{2}$$

Since the sound pressure is attenuated on its way traveling from a virtual person at the r-th row to the microphone in the center, an attenuation factor has to be applied to it. Propagation losses are modeled by the distance law leading to the row-dependent amplitude attenuation factor

$$a_r = \frac{d_0}{d_{\text{mic}} + (r-1)\cdot d_{\text{row}}}, \tag{3}$$

where d_0 represents a reference distance which corresponds to the microphone distance used for the actual applause sample capturing. Note that a_r can also become greater than one, if the distance to a row is smaller than the reference distance.

There are also frequency dependent effects which need to be taken into account. For instance, people on the inner rows act as an obstacle to the inwards propagating sound waves and also the air itself absorbs sound energy. Both results in a frequency dependent attenuation. These shading and absorption effects can be modeled by applying a lowpass or treble shelving filter $h_r(t)$ to the signals. Since the filter characteristics are distance dependent, the filter has a dependency on the row index r.

In this model, we assume the microphone's pickup pattern M_{PU} to be spatially uniform which corresponds to an omnidirectional characteristic:

$$M_{\text{PU}}(\phi) = 1, \tag{4}$$

where $\phi = 0...360^\circ$ denotes the sound waves' incident angles. This is indicated by the blue circle around the microphone in Figure 1.

With the constant pickup pattern, the microphone signal can then be written as

$$M(t) = \sum_{r}^{R}\left(a_r \cdot \sum_{p}^{P_r}\left(M_{\text{PU}}\cdot C_{r,p}(t)\right) * h_r(t)\right), \tag{5}$$

where $*$ denotes a convolution.

3.2. The Layering Algorithm

Applause consists of up to thousands of individual clap signals naturally superimposed on their way traveling from the hands of the clapping person to the ears of the receiver or to a microphone. By

Parameter	Default	Description
$\widehat{P}_\Sigma$	-	desired number of people clapping
d_{mic}	15 m	distance of microphone to first row
d_{row}	1 m	distance between rows
$w_\mathcal{P}$	0.5 m	space a virtual person consumes
δ_{const}	1.15 s	constant relative signal shift
δ_{rand}	0.01 s	random relative signal shift

Table 1: *The applause model's controllable parameters*

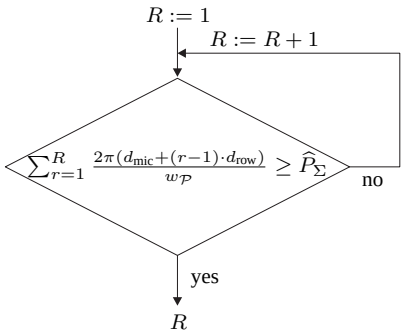


Figure 2: *Determination of the number of rows R.*

layering the captured base signals and applying a simple physical model, the natural behavior is mimicked.

The layering algorithm is designed such that it accepts the desired number of people clapping $\widehat{P}_\Sigma$ as input and produces the corresponding applause signal for the given parameters. A list of the model's controllable parameters is given in Table 1. Starting with the desired number of people clapping, the number of rows has to be determined. This can be done by increasing the number of rows until the expression in (6) comes true, i.e., the overall number of persons P_Σ possible for the given number of rows exceeds or equals the desired number of people clapping. This is also illustrated in Figure 2.

$$\sum_{r=1}^{R} \frac{2\pi(d_{\text{mic}} + (r-1)\cdot d_{\text{row}})}{w_\mathcal{P}} \geq \widehat{P}_\Sigma \tag{6}$$

In the next step, the number of people in each row has to be determined. Except for the last row, this can be done by using Equation (1). Since the last row does not necessarily need to be fully occupied, it has to be treated separately and can be computed using:

$$P_R = \widehat{P}_\Sigma - \sum_{r=1}^{R-1} P_r. \tag{7}$$

To compute the row-dependent attenuation factor a_r, Equation (3) can be applied directly. For the lowpass filter, a first order Butterworth filter with adaptive cut-off frequency was used to match the spectral tilt of an arbitrarily chosen reference applause recording.

