

An Audio-Haptic Aesthetic Framework Influenced by Visual Theory

Angela Chang¹ and Conor O'Sullivan²

¹20 Ames St. Cambridge, MA 02139, USA
anjchang@media.mit.edu

²600 North US Highway 45, DS-175, Libertyville, IL 60048, USA
conor.o'sullivan@motorola.com

Abstract. Sound is touch at a distance. The vibration of pressure waves in the air creates sounds that our ears hear, at close range, these pressure waves may also be felt as vibration. This audio-haptic relationship has potential for enriching interaction in human-computer interfaces. How can interface designers manipulate attention using audio-haptic media? We propose a theoretical perceptual framework for design of audio-haptic media, influenced by aesthetic frameworks in visual theory and audio design. The aesthetic issues of the multimodal interplay between audio and haptic modalities are presented, with discussion based on anecdotes from multimedia artists. We use the aesthetic theory to develop four design mechanisms for transition between audio and haptic channels: *synchronization*, *temporal linearization*, *masking* and *synchresis*. An example composition using these mechanisms, and the multisensory design intent, is discussed by the designers.

Keywords: Audio-haptic, multimodal design, aesthetics, musical expressivity, mobile, interaction, synchronization, linearization, masking, synchresis.

1 Introduction

We live in a world rich with vibrotactile information. The air around us vibrates, seemingly imperceptibly, all the time. We rarely notice the wind moving against our bodies, the texture of clothes, the reverberation of space inside a church. When we sit around a conference table, our hands receive and transmit vibrations to emphasize what is being said or attract attention to the movements of other participants. These sensations are felt by our skin, a background symphony of subtle information that enriches our perception of the world around us.

In contrast, products like the LG Prada phone [22] and the Apple iPhone [1] provide little tactile feedback (figure 1). Users mainly interact with a large touchscreen, where tactile cues are minimal and buttons are relegated to the edges. This lack of tactile feedback causes errors in text entry and navigation [29]. In order to give more feedback, audio cues are often used to confirm tactile events and focus the user's attention [8], e.g. confirmation beeps. However, these audio cues are annoying and attract unwanted attention [16]. Many HCI researchers are now researching how haptics (physical and tactile) can provide a subtler feedback channel [5,13, 27, 28, 29].



Fig. 1. (a) LG Prada phone and (b)Apple iPhone are touchscreen based mobile devices

The use of haptics (particularly vibration) is promising, because it is relatively cost effective and easy to implement [5]. The benefits to a vibrotactile interface, mainly privacy and subtlety, are not new [3,5]. Yet, there is relatively little knowledge on how to aesthetically structure and compose vibrations for interfaces [4,14]. This work addresses the creation of multimodal experiences by detailing our experiences in developing haptic ringtones in mobile phones, through describing an example of audio-haptic stimuli. We share our knowledge and expertise on how to combine audio-haptic information to create pleasing and entertaining multimodal experiences.

2 Background and Motivation

Prior work in HCI has explored mapping vibration to information through the use of scientifically generated media [10,25]. Some work has focused on developing haptic hardware and identifying situations where tactile information can be used, e.g., navigating spatial data [12], multimodal art [15], browsing visual information, and of course, silent alerts [4,6,13,17]. We note the majority of creation techniques for vibrotactile stimuli have been largely designated by device capabilities. O'Modhrain[18] and Gunther [11] have presented works on composing vibrations in multimedia settings, using custom hardware. A key issue has been how to map information to vibration to avoid overload [17]. This work seeks to extend the prior art by suggesting a general aesthetic framework to guide audio-haptic composition.

One inspiration has been the Munsell color wheel [16]. The Munsell color wheel is a tool for understanding how visual effects can be combined. The use of the color wheel gives rise to color theory, and helps graphic designers understand how to create moods, draw attention, and attain aesthetic balance (and avoid overload). We wondered if a similar framework could help vibrotactile designers compose their audio-haptic effects so that there is stimulating, but not overwhelming transition between the audio and haptic modalities to create a complex, but unified multisensory experience.

