ASSESSING POWER IN PUNCHING SOUNDS FROM MAINSTREAM FILMS AND VIDEO GAMES

Hanna Järveläinen, Daniel Hug, and Stefano Papetti

Zurich University of the Arts
Pfingstweidstrasse 96
CH-8031 Zürich, Switzerland
hanna.jarvelainen@zhdk.ch
daniel.hug@zhdk.ch
stefano.papetti@zhdk.ch

ABSTRACT

A preliminary study was conducted with the aim of contributing to a scientifically based sound design for interactive movement sonification, drawing on tacit sound design knowledge. Specifically, the perceived power in punching sounds from mainstream films and video games was assessed through subjective ratings. The same sound clips were analyzed to extract audio features, and correlations with power ratings were examined. Results showed that perceived power in punching sounds is especially influenced by decay time, spectral slope, and features related to Mel-frequency cepstral coefficients (MFCCs). These findings can inform sound design for interactive movement sonification, particularly in the context of action-oriented applications.

1. INTRODUCTION

Audition plays a key role in perceiving bodily movement and its effect on the surrounding environment, making it a channel with high potential for augmenting exercise or rehabilitation activities. Exergames represent an emerging field that is driving the use of movement-related sonification [1]. They are digital games controlled by (whole-body) movements [2] which are becoming increasingly popular for promoting participation in exercise, as well as to improve adherence to exercise and rehabilitation tasks [3, 4]. As such, exergames need to be both effective and attractive for different target groups [5, 6, 7, 1]. However, scientifically based principles and heuristics that could guide the design of sounds for such purposes are still lacking. This is contrasted by the rich tradition of sound design for fictional media in which sonic expression of bodily action plays a central role, be it with realistic foleys or synthetic sound effects. To address this limitation, an interdisciplinary research approach is required that brings together experts in sound design, sound computing, sound perception, as well as sport and movement sciences. One possible starting point in this process is to examine and leverage the strategies used in mass media to convey the qualities of bodily actions [8, 9]. This paper presents an initial exploration in this direction, focusing on the perception of power in punching sounds.

The perception of bodily actions is inherently a multisensory experience — e.g. steps involve hearing, vision and haptic proprioception, and these modalities are in principle interchangeable [10]. Numerous studies have documented the potential and positive impact of auditory feedback on movement learning and execution, such as enhancing movement precision, supporting motor learning and control, and improving the overall experience and motivation in sports and rehabilitation, as well as modulating affective aspects [11, 12].

Tajadura-Jiménez et al. [13] showed that the perception of participant‘s arm length could be changed by varying the spatial properties of tapping sounds. Other research indicated that the effects of sound on body perception also occur in physically demanding situations, and could persist even after the feedback was stopped. Moreover, it was found that sound could contribute to participants' body weight and masculinity/femininity aspirations [14]. In related work, Clausen et al. [15] demonstrated that the perception of body size and strength could be affected using real-time spectral modification of audio captured from actual footsteps. Studies have been conducted to examine the affective and movement-related effects of interactive sonic and tactile feedback at foot level to simulate different types of ground: They explored how augmented footstep sounds could modify gait style and influence the walking experience [16, 17]. These findings support the idea that sounds can be designed to stimulate specific perceptions of bodily action and that the bodily self-image also emerges from sonic feedback from interactions with the world.

Further research addressed the effect of sounds on perception and action during walking [18], or action anticipation and performance in skateboarding [19]. Papetti et al. explored the use of auditory depth cues in a non-visual game to control the adjustment of gestural input [20]. A preliminary evaluation on swimming actions showed a positive effect of ad-hoc sonification created by musicians [21].

In terms of sound design for movement sonification, most current implementations use sounds in a naturalistic, veridical way (e.g., tapping sound for tapping action, walking sound for walking action). However, examples of non-naturalistic sound design for movement-related purposes can be found [11]. In one of the early successful real-world applications, Schaffert used pitch modulated tones to sonify the motion of rowing boats [22]. Experiments to explore alternative mappings of sounds to everyday actions with counter-intuitive interactive sonic feedback showed that sonic reinterpretation may lead to comical results, but the associations worked reliably and motivated sustained user interaction if sound cues were consistent with gestural affordances [23]. Indeed, Stanton and Spence [24] suggests that also nonveridical sounds can influence the perception and execution of action.

Over the years, design approaches for sonification have become increasingly diverse. Schaffert et al. differentiate between various methods for conveying movement-related information through sound, including natural movement sounds, tonal movement sonification (including musification), and rhythmic auditory information [25]. As an example, Rushton and Kantan used FM synthesis timbral changes to indicate gait assymetries in running [26]. Overall, these studies support the hypothesis that sonification of bodily movement can be inspired by sound design and musical expertise, and that meaningful non-naturalistic (fictional) action-sound relations are possible.

