Martin Siefkes & Emanuele Arielli

# An Experimental Approach to Multimodality: How Musical and Architectural Styles Interact in Aesthetic Perception[1]

## 1 Introduction

Multimodality is a rapidly growing field in linguistics and semiotics, and empirical research on interactions between modes forms an integral part of this field. However, research on the interplay of perceptual modes also has a long-standing history in cognitive psychology where it is often investigated with experimental methods. Both traditions come together in experimental aesthetics, an interdisciplinary research field dating back to the beginnings of modern psychology in the 19th century that has recently seen a revival (as witnessed, for example, by the founding of the Max Planck Institute for Empirical Aesthetics in Frankfurt am Main in 2013). Some differences in terminology exist and should be noted: whereas in linguistic multimodality research, the terms *intermodality* and *intersemiosis* are frequently used (cf. Wildfeuer 2012), psychological studies often speak of *crossmodal interactions* (cf., e.g., Vines et al. 2006).

In recent years, approaches to multimodal communication (cf., e.g., Kress/van Leeuwen 2001; O'Halloran/Smith 2011; Bateman/Wildfeuer 2014) have gained rapidly growing attention, sometimes with a specific focus on discourse analysis (cf., e.g., Siefkes/Schöps 2013[2]; Machin 2014). Multimodality research has begun to investigate the interactions between text and image (cf. Hess-Lüttich/Wenz 2006; Bateman 2014), language and gesture (cf. Fricke 2012), and speech and images in film (cf. Bateman/Schmidt 2012; Wildfeuer 2014). This research lays the groundwork for a broader and more precise approach to multimodality and opens new research routes. It is based on analytical and formal approaches founded in structuralism and semiotics, as well as on corpus-linguistic methods;

1 The research presented in this article was partly supported by a Feodor Lynen Return Fellowship for Postdoctoral Researchers by the Alexander von Humboldt Foundation (for M. S.). We wish to thank Kate Maxwell (University of Tromsø) for proofreading, and the editor of this volume, Janina Wildfeuer, for helpful suggestions.

2 This special issue includes four articles on multimodal text and discourse analysis in various modes and media.

although it has brought important theoretical advances and thorough empirical analyses, this research has not, to our knowledge, resulted in experimental studies of intermodal effects.

This paper presents the results of two studies conducted at the University IUAV of Venice and focused on the interactions between the aesthetic perception of architecture and background music in audio-visual presentations. We investigated how two musical styles (modern and baroque) influence the judgment of buildings belonging to the corresponding architectural styles. The two studies therefore connected two different types of artefacts, namely buildings and pieces of music, both of which have conventional meanings and can thus be regarded as semiotic modes, and are also perceived in different perceptual modes, namely visual and auditory perception. The results show that meanings connected with one mode, as measured by ratings on semantic scales, are influenced by another mode.

### 1.1 Intermodal interactions: an interdisciplinary perspective

What we call *intermodality* has been investigated in various disciplines and with different terminology. In cognitive psychology, the term *crossmodality* is used to describe mutual influences between sensorial channels such as auditory, visual or tactile perception, which have, in the last decades, become a thoroughly researched topic in perception studies (an overview of this research is given in Calvert et al. 2004). A much-discussed effect concerns the influence of auditory information on visual perception, and vice versa (cf. McGurk/MacDonald 1976). Research on crossmodal phenomena has typically been restricted to the perception of low-level information, such as in the case of vision to the perception of shape, colour, and movement. Recently, however, studies have begun to show the influence of perceptual information on higher-level cognitive processes (cf. Chapados/Levitin 2008; Davidson 1993; Vines et al. 2006).

Of particular interest is the question of correspondences between modes with regard to their unity and congruence. *Semantic congruence* usually refers to those situations in which pairs of stimuli from different modes are presented as matching. Studies of multisensory perception observe the consequences of presenting matching or mismatching object pictures and sounds (such as a tweeting sound paired with a picture of a dog; cf. Molholm et al. 2004) or of visually presenting letters with matching or mismatching speech sounds (cf., e.g., van Atteveldt et al. 2004). *Synaesthetic congruence*, on the other hand, refers to corresponding experience of stimuli that are reported by a majority of people, such as a high pitched sound corresponding to intense brightness or colour versus a low pitched sound corresponding to lower brightness or colour. It has been argued that stimuli that

are either semantically or synaesthetically congruent will more likely be bound together, a notion that is referred to as the "unity effect" (cf., e.g., Spence 2007; Vatakis/Spence 2008).