Audio-visual theories for cinematography has also influenced this work, particularly, cinematic composer Michel Chion's theories about the relation of audio to vision [7]. *Synchronization* (when audio happens at the same time as visual event), *temporal linearization* (using audio to create a sense of time for visual effects), *masking* (using audio to hide or draw attention away from visual information) and the *synchresis* (use of sound and vision for suggesting or giving illusion to add value onto the moviegoing experience) aroused our interest. The idea behind audiovisual composition is *balance* and *understanding the differences between audio and visual through perceptual studies*.



Fig. 2. Two Audio-haptic displays using embedded (a) MFTs made by Citizen (CMS-16A-07), used in a (b) vibrotactile “puff”, and (c) the Motorola A1000

2.1 Audio-Haptic Media Design Process

The easiest approach to designing vibrotactile interfaces is to use low power pager motors and piezo buzzers. Multifunction transducers (MFTs) enable the development of mobile systems that convey an audio-haptic expressive range of vibration [20], figure 2a. An MFT-based system outputs vibrations with audio much like audio speakers.

Instead of focusing strictly on vibration frequencies (20Hz-300Hz) [26], we recognize that audio stimuli can take advantage of the overlap between vibration and audio (20Hz-20kHz), resulting in a continuum of sensation from haptic to audio called the *audio-haptic spectrum*. In the past, audio speakers were often embedded into the computer system, out of the accessible reach of the user. Mobile devices have allowed speakers to be handheld. By using MFTs instead of regular speakers, the whole spectrum of audio-haptics can be exploited for interactive feedback.

Two example devices were used in our process for exploring the audio-haptic space (figures 2b and 2c). One is a small squishy puff consisting of one MFT embedded in a circular sponge (2b). Another is the Motorola A1000 phone which uses two MFTs behind the touchscreen. Audio-haptic stimuli are automatically generated by playing sounds that contain haptic components (frequencies below 300Hz). If necessary, the haptic component could be amplified using haptic inheritance techniques. In our design process, we used commercially available sound libraries [21,24]. Audio-haptic compositions were made using Adobe Audition [2]. A user holding either the squishy puff or the A1000 phone would be able to feel the vibrations and hear audio at the same time.

2.2 Designing a Visual-Vibrotactile Framework

The Munsell color wheel (figure 3) describes three elements of visual design. One principle element of visual design is the dimension of “warm” or “cool” hues. Warm colors *draw more attention* than “cool colors”. In color theory, warm colors tend toward the red-orange scale, while cool colors tend towards the blue purplish scale. The warm colors are on the opposite side of the color wheel from cool colors. Another component of color theory is value, or the lightness amplitude in relation to neutral. The darker the color is, the lower its value. The final dimension is chroma, or the saturation of the color. The chroma is related to the radial distance from the center of the color wheel.

We wondered whether prior classifications of vibrotactile stimuli could provide a similar framework [4,14,19]. Scientific parameters such as frequency, duration and amplitude (e.g. 100 Hz sine wave for 0.5 seconds) have traditionally been used to describe vibrations in perception studies. Perceiving vibrations from scientifically