With the long-term goal of informing interactive movement sonification by leveraging tacit sound design knowledge contained in mainstream films and video games, we investigated if and how punching sounds from these media convey power through auditory cues, and which audio features are perceptually relevant.

2. PILOT STUDY

A pilot study was carried out in order to define the hypothesis, response variable and test design for the final experiment. A mixed methods approach was employed, combining analytic expert listening with subjective ratings and signal analysis.

A pilot analytic listening test was conducted on jumping and punching sounds from movies and video games. Participants were six expert listeners: advanced sound designers and senior lecturers or senior researchers in the field of sound engineering or electroacoustic music. This preliminary assessment confirmed that such sounds effectively convey movement qualities, even though their design approaches vary. The notion of "power" emerged as a central aspect of the experience.

The perception of power was then studied in an experiment with N=7 participants to explore its possible constituents. The stimuli were manually extracted audio-visual clips from 19 films and 15 games, cleaned of disturbances such as dialogue or music (see Section 3.2). The stimuli were selected from a short list based on the top 50 rated films and games in categories such as "action" or "sports" on major review aggregator sites such as Metacritic, IMDB or Rottentomatoes. Participants were asked to focus on the sound and rate its perceived power, additionally providing short motivation statements. The results aligned with the expert listening assessments to a degree that encouraged a larger experiment on perceived power.

Signal-level features were extracted using audio analysis and music information retrieval techniques. Correlations with perceived power ratings revealed for instance the importance of spectral distribution of energy: strong actions are more likely to be reflected by full frequency spectra and marked timbral brightness, whereas muffled or thin sounds are used for softer actions. These findings confirmed our intuitions about the impact of sound on the perception of movement power and the possible contribution from sound design practice, and motivated a larger study. Indeed, acoustic correlates are known to carry information on the power of impact events, such as changes in loudness, timbral patterns, and decay time [27].

3. EXPERIMENT

3.1. Design and protocol

Due to the exploratory nature of our research question, the experiment was run online. While this compromised full control over the playback settings used by the participants, it was deemed a minor inconvenience in light of the ability to collect a substantial amount of data from participants that would not be available to attend a laboratory study. The test utilized the Web Audio Evaluation Toolkit (WAET) along with the Audio Perceptual Evaluation (APE) multislider interface [28, 29], which allowed to implement a random access rating experiment. Clips from different media (film or video game) were tested separately: For each media, the user interface presented a new page with a horizontal bar representing power ratings from 'very low' (far left) to 'very high' (far right) on a continuous visual analog scale, and green sliders – each corresponding to a different sound clip (see Figure 1). The sliders could be clicked to play back the corresponding sample and dragged horizontally to enter the respective rating. The initial position of the sliders was randomized. Responses were measured as the sliders' final positions relative to the end-points [0,1].

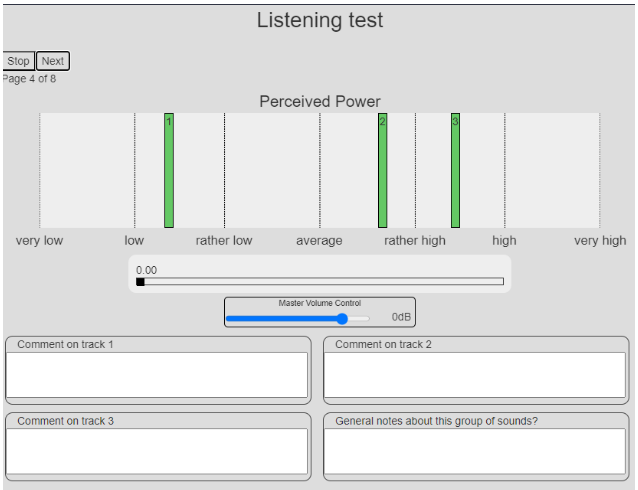

Figure 1: The test user interface.

After a pre-test questionnaire (age, gender, sound/music background, media consumption habits), participants were requested to wear headphones and instructed to listen to groups of punching sounds and rank them according to their perceived power, relative to the other sounds within the group. The volunteers were given the option to provide comments on each sound and rating assigned by them. To acquaint the participants with the task, a practice trial using noise bursts was conducted initially.

A single session comprised seven groups of sounds (i.e., 7 pages) each corresponding to one source (*Media*) and containing 3–7 clips (*Sound*). The pages were presented in randomized order. The total number of Sounds was 28 and participation took 15-20 minutes.

3.2. Stimuli

To limit test duration and to streamline design and analysis, the final experiment concerned punching sounds only. Twenty-eight audio tracks of punching actions from the pilot study were selected, presenting a diverse range of styles, genres, and cultures (see Table 1).

