Applying the EBU R128 Loudness Standard in live-streaming sound sculptures

Marie Højlund
Institute for Communication and Culture, Aarhus University
Helsingforsgade 14
Aarhus, Denmark
musmkh@cc.au.dk

Morten S. Riis
Institute for Communication and Culture, Aarhus University
Helsingforsgade 14
Aarhus, Denmark
mr@cc.au.dk

Daniel B. Rothmann
Institute for Communication and Culture, Aarhus University
Helsingforsgade 14
Aarhus, Denmark
danielrothmann@me.com

Jonas R. Kirkegaard
Sonic College, UC SYD
Lembckesvej 7, Haderslev,
Denmark
jrk@soniccollege.org

ABSTRACT

This paper describes the development of a *loudness*-based compressor for live audio streams. The need for this device arose while developing the public sound art project *The Overheard*, which involves mixing together several live audio streams through a web based mixing interface. In order to preserve a natural sounding dynamic image from the varying sound sources that can be played back under varying conditions, an adaptation of the EBU R128 loudness measurement recommendation, originally developed for levelling non-real-time broadcast material, has been applied. The paper describes the Pure Data implementation and the necessary compromises enforced by the live streaming condition. Lastly observations regarding design challenges, related application areas and future goals are presented.

Author Keywords

NIME, Multimodal, Novel interface, audio, open source, soundscape, audio streaming, web interface, sound art installation.

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing, H.5.2 [Information Interfaces and Presentation] User Interfaces, D.2.6 [Programming Environments] Interactive environments.

1. CONTEXT & RELATED WORK

The loudness-based compressor described in this paper was developed for the project *The Overheard* [16], which is a part of the official programme of Aarhus 2017 – European Capital of Culture [9]. The main objective of The Overheard is to invite everybody to listen more carefully and rediscover the sounds of our everyday surroundings. This is presented by offering several different listening experiences that may be accessed by the public in different ways. This paper addresses a listening experience that allows the audience to enter a website from which they can mix their own soundscape from multiple live audio streams that originate in the sound environments of five public sound sculptures (see Figure 1). Thus, the streamed audio is a mix of the sounds coming from the sound sculptures[1] and the sonographic environments [13] at specific geographic locations.

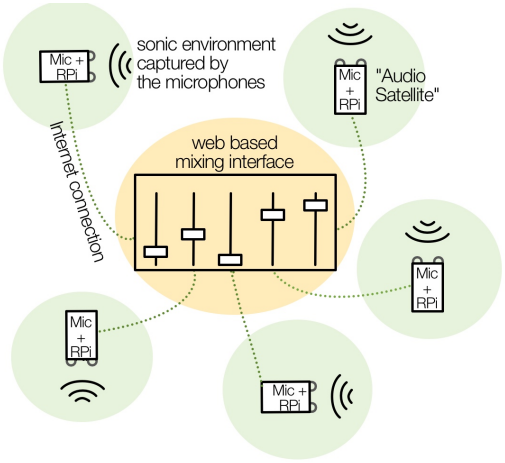

Figure 1: Geographically distributed Audio Satellites streaming to web-based mixing interface

The Overheard has a broad, public impact, and will be encountered by guests visiting the Capital of Culture and/or the website during 2017. The streaming of sound from the sites is enabled by an Audio Satellite [5], which is a device that comprises a Raspberry Pi running Pure Data [17] and DarkIce [8], a stereo audio interface, and two microphones, all collected in a weatherproof box, and requiring only power and an internet connection to work. The captured sound is streamed to the project website [15] where the audience may mix the five audio streams with a simple mixing interface (see Figure 2).

Since it is impossible to have control over both the loudness of the sound sources at each site and the listening conditions of the audience, a system is needed to ensure that each audio stream is within the same loudness range, no matter the sound level, so that they may be mixed together in a meaningful way. This paper suggests a solution for this, with respect to the dynamic quality of the original signal, by applying the EBU R128

[1] Detailed information about the sound sculptures may be found at [16]. The common denominator for all the pieces described is an exploration of the relationship between when the sound-art piece begins, and the natural-sounding environment takes over.

NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark.

loudness measurement in the construction of a dynamic range-mapping compressor based on subjective perceived loudness, rather than a regular compression/limiting algorithm that relies on manipulating the electrical level.

