Leaping Across Modalities: Speed Regulation Messages in Audio and Tactile Domains

Kai Tuuri[1], Tuomas Eerola[2], and Antti Pirhonen[1]

[1] Department of Computer Science and Information Systems
[2] Department of Music
FI-40014 University of Jyväskylä, Finland
{krtuuri,ptee,pianta}@jyu.fi

Abstract. This study examines three design bases for speed regulation messages by testing their ability to function across modalities. Two of the design bases utilise a method originally intended for sound design and the third uses a method meant for tactile feedback. According to the experimental results, all designs communicate the intended meanings similarly in audio and tactile domains. It was also found that melodic (frequency changes) and rhythmic (segmentation) features of stimuli function differently for each type of message.

Key words: audio, tactile, crossmodal interactions, crossmodal design

1 Introduction

When designing applications for mobile or ubiquitous contexts, the means to present information redundantly in audio and tactile domains can provide valuable flexibility. Since the actual context of use is hard to anticipate, it is important that there are options for interacting with the application. Users can also have different preferences about either hearing feedback in audio or feeling it on their skin in a more intimate manner. This is the usual rationale for *crossmodal design approach* [1].

Physical training applications, used in a varying contexts, would benefit of crossmodal design. They use the information from, e.g., heartbeat or Global Positioning System (GPS) sensors, and provide feedback for the user for controlling his or her training performance accordingly. This study focuses on messages that relate to the regulation of running speed. Sounds and vibrations, as gaze-independent presentation modes, are well suited for such an interaction.

As a starting point, this study picks up two existing methods for designing messages for the regulation of running speed: one for sounds [2], and one for vibrotactile feedback [3]. Both methods aim at intuitive interaction, so that sounds and vibrations would communicate their messages as effortlessly as possible – with a minimum requirement for learning. This study examines empirically how these design methods can be applied for crossmodal design, i.e., utilising them in creating both non-speech sounds and vibrotactile stimuli.

This is a preprint of an article published in In Nordahl, R., Serafin, S., Fontana, F. & Brewster, S. (Eds.), Proceedings of 5th International Workshop HAID 2010, Springer, Heidelberg, pp. 10–19, 2010.

1.1 The Quest for Amodality

The concept of crossmodal design is based on an assumed existence of *amodal* content, which can be presented more or less interchangeably in different sensory domains. Communication of such content would then mean the articulation of certain amodal attributes in terms of the chosen presentation modality.

Traditional cognitivist view [4] has seen amodality as information processing on an abstract level, operating independently from modalities, well above sensory domains. But in the light of recent neurostudies, it seems that the thing we call amodality actually refers to close and early interconnections between widely integrated sensory-motor aspects of perception and the very roots of conceptual thinking [5]. According to the embodied perspective to human cognition [6, 7], understanding and thinking indeed are modality dependent, essentially bound with the human body and all of its modalities of interaction. Even the concepts of language are thus strongly dependent on senses and bodily experiencing of the physical world, and on how these experiences are schematised [8, 6, 5].

We must stress that when amodal attributes are referred to in this study, we are talking about certain mental imagery relating naturally to multiple modalities rather than being modality independent (see also [9]). For example, when thinking of an amodal concept of "roughness", one can easily express the mental image with hand gestures, or describe physical attributes that relate to different sensory domains: seeing, hearing and feeling a rough surface as an action-related perception.

According to *image schema theory* [8, 6], "image-like" schematic structures multimodally capture the contours of recurrent sensory-motor experiences of interacting with the environment. They simultaneously act as pre-conceptual, directly meaningful *gestalt* structures of perception, thinking and acting. Despite being called "image", such gestalts have a kinaesthetic character which integrates sensory-motor information. Therefore, in principle, they should be key points of crossmodal design, when trying to find certain sounds and vibrations that evoke similar associations as gestalt completions. Basic image schemas refer to experiences of, e.g., spatial motions or relations, forces and object interactions. In thinking, image schemas are often projected metaphorically.

1.2 Communicative Functions and Design Methods

Within the context of physical training, speed regulation feedback has three main communicative functions: to tell the runner 1) to decrease the pace (*Slow*), 2) to increase the pace (*Urge*) or 3) to keep the current pace (*Ok*). In terms of communication, Slow and Urge functions are directive, as they try to get the runner to undertake a change in speed. The Ok function, in contrast, primarily approves the current state of speed.