Finally, the actual applause signals to be layered have to be generated from the captured base files $c_b(t)$, where $b = 0...B-1$ and B denotes the number of available base signals. Since there is only a limited number of base signals available, they have to be treated in a way such that they can be used multiple times without

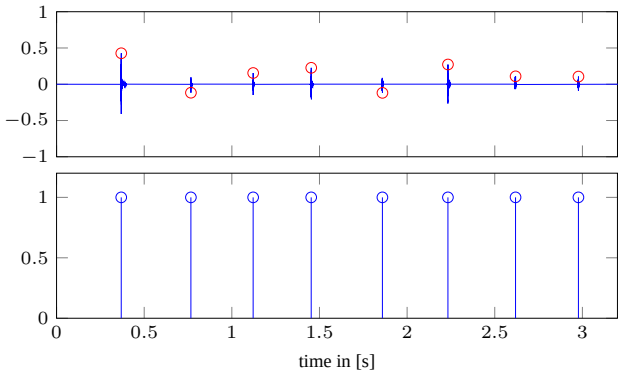


Figure 3: *Annotated base signal and corresponding metadata signal. Top plane: local peak picking; bottom plane: meta data signal*

creating perceivable artifacts due to correlations. This could be accomplished by using circularly time-shifted copies of the base signals. With δ_t denoting the time-shift, the time-shifted applause signals are obtained by

$$C_{\widehat{p}}(t) = \begin{cases} c_b(t + j\delta_t), & \text{if} \quad 0 \leq t < T - j\delta_t \\ c_b\big(t - (T - j\delta_t)\big), & \text{if} \quad T - j\delta_t < t \leq T \end{cases}, \tag{8}$$

where $\widehat{p} = 1...\widehat{P}_\Sigma$, $b = (\widehat{p}-1)\%B$, $j = \text{ceil}(\frac{\widehat{p}}{B} - 1)$, $\%$ denoting the modulo operator, and T denoting the uniform duration of the base signals. Please note, that the time-shifted clap signals were computed using the index $\widehat{p}$ for the sake of simplicity of notation. The relation between the $\widehat{p}$-based and the row- and persons-on-a-row-based indexing is given by $\widehat{p} = \sum_{\hat{r}=1}^{r-1} P_{\hat{r}} + p$. In order to add some variability to the layered applause signals, the time shift δ_t was designed to be a superposition of a constant and a random-based time offset, such that $\delta_t = \delta_{\text{const}} + \Delta_{\text{rand}}$, where Δ_{rand} denotes a function which returns a uniformly distributed random value from the interval $-\delta_{\text{rand}}...\delta_{\text{rand}}$. All random offsets were determined during the algorithm's initialization and invariant during the layering. The maximum magnitude δ_{rand} of the random part of the time-shift can be chosen arbitrarily but should be small compared to the constant part δ_{const} to avoid introducing artifacts. The theoretical maximum number of virtual clapping people possible without risking that the circularly shifted version of a signal is the signal itself again is determined by

$$P_{\max} = \text{floor}\left(\frac{T}{\delta_{\text{const}} + \delta_{\text{rand}}}\right) \cdot B. \tag{9}$$

Please note that this is an estimate for the worst case scenario.

With the microphone's omnidirectional pickup pattern, the determined parameters and signals can be inserted into Equation (5) yielding the microphone signal or final layered applause signal, respectively.