The Overheard project builds on previous explorations of the potential of using interactive sound to get people to engage with living cultural heritage and collective storytelling as a form of community engagement [1, 12]. Furthermore, the project is related to a long history of projects and inventions that, in various ways, aim to connect different geographical locations through sound in real time. Listening through technology goes back as far as the invention of the telephone and radio broadcasting.

In the last decade, owing to the development of easily accessible recording and digital technology, many projects have given access to the soundscapes of places around the world, and today these may be listened to through the internet [7]. Prominent examples includes the *London Sound Survey* launched by Ian M. Rawes in 2008, where the sounds of everyday public life throughout London are collected, in order to document the soundscape of the city of London, and how the sound environment changes [19]. Similarly, *radio aporee* – an open and collaborative platform for research on sound, founded by Udo Noll – allows people to upload sound recordings of everyday sound environments [3].

The perspectives on optimizing the experience of listening to live audio streams also extends to other fields, including live internet radio, live-streaming of concerts, performances, and bio-monitoring, not to mention contexts that combine sound and image. Broader internet bandwidth and faster processors for DSP processing enable higher quality, which commits us to revising all the links in the signal chain, including those related to audio compression and the utilization of dynamics.

2. RENDERED IMPRESSIONS

A listening experience is a multisensory experience [20], because it always involves being present at a specific place and time, and thus it is impossible to reconstruct the fullness of an audible impression of a place by simply playing back a recording of that *place*. Dani Iosafat [13] suggests the term 'psychosonography', which is adopted from psychogeography and seeks to construct an audible *expression* of a multisensory *impression* of the sonic environment of a place. For this purpose Iosafat defines different ways of collecting, treating, and adding to the recorded soundscape of a specific place, in order to render it into a new sonic expression that mimics the impression of the actual place [6].

Iosafat suggests a multitude of ways in which sonographic portraits of different urban places may be presented: (i) sound installations, (ii) recordings of the sound environments enhanced by emphasizing sound objects, (iii) adding sonic signs that are associated with a specific feature of a place, and (iv) adding recordings of musical phrases improvised by musicians to express mood and ambience.

Compared to the recorded, treated, and layered sound of the foregoing approach, *The Overheard* project uses other conceptual methods to mimic and render an expression, including (i) real-time streaming as opposed to pre-recorded audio playback, (ii) streaming of not just *a sound environment* but a carefully designed listening experience in the form of sound sculptures, which for most part generate some sort of musical structure, (iii) carefully selected stereo microphone placement that captures the sounds of the sound installation, the environmental sounds, and the sounds of the people/audience present, and (iv) managing the dynamics by adapting the EBU R128 recommendation, in order to ensure high-quality audio by preserving the dynamics, and hence the detail and a consistent sound level from all five locations (will be developed in section 2.1).

Figure 2: The mixing interface at www.overheard.dk

2.1 Adapting technology

In typical everyday surroundings, listening is shaped and adjusted to the situational and multi-sensory context. For example, this involves the way in which our ears automatically micro-adjust to compensate for expected loud sounds (e.g. anticipated through visual cues) [11]. Since the sites of the five sonic environments of *The Overheard* are unregulated, and potentially include both very loud and very quiet sounds that the user cannot adjust to when listening to them through the website, various design strategies have been used to imitate the psychoacoustic mechanisms of being physically present at the sites.

The main strategy of *The Overheard* is the application of a psychoacoustic concept from the broadcast industry, termed *perceived loudness levelling*, specified in the EBU R128 recommendation [18]. This approach seeks to eliminate perceived sound-level differences in broadcast material, and to avoid the excessive use of dynamic compression in pursuit of making one feature stand out. In other words, it is a solution to a practical problem that not only contributes to a more consistent sound level for the end user, but also results in much better sound quality in terms preserving dynamics [4, 10, 14].

This paper will elaborate on how a loudness-based compressor differs from a regular compressor in that its sensing mode is based on the EBU R128 recommendation, rather than peak or RMS, and adjusts the sound level in a way that is aligned with human perception. This achieves a more perceptually accurate, and thus more 'organic' processing of the sound. Listening tests have shown that the EBU R128 recommendation does indeed predict the perceived loudness with a relatively small error [4], and therefore provides a good basis for our solution.

Adapting the EBU R128 recommendation to this project offers a way to overcome the challenges that arise from the varying sound levels at each of *The Overheard*'s five sonic environments. Alternative solutions include setting a fixed microphone level and hoping for the best, and applying limiters and compressors, compromising dynamics and raising the noise floor.