Audio samples were extracted from the original media and cleaned by an expert sound designer using Izotope RX[1] thus obtaining a single clip per punch. The aim was to isolate the primary impacts from background noise and echoes, and eliminate remaining traces of music or voices. Afterwards, the resulting sounds were meticulously reviewed by expert listeners to ensure they closely resembled the original samples. If any discrepancies were found, the cleaning process was revised.[2] The sounds were normalized to -25 dB LUFS.

3.3. Participants

An invitation with a web link to the experiment was distributed on social media platforms and mailing lists, including both audio expert communities and students of the Zurich University of the Arts. The experiment's website received 342 clicks, while the experiment itself was initiated 125 times. Fifty participants gave responses, including 38 complete cases. The participants were circa 60% male / 40% female, aged on average 38 years (min 18, max 67, sd 13). 80% described their listening skills as average or above. About 45% had amateur or professional musical skills, and circa 25% were sound designers or other sound professionals. Participants reported watching on average 3.3 action, fantasy, or science fiction films and playing on average 2.2 hours of action or fighting games per month.

4. AUDIO ANALYSIS

Audio features were extracted from the stimuli in order to use them as metric predictors of perceived power (see Section 5.2). The analyzed features were selected according to established standards in audio analysis or music information retrieval and computed in Matlab.

The audio samples were initially pre-processed as follows:

1. DC offset removal
2. Normalization (-25 dB LUFS as in the online experiment)
3. Trimming of initial and final silence

Then, the following features were extracted from each sound clip:

- Attack and decay time. To this end, the amplitude envelope was first obtained by applying the Hilbert transform and Gaussian smoothing (using a window size of 1000 samples); next, the envelope was normalized with amplitude ranging [0 1], and attack and decay times were computed based on fixed amplitude thresholds (both set to 0.6).
- Duration
- Dynamic range — computed with respect to the overall RMS level
- Spectral centroid (overall, and over time with mean, std, median) — a measure of the "center of mass" of the spectrum

[1]https://www.izotope.com/en/products/rx.html

[2]The stimuli, the audio features, the response dataset, and analysis code are available at https://doi.org/10.5281/zenodo.7863606

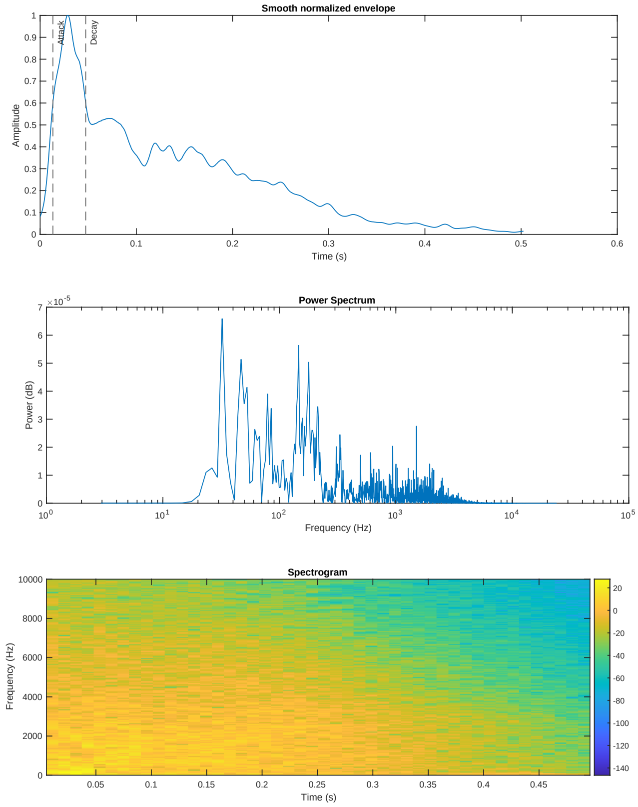

Figure 2: Top: Amplitude envelope (smoothed); bottom: power spectrum and spectrogram of an exemplary clip (kung_fu_hustle_scene_1_04)

- Spectral spread (overall, and over time with mean, std, and median) — i.e. the standard deviation around the spectral centroid
- Spectral flux (overall, and over time with mean, std, and median) — a measure of the variability of the spectrum over time
- Spectral slope (overall, and over time with mean, std, and median) — amount of decrease of the spectrum
- Spectral rolloff (overall, and over time with mean, std, and median) — a measure of the bandwidth of the audio signal
- Mel-frequency cepstral coefficients (MFCCs), with the change in coefficients (delta), and the change in delta values (deltaDelta). In more detail, 13 coefficients were taken into account. Summary statistics were computed, namely min, max, mean, median, variance, skewness, and kurtosis for MFCCs, and mean and variance for delta and deltaDelta.

In addition, A-weighted versions of spectral features were computed to more accurately reflect human auditory perception.

For use as metric predictors in statistical models, the audio features were z-transformed, i.e., centered and scaled. In what follows, such values will be marked with an initial 'z'.