3.3. Metadata

In order to be able to evaluate additional attributes like clap rate, each captured base signal was manually annotated with respect to individual claps. This was done by local peak picking, i.e., a marker was set to every point in time of maximum absolute amplitude of an individual clap. For each base signal a corresponding metadata signal $\gamma_b(t)$ can be generated containing ones at the

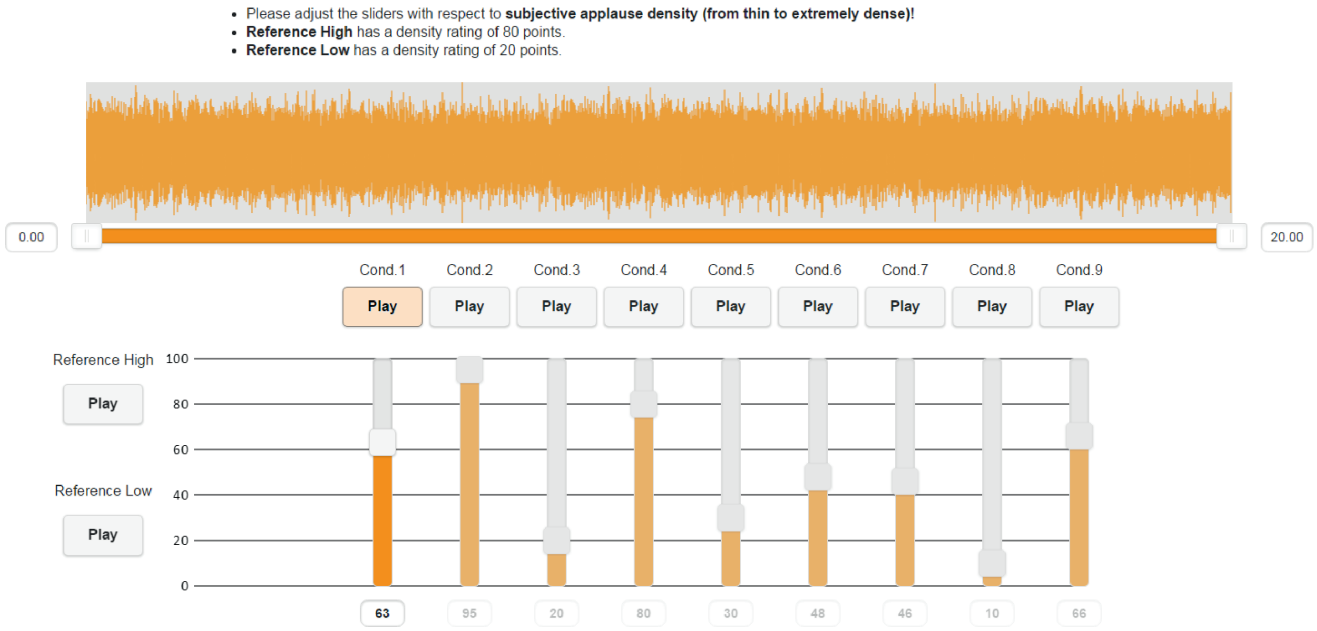


Figure 4: *Customized webMUSHRA [13] graphical user interface.*

marker positions and zeros everywhere else (see Figure 3). This allows for treating the metadata the same way as the base files and therefore layering the metadata almost the same way: attenuation factors and the microphone's directivity pattern can be omitted since the focus is simply on the question whether a clap is present or not and the microphone is assumed to have a constant directivity of one for all incident angles. The layered metadata signals can be described as

$$\Gamma(t) = \sum_{r}^{R} \sum_{p}^{P_r} \gamma_{r,p}(t). \tag{10}$$

The average clap rate of a synthesized applause signal is given by

$$\rho = \frac{1}{T} \sum_{t}^{T} \Gamma(t). \tag{11}$$

4. EVALUATION

4.1. Synthesized applause signals (Test 1)

To assess the perception of applause density, eight mono applause signals with increasing number of virtual people clapping were generated ranging from rather loose to quite dense applause. The individual clap signals were captured in the acoustically optimized sound lab 'Mozart' at Fraunhofer IIS [14]. This room has a mean reverberation time of $\mathrm{T}_m = 0.33\,\mathrm{s}$ and was designed to fulfill the strict recommendations of ITU-R BS 1116-1 [15]. For every single person, three clap signals with a uniform length of $T = 20\,s$ were recorded. Each person was placed individually in a distance of $d_0 = 1\,\mathrm{m}$ in front of a Neumann KM184 directional microphone (cardioid) and successively shown a picture of applauding people at three different venues and asked to clap their hands as if they were part of this crowd. For the layering, the signals were pooled yielding an overall number of $B = 24$ base signals.