When only one sonic environment is streamed, the necessity of handling perceived loudness is diminished, but would still be present, owing to the anticipated variety of playback devices,

ranging from built-in smartphone or laptop speakers to high-quality headphones or speaker systems. But the problems become more evident when the listener is invited to mix the sound from the five sonic environments in the web interface.

2.2 Loudness

As defined by Harvey Fletcher and W.A. Munson, 'loudness' is a psychological term used to describe the magnitude of an auditory sensation [10]. It is a psychoacoustic measurement representing the listeners' perception of volume, not the sound volume as an electrical level or sound-pressure level. When defining loudness, Thomas Lund suggests that the most important parameters at work are sound pressure level, frequency contents, and duration [14]. By a weighted combination, these parameters might approximate the loudness of a given sound, as perceived by an average listener.

Previous contributions to NIME on the subject of soundscape, such as those presented by Miles Thorogood and Philippe Pasquier, have utilized perceptual loudness measurements for audio-feature extraction [21]. However, these measurements focus on analysis, whereas the application of loudness measurements in *The Overheard* focus on processing the output to preserve detail and to keep individual audio streams at coherent levels, in order to ensure an intuitive listening experience through the mixing interface.

During the last decade, the terms *hypercompression* and *loudness war* were often mentioned in discussions of loudness in audio broadcasting. As described by Earl Vickers, 'Loudness war is a term applied to the ongoing increase in the loudness of recorded music, particularly on Compact Discs, as musicians, mastering engineers and record companies apply dynamics compression and limiting in an attempt to make their recordings louder than those of their competitors' [22]. In broadcast television, this has been particularly evident during commercial breaks, where advertisers may attempt to push the volume of their content as high as possible, in hopes of better grabbing the consumer's attention. However, it has been documented that this tendency may generate a number of negative consequences, in terms of both reduced audio quality and increased 'listener fatigue', and also in terms of consumer annoyance owing to jumping sound levels in TV programme flow [22].

In 2010 the European Broadcast Union (EBU) responded to these issues with recommendations for loudness normalization and the permitted maximum level of audio signals, called 'EBU R128'. This defined a new international standard for measuring audio loudness (ITU-R BS.1770-4) [2] along with two new units called 'LU' (Loudness Unit) and 'LUFS' (Loudness Unit Full Scale). This standard aims to solve the problem of volatile volume changes, as it makes it possible to regulate volume in a more organic manner, and enable a fairly consistent perceived loudness.

The EBU R128 recommendation may be summarized as an energy measurement of a filtered set of channels consisting of four main steps:

i) Frequency-weighting filtering, derived from a model of the frequency response of the human anatomy, is applied to the metered audio. This may be thought of as frequency weighting according to a simplification of the equal-loudness contour for the human ear, as determined by Fletcher and Munson [10], taking into account that human hearing is generally more sensitive to frequencies above 1kHz. An advantage of this simplification is that the processing may be done with simple time-domain blocks that have very low computational requirements [2].

ii) The power of the audio signal is calculated as the mean square of the metering period.

iii) All channels are summed together. If the channels represent a surround-sound setup, the rear channels will be weighted to have stronger representation in the measurement, since sounds coming from the rear are often perceived as louder [2].

iv) The sum of channels is optionally gated in 400ms sliding windows. The gating is applied when measuring the 'integrated' loudness, described below, and prevents extended periods of silence from being counted in the segment-wise loudness average.

A supplementary document, *EBU Tech 3341-2016*, suggests three 'EBU Modes'. They define three timescales from which to meter the outputted loudness measurements. The shortest timescale is called 'momentary', the intermediate is called 'short-term', and the programme- or segment-wise timescale is called 'integrated'. These modes suggest three ways of approaching loudness in relation to time: There is the loudness *right now* (momentary), there is loudness *about now* (short-term), and there is loudness as a *gated average over the full duration of a segment* (integrated). In particular, the integrated timescale discourages hypercompression in broadcasts, since a highly compressed segment will be measured as louder, and as a result, regulated by the broadcaster by generating a lower overall volume for the duration of that segment.

When considering the above-mentioned timescales for a live-streaming audio system, it becomes apparent that the integrated timescale is not relevant for the use case under discussion for two main reasons: Firstly, one cannot tell the future, and streaming live audio means that one may not know which sounds will occur next, therefore it is not possible to gauge a full segment for an accurate representation of an integrated loudness measurement; secondly, if one were to measure the integrated loudness from the beginning of a live audio feed up until the present moment, the audio feeds might be continuously active for months – the measured loudness of an audio stream a month, an hour, or even ten minutes ago might not be relevant to the listening experience at the present time.