Figure 2 shows the amplitude envelope with attack and decay time trigger points (top), as well as the power spectrum and spectrogram (bottom) of an exemplary sound clip.

Media	N of sounds	Notes
Films		
Big Hero 6 (2014)	4	Animated superhero film about fictional heavy combat robots.
Kung Fu Hustle (2004)	5	Asian martial arts parody. Expressive fighting sounds boarding comical.
Raging Bull (1980)	7	New Hollywood movie classic. Sounds highly varied, expressive and experimental.
Video games		
God of War (2018)	3	Hack'n'slay set in mythical Greece. Violent, visceral sounds.
Marvel's Spider-Man (2018)	3	Superhero action adventure. Protagonist is nimble and acrobatic.
Super Mario Odyssey (2017)	3	Lighthearted Jump'n'Run, playful, casual combat action.
Ultra Street Fighter IV (2014)	3	Arcade-Style classic fighting game combining physical and energetic hits.

Table 1: List of media and sound characteristics.

5. RESULTS

5.1. Ratings

Figure 3 presents the raw rating data. Due to missing data points by some participants and expected deviations from normality, statistical analyses were carried out by Bayesian inference [30] using R [31] and the brms package [32]. The brms package allows the user to specify Bayesian models using a similar formula syntax as is widely used for linear mixed effects regression models.

Because each Sound belonged to one Media and the task was grouped by Media, differences between the Media were first explored. A simple Gaussian model was fit for the Rating including a common intercept (marked by 1 in the formula below), a nested group effect of Sound within Media ((1 | Media/Sound)), and a random intercept for each Participant ((1|Participant)):

$$\text{Rating} \sim 1 + (1 \mid \text{Media/Sound}) + (1 \mid \text{Participant}) \tag{1}$$

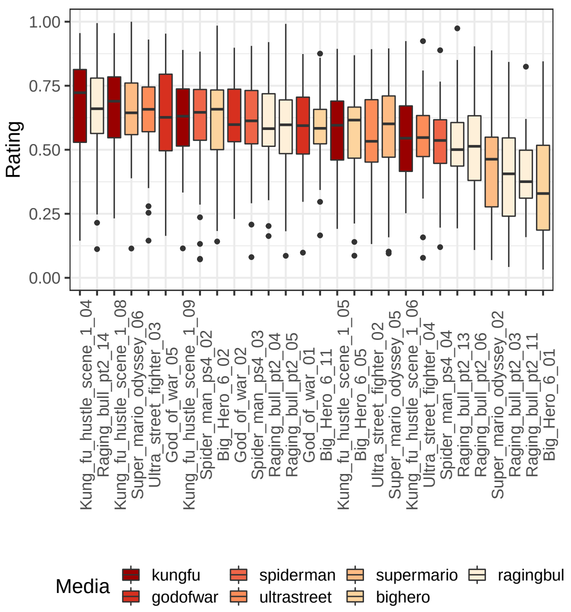

Figure 3: Raw rating data from the experiment.

Figure 4 presents the estimated group-level effect of Media with 95% credible intervals[3]. The overlapping intervals indicate no credible effect, suggesting that differences between Sounds within Media are larger than differences between Media. For this reason, Media was no longer used as a grouping factor in further analyses.

We then investigated differences between individual sounds by modeling Ratings using Sound as a categorical predictor as follows. Because the noise distributions of visual analog slider data deviate from normality due to the limited slider range [0,1], from now on a more suited zero-one inflated beta (ZOIB) distribution was assumed instead of Gaussian [33, 34]. The ZOIB distribution is described by four parameters: mean (μ) and precision (ϕ) of the beta distribution, probability of a binary $\{0, 1\}$ outcome (zoi), and the conditional probability of outcome $\{1\}$ (coi). In Eq. 2, the μ parameter depends on Sound, with an individual intercept for each Participant, and the precision parameter ϕ on Sound. As the raw data contains few zero and one ratings, the coi and zoi parameters were modeled with intercept only:

$$\begin{aligned} \mu\text{Rating} &\sim 1 + \text{Sound} + (1 \mid \text{Participant}) \\ \text{phi} &\sim 1 + \text{Sound} \\ \text{zoi} &\sim 1 \\ \text{coi} &\sim 1 \end{aligned} \tag{2}$$

Estimated Ratings with 95% credible intervals are presented in Figure 7. Differences between successive Sounds in the order of descending means were generally small; hypothesis testing on the model parameters revealed that the differences between the seven lowest-rated Sounds were mostly credibly non-zero. In contrast, the differences between the highest-rated and the thirteen successive Sounds were not.

5.2. Metric predictors

Next, we modeled the Ratings from the audio features described in Section 4. Correlation matrices were computed between Ratings and the z-transformed features. Through an iterative procedure, four predictors were selected that showed the highest correlations with the Ratings but, on the other hand, the lowest mutual correlations (see Figure 5): decay time, mean A-weighted (Aw) spectral slope, kurtosis of MFCC 10, and variance of delta MFCC 12.