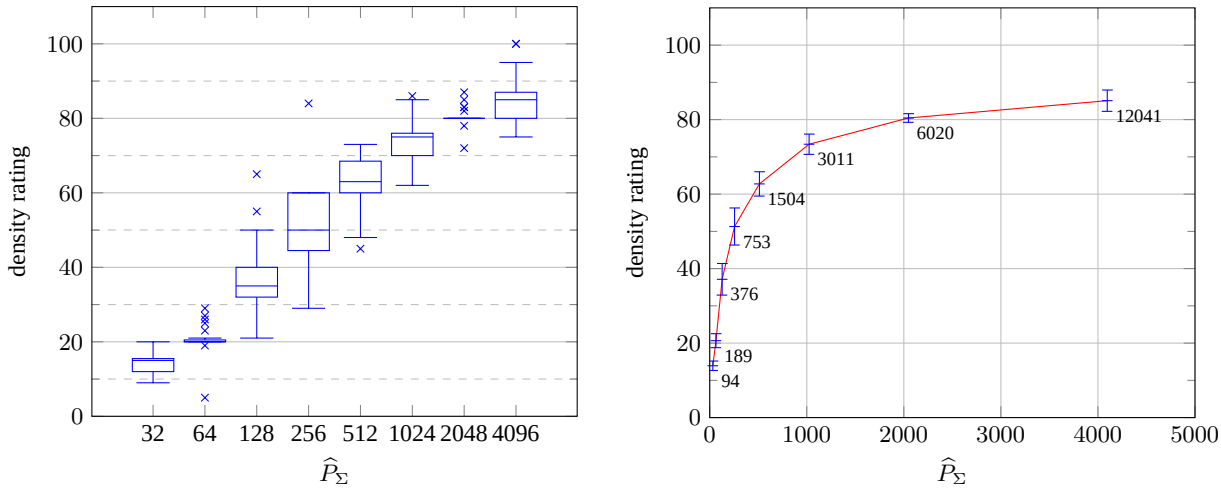
In order to be able to generate applause signals with a higher number of people clapping, the constant time shift δ_{const} had to be decreased compared to the default value in Table 1. The actual values for the constant time shift for a given number of virtual people as well as the corresponding theoretical limit P_{max} according to Equation (9) is given in Table 2. Please note that it was made sure by three persons during informal listening that the generated applause files do not contain artifacts due to the circular time shifting even if the number of virtual people exceeds the theoretical bound. The synthesized applause signals are available at [16].

$\widehat{P}_\Sigma$	32	64	128	256	512	1024	2048	4096
$\delta_{\mathrm{const}}[s]$	1.15	1.15	1.15	1.15	1.15	0.599	0.3	0.15
P_{max}	384	384	384	384	384	672	1200	1920

Table 2: *Parametrization of the layering algorithm for a given number of virtual people and theoretical limit P_{max} according to Equation (9).*

The stimuli were presented to the listening test participants side by side in a multi-stimulus test and as blinded conditions in randomized order. They were to be judged according to their perceived subjective applause density on a scale ranging from 0 to 100 density points and relative to each other. The conditions corresponding to $\widehat{P}_\Sigma = 64$ and $\widehat{P}_\Sigma = 2048$ were used as references corresponding to density levels of 20 and 80 density points, respectively. They were placed next to these values on the density scale and could be listened to any time for reference. However, they were also hidden among the test conditions. This means, the participants also had to identify these hidden references among the stimuli as equally dense compared to the respective reference signals and put them to the corresponding density values.