2.3 Defining 'desirable level'

In order to adjust the loudness of a signal to a desirable level, the desirable level must first be defined. According to studies by Thomas Lund, typical audio consumers have certain identifiable dynamic range tolerances for different listening situations, which allow for comfortable and clearly distinguishable audio consumption [14]. Such listening situations may include the cinema, a living room, a bedroom, the street, or in a car. In situations with significant background noise, such as in transport or urban environments, a wide dynamic range is not viable, since sounds may be either too soft and difficult to distinguish, or uncomfortably loud, distorted, and possibly causing hearing damage. Lund's studies show that for most *real-world* listening situations, a quite narrow dynamic range is more desirable – about a 12 dB range with 8 dB headroom for bedroom listening, for example [14]. Since audio from the five sonic environments of *The Overheard* will be live-streamed to an unknown number of users in unknown listening situations, a compromise in dynamic range had to be made to accommodate a wide variety of users. In line with these

considerations, the algorithm we applied was designed to accomplish the following stated design goals:

1. It must automatically re-map the dynamics of an audio source, so that it is almost always at a clear and distinguishable volume for most listening situations;
2. It should do so based on the perceived loudness of an audio source rather than electrical level, utilizing the EBU R128 loudness standard;
3. It should work correctly with a live signal, without knowing which sounds might come next;
4. It should be secured from potential clipping or overloads;
5. It must run efficiently on a Raspberry Pi (the computer embedded in the Audio Satellite)
6. This algorithm's parameters must be changeable from another computer via a local network

To attain these goals, three steps of Dynamic Range Translation are implemented by the Audio Satellites. Outlining this algorithm, [Figure 3], its three main stages may be described as follows: Firstly, a short-term EBU R 128 loudness measurement is made, secondly, the dynamics are translated by two loudness-based compressors (one upward, one downward), and thirdly, an *overload safety net* is applied, consisting of an RMS compressor and a limiter that alters only the very top of the dynamic range. In the case of very quiet and subtle incoming audio, a lower lift threshold of -50 LUFS is enforced. From this point down the volume will not be raised further, though the amount of gain increase until that point will still be applied. The lower lift threshold is implemented as a measure to avoid raising possible noise generated by the recording device, such as microphone, preamp, or bit quantization noise. The effects of the loudness-based Dynamic Range Translation stages are illustrated in figure 4.

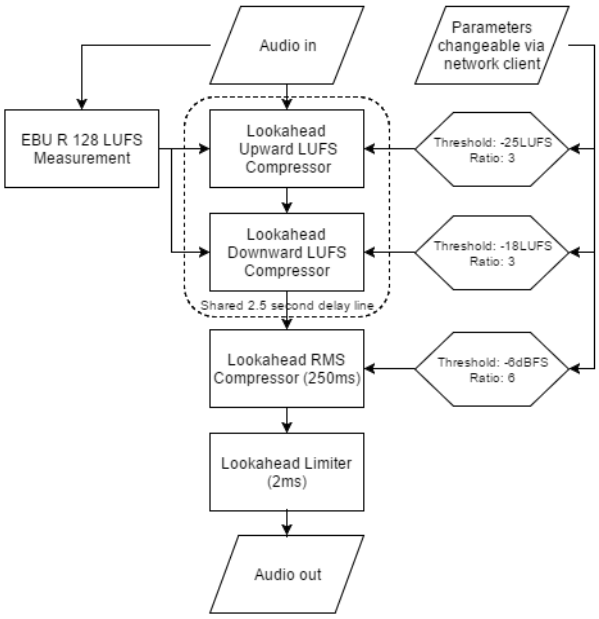

Figure 3: A diagram of the main components in the loudness-based dynamic translation algorithm

A significant aspect of the applied algorithm is its application of timescales. Since the latency of streaming audio from the Audio Satellite to the consumer is already 5–10 seconds (depending on network), a small added latency does not subtract from the listening experience. Therefore, all dynamic processing is designed to work with two seconds of *lookahead*. This means that the compression calculated for the current signal will instead be applied to a delayed signal, effectively making it possible for the compressor to alter the volume slightly *before* or *going into* a dynamic event in the audio. This helps reduce the over-exaggeration of transients, and allows for smoother volume adjustment before significant dynamic events.