Several models were fit including various combinations of these predictors. They were then compared in terms of their expected predictive accuracy by leave-one-out cross-validation; the

[3]Credible intervals and credible effects are the Bayesian counterparts of confidence intervals and significant effects in frequentist analysis

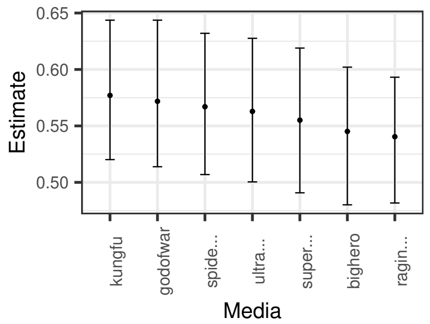

Figure 4: Estimated differences between Media with 95% credible intervals.

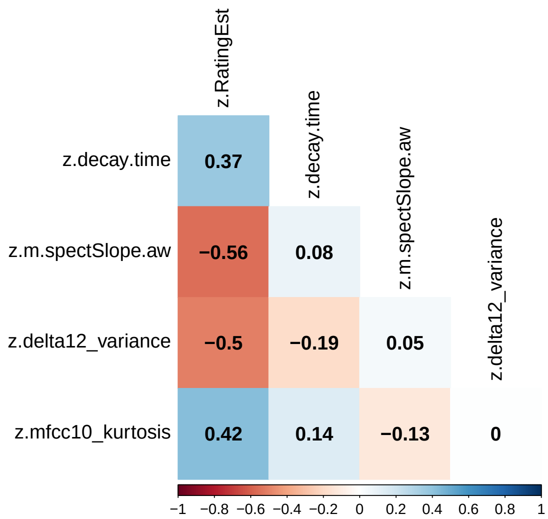

Figure 5: Correlations between the selected metric predictors and Ratings.

best fit was obtained by modeling μ on all four metric predictors and ϕ on the intercept only:

$$\begin{aligned} \mu\text{Rating} \sim & \text{z.Decay time} + \text{z.mean spectral Slope (Aw)} + \\ & \text{z.Delta12 variance} + \text{z.MFCC10 kurtosis} + \\ & (1 \mid \text{Participant}) \\ \text{phi} \sim & 1 \\ \text{zoi} \sim & 1 \\ \text{coi} \sim & 1 \end{aligned} \tag{3}$$

The resulting parameter estimates are presented in Figure 6(top). Note that the ZOIB-model estimates logit-transformed parameter values[4] and therefore the parameter estimates cannot be

[4]The logit function maps values from $p \in [0, 1]$ to $x \in [-\infty, \infty]$ according to $x = \log(\frac{p}{1-p})$; the inverse mapping is given by the logistic function $p = \frac{1}{1+e^{-x}}$

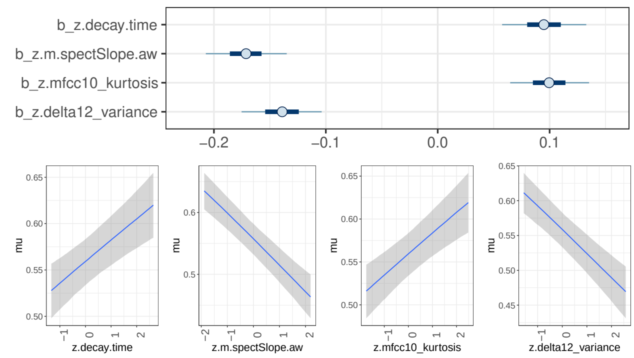

Figure 6: Parameter estimates from model 3 on logit-scale (top) and respective conditional effects on original scale (bottom).

directly interpreted in terms of ratings. However, the bottom panels of Figure 6 present the respective conditional effects in the original range. The strongest effect is that of spectral slope, where an increase by 1 standard deviation causes a decrease of 4.2 percentage points in the mean estimate. A similar change in delta MFCC 12 variance causes a drop of 3.5 percentage points. Decay time and MFCC 10 kurtosis have positive effects of 2.3 and 2.4 percentage points, respectively. Figure 7 presents the estimated means from model 2, and the average fitted values based on audio features from model 3. The predictions fall mostly within the estimated 95% credible intervals with only two exceptions, even though the metric model somewhat overestimates low-ranking sounds and underestimates high-ranking sounds.

To further investigate the seemingly confusing negative effect of mean spectral slope, the entire evolution of this feature over time was examined. The respective time series – seen in the bottom panel of Figure 8 – show one or several sudden drops, after which the spectral slope increases again. High-rated sounds (green: Rating estimate at least 0.5 std above average) show typically lower global minima than low-rated Sounds (red: Rating estimate at least 0.5 std below average).