For these functions, a *prosody-based* (PB) method for designing non-speech sounds has been proposed [2]. This method aims to utilise "speech melodies" (i.e., intonation) of short vocal expressions which are spontaneously produced for each communicative function. It has previously been found that independently of

verbal content, humans rely on intention-specific forms of intonation in communication, especially when interacting with infants [10]. Studies on prosody-based sound design [2, 11] have found intonation patterns specific to Slow, Urge and Ok functions (along with an additional Reward function), which can be utilised as "musical" parameters in design.

The other design approach is a method which simply utilises the *direct analogy* (DA) between changes in frequency and the corresponding messages of "decelerate", "accelerate" and "keep it constant" [3]. It has been utilised in the design of vibrotactile stimuli, in which the vibration frequency decelerates for the Slow function, accelerates for the Urge function and keeps unchanged for the Ok function.

At the physical level of implementation, both methods basically concern simple frequency-related features along the temporal dimension. In this study, we test their crossmodal functionality by using the physical vibrations as stimuli for different sensory domains.

1.3 Research Questions

1. *How effectively do the different designs serve the intended communicative functions?* Both design methods (PB and DA) have already proven their usefulness in their original domains [12, 3]. Although not being the main focus of this study, it is interesting to see how the designs based on vocal gestures of human expressions compare to the "mechanically" straightforward direct analogy designs.

2. *How does communication vary across the audio and tactile domains?* Both design methods are based on principles which suppose the attribution of certain amodal meanings to physical cues presented in a temporal continuum. In the PB method, such meanings refer to kinaesthetic imagery of projected gestural forms which reflect "bodily affect" in vocalisation [13]. In the DA method, amodal meanings also refer to kinaesthetic imagery such as "decelerating" or "falling". We hypothesise that the designs would work across the domains; i.e, stimuli would still crossmodally resonate with similar "embodied gestalts" and evoke the corresponding spatio-motor mental imagery. We are also interested to see how crossmodal attributions vary across design bases and functions.

3. *What is the role of melodic and rhythmic factors (i.e., frequency changes and segmentation) in communicating the intended functions?* We want to explore how important are the roles these features serve in communication, and whether these roles are weighed differently across functions, domains and different design bases. We especially want to see what kind of effect the melodic factors have within tactile domain, which is not commonly thought of as being in compliance with "melody".

2 Method

2.1 Apparatus

Two Engineering Acoustic C-2 vibrotactile actuators (http://www.eaiinfo.com/) were used for tactile presentation and high-quality active speakers were used for

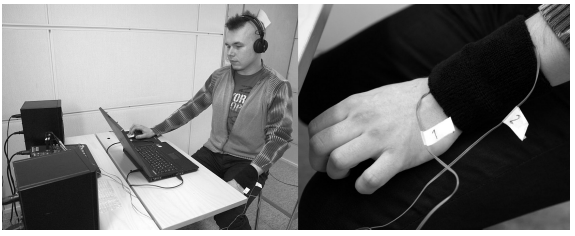

Fig. 1. Experimental setting with a participant performing the tactile task.

audio. As actuators are driven by audio signal (usually in sine waves), it was possible to use the same sound file as a source for both the audio and tactile stimulus. Optimal signal levels were set separately for both domains. To enhance tactile sensation, two actuators were used concurrently, both attached under a wristband in the backside of a left wrist (see Figure 1).

2.2 Stimuli

Three different design bases were prepared, two of them utilising the PB method. The first design base (*PB1*) consists of the same intonation contours for Slow, Urge and Ok functions that were used in the previous evaluation study [12]. These contours were "designerly" chosen from the bulk of 40 source utterances for each function. In contrast, the second design base (*PB2*) uses intonation contours chosen by a statistical classification model based on function-specific prosodic characteristics [11]. The third design base of this experiment represents the DA principle. From the previous study [3], we chose the stimuli designs of 1750 ms duration as they were highly rated and are also better in accordance with the stimuli durations of other design bases.

The stimuli of the DA base were already optimised for C-2 actuators [3], but all source contours of the PB bases were preprocessed to conform with the tactile presentation technology. Intonation contours were first centered to 250 Hz within each source utterer to remove the pitch differences caused, e.g., by the utterer's gender. This center frequency is the recommended operating frequency of C-2. A pilot testing revealed that the original pitch ranges of contours were too ample for the actuator's optimal range. Also, in terms of temporal sensitivity of touch, the contours felt too quick. Therefore the pitch ranges were scaled down by a factor of 0.75 and the contour lengths were scaled up by a factor of 1.25. Excess fluctuations in pitch were finally smoothed out. All modifications were subtle enough to retain the original characteristics of the contours for audio domain.

