DOI: http://dx.doi.org/10.14236/ewic/EVA2017.39

The Parallels between the Study of Cross-modal Correspondence and the Design of Cross-Sensory Mappings

Augoustinos Tsiros
Centre for Interaction Design
Edinburgh Napier University
10 Colinton Road, EH10 5DT
Scotland
a.tsiros2@napier.ac.uk

The aim of this paper is to examine how recent research findings and methodologies from the field of cognitive science could be utilised to inform the design and evaluation of cross-modal mappings for multimodal user interaction. In this paper we argue that by using empirical methods to enact embodied knowledge about cross-modal correspondences, we can form an adequate empirical framework for the design of multimodal mappings that successfully align with prior perceptual knowledge. This alignment can significantly improve the human computer dialogue and the analytical, creative and pedagogical value of user interfaces.

Display and interaction design. Mappings. Perceptual learning. Crossmodal correspondence.

1. INTRODUCTION

Effective interface design is important in allowing users to efficiently interpret and interact with information. As information displays and user interfaces become more elaborate, interface developers need to consider perceptual and cognitive factors involved in human sensory processing and communication to allow for the effective transfer of information and natural interaction between the user and the system. This is paramount in multimodal interface that utilise multiple forms of sensing for interaction and/or rendering of sensory information in non-linguistic formats, such as interfaces which purpose is to provide cross-sensory mappings (e.g., visualisation, sonification, physicalisation and non-linguistic sensorimotor user interaction). In such interaction scenarios, digital technologies allow flexibility and freedom on how the mapping between input and output parameters will be implemented.

Developers and researchers creating these types of interfaces have to make decisions on how input-output parameters are associated, for example systems such as a gestural controller for navigation and interaction within a virtual environment, a speech visualisation software, or a motion sensing musical interface. Although it has long been recognised that a good mapping between user input and system output which conforms to the user's expectations is critical for effective control and interpretation of a given system's output (Lidwell, Holden, & Butler 2003, Norman 1983, Norman 1999), most developers still adopt an ad-hoc understanding of "good design".

A great deal of research efforts has been put to understand and develop perceptually informed interface designs and interaction paradigms, see (Hermann, Hunt, & Neuhoff 2011, Ware 2000).The focus has been primarily on developing design frameworks that are considerate of the subtleties involved in a single mode of perception and make efficient use of the unimodal capabilities offered, such as attention, sensory stream segregation and bandwidth. However, less emphasis has been given on how multiple senses interact and on considering how to best utilise humans' multisensory capabilities in the design of interfaces.

Understanding cognitive processes and the involvement of higher level cognitive function across all sensory modalities is important for developing multimodal interfaces that are considerate of the subtleties involved in human perception and cognition and make efficient use of

humans' multisensory capabilities. The aim of this paper is to examine how recent research findings and research methods from the field of cognitive science could be utilised to inform the design and evaluation of cross-modal mappings for multimodal user interaction and information display.

In this paper, we argue that by using empirical methods to enact embodied knowledge about cross-modal mental models, we can form an adequate empirical framework for the design of multimodal mappings that successfully align with prior perceptual knowledge. This alignment can significantly improve the human computer dialogue and the analytical, creative and pedagogical value of user interfaces. Exploring the significance of multisensory phenomena in human-computer interaction, and specifically in multimodal interface design as applied to pedagogy, analytics, assistive and creative applications, may lead to the development of interfaces that are more intelligent, adaptive, and ultimately more useful to the user. In this paper, we discuss the literature from the field of cognitive psychology and more specifically research on cross-modal correspondence in order to (i) investigate the relevance of the findings that derived from these studies in the problem of multimodal representation (e.g., visualisation, sonification), (ii) inform the mapping between users' sensorimotor input and sensory related parameters (e.g., gestural interaction with sound, graphics and virtual environments), and (iii) assist in paving the way for future research in multimodal interaction.