Although spectral centroid was not included in model 3 due to moderately high negative correlation with spectral slope ($r = -0.45$), the evolution of spectral centroids over time was analyzed: the top panel of Figure 8 shows higher spectral centroids for high-rated Sounds.

6. DISCUSSION

As mentioned in Section 1, the concept of "power" in the context of action sounds is multifaceted and can be difficult to define. In order to focus and streamline our investigation, we chose to limit our analysis to punching sounds that are commonly found in mainstream films and games. It is important to note that the sounds within our dataset were originally created to accompany visual representations of punching actions. Another noteworthy aspect is that, consistent with action movies and games, the majority of the sounds in our dataset tend to fall within the medium-high power range (see Figure 3). Additionally, even clips originating from the same source could have significant variations in their sonic signature, making direct comparisons difficult. However, this choice was intentional in order to capture a wide range of perceived power. While this approach possibly made it challeng-

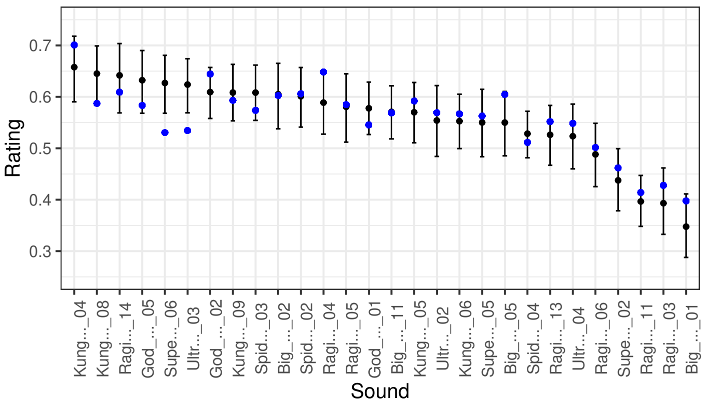

Figure 7: Estimated means and 95% credible intervals for ratings from model 2 (black) versus fitted values based on metric predictors from model 3 (blue).

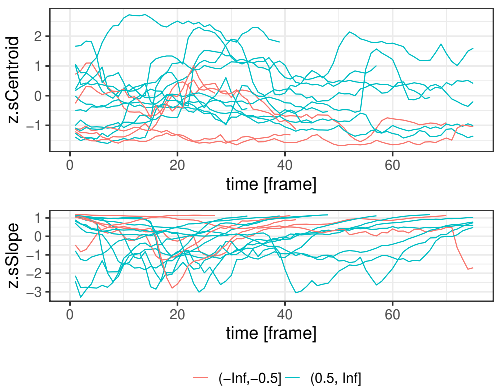

Figure 8: Spectral centroid (top) and spectral slope (bottom) time-series for low-rated (red) and high-rated Sounds (green).

ing to compare individual sounds, grouping them together helped manage cognitive load.

The recorded ratings show that differences between Sounds within Media are more significant than those observed between Media. This suggests the possibility of a continuum of perceived power which is independent of contextual factors, such as style, and that there exist "transcendent" properties of the sonic power experience.

In this perspective, Figure 8 shows that high-rated punching Sounds have a deeper drop in spectral slope after the attack and a higher spectral centroid during decay. However, our endeavors to predict the ratings using audio features did not yield results that were as straightforward to interpret. Figure 6 indicates that manipulating decay time and spectral slope can be an effective sound design technique for attaining a desired perceived power. However, results pertaining to MFCCs are not as easily applied. Indeed, MFCCs are a compact representation of the spectral envelope of an audio signal, and as such they do not contain all the information needed to reconstruct the original signal. In general, each MFCC represents a different aspect of the spectral information of the audio signal, with lower coefficients typically capturing information about the overall spectral envelope of the signal, and higher coefficients capturing more detailed spectral information. At this point, it is important to note that our study is still exploratory, and that a more extensive dataset would be required to provide sufficient information to guide these processes.

Previous research on realistic sounds has shown that in addition to the movement or action itself, the sound conveys information on the resulting impact and involved objects, such as material, size, shape, hardness, and force and location of impact, and that these cues are often hard to separate from each other and utilized differently between individuals [27, 35]. In isolation, judgments of force impact were based on previous long-term experience and individual decision rules which deviated significantly from ideal use of available physical cues [35]. Our findings reflect the importance of impact properties through the high significance of decay-related features. The effects of long-term experience and individual decision rules could not be examined presently but motivate future studies focusing on specialized gaming communities, for instance.

Regarding the experimental protocol, data collected online are expected to be noisier than those collected in a laboratory setting. This is mostly due to a lack of control over factors such as headphones settings, the environment in which the experiment was conducted, and the level of focus of the participants. A significant portion of users did not complete the test, possibly due to technical difficulties or loss of interest. Indeed, from over 100 logged clicks, we could collect data from 50 persons, and this limited the complexity and accuracy of our statistical models. To further refine them, alternative approaches such as using non-linear models or incorporating functional predictors such as spectral slope and spectral centroid could be considered.

