SONIC CHARACTERISTICS OF ROBOTS IN FILMS

Adrian B. Latupeirissa
Sound and Music Computing
KTH Royal Institute of Technology
ablat@kth.se

Emma Frid
Sound and Music Computing
KTH Royal Institute of Technology
emmafrid@kth.se

Roberto Bresin
Sound and Music Computing
KTH Royal Institute of Technology
roberto@kth.se

ABSTRACT

Robots are increasingly becoming an integral part of our everyday life. Expectations on robots could be influenced by how robots are represented in science fiction films. We hypothesize that sonic interaction design for real-world robots may find inspiration from sound design of fictional robots. In this paper, we present an exploratory study focusing on sonic characteristics of robot sounds in films. We believe that findings from the current study could be of relevance for future robotic applications involving the communication of internal states through sounds, as well for sonification of expressive robot movements. Excerpts from five films were annotated and analysed using Long Time Average Spectrum (LTAS). As an overall observation, we found that robot sonic presence is highly related to the physical appearance of robots. Preliminary results show that most of the robots analysed in this study have "metallic" voice qualities, matching the material of their physical form. Characteristics of robot voices show significant differences compared to voices of human characters; fundamental frequency of robotic voices is either shifted to higher or lower values, and the voices span over a broader frequency band.

1. INTRODUCTION

Robots are increasingly becoming an integral part of modern society. With an increased presence of social robot interfaces comes increased demands on robots to effectively communicate with their human counterparts. The work presented in this paper is conducted within the context of the SONAO project, previously described in [1]. The SONAO project aims to improve the comprehensibility of robot non-verbal communication (NVC) through an increased clarity of robot expressive gestures and non-verbal sounds. Previous research conducted within the SONAO project has focused on developing re-targeting techniques for a NAO [1] robot based on findings from virtual character animation research [2] and perception of mechanical sounds inherent to expressive gestures of a NAO robot [1]. Future work in the SONAO project includes the use of movement sonification to increase the comprehensibility of robotic gestures and emotional states. In the current study, we shift the focus from physical robots such as the NAO, to fictional robot characters in films and the sonic representations thereof. The main aim of the work is to gain insight into how Foley artists have tackled the task of designing robot sounds. We believe that findings from this exploratory work could be relevant for sound designers focusing on robotic interfaces, particularly for future implementations involving sonification of robot movements.

Previous work focusing on developing sounds for the robot NAO includes e.g. [3–5]. However, even if some previous studies have focused on sounds for communication and emotional expression in Human Robot Interaction (HRI), the sounds used in such work has often been based on simple sound synthesis methods. For example, sonification has only been used to a very limited extent in HRI (see e.g. [6, 7]). Moreover, previous work have often lacked detailed descriptions of mapping strategies or motivations of design decisions.

Our hypothesis is that sonic representations of robots in films could influence the expectations on sounds produced by real-world robots, thus affecting human robot interaction. In previous work, it has been reported that participants in sound design workshops referred to sound in movies when asked to describe sonic interaction experiences [8]. On a general note, it has been suggested that interfaces from science fiction films offer lessons to interaction designers, as science fiction interfaces reflect current interface understandings in terms of expectations from users [9], and that our concept of robots is influenced by the image of robots from science fiction [10]. In the current paper, we present an exploratory study focusing on sonic characteristics of robot sounds in films. We believe that findings from this study could be relevant for sound design in the field of HRI.

2. BACKGROUND

The term *"non-verbal communication"* refers to utterances that do not involve semantics in natural spoken language but may still facilitate rich communication and expression. Non-verbal communication can be organized into four categories: Gibberish Speech (GS), Non-Linguistic Utterances (NLUs), Musical Utterances (MUs) and Paralinguistic Utterances (PUs), all of which are brought together under the umbrella term Semantic-Free Utterances (SFUs) [11]. SFUs can be described as auditory communication or interaction means for machines that allow

[1] https://www.softbankrobotics.com/emea/en/nao

emotion and intend expression, composed of vocalizations and sounds without semantic content [11]. Previous research has shown that NLUs can convey affect and that people show categorical perception at a level of inferred affective meaning when listening to robot-like sounds [12]. In the current paper, we examine both non-verbal (e.g. sounds emanating from the body of the robot; robot movement sounds) and verbal (robot speech) sounds of fictional robots.