2. THE EFFECTS OF DIFFERENT TYPE OF SENSORY REPRESENTATIONS

We can broadly distinguish two types of representations humans use to communicate information i.e., arbitrary and sensory (Ware 2000). Following Ware's classification, in this paper, the word arbitrary is used to refer to representations that have to be learned because they have an arbitrary relationship with the concept that they represent. Examples of arbitrary forms of representation include spoken and written language, mathematical symbols, music notion or arbitrary icon that symbolises various concepts in computer interfaces. The word sensory is used to refer to representations that derive their communicative power from some form of structural similarity to the concept they stand for. An example of a sensory representation would be an MRI representation that shows the activation of brain structures or an audio spectrogram that shows in visual form the energy distribution of the frequency spectrum of a sound. Because sensory representations are grounded on sensory resemblance, they can be easier to learn, interpret and remember in comparison to arbitrary forms of representation.

In order to understand better why sensory representations are a more effective way of representing information, it is helpful to consider Kahneman's model of two cognitive systems (Kahneman 2011). Kahneman distinguishes between two modes of thinking, Fast (intuition) and Slow (reason). According to Kahneman's conception, intuition has a very different set of attributes from reason. Some of these differences include: reason is slow, serial in operation and effortful, while intuition is fast, parallel and effortless.

On the one hand, perceptually informed sensory representations are by far more likely to be processed by intuition. A good example of a type of system that taps into intuition would be a virtual reality headset. Users need no training to interact with this technology, since the technology fully complies with the users' prior perceptual knowledge about how the auditory and visual scenes are affected when moving their head in the natural world. In this example we have an instant transfer of perceptual skills from the natural to the virtual environment and an otherwise complex multiparametric interface can be operated without much effort or previous training. This example demonstrates well the advantages of effective utilisation of the users' intuitions in the control of a complex computer interface.

On the other hand, arbitrary representations have to be learned and a person has to memorise the symbols and establish associations between the representation and concept being represented. Hence these representations are likely to be handled by reason initially. Once somebody has become fluent in an arbitrary system, this very same representation will be handled by intuition, but the transition of information processing from reason to intuition might take a considerable amount of time.

A control system or information representation that initially utilise the attributes of reason might take a long time to master and use the more desirable attributes of intuition. This raises the question, how designers of computer interface and information displays can effectively utilise prior perceptual knowledge? In this paper, the answer to this question is explored by quantifying the sensorimotor skills and perceptual models which users have developed through a lifetime of experiencing causal relationships in the environment and applying them into the design of cross-sensory mappings of computer interfaces.

The study of crossmodal correspondence has already began developing an empirical framework that could be useful in the quantification of cross-sensory associations. The understanding that is

emerging from this branch of research is of high relevance to the design of intuitive mappings.

3. SENSORY STIMULATION AND PERCEPTUAL LEANING

When complex sensorimotor skills have to be developed, such as learn how to drive or play a musical instrument, the transfer time from these newly acquired skills from reason to intuition (that is the ultimate goal) can be very long and effortful. These skills require that humans develop sensorimotor correspondences and causal relationships between sensory input and motor action. This is another domain in which perceptually grounded sensory representation could aid. Research in perceptual learning suggests that multisensory stimulation is extremely important in the context of learning, (Goldstone 1998, Shams & Seitz 2008). It has been argued that the integration of multisensory information is essential to construct a meaningful representation of the behaviours of a system found in the natural environment. Furthermore, the ability to construct meaningful internal representations of the environment, actually depends on integrating and segregating between multisensory stimuli received from the environment (Shimojo 2001, Brandwein et al. 2011).