For each function within each design base, three different versions of stimuli were prepared: one with frequency changes and segmentation (*FC+Seg*) and other ones without any frequency changes (*NoFC*) or segmentation (*NoSeg*). All FC+Seg versions are illustrated in Figure 2. The segmentation in the PB bases is derived from the original utterances. As the DA pitch contours originally had no segmentation, it was implemented by inserting short gaps of silence to

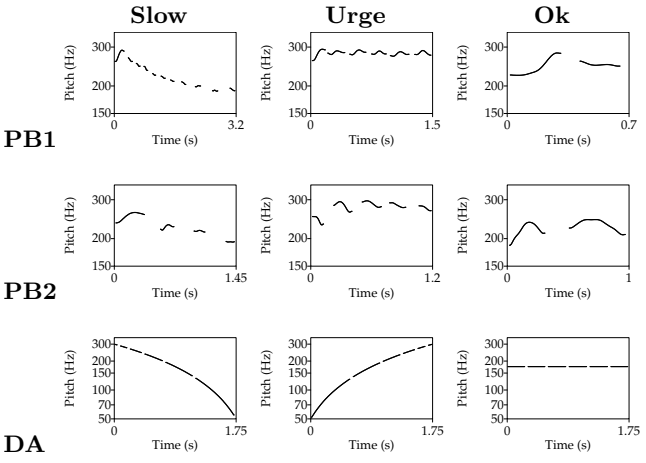

Fig. 2. Visualisations of stimuli containing frequency changes and segmentation.

the FC+Seg and NoFC contours in accordance with the DA principle: for Slow function the onset time intervals of consecutive segments decelerate, for Urge they accelerate and for Ok they remain even. All NoFC versions retain the segmentation but their pitch is flattened to the mean frequency (in Hz) of the contour. Within the DA base, the NoFC versions have the same pitch level (175 Hz) for all functions, but in other cases flattened pitch levels varied across functions. For prosody based NoSeg versions, gaps in the contour were filled by using interpolation. Short rising and falling ramps of intensity envelopes were added to all segment onsets and offsets respectively, to prevent audible "pops".

All 27 stimuli (3 design bases × 3 functions × 3 versions) were finally synthesised as sine waves. Intensity levels were the same for all. Despite the optimisations for tactile domain, they all are also in a comfortable hearing range. The preparation of stimuli was made with Praat software (http://praat.org/).

2.3 Participants and procedure

Twenty-two students of our university took part in the experiment. They were from different departments, representing many different major subjects. Of the participants, 8 were male and 14 were female. The average age in the group was 25.5 years. All participants reportedly had normal hearing and sense of touch.

Each participant performed rating tasks for both the audio domain and the tactile domain. Half of the participants did the tasks in reverse order, in order to counterbalance the learning effect. In both tasks, all stimuli were presented twice. The participants rated the amount of each of the three functions that were represented by the 54 stimuli (2 × 27). The ratings were carried out in a random order and using a five-level Likert scale (0-4). In order to block the audible sounds produced by the actuators during the tactile task, the participants wore closed earphones through which white noise was played at a comfortable level.

Before the first task, instructions were read aloud to each participant. Participants were encouraged to rely on their intuition and give ratings relatively quickly. Three novel training stimuli were presented before each task to help the participants to adapt themselves to the type of stimuli. After both tasks, they were asked to freely comment on the experiment.

3 Results

3.1 Effectiveness of Design Bases

For each rating scale (Slow, Urge, and Ok), a separate two-way within-subjects ANOVA was conducted with the within-subjects factors being function (3 levels, Slow, Urge, Ok) and design (3 design bases: PB1, PB2, DA) and the dependent variable being the ratings across the two domains, repetitions and pitch and segmentation variants. The means for all three ratings across the two variables are displayed in Figure 3. It is noticeable that within the ratings for each function, the mean ratings for correct target function were clearly the highest ones.

For the ratings of Slow, ANOVA yielded a significant effect of function, $F(2,42) = 83.05$, $p<.001$ and design, $F(2,42) = 19.44$, $p<.001$, as well as a significant interaction between the two, $F(4,84) = 14.03$, $p<.001$. The ratings of Slow for the correct function (Slow) were clearly separated from the other two functions. The design bases worked significantly differently from each other, PB1 producing the highest overall ratings, followed by the DA base.