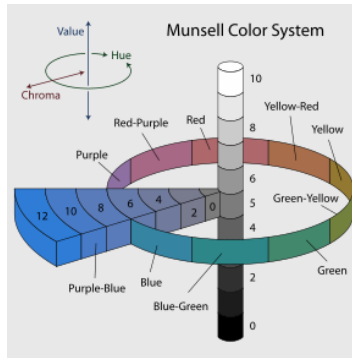


Fig. 3. Munsell Color System showing Hue, Value, and Chroma¹

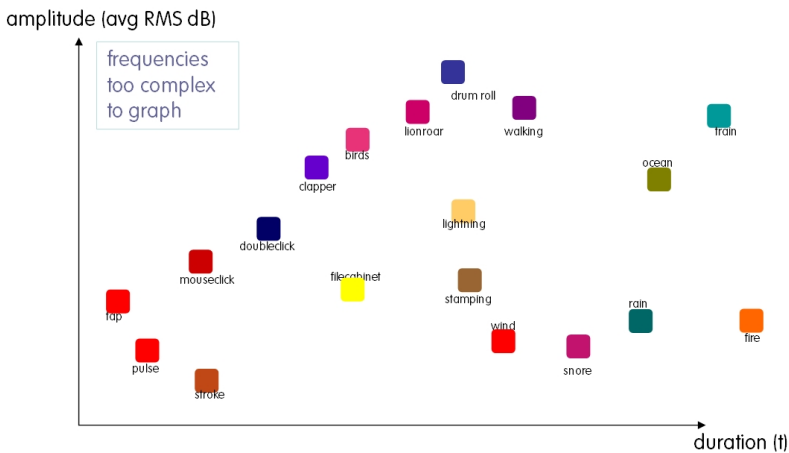


Fig. 4. Audio-haptic stimuli plot based on amplitude and vibration (scatter plot)

generated stimuli has some flaws, particularly since they are detached from everyday experience. These synthetic vibrations are unrelated to the experience of human interaction with objects. Users often have to overcome a novelty effect to learn the mappings. In contrast, normal everyday interactions with our environment and objects result in vibrations and sound. In this inquiry, we have elected to select stimuli based on sound stimuli from commercially available sound libraries [21, 24].

As a starting point, approximately 75 audio-haptic sounds were selected based on their audio-haptic experience. When laid out on a grid consisting of frequency, duration and amplitude, it was hard to organize these complex sounds based on frequency. Graphing the duration and amplitude produced a scatterplot, and did not suggest any aesthetic trends (figure 4 shows a plot with less sounds than our actual plot).

¹ Munsell Color System, <http://en.wikipedia.org/wiki/Munsellcolorsystem> on June 23, 2008.

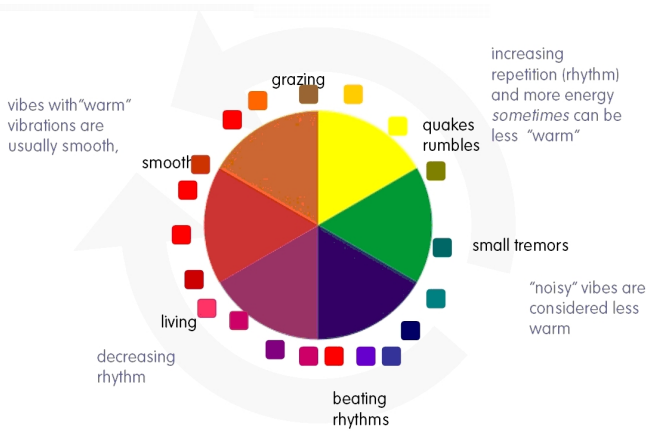


Fig. 5. Activity Classification for audio-haptics shows a number of problems: conflicting warmth trends, multiple classifications for similar sounds

Another prior framework for classifying audio-haptic media is activity classification [20], which characterizes audio-haptic media by context of the activity that may generate the stimuli. While the activity classification system is a good guide for users to describe the qualitative experience of the stimuli, it is hard for designers to use as a reference for composition. The main problem with this mapping was that the stimuli could belong to more than one category. For example, “fire” or “wind” sound could belong to both surface and living categories. The same stimuli could be considered complex or living. A new framework should contain dimensions that are quantitatively distinguishable. The stimuli were characterized according to the activity classification and graphed onto a color wheel to determine if there were any aesthetic trends, figure 5. There were conflicting trends for warmth or cool that could be discerned. Smooth sounds, such as pulses or pure tones could be considered warm, but “noisy” stimuli such as quakes or beating sounds could also be considered warm (drawing attention).

The Munsell color wheel suggests that energy and attention are perceptual dimensions for distinguishing stimuli. In the audio-haptic domain, a temporal-energy arrangement scheme was attempted by using ADSR envelopes. ADSR (Attack-decay-sustain-release) envelopes are a way to organize the stimuli based on energy and attention [23], figure 6. The attack angle describes how quickly a stimuli occurs, the decay measures how quickly the attack declines in amplitude, the sustain relates to how long the stimuli is sustained, and the angle of release can correspond to an angle of warmth of ambience, resulting in an “envelope” (figure 6a). Some typical ways to create interesting phenomena is to vary the ADSR envelope are shown (figure 6 b).