In linguistics, a related discussion under the label *intermodality* began about a decade ago (cf. Marsh/White 2003; Martinec/Salway 2005). Siefkes (forthcoming) proposes a formal approach that integrates intermodal interactions into Segmented Discourse Representation Theory (SDRT). Generally speaking, multimodal texts are usually assumed to presuppose not only the participation of different modes, but also some degree of intermodal integration; thus, in every multimodal text interpretation, intermodal effects play a certain role.

## 1.2 Intermodal influences between styles

In this article, we introduce a first approach to intermodal effects between matching and mismatching styles presented in different modes. Specifically, we assume that styles in different cultural areas that are designated with the same category name (e.g., "baroque" music and "baroque" architecture) have some features in common[3]; when we regard music and architecture as semiotic modes, we can investigate the intermodal congruence and incongruence of styles as a special case of intermodality, which in turn is an important aspect of multimodality research. The methods presented in this paper, which are mostly from the area of experimental psychology, are therefore directly relevant for multimodality research, since they enable the experimental verification of hypotheses about how modes work together in multimodal texts.

Style is a principle of cognitive organisation that is used to recognise repeating patterns in artefacts and behaviour, independently of function, specific content, and context (cf. Siefkes 2012: 38–43; 2013: 4–5). When, for example, an architect designs buildings with different functions and sizes, under different conditions of location, terrain, contractual demands, and budget, we are still sometimes able to recognise the architect's characteristic style. Styles can be defined across domains and have been described in many areas of culture and daily behaviour.[4]

Though styles exist that are domain-specific, some styles can be defined as sets of general stylistic features that are applicable (sometimes with domain-specific adjustments) to different domains of culture such as music, architecture, art, and literature. Such styles can therefore be applied to establish intermodal congruence

3 See the explanation of epochal styles in Siefkes 2012: 415–419.

4 See Siefkes 2012 for an integrative approach to style in different domains; an introduction is given in Siefkes 2013.

or incongruence. In the studies reported here, we take advantage of this aspect of style and use the domain-crossing styles 'baroque music/architecture' and 'modern music/architecture'.

In a pilot study planned and executed in 2012, we wanted to test the possibility that styles interact across perceptual modes. We investigated this specifically for the influence of musical style on the perception of architectural style and formulated the general hypothesis that music played inconspicuously in the background would influence judgments of architectural style. This pilot study was implemented at the University IUAV of Venice with students as participants. It confirmed the general hypothesis but also pointed towards an even more interesting possibility: Some scales seemed to be applied involuntarily to the congruence/incongruence-relation between the modes themselves. The study hinted at the possibility that context factors as well as characteristics of participants might influence the results; therefore, we decided to conduct two studies with the same materials but different participant pools, one of them an online study.

## 2 The studies: Hypotheses and predictions

We wanted to test the interaction of baroque and modern architecture with baroque and modern music in a 2×2 factorial design (for the factors background music and congruence/incongruence). In our studies, the semantic differential approach (cf. Osgood/Suci/Tannenbaum 1957) was used with the intention of measuring semantic values ascribed to styles. On the basis of previous work on style (cf. Siefkes 2011, 2012), which considers style as a multimodal phenomenon, we devised a list of semantic scales that we assumed might be relevant both for architectural and musical styles.

Based on previous research on the perception of architectural styles (cf., e.g., Mastandrea et al. 2011) and the pilot study, we made the following predictions:

> H1: Semantic scales measure stylistic differences: Different architectural styles are differently rated on some (adequately selected) semantic scales, which capture perceptual qualities and/or semantic associations that viewers connect with these styles.

We use the term *semantic scale* to refer to rating scales that are defined by a pair of bipolar adjectives such as *high – low*, or nouns such as *reason – feeling*. A *semantic differential* consists of a number of such scales, and viewers are asked to rate stimuli on these scales, allowing the researchers to determine if the stimulus in question is associated with one or the other of the polar terms, and how strong the associations are. The first hypothesis states that viewers of architecture are able to express some of the stylistic qualities they experience with the help of semantic scales.

H2: Stylistic cross-over effect: Music affects the stylistic judgment of architectural buildings on the basis of perceptual qualities and/or semantic associations connected with the musical style.

This hypothesis implies that viewers asked to judge architectural styles are not able to completely block out the stylistic qualities of music they hear. A sad musical piece, for instance, might transmit the quality "sadness" to the evaluation of the architecture.