As for all products involving some kind of sound design, sounds can play a role in our aesthetic, quality, and emotional experience [13]. In [13], the authors make a distinction between sounds that are generated by operating of the product itself, and sounds that we intentionally add to a product. In the context of HRI, we need to consider both intentional sounds that are specifically designed to communicate certain emotional reactions or intentions, and consequential sounds inherent to the robot's movements. A study focusing on consequential sounds for servo motors commonly used to prototype robotic movement was presented by Moore et al. in [14]. Results suggested both anthropomorphic associations with sounds and negative impressions of motor sounds overall. In the current paper, nonverbal communication and sounds used to augment particular emotions through movement can be considered intentional sounds. One of the benefits of working with fictional robots in films is that a Foley artist can design all sounds produced by a robotic character, which is usually not the case for actual mechanical robots in real life (their movements often automatically produce sounds which are not specifically designed).

In [3], a library of emotional expressions consisting of gestures and sounds was presented. However, the authors did neither describe the sound design in detail, nor the mapping strategies used. In [4], different sounds defined to express robot emotions were evaluated using recognition ratios. Sounds ranged from gibberish speech, alienated human voices, "bleeps" to animal sounds. In [5], authors introduced BEST (Bremen Emotional Sound Toolkit) [2] , a validated set of 408 short (700ms to 16s) electronic sound emblems, created to augment the nonverbal capabilities of the NAO robot.

Up to this point, relatively little work has focused on how Foley artists design robot sounds. In [15], authors discuss the use of non-verbal sounds for communication of affect in interaction with robots, mentioning the sound designer Ben Burtt, who produced the sounds for the R2-D2 robot in Star Wars and Wall-E, as a source of inspiration. The story of how Burtt struggled for months before finding a R2D2 sound with credibility and character is described in detail in [16]. Burtt started experimenting with various synthesizers (Moog and ARP) to produce electronic beeps and tonalities. However, these sounds lacked emotional meaning, and Burtt therefore started blending the electronic sounds with mechanically generated sounds ("emotional" acoustic noises such as e.g. whistling sounds and expressive squeaks produced by bits of metal touching dry ice). Finally, he produced baby babble using his own voice and intercut it with electronic tones. The final version of R2-D2 involved a method in which Burtt played the synthesizer simultaneously as he recorded his voice, which in turn triggered electronic sounds and simultaneously shaped envelopes and pitches.

In [10], seven different musical sounds, five of which expressed intention and two that expressed emotion, were designed for the robot Silbot. In order to identify sound design considerations, sounds of the robots R2D2 and Wall-E were initially analysed. A total of 175 sound samples from Star Wars and 100 sounds from Wall-E were categorised into two different groups: intention sounds (conveying meaning/emphasizing a situation) versus emotional sounds (expressing feelings). Authors found that intonation, pitch and timbre were dominant musical parameters to express intention and emotion.

3. METHOD

This study aims to analyse robot sounds in films, thereby creating a basis of knowledge for future studies in the SONAO project. As mentioned above, our hypothesis is that sonic portrayal of robots in films could have an influence on expectations on sounds produced by robots. Specifically, we are looking into robot's sonic presence (i.e. sounds that signify the presence of a robot in a scene), auditory expression (i.e. sounds that signify the display of emotion), and spectral characteristics of robot speech. Results will inform the design of future sonic representation of real-world robots in the SONAO project.

3.1 Film Selection

Five films were selected to be analysed in the current study. Main criteria for inclusion was that there was a presence of a humanoid robot with human-like behaviour in the film. This selection was done since the focus of the SONAO project is mainly on humanoid robotic interfaces. Furthermore, the inclusion criteria was defined so as to limit the total number of investigated robotic interfaces. The defining factor of the behaviour in this context was that the robot was capable of establishing an empathetic conversation with human characters in the film. To narrow down the selection, only non-animated films involving English-speakers were considered. In addition, it was important that the robot had sufficient screen time with no noticeable background music or noise, as the robot sounds would otherwise have to be separated from other sounds using source-separation methods. With these criteria in mind, one film was selected from each decade from the 1970s to 2010s. The selected films are *The Black Hole* [3] (1979, produced by Walt Disney Production); *Short Circuit* [4] (1986, TriStar Pictures, et al.); *Bicentennial Man* [5] (1999, 1492 Pictures, et al.); *I, Robot* [6] (2004, 20th Century Fox, et al.); and *Chappie* [7] (2015, MRC, et al.).