The ratings of Urge showed a significant effect of function, $F(2,42) = 118.9$, $p<.001$, but not design, $F(2,42) = 1.13$, $p=.33$. However, a small but significant interaction between the two exists, $F(4,84) = 3.86$, $p<.01$. In other words, overall, the design bases produced equal results, but within the correct communicative function the DA base conveyed the intended meanings more effectively.

Finally, an analysis of the ratings of Ok, ANOVA indicated a significant effect of function, $F(2,42) = 81.70$, $p<.001$, and design, $F(2,42) = 12.47$, $p<.001$, as well as a significant interaction between the two, $F(4,84) = 6.99$, $p<.001$. The ratings for the correct function were distinct from the other target functions (Slow and Urge), and the direct analogy (DA) produced the highest overall ratings. The prosody-based designs were not statistically significantly different from each other.

3.2 Domain-Related Differences

To explore the effect of domain on the ratings, separate three-way ANOVAs were conducted for each rating scale. This time the within-subjects factors consisted of domain (2: audio and tactile), design (3 design bases: PB1, PB2, DA), and function (3 communicative functions). ANOVAs yielded a non-significant effect of domain, $F(1,21) = 0.07, 2.38, 0.05$, respectively for the Slow, Urge and Ok ratings (all $p>0.20$). However, there were few interactions between the domain and the other two factors, especially in the ratings of Slow. The interactions between

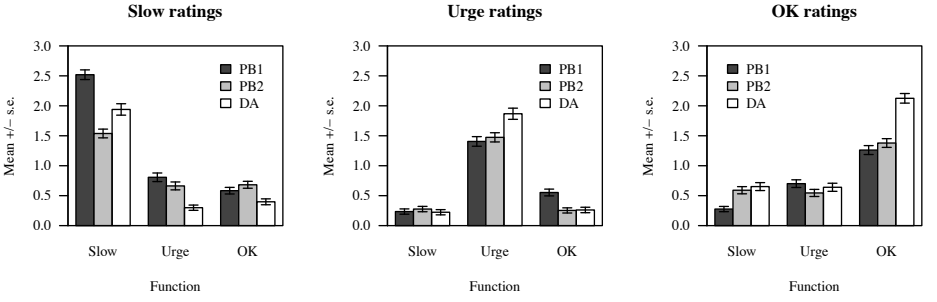

Fig. 3. Mean ratings of Slow, Urge and Ok across design bases and communicative functions.

Table 1. Recognition rates across domains, functions and design bases.

Domain	*Audio*			*Tactile*		
Design	PB1	PB2	DA	PB1	PB2	DA
Slow	0.82	0.64	0.75	0.82	0.64	0.55
Urge	0.47	0.50	0.68	0.46	0.52	0.55
Ok	0.64	0.61	0.81	0.47	0.60	0.80

domain and function were significant, $F(2,42) = 6.82$, $p<.01$, and the interactions between domain and design were also significant, $F(2,42)=10.98$, $p<.001$. In Urge ratings, the domain and design interaction was significant, $F(2,42) = 4.39$, $p<.05$. The sources of these interactions seem to relate to better functionality of the direct analogy (DA) for Slow and Urge within the audio domain (see Table 1). In all, differences due to the domain were surprisingly small.

To illustrate the differences between the domains and other factors, the ratings of the three communicative functions were converted into recognition rates. In this, the highest rating across the three rating scales was compared with the corrected intended function for each example. If the highest rating and the target function matched, the item received a value of 1 (correct) and mismatching items received a value 0 (incorrect). This individual classification was aggregated across participants, domains, functions and design bases, and the mean recognition accuracy is shown in Table 1. These numbers illustrate the sparsity of the domain effect in recognising the functions. The overall recognition rate was somewhat better with the audio (66%) than with the tactile stimulation (60%), and this difference was statistically significant with Kruskal-Wallis test, $\chi^2 = 8.83, p < .01$. The fact that the DA design base, in particular, seemed to work better in the audio domain was surprising.