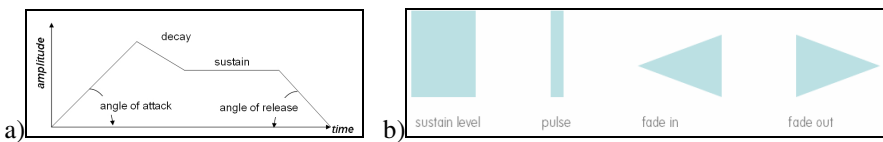


Fig. 6. a)ADSR envelope for audio design composition b) typical ADSR envelopes

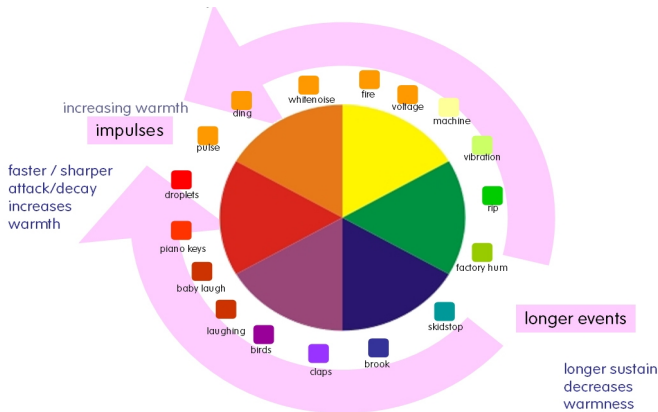


Fig. 7. The visual-vibrotactile audio-perception map temporal trend for the attack/decay and duration of a stimuli, increased sharpness and length attract attention.

We analyzed the average RMS value of the audio-haptic sounds, and gave a numerical value for the length of the stimuli [23]. A sound analysis tool, Adobe Audition [23], was used to analyze the envelope of each audio-haptic stimulus. Each stimulus was plotted atop the color wheel such that warmth (intensity) corresponded to attack/decay and amplitude. Higher intensity draws more user attention to the stimuli. A resulting visual-vibrotactile audio-perception map was generated (figure 7).

The correspondence between temporal-energy suggested an aesthetic trend based on amplitude and duration. The perception map was informally evaluated by 10 media artists using a squishy puff device connected to an on screen presentation. The artists were asked to select the mapped stimuli and evaluate the mapping out loud. General feedback on the visual-vibrotactile perception map was assessed to gauge how “natural or believable” the stimuli seemed.

Here are some aesthetic trends that were observed:

- Amplitude corresponds to saturation. The higher the amplitude, the more noticeable it is. In general, the amplitude effect dominates the stimuli.
- Attack/decay duration corresponds to intensity or haptic attention. Faster (sharper) attack/decays are more noticeable than smoother attack/decays. The smoother attacks are more subtle in haptic feel. However, there is a minimum length (10 milliseconds) where the skin cannot feel the vibration at all and the stimuli are perceived as pure audio and may be missed.
- Longer events, such as rhythmic or sustained stimuli, are also more noticeable (audibly “warmer”), but can be ignored over time as the user attenuates the stimuli. The skin can lose sensitivity to sustained sounds.

By noting the elements of composition as amplitude, attack/decay duration and sustain, **two main compositional effects** can be inferred from the visual and cinematic design literature:

1. *Balance*: Balancing interaction between high and low amplitude audio-haptic stimuli.
2. *Textural variation*: Alternating between impulse and sustained stimuli, and playing with the ADSR envelopes to create textural variation.

These two observed effects can be used as guidelines to create dramatic effects through manipulating attention between the two modalities.

3 An Audio-Haptic Composition Example

We present an audio-haptic ringtone to describe some composition mechanisms used to balance and textural variation, separately and concurrently (figure 8).