Nonetheless, this study provides an exploratory step towards predicting perceived power from audio features, which can serve as motivation for future research and design efforts.

7. CONCLUSION

This study aimed to establish a scientifically based approach to designing sounds for interactive movement sonification by investigating the perceived power of punching sounds in a small selection of films and video games. The results indicate that perceived power is influenced by decay time, spectral slope, and a choice of MFCC-related features. Despite the limitations of online data collection and a small sample size, such findings provide valuable insights for designing sounds with a target power signature. Overall, the proposed approach has significant potential in action-oriented applications such as exergames, where perceived power plays a critical role in the user's engagement with the system.

Although the mapping of MFCC-related features to sound design strategies presents challenges, techniques such as codebook-based audio synthesis and generative models can be utilized to synthesize audio from them.

Building on the foundation of this preliminary study, it is hoped that further research will enable sound designers to predictably create variations of a reference sound in terms of power, such as "weak", "medium", and "strong". Once sounds are synthesized to convey a target power perception, they should undergo evaluation to ensure they achieve their intended effect and to close the loop of the design process.

Future research could also explore alternative experimental protocols, qualitative listening assessments, additional audio features, and the effects of contextual factors like visual stimuli and user characteristics on perceived power.

8. ACKNOWLEDGMENT

The authors would like to express their gratitude to Jan Godde, expert sound designer, for his contribution in cleaning up the original audio files.

9. REFERENCES

[1] A. L. Martin-Niedecken, E. Márquez Segura, K. Rogers, S. Niedecken, and L. Turmo Vidal, "Towards Socially Immersive Fitness Games: An Exploratory Evaluation Through Embodied Sketching," in *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts - CHI PLAY '19 Extended Abstracts*. Barcelona, Spain: ACM Press, 2019, pp. 525–534.

[2] Y. Oh and S. Yang, "Defining exergames & exergaming," in *Proceedings of Meaningful Play 2010*, East Lansing, MI, USA, 2010, p. 17.

[3] J. W. Burke, M. D. J. McNeill, D. K. Charles, P. J. Morrow, J. H. Crosbie, and S. M. McDonough, "Optimising engagement for stroke rehabilitation using serious games," *The Visual Computer*, vol. 25, no. 12, p. 1085, Aug. 2009.

[4] N. Skjæret, A. Nawaz, T. Morat, D. Schoene, J. L. Helbostad, and B. Vereijken, "Exercise and rehabilitation delivered through exergames in older adults: An integrative review of technologies, safety and efficacy," *International Journal of Medical Informatics*, vol. 85, no. 1, pp. 1–16, Jan. 2016.

[5] A. S. Lu, H. Kharrazi, F. Gharghabi, and D. Thompson, "A Systematic Review of Health Videogames on Childhood Obesity Prevention and Intervention," *Games for Health Journal*, vol. 2, no. 3, pp. 131–141, June 2013.

[6] E. J. Lyons, "Cultivating Engagement and Enjoyment in Exergames Using Feedback, Challenge, and Rewards," *Games For Health Journal*, vol. 4, no. 1, pp. 12–18, Feb. 2015.

[7] G. McRae, A. Payne, J. G. E. Zelt, T. D. Scribbans, M. E. Jung, J. P. Little, and B. J. Gurd, "Extremely low volume, whole-body aerobic-resistance training improves aerobic fitness and muscular endurance in females," *Applied Physiology, Nutrition, and Metabolism = Physiologie Appliquee, Nutrition Et Metabolisme*, vol. 37, no. 6, pp. 1124–1131, Dec. 2012.

[8] D. Hug, "Investigating Narrative and Performative Sound Design Strategies for Interactive Commodities," in *Auditory Display*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, vol. 5954, pp. 12–40.

[9] ——, "<<CLTKTY? CLACK!>> Exploring Design and Interpretation of Sound for Interactive Commodities," Ph.D. dissertation, Kunstuniversität Linz, Linz, 2017.

[10] L. D. Rosenblum, J. W. Dias, and J. Dorsi, "The supramodal brain: Implications for auditory perception," *Journal of Cognitive Psychology*, vol. 29, no. 1, pp. 65–87, Jan. 2017.

[11] O. Höner, A. Hunt, S. Pauletto, N. Röber, T. Hermann, and A. O. Effenberg, "Aiding Movement with Sonification in "Exercise, Play and Sport"," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Berlin: Logos Verlag, 2011, pp. 525–553.

[12] V. van Rheden, T. Grah, and A. Meschtscherjakov, "Sonification approaches in sports in the past decade: A literature review," in *Proceedings of the 15th International Conference on Audio Mostly*. Graz Austria: ACM, Sept. 2020, pp. 199–205.