3.3 Effects of Melodic and Rhythmic Manipulations

The roles of frequency changes and segmentation were investigated with a series of three-way ANOVAs. Frequency change factor (two levels: FC and NoFC), segmentation factor (two levels: Seg and NoSeg) and communicative function

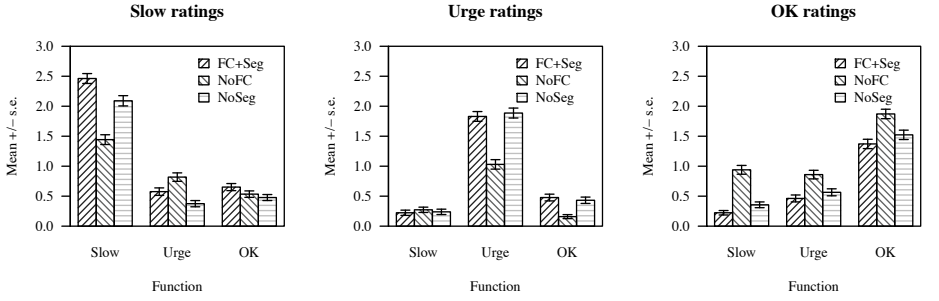

Fig. 4. Mean ratings of Slow, Urge and Ok across pitch and rhythm manipulations and communicative functions.

were the within-subject factors for the three ratings given by the participants. To constrain the number of interactions, only frequency change × function and segmentation × function were tested in this analysis since the previous analyses contain most of the other possible interactions.

In Slow, both the frequency change and segmentation factors achieved statistical significance, $F(1,21) = 4.55$, and 14.44 $p<.05$ and .001, respectively. Also the interactions between the function and the frequency change and segmentation factors were statistically highly significant, $F(2,42) = 74.2$ and 47,9, both $p<.001$. These interactions are interestingly demonstrated in Figure 4: non-segmented versions seem to convey the intended meaning of the message more effectively than the segmented version with no frequency changes. Therefore, leaving out the pitch information would harm the communicative function more than leaving out the segmentation. The best rated stimuli for Slow function, in all design bases and in both domains, were the versions that contained both features (FC+Seg).

The Urge ratings were different across frequency changes, $F(1,21) = 28.8$, $p<.001$, but not across segmentation despite the interactions between function and frequency changes, $F(2,42) = 21.7$, $p<.001$, and function and segmentation, $F(4,84) = 97.4$, $p<.001$. As can be observed in Figure 4, the frequency changes seems to be the most effective feature for the Urge function, and leaving out the segmentation does not seem to harm the communication. For the Urge function, the best rated stimuli in the DA and PB1 bases indeed were the ones with no segmentation (NoSeg), while in PB2 it contained both features (FC+Seg).

In the Ok ratings, only a significant main effect of frequency changes was observed, $F(1,21) = 21.9$, $p<.001$, although an interaction effect between function and segmentation was also observed, $F(4,84)=65.1$, $p<.001$. The interpretation, evident from the Figure 4 as well, points out that the apparently best communication of this function was without frequency changes (NoFC). The best rated stimuli for the Ok function also lacked frequency changes within all designs and in both domains. It must be noted that the DA principle for the Ok function did not permit frequency changes (FC+Seg and NoFC were identical). The lack of this feature might partly explain the superior success of the DA base for the

Ok function, illustrated in Figure 3 and Table 1. Figure 4 also shows that NoFC stimuli for other functions were given relatively high ratings in the Ok scale.

With two-way ANOVAs, we finally tested if the domain factor had any interactions with either the frequency change factor or the segmentation factor. The only statistically significant interaction was found between domain and frequency changes in the ratings of Urge, $F(3,63) = 17.2$, $p<.001$. In other words, the melodic and rhythmic features generally functioned similarly for each function, regardless of domain. The tactile domain thus did not have any apparent handicap with respect to the usage of melodic features.

4 Conclusions and Discussion

All design bases performed well in terms of communicating the intended meaning, bearing in mind that the ratings were given on the basis of intuitive associations rather than any learnt or accustomed coding. Due to its straightforward nature, the DA design base generally seemed to function best. However, both PB design bases functioned effectively as well, especially PB1 which scored the best ratings in communicating the Slow function. The affect-based character of PB designs was evident in the spontaneous expressions of some of the participants, stating that certain stimuli "...just felt like someone were telling you to slow down", for example.