H3: Stylistic congruence effect: The stylistic congruence or incongruence between architectural and musical styles has a specific effect on some perceptual qualities or semantic associations of the architecture, resulting in an influence on certain rating scales which are suitable for measuring congruence or incongruence between styles.

Hypothesis H3 entails two separate, more specific hypotheses: (a) subjects have a unitary perception of the meaning of general stylistic categories such as *baroque* across perceptual modes (visual vs. auditory stimuli) as well as across semiotic modes (architecture vs. music); (b) subjects transfer the intermodal perception of congruity or incongruity to their ratings of architecture – in other words, they are not able to completely separate the intermodally experienced congruence or incongruence from the rating of a single mode.

The pilot study had indicated that the semantic scales *balanced – unbalanced* and *incomplete – complete* might be susceptible to the effect stated in H3. We additionally introduced a third rating scale which should also register the effect *coherent – incoherent*. We postulated that these three scales are applicable to intermodal congruence or incongruence; if H3 is valid, the effect will show in these three scales. The other scales that we tested (see Section 2.1) were not supposed to be susceptible to this effect.

The two studies (I and II) were conducted in two completely different settings, in order to identify factors that could lead to results that are context specific. In the first study, we made use of crowdsourcing as a method of online research (see below for details of the study design). Crowdsourcing can be defined as a job given to an undefined group of people in the form of an open call on the World Wide Web. Evaluations of crowdsourcing participant pools (cf. Buhrmester et al. 2011) have shown that subjects show a broader age range, higher average age, and more diverse socio-economic backgrounds than conventional subject pools at universities or research centres. The validity of scientific results garnered in crowdsourcing has repeatedly been demonstrated: Paolacci/Chandler/Ipeirotis (2010) replicated some classical cognitive experiments on reasoning and found out that results were identical to those from the laboratory experiments. Buhrmester et al. (2011) compared studies conducted

on Mechanical Turk (one of the largest platforms for crowdsourcing) with respect to several psychometric scales and found no meaningful differences between the populations.

## 2.1 Study I

626 subjects participated in this study. All participants came from the USA, ranging over most states and a large number of different cities.[5] The stimulus material used in this study consisted of four videos, each 25 seconds in length. Altogether, 10 pictures of buildings and two musical pieces were used (see Table 13.1).

Four videos were created from these stimuli. Video 1 combined baroque architecture with baroque music (combining stimuli that were intermodally congruent in regard to style). Video 2 combined baroque architecture with modern music (intermodally incongruent). Video 3 combined modern architecture with baroque music (intermodally incongruent), and video 4 combined modern architecture with modern music (intermodally congruent).

5 By using crowdsourcing, we could target English native speakers for our study; furthermore, we could reach a more diverse population (especially regarding socio-economic background) in a different country (USA). Our job was evaluated by the participants, in an optional anonymous feedback given on the Crowdflower website, with an overall rating of 4.1 on a scale from 1 to 5 (where five was best); ca. 50% of participants gave feedback. Effective controls against multiple answers (subjects entering their results under two different entities) were implemented; options were enabled that prevented multiple answers on the basis of IP address and other data.