[13] A. Tajadura-Jiménez, T. Marquardt, D. Swapp, N. Kitagawa, and N. Bianchi-Berthouze, "Action Sounds Modulate Arm Reaching Movements," *Frontiers in Psychology*, vol. 7, p. 1391, 2016.

[14] A. Tajadura-Jiménez, J. Newbold, L. Zhang, P. Rick, and N. Bianchi-Berthouze, "As Light as You Aspire to Be: Changing Body Perception with Sound to Support Physical Activity," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–14.

[15] S. Clausen, A. Tajadura-Jiménez, C. P. Janssen, and N. Bianchi-Berthouze, "Action Sounds Informing Own Body Perception Influence Gender Identity and Social Cognition," *Frontiers in Human Neuroscience*, vol. 15, p. 688170, July 2021.

[16] R. Bresin, A. de Witt, S. Papetti, M. Civolani, and F. Fontana, "Expressive sonification of footstep sounds," in *Interact. Sonification Work.*, R. Bresin, T. Hermann, and A. Hunt, Eds., Stockholm, Sweden, 2010.

[17] A. Lécuyer, M. Marchal, A. Hamelin, D. Wolinski, F. Fontana, M. Civolani, S. Papetti, S. Serafin, and Others, "Shoes-your-style: Changing sound of footsteps to create new walking experiences," Alghero, Italy, pp. 13–16, 2011.

[18] S. Papetti and F. Fontana, "Effects of audio-tactile floor augmentation on perception and action during walking: Preliminary results," in *Sound Music Comput. Conf.*, 2012.

[19] P. Cesari, I. Camponogara, S. Papetti, D. Rocchesso, and F. Fontana, "Might as Well Jump: Sound Affects Muscle Activation in Skateboarding," *PLoS ONE*, vol. 9, no. 3, p. e90156, Mar. 2014.

[20] S. Papetti, D. Devallez, and F. Fontana, "DepThrow: A physics-based audio game," in *Int. Conf. Audit. Disp.*, Paris, France, 2008.

[21] G. Seibert, D. Hug, and M. Cslovjecsek, "Towards an Enactive Swimming Sonification: Exploring Multisensory Design and Musical Interpretation," in *Audio Mostly - 10th Conference on Interaction with Sound*. New York: ACM Press, 2015, pp. 1–8.

[22] N. Schaffert, K. Mattes, and A. O. Effenberg, "A Sound Design for Acoustic Feedback in Elite Sports," in *Auditory Display*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, S. Ystad, M. Aramaki,

R. Kronland-Martinet, and K. Jensen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, vol. 5954, pp. 143–165.

[23] P. Polotti, S. Delle Monache, S. Papetti, and D. Rocchesso, "Gamelunch: Forging a dining experience through sound," in *CHI '08 Ext. Abstr. Hum. Factors Comput. Syst.* New York, New York, USA: ACM Press, 2008, p. 2281.

[24] T. R. Stanton and C. Spence, "The Influence of Auditory Cues on Bodily and Movement Perception," *Frontiers in Psychology*, vol. 10, p. 3001, Jan. 2020.

[25] N. Schaffert, T. B. Janzen, K. Mattes, and M. H. Thaut, "A Review on the Relationship Between Sound and Movement in Sports and Rehabilitation," *Frontiers in Psychology*, vol. 10, Feb. 2019.

[26] T. A. Rushton and P. R. Kantan, "A real-time embodied sonification model to convey temporal asymmetry during running," p. 8, 2022.

[27] D. Rocchesso and F. Fontana, Eds., *The Sounding Object*. Firenze: Edizioni di Mondo Estremo, 2003.

[28] N. Jillings, B. De Man, D. Moffat, J. D. Reiss, and R. Stables, "Web Audio Evaluation Tool: A framework for subjective assessment of audio," in *Proceedings of Web Audio Conference*, Atlanta, USA, 2016.

[29] B. De Man and J. D. Reiss, "APE: Audio perceptual evaluation toolbox for MATLAB," in *Audio Engineering Society Convention 136*, Apr. 2014.

[30] J. K. Kruschke, *Doing Bayesian data analysis - A tutorial with R, JAGS, and Stan*, 2nd ed. Academic Press, 2014.

[31] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021.

[32] P.-C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *Journal of Statistical Software*, vol. 80, no. 1, pp. 1–28, 2017.

[33] F. Liu and Y. Kong, "zoib: An R Package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression," *R J.*, vol. 7, no. 2, pp. 34–51, 2015.

[34] R. Ospina and S. L. P. Ferrari, "Inflated beta distributions," *Stat Pap.*, vol. 51, pp. 111–126, 2010.

[35] R. A. Lutfi, C.-J. Liu, and C. N. J. Stoelinga, "Auditory discrimination of force of impact," *J. Acoust. Soc. Am.*, vol. 129, p. 1017, 2011.