The test was conducted using Sennheiser HD 650 headphones. As graphical front end, a customized version of the webMUSHRA



(a) *Boxplot of density ratings depending on the number of virtual people clapping ($\widehat{P}_\Sigma = 64$ and $\widehat{P}_\Sigma = 2048$ correspond to the two references).*

(b) *Mean and t-distribution-based 95% confidence intervals. Corresponding clap rate $\rho\left[\frac{claps}{s}\right]$ is written next to confidence intervals.*

Figure 5: *Visualization of participants' responses for the listening test using synthesized applause signals.*

[13] tool was used. A screen shot of the graphical user interface is depicted in Figure 4.

4.2. Naturally recorded applause signals (Test 2)

Applicability to naturally recorded applause signals was assessed in a second listening test. The general test procedure was kept as in the first test but naturally recorded signals with different levels of applause density served as test and reference conditions. The following items were used, where *BBC Applause* and *ARL Applause* were part of the MPEG Surround item test set [17], *Klatschen* was part of the item set used in [18], *SmallCrowdClapping 2* [19] and *Initial Applause* [20] were taken from the freesound web site, *Intro3* and *17Exerc7* were taken from Frank Zappa's 'The Yellow Shark' where the first is an excerpt of the last applause at the end of the first track and the latter is an excerpt of the applause at the end of the 17th track of that record. Applause signals were looped and passively downmixed where necessary to obtain mono signals with a uniform length of 20 s. Additionally, two synthetically generated applause signals with $\widehat{P}_\Sigma = 32$ and $\widehat{P}_\Sigma = 1024$ where included to establish a connection between both listening tests. An overview of the used items and the condition number mapping is given by Table 3. In this test, conditions 2 and 7 served as 20 and 80 density point references, respectively.

Condition	Name
1	synthesized ($\widehat{P}_\Sigma = 32$)
2	klatschen
3	SmallCrowdClapping 2
4	Initial Applause
5	ARL Applause
6	synthesized ($\widehat{P}_\Sigma = 1024$)
7	BBC Applause
8	Intro3
9	17Exerc7

Table 3: *Mapping of item name and condition number of the natural applause signals.*

Test 1									
$\widehat{P}_\Sigma$	32	64	128	256	512	1024	2048	4096	
mean	13.91	20.65	37.13	51.30	62.74	73.39	80.43	85.09	
sd	2.94	4.35	9.84	11.53	7.57	6.31	2.69	6.63	
Test 2									
cond	1	2	3	4	5	6	7	8	9
mean	16.65	20.12	30.82	56.94	58.00	70.94	80.35	91.41	93.18
sd	8.10	2.89	17.51	12.34	14.21	9.16	3.90	8.24	12.80

Table 4: *Mean ratings and standard deviations for synthesized signals (upper table) and natural recordings (bottom table) including corresponding number of people clapping or condition number, respectively.*

5. RESULTS

5.1. Synthesized applause signals

In the first listening test, 23 participants, among which 20 male and 3 female, with an average age of 26.6 years (SD[1]=8.1) ranging from 18 to 53 years took part. Figure 5a shows boxplots of the raw data grouped by the number of people clapping. The very small to non-existent inter-quartile ranges of the reference conditions ($\widehat{P}_\Sigma = 64$ and $\widehat{P}_\Sigma = 2048$) indicate that most participants could easily identify the references and put the slider exactly to the desired value. The apparently high number of outliers for these conditions might result from the slider's measurements which itself cover a range of about 10 density points and therefore add some inaccuracies. If responses which lie within a ± 10 density point tolerance region are considered to have met the reference condition, only one response for $\widehat{P}_\Sigma = 64$ is to be considered as a true miss. An approximately logarithmic increasing perception

[1] SD = standard deviation

of applause density can be observed (approximately linear in the logarithmic presentation of the people clapping).