When compared with the previous studies concerning DA and PB1 designs [12, 3], the new results accords with some earlier findings. For example, the NoSeg DA version for Urge (88% recognition) and the FC+Seg PB1 version for Slow (89% recognition) performed especially well. In the previous evaluation of the PB1 design base [12], some participants interpreted the "agitating" imagery associated with Urge samples as warning against going too fast. A similar recognition ambiguity was found in this experiment as well, weakening the ratings for the Urge function. One participant pondered this issue spontaneously: "...it felt like rushing, but it was similar to the warnings in heart-rate meters".

Although there are similarities in the function-specific features between the DA and PB designs, they also differ in many aspects. This indicates that the coupling between the features and the related attributions is not exclusionary. Thus, it should be possible to combine the features relating to the same function. For example, the ascending pitch could be applied to PB Urge designs to potentially reduce the ambiguity in interpretation. Similarly, DA designs could benefit from affect-related features of PB designs.

The most important finding of this study is that domains indeed seem to function in an interchangeable manner, thus supporting the hypothesis. This finding suggests that, regardless of the original usage of any design principle or presentation feature, it might be worth exploring their applicability across modality domains. Many of the participants expressed that "...understanding was easy to 'catch' in both domains", and that "...both domains felt comprehensive" or "...in tactile domain, I played the rhythm in my mind". The audio domain, however, was preferred by the majority of the participants.

In the experiment, the same stimuli were used directly in both domains. This might not be the optimal usage for real-life designs. Of course, we would recommend better utilisation of the domain-related strengths and restrictions: for instance, using the most suitable pitch register and timbre for audio. When audio and tactile stimuli are presented concurrently, the "fused" perception (i.e., *synchresis* [14]) can be something different from the sum of its "parts". Therefore we also recommend creative uses of crossmodal attributes, which would not only be justified as a modality option but also as a multimodal enrichment in supporting the contextually appropriate perception.

Acknowledgments. This work is funded by Finnish Funding Agency for Technology and Innovation, and the following partners: GE Healthcare Finland Ltd., Suunto Ltd., Sandvik Mining and Construction Ltd. and Bronto Skylift Ltd.

References

1. Hoggan, E. & Brewster, S.: Designing audio and tactile crossmodal icons for mobile devices. In Proc. of the 9th International Conference on Multimodal Interfaces. NY: ACM, 162–169 (2007)
2. Tuuri, K. & Eerola, T.: Could function-specific prosodic cues be used as a basis for non-speech user interface sound design? In Proc. of ICAD 2008, Paris: IRCAM (2008)
3. Lylykangas, J., Surakka, V., Rantala, J., Raisamo, J., Raisamo, R. & Tuulari, E.: Vibrotactile Information for Intuitive Speed Regulation. In Proc. of HCI 2009. 112–119 (2009)
4. Fodor, J. A.: The language of thought. Cambridge, MA: Harvard University Press (1975)
5. Gallese, V. & Lakoff, G.: The brain's concepts: The role of the sensory-motor system in reason and language. Cognitive Neuropsychology, 22, 455–479 (2005)
6. Johnson, M. & Rohrer, T.: We are live creatures: Embodiment, American pragmatism and the cognitive organism. In: J. Zlatev, T. Ziemke, R. Frank, & R. Dirven (Eds.), Body, language, and mind, vol. 1. Berlin: Mouton de Gruyter, 17–54 (2007)
7. Leman, M.: Embodied Music Cognition and Mediation Technology. Cambridge, MA: MIT Press (2008)
8. Johnson, M.: The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. Chicago, IL: University of Chicago (1987)
9. Pirhonen, A. & Tuuri, K.: In Search for an Integrated Design Basis for Audio and Haptics. In Proc. of HAID 2008, LNCS 5270. Springer-Verlag, 81–90 (2008)
10. Fernald, A.: Intonation and communicative intent in mothers' speech to infants: Is the melody the message? Child development, 1497–1510 (1989)
11. Tuuri, K. & Eerola, T.: Identifying function-specific prosodic cues for non-speech user interface sound design. In Proc. of the 11th International Conference on Digital Audio Effects, 185–188 (2008)
12. Tuuri, K., Eerola, T. & Pirhonen, A.: Design and Evaluation of Prosody Based Non-Speech Audio Feedback for Physical Training Application. (journal submission)
13. Tuuri, K.: Gestural attributions as semantics in user interface sound design. In: Kopp, S., Wachsmuth, I. (Eds.), Gesture in Embodied Communication and Human-Computer Interaction, LNAI 5934. Springer-Verlag, 257–268 (2010)
14. Chion, M.: Audio-vision: Sound on screen. NY: Columbia University Press (1990)