Humans and other animals inhabit a multisensory environment and are equipped with multiple senses in order to sense the information available. Therefore, it is likely that the human brain has evolved to develop, learn and operate optimally in multisensory settings. For example, studies on selective attention of infants during speech listening have shown that, in the first six months infants tend to focus their attention more to the eyes region of the speaker than the mouth, (Lewkowicz & Hansen-Tift 2012). Interestingly, when infants reach about 6 months of age, which is the age when they begin to bubble, they are more motivated by the visual cues provided by the mouth of the speaker. Lewkowicz et al. suggest that the sudden shift of infants' attention from the eye region to the mouth region has been attributed to the fact that the audio-visual cues provided by listening to a speaker in combination with looking at the lips and the mouth's movement, provide a richer informational content which is beneficial for the learner. Research findings suggest that during multisensory stimulation it is not only the respective sensory structures that are activated but a larger set of processing structures of the brain (Shams & Seitz 2008). According to the finding of Shams & Seitz, the activation of a larger set of cognitive structures results in better and faster acquisition and retention of information; other studies also found that multisensory settings are more effective than unisensory ones for acquiring skills, see (Hardison 2003 2004, Sueyoshi & Hardison 2005).

The predisposition humans exhibit to exploit the multisensory information available to them in order to develop their own sensory motor capabilities, raises the question: could perceptually informed artificial sensory stimulation aid humans in comprehension and skill acquisition? If so, then how can we best create multiple manifestations of a concept to provide a rich multisensory experience to support learning and comprehension?

3.1 Mechanisms of Perceptual Learning

According to Goldstone (1998), perceptual learning involves mainly four mechanisms: attention weighting, imprinting, differentiation, and unitisation. In attention weighting, perception becomes adapted to tasks and environments, by increasing the attention paid to important dimensions and features. In imprinting, receptors are developed that are specialised for stimuli or parts of stimuli. In differentiation, stimuli that were once indistinguishable become psychologically separated. In unitisation, tasks that originally required detection of several parts are accomplished by detecting a single constructed unit representing a complex configuration.

In an attempt to understand how a computer interface could support users in the process of acquiring sensorimotor skills, let's consider the example of a person wishing to acquire pronunciation skills in a second language, and how the four components of learning as suggested by Goldstone, (1998) could be associated to computer aided learning application for pronunciation training.

- *Attention weighting:* The learners need aid to guide their attention to important acoustic and articulatory features of speech. Appropriate sensory feedback provided by the system should guide the learners' attention to the important acoustic and articulatory features of the second language.
- *Imprinting:* The system should provide engaging and informative production training through rich sensory stimulation to help the learner developed specialised receptors for identifying these acoustic and articulatory features from a stream of utterances.
- *Differentiation:* The language's articulatory and phonetic features that initially were difficult for the learner to differentiate, become salient after training. Sensory feedback provided by the system should enhance the learners' ability to differentiate the articulatory as well as acoustic features of words that are acoustically similar.

- *Unitisation:* As low level features of speech consist of two elements (i.e., acoustic and articulatory parameters), a training system should provide appropriate feedback to help the learner understand the correspondence between these two dimensions. Unitisation occurs when after motor training, the learner has internalised the correspondence between acoustic and articulatory features to the extent where the phonetic features of previously known words as well as that of novel words can be produced effortlessly and by devoting minimal attention to the motoric actions required for correct production (i.e., becoming intuitive and automatic).

Developing rich multisensory stimulation could be beneficial for comprehension, interpretation and retention of information. Hence, understanding how to best utilise multisensory processes into the design of interfaces is of great importance. While digital technologies allow to create rich multimedia environments that convert the sensory stimulus of one sense to another, the problem of how to develop effective mappings across sensory modalities remain. The following sections discuss research findings from the field of cognitive science that could inform cross-sensory mappings for applications which aim to aid comprehension, learning and assist on the pursuit of 'natural' interfaces between user and computer systems.

3. CROSSMODAL CORRESPONDENCE AND CROSS-SENSORY MAPPINGS

So far we have established an understanding of why multiple forms of sensing enable us to construct a more precise and robust representation of the external world. Computer interfaces and information displays that tap into these mental models could benefit in terms of accommodating users' sensorimotor skills and support comprehension and retention of information. The subsequent issue is the effective design of interfaces that tap into the mental models. The specific problem that is of primary concern, is the mapping between low-level sensory stimulus properties in information displays and human computer interfaces.