This piece explores the relationship between the musical ringtone and the vibrotactile alert. These mechanisms are introduced gradually, then developed - separately and overlaid - in the following short segments. Overall a sense of urgency is generated to fulfill the primary purpose of mobile alerting mechanism.

The piece begins with a haptic texture that has been created using the haptic inheritance method [5], where the haptic texture has been derived from the percussive instrumental elements. The texture can be classified as a type of pulse with a soft tail. The texture enters in a rhythmic manner, increasing somewhat in strength over the course of its solo 2 bar introduction. Then as the phrase repeats, the audible percussive element fades in a rhythmic unison.

When the percussive phrase is completely exposed (spectrally, dynamically and musically), a 2 bar answering phrase enters. This answering phrase contains no significant haptic texture but generates content in the broader frequency spectrum with the introduction of a musical sequence. The woody pulse that punctuates the phrase is “warm” in drawing attention to the rhythmic audio crescendo. Apart from the melodic and rhythmic phrase this sequence also contains a melodic decoration in the vein of a traditional auditory alert.

In the final section the haptic rhythm re-enters solo, apart from the appearance of the melodic alert decoration which is played at the same time in the rhythmic sequence as before.

The order of exposition is as follows:

1. Haptic Rhythm
2. Haptic Rhythm + Percussive Instruments
3. Percussive Instruments + Musical/Melodic Elements
4. Haptic Rhythm + Melodic Alert

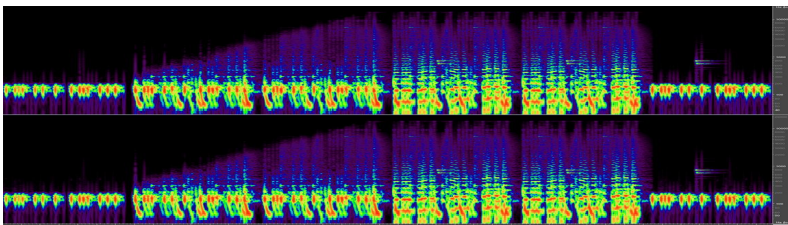


Fig. 8. Spectrogram of audio-haptic ringtone example

As such the envelope created is a type of privacy envelope. This can be inferred from the attached frequency spectrum of the piece; when the energy is concentrated in the lower range the privacy is greater, lessening over time, before a final private reminder is played.

4 Discussion

4.1 Aesthetic Perception Map Feedback

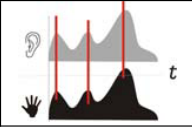
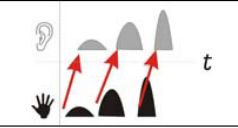
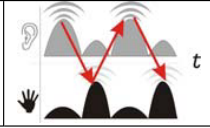
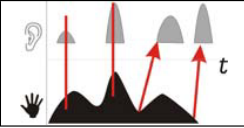
Overall, the feedback on our perception map was positive. Many artists commented that the effects of synchronizing haptics and audio were compelling, and that they could readily imagine the textures of the materials or ambience of the stimuli in the perception map. This suggests that synchresis (synchronizing and synthesizing) artificial audio-haptic stimuli could create realistic experiences that do not exist in the real world, a very interesting phenomenon [9]. Another common observation was that higher haptic amplitude present in the stimuli created a feeling of realism, particularly for impacts. The effect of audio reverb, combined with sustained and synchronized haptics seemed “more realistic” than without haptic feedback. For example, combined modal experiences allowed users to gauge the differences between metal objects hitting each other, vs. a hammer hitting an anvil. Another example is being able to distinguish a splashing “large rock in a stream” compared with a “drop of water hitting a large pool”. We believe the haptic component helped users imagine or identify the underlying material and environmental information in the stimuli.