*Table 13.1: List of stimuli and sources used in both studies*

| Style | Mode | Stimulus | Source |
|---|---|---|---|
| Baroque | Architecture | San Carlo alle Quattro Fontane, Rome, Italy (*Francesco Borromini*) | http://commons.wikimedia.org/wiki/File:SCarloQuattroFontaneRome2.jpg, © Welleschik |
| | | Einsiedeln Abbey, Einsiedeln, Switzerland (*Caspar Moosbrugger*) | http://commons.wikimedia.org/wiki/File:Kloster_Einsiedeln_Frontansicht.jpg, © Hofec |
| | | Palais du Luxembourg, Paris, France (*Salomon de Brosse, Alphonse de Gisors*) | http://commons.wikimedia.org/wiki/File:Palais_Luxembourg_Sunset_Edit.JPG, © Benh Lieu Song, modified by Sanchezn |
| | | Theatinerkirche, Munich, Germany (*Agostino Barelli, Enrico Zuccalli*) | http://commons.wikimedia.org/wiki/File:MünchenTheatinerkirche_a.jpg, © Uwe Barghaan |
| | | Stift Melk, Melk, Austria (*Jakob Prandtauer*) | http://de.wikipedia.org/wiki/Datei:Stift_Melk_001.jpg, © User:Fb78 (commons.wikimedia.org) |
| | Music | *Georg Philipp Telemann*, Concerto grosso in D major, TWV 54:D3 (1716), Intrada – Grave (interpreted by *Trevor Pinnock / The English Concert*, 0:00–1:12) | http://www.youtube.com/watch?v=uZ80V-YWEnc, © Deutsche Grammophon |
| Modern | Architecture | SAS Royal Hotel, Copenhagen, Denmark (*Arne Jacobsen*) | http://commons.wikimedia.org/wiki/File:SAS_Royal_Hotel,_Copenhagen,_1955–1960.jpg, © seier+seier |
| | | Aalto-Hochhaus, Bremen, Germany (*Alvar Aalto*) | http://commons.wikimedia.org/wiki/File:AaltoVahr-01.jpg, © Jürgen Howaldt |
| | | Wohnstadt Carl Legien, Berlin, Germany (*Bruno Taut*) | http://commons.wikimedia.org/wiki/File:Berlin_C_Legien_Trachtenbrodtstr_26.jpg, © Doris Antony |
| | | Georgia Pacific Building, Atlanta, Georgia, USA (*Skidmore, Owings & Merrill*) | http://commons.wikimedia.org/wiki/File:Georgia_pacific_building_atlanta_01.jpg, © Ashley Moore |
| | | Neue Nationalgalerie, Berlin, Germany (*Ludwig Mies van der Rohe*) | http://commons.wikimedia.org/wiki/File:Neue_Nationalgalerie_Berlin.jpg, © Manfred Brückels |
| | Music | *Philip Glass*, Soundtrack to the film *Koyaanisqatsi*, track 3: "Cloudscape" (0:00–1:12) | http://www.youtube.com/watch?v=5SUpWrSVaIg, © Philip Glass |

Subjects had to evaluate the style of the buildings shown in the videos, giving their general aesthetic preference, and rating them on different semantic scales describing expressive, stylistic and aesthetic qualities. The scales were presented as sliders which were movable by mouseclick (with numerical values from 0 to 101; sliders were preset to 50). The first scale measured general aesthetic preference and was labelled with *not at all – very much*. Subjects were thus asked to answer the question "Do you like the style of the buildings in the video?" In a next step, they were then instructed: "Please judge the style on the following dimensions", using the scales: *introverted – extraverted, unbalanced – balanced, bright – dark, incoherent – coherent, grave – agitated, modest – bold, reason – feeling,* and *complete – incomplete*.

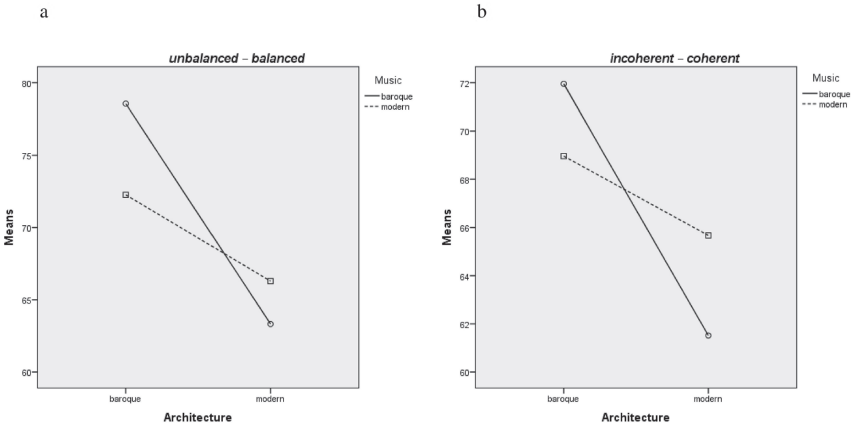
Judgments (n = 626) were ordered and paid via Crowdflower.com; the study was hosted on the online study service Fluidsurveys.com. A validator was given at the end of the study, which had to be entered on the Crowdflower website. Incompletely filled-out surveys were not evaluated. The study logic was programmed to randomly assign participants, after a general instruction at the beginning of the study, to one of four conditions. In each condition, one of the four videos was shown. Instructions and all rating scales were given directly below the video. Videos were uploaded on YouTube and embedded in the survey; they did not run automatically, but were started by clicking on them (visualised by a large arrow as start symbol) and could be paused or navigated by the participant via a navigation bar under the video. After submitting their answers by clicking "Next" at the bottom of the page, subjects were asked "Did you hear the background music?" (Yes/No). There was no "Back" button in the survey, and later editing of the results was disabled. 125 subjects that answered "No" were excluded from analysis. The 501 subjects that answered "Yes" took a mean of 01:31.3 min (*SD* = 01:17.4) for the study (measured by the Fluidsurveys website). For each of the nine sliders as dependent variables, a 2×2 ANOVA was conducted with *architecture* and *music* as factors.