The different senses provide information of differed level of precision. The information we receive at any given moment are combined to estimate the most likely state of the world. The information we perceive through the different senses can be related and unrelated and redundant or complementary (Parise 2015). Redundant cues refer to sensory information perceived through different senses which describe the same feature of the physical world. For instance, the size of an object can be perceived through vision and touch. While complementary cues, refer to the features of the physical world which can be experienced only by one sensory modality (Parise & Spence 2013). For example, an object's color can only be perceived through vision and the sound's timbre through listening. So it could be argued that, when designing interfaces where redundant cues are available, making decisions about how the different parameters are associated to the interface controls is more straightforward than designing the same relationships for complementary cues. As in the former case, we have robust mental models to drive our design decisions while in the latter case the mental models 'do not exist'.

However, people often map stimulus properties from different sensory modalities onto each other in a manner that is surprisingly consistent. For instance, listeners consistently have reported high levels of perceived congruence between auditory pitch and light intensity (higher auditory pitches for higher luminosity); visual size and pitch (lower pitches for bigger visual shapes and vice versa), visual size to loudness(louder sound for bigger visual shapes and vice versa), spatial height to pitch (higher pitches for higher spatial location), see (Eitan 2013, Spence 2007 & 2011) for review .

Research findings suggests that correspondences of stimulus properties maybe innate or learned in very early stages of a person's life, (Jeschonek, Pauen, & Babocsai 2012, Wagner, Winner, Cicchetti, & Gardner 1981, Walker et al. 2010). For example, studies have shown that pitch to spatial height correspondence affect the attention of infants (less than 6 months of age), (Dolscheid, Hunnius, Casasanto, & Majid 2014). While, 6 month old babies and 3 year-olds consistently matched visual arrows pointing up or down to rising and falling pitch curves, (Mondloch & Maurer 2004, Wagner et al. 1981). Furthermore, 9 years-old consistently mapped pitch height to visual size (Marks, Hammeal, & Bornstein 1987). Other studies suggest that cross-modal associations such as that of luminance (visual brightness) to pitch are common between humans and chimpanzees, which supports the view that the phenomenon of cross-modal correspondence is not culturally learned or a linguistic phenomenon, but instead is a fundamental cognitive feature shared perhaps by all primates, (Ludwig, Adachi, & Matsuzawa 2011). Other studies have shown that consistent associations between other modalities are exhibited by children as well, such as vision and touch, audition and body movement, audition and taste/flavour, (for full list of references see Spence (2011)). Crossmodal correspondences, according to many researchers, can originate at three different hierarchical levels of sensory information processing including psychophysical, perceptual and post-perceptual/cognitive (Landy, Banks, & Knill 2012,

Marks 1974, Marks 1989, Martino & Marks 1999, Shams & Beierholm 2012).

The first type of crossmodal correspondences have been named structural correspondences and it has been hypothesised that they occur due to the similarities of the transformation of sensory information into perceptual information, (Handel 2006, Parise & Spence 2013, Walsh 2003). The second type of crossmodal feature correspondences originates from statistical or functional regularities that can be commonly observed in the physical environment. For example a common statistical regularity is the size of physical body and its resonating frequencies, the changing distance to a sound source and its relative perceived amplitude level. The brain uses these learned statistical regularities to assess which cross-modal cues to combine and which to segregate. These statistical regularities are likely to be independent of culture. It could be argued that the degree of perceived correspondence between two cross-modal cues by an observer depends on the probability of these associations to occur in the environment. Observers can be trained through exposure to artificially compose cross-modal correspondences to learn new correspondences; however these newly learned feature associations will not be universal (i.e., shared by all humans like natural correspondences).

The third type of correspondences are semantically mediated correspondences. For example, human perceive different colour hues as being cold and warm, auditory pitch as being high and low, and associate moods states with different colour hues. While in the case of statistical correspondence, the perceived similarity between two objects occurs because the two objects share structural or functional features, in semantically mediated correspondence similarity occurs when two objects are linked conceptually.