[2] http://gaips.inesc-id.pt/emote/best-bremen-emotional-sound-toolkit/

[3] https://www.imdb.com/title/tt0078869/

[4] https://www.imdb.com/title/tt0091949/

[5] https://www.imdb.com/title/tt0182789/

[6] https://www.imdb.com/title/tt0343818/

[7] https://www.imdb.com/title/tt1823672/

In the current study, all films were produced in USA, and all of the robot characters were male. The issue of representation in this context does not go unnoticed. Future studies factoring differences in culture and gender will be conducted in later stages of the current project, with work including female robot characters that fit the inclusion criteria e.g. Ava in *Ex Machina* [8] (2014, Universal Pictures International, et al.) and L3-37 in *Solo: A Star Wars Story* [9] (2018, Lucasfilm, et al.).

3.2 Analysis

For each selected film, video excerpts displaying respective robot and a human counterpart were isolated. The excerpts are short (ranging from 1 second to 2 minutes), containing dialogue between the characters, sonic display of emotion, or movement sound effects. Between 10 to 20 video excerpts were isolated from each film in order to be used in the analysis. All video clips that are discussed in section 4 are available as supplementary material [10]. For the analysis of audible sonic presence and auditory expression, each video clip was annotated and analysed from spectrograms. The results were also compared to their relation to the physical appearance and action performed by the robot. Key findings from this analysis are presented in section 4.

For the purpose of speech analysis, short video excerpts of robotic speech were isolated. For comparative purposes, video excerpts with the speech of the main human character (same gender) were also isolated. A special case was present for the film *Bicentennial Man*, where the robot, Andrew Martin (played by Robin Williams), transitioned from having a fully mechanised appearance in the beginning of the film into having a human-like appearance towards the end. In this film, we also compared the speech spectra between the robot Andrew and the human Andrew. The sound files were first analysed in Praat [17] to determine the fundamental frequency of the speech using the f0 detection scripts developed by De Looze [18] [11]. The sound files were then analysed using the Long-term Average Spectrum (LTAS) function `iosr.dsp.ltas` from the IoSR MatLab Toolbox [12] in MATLAB. Highlights of the results are presented in section 4.

By comparing speech spectra between characters from the same film, we could ensure the same quality of the sonic feature (since different films most likely will have used different approaches when it comes to sound mastering). For simplicity, the current study only focused on male characters (humans and robots). To be more precise, comparisons were made between a human (typically the main character) versus a robot in the same film.

[8] `https://www.imdb.com/title/tt0470752/`
[9] `https://www.imdb.com/title/tt3778644/`
[10] `https://kth.box.com/v/robotmovies`
[11] `http://celinedelooze.com/Homepage/Resources.html`
[12] `https://github.com/IoSR-Surrey/MatlabToolbox/`

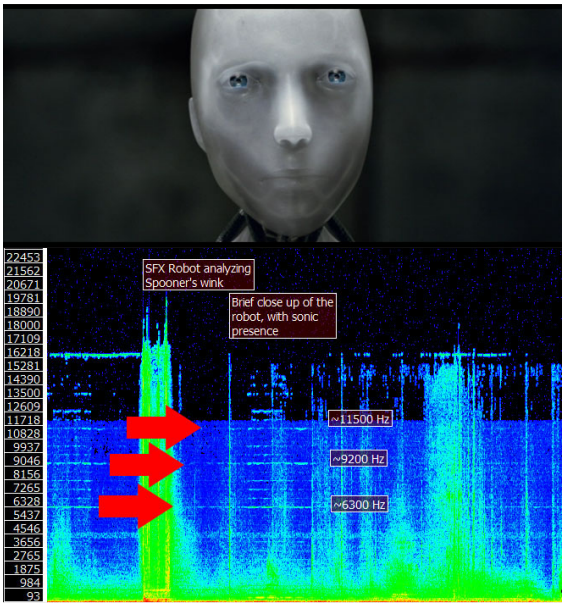

Figure 1. A close up view of the robot Sonny, accompanied by high-frequency tones.