Figure 5b depicts means and t-distribution-based 95% confidence intervals of the participants' responses in a linear domain. All confidence intervals are non-overlapping. Also the distance from the upper confidence interval limit for $\widehat{P}_\Sigma = 2048$ to the lower limit for $\widehat{P}_\Sigma = 4096$ is 0.6 density points. This indicates that all conditions were perceptually well distinguishable. The red line, connecting the mean values, illustrates the quasi-logarithmic connection of people clapping and perceived density and also indicates that density cannot increase infinitely but saturates at some level. As a reference, the average clap rate is written next to the confidence intervals. It provides information of how many discrete events per second an applause signal consists of. The clap rate increases roughly linearly with the number of people clapping. In general, the results show that applause density can be rated consistently and density perception grows roughly logarithmically with increasing number of people.

5.2. Naturally recorded applause signals

Figure 6 shows boxplots of the responses of the listening test with naturally recorded applause signals. 17 participants among which 14 male and 3 female with an average age of 27.6 years (SD=9.1) ranging from 19 to 53 years were asked to rate applause density of naturally recorded applauses. Except for one response, the reference conditions (conditions number 2 and 7, respectively) can be considered to be recognized within the same ± 10 density points range of tolerance. The two synthesized applause signals correspond to condition 1 ($\widehat{P}_\Sigma = 32$) and 6 ($\widehat{P}_\Sigma = 1024$), respectively. The plot shows that it is possible, based on participants' responses, to produce a plausible ranking of the stimuli according to mean perceived density. It also shows that the spread of density ratings per stimulus is much wider if rating naturally recorded applause signals instead of synthesized ones, which indicates higher uncertainty of the participants. This is also supported by considering the higher standard deviations of the density ratings in the second listening test as given in Table 4. Between the first and the second test, the mean standard deviation increases from 6.48 to 9.91. Also, the average time a participant needed to pass through the test increased around 60% from 2.52 min (SD=1.93) to 4.17 min (SD=2.30) although the second test had only one additional item. This leads to the conclusion that applause density can also be rated consistently for naturally recorded applause signals but participants need more time, i.e., it appears to be harder, and the responses include more uncertainty.

6. CONCLUSION

This paper investigated the perceptual property that may be attributed to sound textures consisting of dense pseudo-random transient events like applause signals. Specifically, we proposed a novel perceptual attribute 'density' to characterize such sound textures. In order to exemplary verify the viability of this attribute on applause signals, a set of applause stimuli of varying densities were created. The generating procedure based on layering started from a set of dry clap recordings and placed these signals in a simple model of a virtual space where virtual clapping people are located in circular rows centered around the virtual microphone considering distance dependent level and timbre. A dedicated listening test methodology was designed based on a multi-stimulus

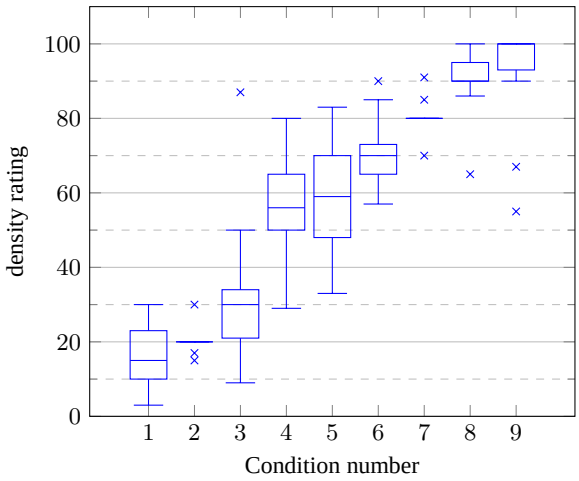


Figure 6: *Boxplots of the participants' responses in the listening test using naturally recorded applause signals (condition 2 and 7 correspond to the 20 and 80 density point references).*

test. Two tests were performed for assessment of the subjective applause density, one using the layered stimuli and the other testing natural applause recordings. The results showed that listeners reacted consistently and were able to reliably distinguish between different applause densities. It was shown that the increase in subjective density points roughly follows the logarithm of the number of clapping people. Therefore, it is evident that density is indeed a psychoacoustic property closely related to the physical measure of an impact rate, but more meaningful in terms of perception. Moreover, the applicability of the applause density attribute to naturally recorded applause signals was confirmed in a second listening test.