For all dependent variables, a significant main effect (for $\alpha = 0.05$) was found for the factor *architecture*, with the exception of *grave – agitated*. These results indicate that the subjects were able to meaningfully apply the given scales to the presented stimulus material, expect possibly for the scale *grave – agitated*. Either the participants were not able to apply this judgment scale to architecture, which might be explained with the fact that it is more often used in the context of music than of architecture. However, the possibility cannot be excluded that both styles were simply experienced as very similar on this scale. However, given the clear differences between the two styles on all other scales, the explanation that this scale was not successfully applied to the stimulus material seems more likely.

No main effect was found for the factor *music*. For the interaction *architecture* × *music*, effects were found for two of the nine ANOVAs. For the scale

*unbalanced – balanced* as dependent variable, as well as the main effect for *architecture*, the interaction *architecture* × *music* was significant (see Figure 13.1a). For the scale *incoherent – coherent* as dependent variable, as well as the main effect for *architecture*, the interaction *architecture* × *music* was significant (see Figure 13.1b).

*Figure 13.1: Selected results of the first study. (a) Mean value diagram for the semantic scale* unbalanced – balanced *in study I. Baroque architecture is judged as more balanced (F(1, 497) = 32.510, p = 0.000; baroque arch. M = 75.41, SD = 1.37, modern arch. M = 64.81, SD = 1.26). In the interaction effect architecture × music, architecture was judged as more balanced when it was accompanied by music in the same style (F(1, 497) = 6.208, p = 0.013; baroque arch./baroque music M = 78.55, SD = 1.81, baroque arch./modern music M = 72.26, SD = 2.05, modern arch./baroque music M = 63.32, SD = 1.83, modern arch./modern music M = 66.29, SD = 1.73).*
*(b) Mean value diagram for the semantic scale* incoherent – coherent *in study I. Baroque architecture was judged as more coherent (F(1, 497) = 14.935, p = 0.000; baroque arch. M = 70.46, SD = 1.31, modern arch. M = 63.59, SD = 1.20). In the interaction effect architecture × music, architecture was judged as more balanced when it was accompanied by music in the same style (F(1, 497) = 4.047, p = 0.045; baroque arch./baroque music M = 71.95, SD = 1.73, baroque arch./modern music M = 68.96, SD = 1.96, modern arch./baroque music M = 61.52, SD = 1.74, modern arch./modern music M = 65.67, SD = 1.66)*




On the basis of hypothesis H3, we would expect two results: firstly, that the three semantic scales *unbalanced – balanced, incoherent – coherent* and *complete – incomplete* would vary with the stylistic congruence or incongruence between

modes. Secondly, that the direction of this effect would be as follows: for stylistic congruence (baroque stimuli in both modes, or modern stimuli in both modes), the judgments would deviate towards *balanced*, *coherent*, and *complete*. This effect was confirmed for two of the three scales (see Figure 13.1). For *complete – incomplete*, although the means displayed differences in the expected directions, no significant effect was found.

## 2.2 Study II

The second study was conducted at the University IUAV of Venice. It used participants drawn from a subject pool available at the university. A total of 77 students from the degree programmes in architecture and fashion design participated in the study. None reported any impairments in colour vision. The semantic scales of the first study were used again in the second study, with the following technical differences: The participants used a 7-point Likert scale for rating, and they rated on paper, crossing one of seven boxes for the scale points.

The same four videos were used as in study I, but with one difference: each participant rated both architectural styles. Participants were thus assigned to one of two groups, which were tested separately: one group was shown both architectural styles with baroque music, the second both styles with modern music. It was explained that the videos were taken from the Internet, but no mention of the music was made.[6] In the instructions, participants were asked to look at the pictures and rate the architectural style of the buildings that were shown. The subjects could read the semantic scales before the presentation of the video. Each video was then shown twice to the participants. Since each participant was shown two videos and judged both architectural styles, study II employed a mixed design with *architecture* as within-subjects factor and *music* as between-subjects factor. The within-subjects design has statistical advantages, since it allows the factoring out of individual differences by directly comparing the judgments each subject gives for both styles.