Doing a systematic review of research findings from the study of crossmodal correspondence could lead to an empirical framework that designers and developers can utilise for the design of multimodal mappings for sensorimotor interaction, information display and multimedia design. For instance, Figure 1 shows the frequency of occurrences of audio-visual stimulus properties associations reported in the literature. This graph serves as an example and is not an exhaustive account of occurrences that can be found in the body of literature. **Figure 1**, shows how much evidence there is in the literature about the correspondence between different auditory and visual feature dimensions. There are many other factors that could be taken into consideration such as conflicting evidence, evidence of strength of associations, study sample size and reliability of the method used in the study etc. Additionally, to further enhance this framework, it would be useful to include evidence from studies that have used complementary polar features and report on their effectiveness in applied contexts such as interaction design, educational applications, and information retention. Currently, there are not many studies providing this type of evidence.

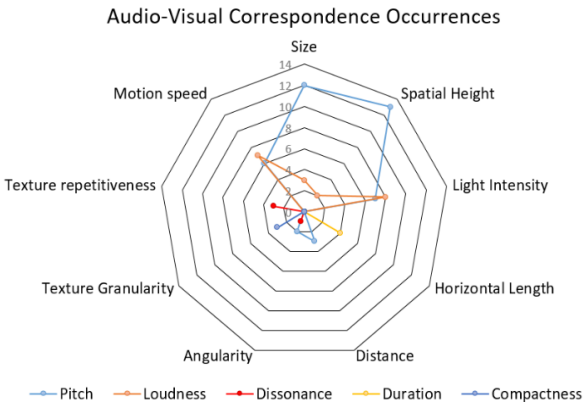

Figure 1: *Number of occurrences of crossmodal associations reported in the literature. For complete list of references used to produce the graph, follow the link.*

3.1 Standing Issues & Practical Suggestions

Having given an overview of what crossmodal correspondences are and why we should utilise them, the remainder of this paper will discuss standing problems, their implications in the effective utilisation of crossmodal correspondences in the context of human computer interaction and provide some guidelines to help solving the questions that arise. Literature in crossmodal correspondence points to a number of issues that could be seen as concerning for the direct application of crossmodal associations to the design of cross-sensory mappings. First, it should be noted that complementary features of one modality can map equally well to more than one features of another, (Eitan 2013). For instance, auditory pitch and loudness are both good correlates of physical size, spatial height, distance and speed. Using simple and parametrically manipulated stimuli, designers can test which mapping is more effective for a given application. If the mapping of the intended application is multidimensional, then tests could be performed to test the effectiveness of the mapping as a whole (i.e., when all feature associations are present). For instance, for seeking the best visual correlate to auditory loudness, it would be useful to test all corresponding visual cues both individually through parametric manipulation for of a single dimension (e.g., loudness-size, loudness-height etc.) and then test again in the context of the mapping. Furthermore, research findings suggest that interactions between different dimensions of sensory cues can occur (Eitan 2013). For example, an increase in auditory tempo rate is often correlated to

an increase in visual motion speed; however when an increase in auditory tempo is accompanied by a decreasing loudness, the correspondence between tempo and speed is weakened. These types of interactions between sensory cues could be seen as an issue in utilising crossmodal correspondences. Hence, the effectiveness of cross-sensory mappings that consist of multiple dimensions has to be tested to ensure that there are no interactions between the differed dimensions of the mapping when multiple parameters co-vary. For instance, if we consider a visualisation of speech representation for a pronunciation training application, where the mapping consist of multiple dimensions, such as pitch – spatial height, loudness-physical size and auditory noisiness to texture entropy: in this type of system it will be important to test both the effectiveness of each individual dimension but also their effectiveness when multiple parameters vary simultaneously.