7. ACKNOWLEDGMENT

The authors would like to thank Michael Schoeffler for his technical support while customizing the webMUSHRA listening test environment.

8. REFERENCES

[1] B. H. Repp, "The Sound of two hands clapping: An exploratory study," *Journal of the Acoustical Society of America*, vol. 81, no. 4, pp. 1100–1109, 1987.

[2] L. Peltola, C. Erkut, P. R. Cook, and V. Välimäki, "Synthesis of Hand Clapping Sounds," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 1021–1029, 2007.

[3] A. Jylhä and C. Erkut, "Inferring the Hand Configuration from Hand Clapping Sounds," in *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finnland, 2008.

[4] D. Schwarz, "State of the Art in Sound Texture Synthesis," in *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*, Paris, France, 2011, pp. 221–231.

[5] J. H. McDermott and E. P. Simoncelli, "Sound Texture Perception via Statistics of the Auditory Periphery: Evidence

from Sound Synthesis," *Neuron*, vol. 71, no. 5, pp. 926–940, 2011.

[6] W.-H. Liao, A. Roebel, and A. W. Su, "On the Modeling of Sound Textures Based on the STFT Representation," in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, 2013.

[7] J. Brümmerstedt, R. McWalter, and T. Dau, "Analysis and synthesis of environmental sound based on auditory principles," in *DAGA*, Nuremberg, Germany, 2015, pp. 1452–1455.

[8] S. Siddiq, "Morphing of Granular Sounds," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15)*, Trondheim, Norway, 2015, pp. 4–11.

[9] Z. Néda, E. Ravasz, T. Vicsek, Y. Brechet, and A.-L. Barabási, "Physics of the rhythmic applause," *Phys. Rev. E*, vol. 61, no. 6, pp. 6987–6992, 2000.

[10] Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, and A.-L. Barabási, "The sound of many hands clapping: Tumultuous applause can transform itself into waves of synchronized clapping," *Nature*, vol. 403, pp. 849–850, 2000.

[11] A. Masurelle, "Towards a gestural control of environmental sound texture synthesis," Master's thesis, Pierre and Marie Curie University (UPMC), Sorbonne Universités, Paris, France, 2011.

[12] K. Kawahara, Y. Kamamoto, A. Omoto, and T. Moriya, "Evaluation of the Low-Delay Coding of Applause and Hand-Clapping Sounds Caused by Music Appreciation," in *138th Convention of the AES*, Warsaw, Poland, 2015.

[13] M. Schoeffler, F.-R. Stöter, B. Edler, and J. Herre, "Towards the Next Generation of Web-based Experiments: A Case Study Assessing Basic Audio Quality Following the ITU-R Recommendation BS.1534 (MUSHRA)," in *1st Web Audio Conference*, Paris, France, 2015.

[14] A. Silzle, S. Geyersberger, G. Brohasga, D. Weninger, and M. Leistner, "Vision and Technique Behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory," in *126th Convention of the AES*, Munich, Germany, 2009.

[15] ITU-R BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," 1997.

[16] Audio Examples of Synthesized Layered Applause Signals. [Online]. Available: https://www.audiolabs-erlangen.de/resources/2016-DAFx-ApplauseLayering/

[17] J. Breebaart, J. Herre, C. Faller, J. Rödén, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K. Kjörling, and W. Oomen, "MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status," in *119th Convention of the AES*, New York, USA, 2005.

[18] A. Kuntz, S. Disch, T. Bäckström, and J. Robilliard, "The Transient Steering Decorrelator Tool in the Upcoming MPEG Unified Speech and Audio Coding Standard," in *131st Convention of the AES*, New York, USA, 2011.

[19] Small Crowd Clapping 2. [Online]. Available: https://freesound.org/people/snakebarney/sounds/138114/

[20] Initial Applause. [Online]. Available: https://freesound.org/people/unfa/sounds/207728/