The EBU Mode of the loudness measurement has been altered to produce an averaged loudness value each second, which is then interpolated over the duration of another second. Although not fully EBU Mode compliant, these two seconds of dynamic processing and interpolation time mean that the timescale of this alteration may be considered similar to EBU's definition of *short-term*. It alters the dynamics of *about now*, leaving the momentary signal, and, essentially, most sound events lasting less than a second, unaltered. Subtle dynamic details of the momentary signal may be conserved, whereas the overall loudness of the signal will be regulated in the short term. This way, the loudness of a signal may be controlled without significant reduction in momentary audio quality.

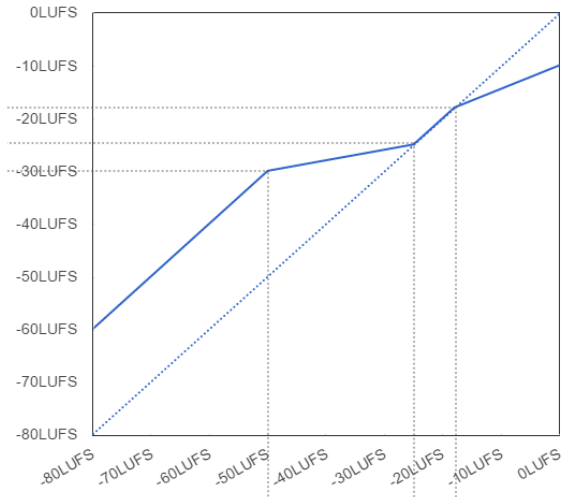

Figure 4: An illustration of loudness-based dynamic range translation

2.4 Implementation

Pure Data is the main software platform used to run the algorithm on the Audio Satellite. Pure Data was chosen for its open source nature, compatibility, and speed of prototyping. Since the program runs on a Raspberry Pi, which features an ARM processor, instead of the Intel processors usually found in PC systems, the Pure Data implementation was designed without the use of *externals*, for optimal ARM processor compatibility. This situation limited the number of advanced functional objects available, requiring us to program many of these functions through the use of basic objects.

The Pure Data implementation was developed through several iterations, alongside the design of the algorithm, prototyping, and testing ideas. The final program consists of the general steps described by the algorithm, each step implemented as a 'sub-patch' for better overview, with an added section for controlling essential parameters via local network.

Inside each sub-patch the desired functionality is achieved by combining Pure Data objects with advanced functionality and customized functionality from basic programming objects. For example, the frequency weighting filter for the loudness

measurement sub-patch is simply constructed from two 'biquad~' objects for each channel, initialized with the coefficients specified by the R128 technical documents. In contrast, the averaging of loudness values over the duration of a second required a larger number of basic objects in more complex configurations, since functionalities such as counters and accumulators had to be built from scratch.

The compressor implementation is based on a simplified hard knee compressor [23]. In the case of the downward compressor, if the signal is below the loudness threshold, the gain computation may be described as follows:

output = input

If the signal is above the loudness threshold, then it will be:

output = threshold + (input − threshold) / ratio

In our case, the input, output, and threshold parameters are based on loudness measurements, the input being the interpolated short-term measurement, and the threshold being a static one. The gain computation is followed by a 'lop~' lowpass filter, that simultaneously determines the speed of the compression attack and release. The incoming audio signal is fed into the gain computation algorithm followed by a ~2-second delay line, and this delayed signal is processed by the compressors, constituting the lookahead functionality.

For performance reasons, the Raspberry Pi in the Audio Satellite runs in 'headless mode', and thus no graphical user interface is available. This means that the user can only adjust the parameters in the Pure Data patch through command-line instructions. Therefore, a parameter control solution is also provided in the form of another Pure Data program designed to run on a client PC, from which parameters may be adjusted via local network. The user connects to the Audio Satellite by entering its local network IP address – When the 'SAVE' button is pressed, all parameters of the control program are transmitted to the Pure Data program running on the Audio Satellite, which is programmed to automatically forward received parameters to the correct destinations. The parameters are stored in a text file on the Raspberry Pi, which will be loaded on next start-up. In this way, parameters for the volume control algorithm on an Audio Satellite may be permanently changed, in order to more flexibly achieve specific dynamic ranges for particular situations, without having to further alter the program running on the device.