In addition, it has been argued that correspondences can be affected depending on whether the stimuli used for testing vary in a discreet or continues manner, (Eitan 2013, Granot & Eitan 2011). For instance, while small physical size is often associated with high auditory pitch sounds and larger size with lower auditory pitch, a shrinking shape is associated with a fall in pitch and an expanding shape is best associated with a rise in pitch (Eitan, Schupak, Gotler, & Marks 2011). This suggests that that future research efforts should aim to assess the validity of previous experimental results in contexts where parametric manipulation is discreet and continuous.

Finally, the sensory stimuli in most of the studies commonly undertaken tend to be limited to simple synthetic stimuli. Considering that most natural sensory input are more complex than uni-dimensional stimuli, it is uncertain whether these results still apply to complex stimuli. For instance, confounds (less salient features that are not part of the mapping) might hamper the effectiveness of the mapping, by reducing the ability to interpret the information, discriminate individual associations of the mapping, or increase cognitive effort required to interpret or control.

3.2 Methodological Considerations

We could broadly distinguish two main experimental paradigms for the study of cross-modal correspondence i.e., speeded and non-speeded, (Marks 2004). The speeded classification paradigm which is currently the most commonly used method for the study of selective attention of multidimensional varying stimuli was first proposed by Gerner in the 1960s. The speeded classification/identification method relies on subjects identifying or classifying particular characteristics of a stimulus as quickly as possible. If subjects' reaction times of the selective attention task is greater than in the baseline task due to the variation of irrelevant feature dimension, then it is assumed that there is an interference between these feature dimensions known as *Gerner's interference*.

Gerner's interference methods show that some cross-sensory associations are more effective than others in activating early sensory information processing and selective attentions. This suggests that creating artificial mappings using associations established through prior perceptual knowledge should result in interfaces and information displays that are easier to comprehend and are more robust to external noise, i.e., sensory stimuli by less salient features that might be present in the users' environment. By providing information in formats that are easier for people to trace and understand, greater cognitive resources can be allocated to thinking about information, interpretation, linkages and decision-making.

An even older method used to study crossmodal correspondences relies on non-speeded tasks. There is a number of different unspeeded methods for studying crossmodal correspondences, including stimulus tracing tasks, pairwise similarity judgments, forced matching, multiple items arrangement and confusion tasks; for a description and the pros and cons of each method, see (Giordano, Susini, & Bresin 2013, Goldstone 1994, Kriegeskorte & Mur 2012). Furthermore, the method that has been most commonly used in music research is free tracing tasks. Free tracing was first used in mental imaging research. In free tracing experiments, participants are asked to perform a sensorimotor response (e.g., a gesticulation, draw a sketch) to an audio stimulus. Godoy was one of the first researchers to conduct this type of research in musicology (Godøy, Haga, & Jensenius 2006a & 2006b, Godøy 2006). This experimental method was then adopted by a number of other researchers (Caramiaux, Francoise, Bevilacqua, & Schnell. 2014, Baptiste Caramiaux, Bevilacqua, & Schnell 2010a 2010b, Kussner & Leech-Wilkinson 2013, Küssner, Tidhar, Prior, & Leech-Wilkinson 2014, Küssner 2014).

Speeded methods are particularly suitable for studying correspondences in pre-attentive perception (correspondences at a psychophysical level). However correspondences can also occur at structural, statistical and semantic levels. In unspeeded tasks, conceptual properties and higher level cognitive processes (such as semantic or psycholinguistic levels) become more influential in the subjects' responses, (Goldstone 1994). Therefore, unspeeded tasks are more suitable for studying higher level cognitive processes. Moreover, studying cross-modal correspondence in

applied context (e.g., cross-sensory mappings for musical interaction) could have different requirements from studying early stages of sensory processing. Therefore, it could be argued that speeded and unspeeded methods are not mutually exclusive but complementary. Hence, the appropriate method for assessing the effectiveness of a given cross-sensory mapping should be determined by the type of problem which an interface or an information display is attempting to address.