4. RESULTS

An overall observation that we made after watching the films and analysing their sounds is that robot sonic presence is highly related to the physical appearance of the robot itself. Whirring sound of motors are commonly used for mechanical robots such as *Bicentennial Man*'s Andrew Martin and Chappie. These sounds are used to emphasize movements. Some of the movement sounds are also used to emphasize emotions. For example, the sound of Andrew's head movements is used to express sadness. When Andrew's head faces downwards, a mechanical sound characterized by a falling pitch is used. For Chappie, his two ears are used to emphasize his emotion; they go up or down, which is accompanied by a sound effect characterized by a rising or falling pitch.

A different approach is used for the robot Sonny in *I, Robot*. Sonny's futuristic physical appearance is much more flexible than the other robots in the current study, and this appearance is accompanied by more fluid and less mechanical sounds to emphasize his movements. Sonny's presence on the scene can be recognized by high-frequency sounds presumably emitted by his body. This is evident in the interrogation scene; as detective Spooner enters the room, the scene shows a brief close-up of Sonny's face accompanied by three simultaneous high-frequency tones centered at around 6300 Hz, 9200 Hz, and 11500 Hz respectively (see figure 1). In a later scene, where detective Spooner and Dr. Calvin enter a room to talk to Sonny, similar tones are also audible as the two human characters walk toward the robot (see figure 2). The only similar sonic presence observed for the other films in the current data set was for *Bicentennial Man*, in which the robot Andrew breaks his body after falling out of a window. This scene is accompanied by a continuous sound of broken machinery.

Analysis of the sounds of Andrew Martin as a robot ver-

Film	Character	f0-min	f0-max	Key	Range
Bicentennial Man	Andrew Martin (robot)	81	239	112	1.558
	Andrew Martin (human)	60	185	96	1.616
	Richard Martin (human)	70	293	118	2.067
Short Circuit	Number 5 (robot)	110	332	187	1.594
	Newton Crosby	125	439	199	1.808

Table 1. Highlights from speakers' register analysis from Praat: the bottom line and top line (f0-min and f0-max), Key, and Range. The f0-min, f0-max, and key are given in linear scale (Hertz), range in a logarithmic scale (octaves).

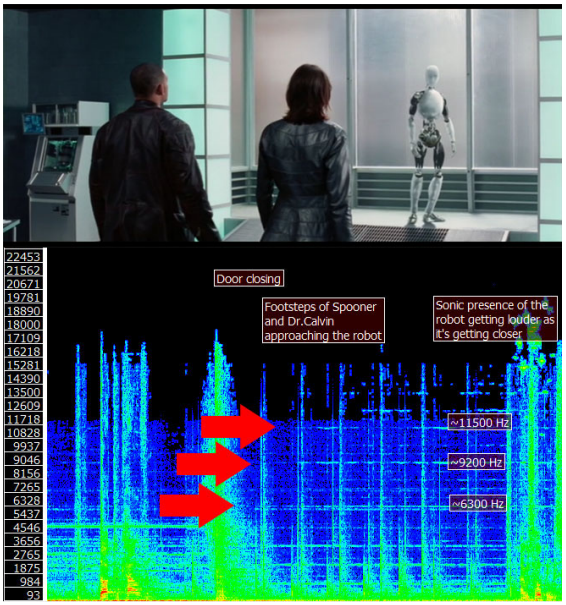

Figure 2. Similar tones are also present in other scene.

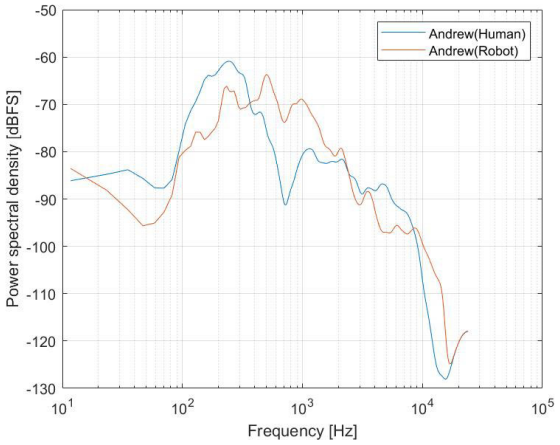

Figure 3. LTAS comparison between the two forms of Andrew Martin, human and robot.