One artist commented that perhaps amplitude was analogous to color saturation, drawing attention and perhaps even overwhelming the senses. She expressed some doubt that there could be three dimensions for audio-haptic perception. Rather, she suggested that instead of a color wheel, it could be a grayscale wheel that may be most appropriate. Another interesting issue is that visual warmth is achieved through reddish tones, which grab attention, but audio warmth is achieved through sustain duration, which fades as time progresses. Whether tactile warmth is affected by sustain duration is an interesting question. In general, feedback for the perceptual map was well perceived, but future work will need to verify its accuracy through multidimensional scaling.

4.2 Audio-Haptic Modality Transfer Discussion

Our example composition makes use of the four audio-haptic mechanisms between the dual modalities (Table 1). Here *temporal linearization* of the haptic elements create suspense, and *synchronization* with (percussion) audio allows transfer of the attention to the audio track. *Masking* is achieved by increasing the audio amplitude and spectral range over the haptic elements in the main body of the piece. *Synchresis* is created by the reverb of the audio elements, creating a sense of ambient resonance. We propose that distributing both warm and cool stimuli across the two modalities helps attain compositional *balance*. Similarly, *textural variation* is achieved by contrasting the sharpness of the synchronized syncopated percussion and haptic rhythm with the more sustained musical elements.

Table 1. Audio-haptic mechanisms used in achieving balance and textural variation

Synchronization	Temporal Linearization	Masking	Synchresis
Haptic and audio stimuli are united in <i>time</i> to draw attention to the unity of the stimuli.	Haptic or audio sequencing to create a connection between multisensory stimuli, to create a sense of causality. Usually haptic will precede audio.	The hiding of a (typically haptic) stimulus by another stimulus of higher amplitude when they are presented in together.	Creation using audio and tactile to create an association of an effect. Makes use of linearization and synchronization to create a mental union of two modalities, creating a distinct new association..
			

It should be noted that we are reporting the state of the art of how we practice audio-haptic design. We hope that this tutorial can be useful to help others struggling with audio-haptic composition. Discussion of the resulting artifact through use of the mechanisms was subjective to the ringtone designers, and has not been empirically tested. We look forward to contributing to a collaborative discussion on how to achieve audio-haptic balance and textural variation.

5 Conclusion

This paper presents the evaluation of audio-haptic stimuli from sound libraries as a starting point toward realizing the potential for audio-haptic composition. We discuss an approach to arranging the stimuli based on temporal-energy distributions, in a framework influenced by visual and cinematic design principles. Several perceptual features were observed by artists testing a perceptual map of audio-haptic stimuli. These observations resulted in the creation of guidelines for achieving balance and textural variation. These perceptual findings need to be supported by further studies using quantitative evaluation.

From the compositional guidelines we have developed four principle mechanisms for audio-haptic interaction design: *synchronization*, *temporal linearization*, *masking* and *synchresis*. We present an example of an audio-haptic ringtone composition that utilizes these mechanisms and discuss the multimodal effects achieved. We describe how these four approaches can be used to manipulate attention across modalities. Our example showcases a practical example of these mechanisms. In summary, we describe how compositions of carefully designed audio-haptics stimuli can convey a richer experience through use of the audio-haptic design mechanisms presented.

Acknowledgements

The authors thank Prof. John Maeda for the questions that started this research and Prof. Cynthia Breazeal for her advice. The authors would also like to express gratitude to Motorola’s Consumer Experience Design group for their support of this work.