4. CONCLUSIONS

This paper examined the mapping between low-level sensory stimulus properties in information displays and human computer interfaces. Through a number of examples, we established an understanding of why computer interfaces that utilise cross-sensory mappings can be more effective. Mappings which are considerate of the users' prior perceptual knowledge and the subtleties of multisensory aspects of perception, could benefit in terms of accommodating users' sensorimotor skills and support comprehension and retention of information. Hence I suggest that empirical findings and methods used to study crossmodal correspondence could help design interfaces and displays that tap into users' mental models.

Finally, I pointed to a number of potential issues related to the direct application of crossmodal correspondences in human computer interaction and information displays and I provided some practical guidelines for the evaluation of cross-sensory mappings that might help to tackle these issues. Future work includes making a more thorough systematic review of the literature of crossmodal correspondence across all sensory modalities. Other directions for future research involves (i) testing the effectiveness of crossmodal mappings in different contexts including assistive, educational and creative applications, (ii) assess the effectiveness of cross-sensory mappings in multidimensional contexts, such as when mapping is complex, (iii) assess the effect of confounding variables on the corresponding associations, and (iv) test associations using static and dynamic stimuli.

5. REFERENCES

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., and Molholm, S. (2011) The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(5), 1042–55.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010a) Study on Gesture-Sound Similarity. *3rd Music and Gesture Conference*, 1–2.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010b) Towards a gesture-sound cross-modal analysis. In I. Kopp, S., Wachsmuth (Ed.), *Gesture in Embodied Communication and Human-Computer Interaction* (pp. 158–170). Springer, Heidelberg.

Caramiaux, B., Francoise, J., Bevilacqua, F., and Schnell., N. (2014) Mapping Through Listening. In *Computer Music Journal* (Vol. 38, pp. 1–30).

Dolscheid, S., Hunnius, S., Casasanto, D., and Majid, a. (2014) Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures. *Psychological Science*, (April), 1–6.

Eitan, Z. (2013) How pitch and loudness shape musical space and motion: new findings and persisting questions. In S.-L. Tan, A. J. Cohen, S. D. Lipscomb, & R. A. Kendall (Eds.), *The Psychology of Music in Multimedia* (pp. 161–187). Oxford University Press.

Eitan, Z., Schupak, A., Gotler, A., and Marks, L. E. (2011) Lower pitch is larger, yet falling pitches shrink: Interaction of pitch change and size change in speeded discrimination. *Proceedings of Fechner Day*.

Giordano, B. L., Susini, P., and Bresin, R. (2013) Perceptual Evaluation of Sound-Producing Objects. In S. S. and K. Franinović (Ed.), *Sonic Interaction Design* (pp. 151–198). London, England: MIT Press.

Godøy, R., Haga, E., and Jensenius, A. (2006) Exploring music-related gestures by sound-tracing: A preliminary study. In *2nd ConGAS International Symposiumon Gesture Interfaces for Multimedia Systems*.

Godøy, R. I. (2006) Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, *11*(2), 149.

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006) Playing " Air Instruments ": Mimicry of Sound-producing Gestures by Novices and Experts. *Gesture in HumanComputer Interaction and Simulation*, *3881*(6801), 256–267.

Goldstone, R. (1994) An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 381–386.

Goldstone, R. L. (1994) The role of similarity in categorization: providing a groundwork. *Cognition*, *52*(2), 125–57.

Goldstone, R. L. (1998) Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

Granot, R. Y. and Eitan, Z. (2011) Musical tension and the interaction of dynamic auditory parameters. *Music Perception*, *28*, 219–245.

Handel, S. (2006) *Perceptual Coherence: Hearing and Seeing*. Oxford University Press.

Hardison, D. M. (2003) Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522.

Hardison, D. M. (2004) Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, *8*(1), 34–52.

Hermann, T., Hunt, A., and Neuhoff, J. G. (2011) *The Sonification Handbook*. *The Sonification Handbook*.