3. CONCLUSION & FUTURE DEVELOPMENT

This paper describes the development of the Dynamic Range Translation (based on the EBU R 128 loudness standard) implemented as a part of the live-streaming device Audio Satellites. The design goal for the Audio Satellite is to build a device that can handle live-streaming of high quality audio resembling the task of listening to complex everyday sound environments. Taking this as a starting point, *psychosonography* drives the overarching aesthetic choices of how to render the sonic environment through technology. A *psychosonographic* rendering of a site, listened to from another location, requires many choices and considerations regarding elements such as the acoustic properties of the sites, and the placement and quality of microphones. This paper has focused on the element that adapts the perceived loudness within a Raspberry Pi running Pure Data, introducing real-time streaming in terms of important *psychosonographic* rendering. Future investigations will expand the *psychosonographic* rendering to include aspects such as how the different live streams interact. Future work will focus on how people respond, engage with, and interact through the website, as we wish to investigate how *The Overheard* platform may be developed as a new way of experiencing sound art remotely across a larger geographical area.

REFERENCES

[1] Aamund D., Breinbjerg M. and Fritsch J. Ekkomaten – exploring the echo as a design fiction concept. *Digital Creativity*, *24*(1). 60–74.

[2] Algorithms to measure audio programme loudness and true-peak audio level. Retrieved January 31, 2017, from ITU-R: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-4-201510-I!!PDF-E.pdf

[3] Aporee, 2016. Retrieved January 31, 2017 from http://aporee.org/maps/

[4] Begnert, F., Ekman, H. and Berg, J. Difference between the EBU R-128 meter recommendation and human subjective loudness perception. in *131st Audio Engineering Society Convention* (New York, USA, 2011), Red Hook, NY: Curran Associates, Inc., 2012, Vol. 1, 40-49.

[5] Breinbjerg, M., Højlund, M., Riis, M., Fritsch, J. and Kirkegaard, J.R., Audio Satellites: Overhearing Everyday Life. In *COOP 2016: Proceedings of the 12th International Conference on the Design of Cooperative Systems*, (Trento, Italy, 2016), Springer International Publishing, 297-302.

[6] Chion, M. *Audio Vision.* Columbia University Press, New York, 1994.

[7] Cyberforest. Retrieved January 31, 2017, from http://www.cyberforest.jp/

[8] DarkIce, 2016. Retrieved January 31, 2017, from http://www.darkice.org/

[9] European Capital of Culture, Aarhus 2017. Retrieved January 31, 2017, from http://www.aarhus2017.dk/en/

[10] Fletcher, H and Munson, W. A. Loudness, Its Definition, Measurement and Calculation *The Bell System Technical Journal 12* (4). 377 - 430.

[11] Fox, S. *Human Physiology (ninth ed.)*. McGraw-Hill, New York, pp. 267–9, 2006.

[12] Højlund M, Riis M. Wavefront aesthetics – attuning to a dark ecology. *Organised Sound*, *20* (Special Issue 02). 249–262.

[13] Iosafat, D. On Sonification of Place: Psychosonography and Urban Portrait. *Organised Sound 14*, 1. 47–55.

[14] Lund, T. Control of Loudness in Digital TV. in *Broadcast Engineering and Information Technology Conference* (Las Vegas, NV, 2006), 57-65.

[15] Overheard.dk, 2017 (will be online from march 1st 2017). Retrieved January 31, 2017, from http://www.overheard.dk/

[16] Overheard, 2017. Retrieved January 31, 2017, from http://www.aarhus2017.dk/da/program/the-overheard/

[17] Pure Data, 2017. Retrieved January 31, 2017, from https://puredata.info/

[18] R 128 - Loudness Normalisation And Permitted Maximum Level Of Audio Signals, 2014. Retrieved January 31, 2017, from EBU: https://tech.ebu.ch/docs/r/r128-2014.pdf

[19] Soundsurvey, 2016. Retrieved January 31, 2017, from http://www.soundsurvey.org.uk

[20] Thibaud, J.P. The Sensory Fabric of Urban Ambiances. *The Senses & Society, 6,* 2. 203-215.

[21] Thorogood, M., and Pasquier, P. Impress: A Machine Learning Approach to Soundscape Affect Classification for a

Music Performance Environment. in *Proceedings of Proceedings of the International Conference on New Interfaces for Musical Expression,* (Daejeon, Republic of Korea, 2013), 256-260.
[22] Vickers, Earl. The Loudness War: Background, Speculation and Recommendations. in *AES 129th Convention* (San Francisco, CA, USA, 2010), 1-27.
[23] Giannoulis, D., Massberg, M., Reiss, J.D. Digital Dynamic Range Compressor Design - A Tutorial and Analysis. in *Journal of the Audio Engineering Societ*y (2012), 399-407