References

1. Apple iPhone, <http://www.apple.com/iphone/>
2. Adobe Audition 3, <http://www.adobe.com/products/audition/>
3. Brewster, S., Chohan, F., Brown, L.: Tactile feedback for mobile interactions. In: Proc. of the CHI 2007, pp. 159–162. ACM Press, New York (2007)
4. Brewster, S.A., Wright, P.C., Edwards, A.D.N.: Experimentally derived guidelines for the creation of earcons. In: Proc. of HCI 1995, Huddersfield, UK (1995)
5. Chang, A., O’Modhrain, M.S., Jacob, R.J., Gunther, E., Ishii, H.: ComTouch: a vibrotactile communication device. In: Proc. of DIS. 2002, pp. 312–320. ACM Press, New York (2002)
6. Chang, A., O’Sullivan, C.: Audio-Haptic Feedback in Mobile Phones. In: CHI 2005 Extended Abstracts, pp. 1264–1267. ACM Press, New York (2005)
7. Michel, C., Gorbman, C., Murch, W.: Audio-Vision Sound on Screen. Columbia University Press, New York (1994)
8. Gaver, W.: Chapter from Handbook of Human Computer Interaction. In: Helander, M.G., Landauer, T.K., Prabhu, P. (eds.) Auditory Interface. Elsevier Science, Amsterdam (1997)
9. Geldard, F., Sherrick, C.: The cutaneous rabbit: A perceptual illusion. *Science* 178, 178–179 (1972)
10. Geldard, F.A.: *Body English*. Random House (1967)
11. Gunther, E.: *Skinscape: A Tool for Composition in the Tactile Modality*. MIT MSEE Thesis (2001)
12. Jones, L., Nakamura, M., Lockyer, B.: Development of a tactile vest. In: Proc. of Haptic Symposium 2004, pp. 82–89. IEEE Press, Los Alamitos (2004)
13. Luk, J., Pasquero, J., Little, S., MacLean, K., Levesque, V., Hayward, V.: A role for haptics in mobile interaction. In: Proc. of CHI 2006, pp. 171–180. ACM Press, New York (2006)
14. MacLean, K.E.: *Designing with Haptic Feedback*. In: Proc. of IEEE Robotics and Automation (ICRA2000), San Francisco (2000)
15. Merrill, D., Raffle, H.: The sound of touch. In: CHI 2007. Extended Abstracts, pp. 2807–2812. ACM Press, New York (2007)
16. Munsell, A.H., Munsell, A.E.O.: *A Color Notation*. Munsell Color Co., Baltimore (1946)
17. Nelson, L., Bly, S., Sokoler, T.: Quiet calls: talking silently on mobile phones. In: Proceedings of CHI 2001, pp. 174–181. ACM Press, New York (2001)
18. O’Modhrain, S., Oakley, I.: Adding Interactivity. In: Proc. of the Int. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Sys. (HAPTICS 2004), pp. 293–294. IEEE Press, Los Alamitos (2004)
19. O’Sullivan, C., Chang, A.: An Activity Classification for Vibrotactile Phenomena. In: McGoekin, D., Brewster, S.A. (eds.) HAID 2006. LNCS, vol. 4129, pp. 145–156. Springer, Heidelberg (2006)
20. O’Sullivan, C., Chang, A.: Dimensional Design; Explorations of the Auditory and Haptic Correlate for the Mobile Device. In: Proc. of ICAD 2005, Limerick, Ireland, pp. 212–217 (July 2005)
21. Pinnacle Systems. Pinnacle Studio 10 Sound Library. [Mountain View, CA] (2005)
22. Prada phone by LG, <http://www.pradaphonebylg.com/>
23. Riley, R.: *Audio Editing with Cool Edit*. PC, Tonbridge (2002)
24. *Sound Library Sonothèque = Sonoteca*. [France]: Auvidis (1989)
25. Tan, H.Z.: Perceptual user interfaces: haptic interfaces. *Communications of the ACM* 43(3), 40–41 (2000)

26. Verrillo, R.T., Gescheider, G.A.: Tactile Aids for the Hearing Impaired. In: Summers, I. (ed.) *Perception via the Sense of Touch*, ch. 1. Whurr Publishers, London (1992)
27. Vogel, D., Baudisch, P.: Shift: a technique for operating pen-based interfaces using touch. In: *Proc. of the CHI 2007*, pp. 657–666. ACM Press, New York (2007)
28. Williamson, J., Murray-Smith, R., Hughes, S.: Shoogle: excitatory multimodal interaction on mobile devices. In: *Proc. of the CHI 2007*, pp. 121–124. ACM Press, New York (2007)
29. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., Baudisch, P.: Earpod: eyes-free menu selection using touch input and reactive audio feedback. In: *Proc. of CHI 2007*, pp. 1395–1404. ACM Press, New York (2007)