Jeschonek, S., Pauen, S., and Babocsai, L. (2012) Cross-modal mapping of visual and acoustic displays in infants: The effect of dynamic and static components. *Journal of Developmental Psychology.*, *10*, 337–358.

Kahneman, D. (2011) *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Kriegeskorte, N. and Mur, M. (2012) Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Frontiers in Psychology*, *3*(July), 245.

Küssner, M. B. (2014) *Shape, drawing and gesture: cross-modal mappings of sound and music*. King's College London.

Kussner, M. B. and Leech-Wilkinson, D. (2013) Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm. *Psychology of Music*, *42*(3), 448–469.

Küssner, M. B., Tidhar, D., Prior, H. M., and Leech-Wilkinson, D. (2014) Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology*, *5*(July), 789.

Landy, M. S., Banks, M. S., and Knill, D. C. (2012) Ideal-Observer Models of Cue Integration. In and M. S. L. Julia Trommershäuser, Konrad Kording (Ed.), *Sensory Cue Integration* (pp. 1–35). Oxford Scholarship Online.

Lewkowicz, D. J. and Hansen-Tift, A. M. (2012) Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431–6.

Lidwell, W., Holden, K., and Butler, J. (2003) *Universal Principles of Design*. Rock Port.

Ludwig, V. U., Adachi, I., and Matsuzawa, T. (2011) Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20661–5.

Marks, L. E. (1974) On the Associations of Light and Sound. The Mediation of Brightness, Pitch, And Loudness. *American Journal Of Psychology*, *87*(1–2), 173–188.

Marks, L. E. (1989) On Cross-Modal Similarity : The Perceptual Structure of Pitch , Loudness , and Brightness. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 586–602.

Marks, L. E. (2004) Cross-modal interactions in speeded classification. In Gemma Calvert, C. Spence, & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 85–106). MIT press.

Marks, L. E., Hammeal, R. J., and Bornstein, M. H. (1987) Perceiving similarity and comprehending metaphor. *Monogram Society Research Children Development*, *52*, 1–102.

Martino, G. and Marks, L. E. (1999) Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*, 903–923.

Mondloch, C. J. and Maurer, D. (2004) Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133–136.

Norman, D. (1999) Affordances, Conventions and Design. *Interactions*, *6*(3), 38–42.

Norman, D. A. (1983) Some observations on mental models. In D. G. and A. L. Stevens (Ed.), *Mental models* (pp. 7–14). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers Inc.

Parise, C. and Spence, C. (2013) Audiovisual cross-modal correspondences in the general population. *The Oxford Handbook of Synesthesia*, 790–815.

Parise, C. (2015) Crossmodal Correspondences: Standing Issues and Experimental Guidelines. *Multisensory Research*, *29*(October), 7–28.

Shams, L. and Beierholm, U. (2012) Humans' Multisensory Perception, from Integration to Segregation, Follows Bayesian Inference. In and M. S. L. Julia Trommershäuser, Konrad Kording (Ed.), *Sensory Cue Integration* (pp. 1–16). Oxford Scholarship Online.

Shams, L. and Seitz, A. R. (2008) Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–7.

Shimojo, S. and Shams, L. (2001) Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 505–509.

Spence, C. (2007) Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61–70. https://doi.org/10.1250/ast.28.61

Spence, C. (2011) Crossmodal correspondences: a tutorial review. *Attention, Perception & Psychophysics*, *73*(4), 971–95.

Sueyoshi, A. and Hardison, D. M. (2005) The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699.

Wagner, S., Winner, E., Cicchetti, D., and Gardner, H. (1981) Metaphorical" mapping in human infants. *Child Development*, *52*, 728–731.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., and Johnson, S. P. (2010) Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21–5.

Walsh, V. (2003) A theory of magnitude: Common cortical metrices of time, space and quality. *Trends in Cognitive Sciences*, *7*, 483–488.

Ware, C. (2000) *Information Visualization: Perception for Design*. Morgen Kaufman.