sus a human proved to be particularly interesting, based on the spectral analysis results. The human-like versus robot-like Andrew not only differed in terms of visual appearance, but also in terms of voice. Robot Andrew has a "metallic" quality in his voice, matching the material of his physical form, while the human version of Andrew retains the actor's natural voice characteristics. Table 1 and figure 3 show significant differences between the two voices. Fundamental frequency of robot Andrew has been shifted to higher values. In addition, the robot's voice is characterized by a broader frequency band compared to his human counterpart. In the same film, the voice of the other main human character (Richard Martin) is characterized by a narrower frequency band, compared to the robot version of Andrew (see figure 4). Similarly, in the film *Short Circuit*, the voice of the main human character (Newton Crosby) is also characterized by a narrower frequency band compared to robot Number 5 (see figure 5). The difference between *Short Circuit* and *Bicentennial Man* is that the fundamental frequency of robot character in *Short Circuit* is shifted to lower values.

5. DISCUSSION AND CONCLUSIONS

Of course, one may argue that other auditory features than LTAS might provide interesting information for the characterizations of robot sounds. Other auditory features might be of larger importance in this context, and this will be investigated in future experiments making use of voice sketching [19] for depicting robot actions and intentions. Nevertheless, we have shown that LTAS can be used for characterizing robot sounds and that robots in films are portrayed using broader frequency bands and other formants, compared to humans. Moreover, sound characteristics of the robots appear to vary both with the robot's movements as well as its physical appearance. This information can be used in the design of future sonic renderings of robot movements and of their non-verbal sounds when interacting with humans, in combination with the manipulation of acoustical cues for rendering different emotions as shown in the research field of emotional expression in speech and music [20, 21].

For simplicity, the current study has focused only on films in which the main language was English. One may argue that sound design in HRI should be characterized by intercultural diversity, in the sense that the sounds should be interpreted similarly independently by language of origin of the listener. Still, sound design in popular films creates expectations about how a robot should sound in reality, and robots presented in films are usually associated to a specific country of origin (i.e. Japan, USA, and Germany). Therefore, the selection of films used in the present study can be considered of importance in this context. Nevertheless, as mentioned in the Method section, the number of films on which we base our analysis will be expanded in future

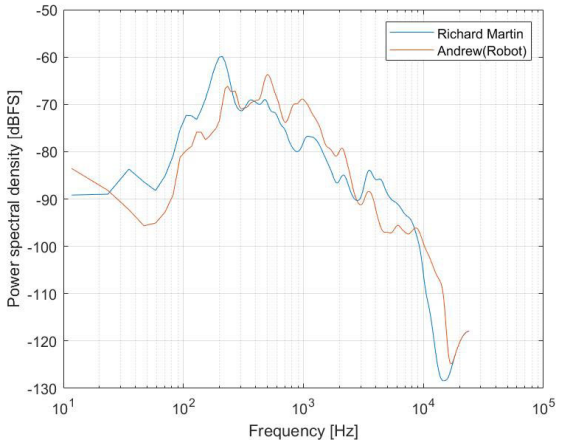

Figure 4. LTAS comparison between Richard Martin and the robot form of Andrew Martin.

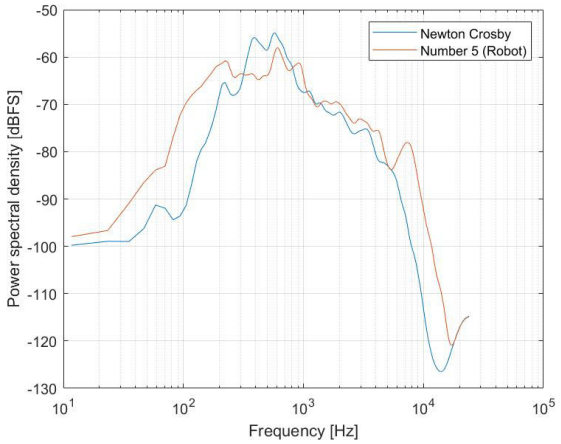

Figure 5. LTAS comparison between Newton Crosby and the robot Number 5.

work. In future data sets, female robot characters as well as robots from different countries will be represented.

Acknowledgments

We thank the three anonymous reviewers for very helpful comments that contributed to improve the quality of our paper. This project was funded by Grant 2017-03979 from the Swedish Research Council and by NordForsks Nordic University Hub "Nordic Sound and Music Computing Network—NordicSMC", project number 86892.

6. REFERENCES

[1] E. Frid, R. Bresin, and S. Alexanderson, "Perception of Mechanical Sounds Inherent to Expressive Gestures of a NAO Robot-Implications for Movement Sonification of Humanoids," in *Proceedings of the Sound and Music Computing Conference*, 2018, pp. 43–51.

[2] J. B. A. Elanjimattathil Vijayan, S. Alexanderson and I. Leite, "Using Constrained Optimization for Real-Time Synchronization of Verbal and Nonverbal Robot Behavior," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 2018, pp. 1955–1961.

[3] J. Monceaux, J. Becker, C. Boudier, and A. Mazel, "First Steps in Emotional Expression of the Humanoid Robot NAO," in *Proceedings of the 2009 International Conference on Multimodal Interfaces*. ACM, 2009, pp. 235–236.

[4] M. Häring, N. Bee, and E. André, "Creation and Evaluation of Emotion Expression with Body Movement, Sound and Eye Color for Humanoid Robots," in *RoMan, 2011 IEEE*. IEEE, 2011, pp. 204–209.

[5] A. Kappas, D. Küster, P. Dente, and C. Basedow, "Simply the BEST! Creation and Validation of the Bremen Emotional Sounds Toolkit," in *International Convention of Psychological Science*, 2015.

[6] R. Zhang, M. Jeon, C. H. Park, and A. Howard, "Robotic Sonification for Promoting Emotional and Social Interactions of Children with ASD," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, 2015, pp. 111–112.

[7] J. Bellona, L. Bai, L. Dahl, and A. LaViers, "Empirically Informed Sound Synthesis Application for Enhancing the Perception of Expressive Robotic Movement," in *Proceedings of the International Conference on Auditory Display*. Georgia Institute of Technology, 2017, pp. 73–80.

[8] D. Hug, "CLTKTY? CLACK! Exploring design and interpretation of sound for interactive commodities," Ph.D. dissertation, University of Linz, 2017.

[9] N. Shedroff and C. Noessel, "Make it so: Learning from sci-fi interfaces," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '12. New York, NY, USA: ACM, 2012, pp. 7–8.

[10] E.-S. Jee, Y.-J. Jeong, C. H. Kim, and H. Kobayashi, "Sound Design for Emotion and Intention Expression of Socially Interactive Robots," *Intelligent Service Robotics*, vol. 3, no. 3, pp. 199–206, 2010.

[11] S. Yilmazyildiz, R. Read, T. Belpeame, and W. Verhelst, "Review of Semantic-Free Utterances in Social Human-Robot Interaction," *International Journal of Human-Computer Interaction*, vol. 32, no. 1, pp. 63–85, 2016.

[12] R. Read and T. Belpaeme, "People interpret robotic non-linguistic utterances categorically," *International Journal of Social Robotics*, vol. 8, no. 1, pp. 31–50, 2016.

[13] L. Langeveld, R. van Egmond, R. Jansen, and E. Ozcan, "Product sound design: Intentional and consequential sounds," in *Advances in industrial design engineering*. InTech, 2013, p. 47.

[14] D. Moore, H. Tennent, N. Martelaro, and W. Ju, "Making noise intentional: A study of servo sound perception," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 12–21.

[15] C. L. Bethel and R. R. Murphy, "Auditory and Other Non-Verbal Expressions of Affect for Robots," in *2006 AAAI Fall Symposium Series, Aurally Informed Performance: Integrating Machine Listening and Auditory Presentation in Robotic Systems, Washington, DC*, 2006, pp. 1–5.

[16] B. Burtt, *Star Wars Galactic Phrase book & Travel Guide: Part II - Behind the Sounds*. Del Rey, 2001.

[17] P. Boersma and D. Weenink. (2019) Praat: Doing Phonetics by Computer [Computer Program]. [Online]. Available: http://www.praat.org/

[18] C. De Looze and D. Hirst, "Detecting changes in key and range for the automatic modelling and coding of intonation," in *Speech Prosody*, 2008, pp. 135–138.

[19] S. Delle Monache, D. Rocchesso, F. Bevilacqua, G. Lemaitre, S. Baldan, and A. Cera, "Embodied sound design," *International Journal of Human-Computer Studies*, vol. 118, pp. 47 – 59, 2018.

[20] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological bulletin*, vol. 129, no. 5, p. 770, 2003.

[21] R. Bresin and A. Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.