In the two-factorial design, two factors and one interaction were tested for ratings on nine scales as dependent variables. It is noteworthy that four main effects

6 It was assumed that this would make the music less conspicuous, since many user-created videos on Internet video sites such as YouTube or Vimeo show sequences of images with musical background.

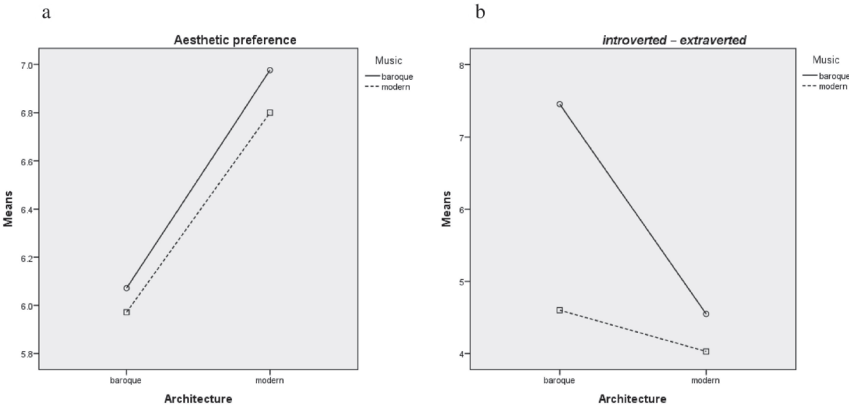
for the factor "architecture" from the online study were confirmed in the second study: baroque architecture is judged to be more *extraverted*, modern architecture is judged as *darker*, *bolder*, and connected more with *feeling*. For the rating scale *grave – agitated*, neither the first nor the second study found a significant main effect. An interesting difference between the studies concerns general aesthetic preference: whereas the participants of the online study had preferred baroque architecture, the study subjects in the second study preferred modern architecture (see Figure 13.2a). This difference might be explained by the fact that study II was conducted at a university of design and architecture. The students that participated in this study were not laypersons with regard to architecture; although some of them studied fashion design, they were probably influenced by a professional bias towards modern architecture.

Turning to intermodal influences, we found two main effects for *music*: baroque music influences judgments of the architecture it accompanies towards *extraverted* (see Figure 13.2b), and modern music influences judgments towards *dark*. Since the instructions clearly asked participants to judge the architecture, but did not mention the music, we call such effects 'carry-over effects': the music significantly influenced the judgment of the architecture on these two scales. These general carry-over effects indicate that qualities of the background music were inadvertently applied to judgments of the architectural style.

In addition, a second type of intermodal effect was found, where the relation between architectural and musical styles influenced the ratings of the architecture. In the chosen study design, effects such as these show up in the form of an interaction effect *architecture* × *music*, indicating that the two factors influenced each other. In study II, interaction effects were found for all three scales where they had been expected (see above), namely for the scales *unbalanced – balanced* (see Figure 13.2c), *incoherent – coherent* (see Figure 13.2d), and *complete – incomplete*. For all three semantic scales, group means varied in the semantically plausible direction: in intermodally congruent presentations (i.e., where stimuli in both modes had the same style), the architecture was perceived as more *balanced*, more *coherent*, and more *complete*.

However, we should consider that three further interaction effects were found: architecture accompanied by music in the congruent style was also judged to be more *extraverted*, *agitated*, and *bold*. For the scale *introverted – extraverted* (see Figure 13.2b), we interpret this effect by positing that the carry-over effect (the main effect for *music*) applied only to the baroque music: the mean values for modern architecture are very close together, indicating that only baroque and not modern architecture was experienced as more *extraverted* with baroque music playing in the background.

*Figure 13.2: Selected results of the second study*
*(a) Mean value diagram for the rating scale Aesthetic preference in study II. Modern architecture was preferred (F(1, 75) = 4.930, p = 0.029; baroque arch. M = 6.03, SD = 2.05, modern arch. M = 6.90, SD = 2.36). To allow comparison with the other figures, the diagram shows the mean group values split for the factors* Architecture *and* Music.
*(b) Mean value diagram for the semantic scale* introverted – extraverted *in study II. For this scale, main effects for architecture and music were found, as well as an interaction effect. Baroque architecture was judged as more extraverted (F(1, 75) = 24.362, p = 0.000; baroque arch. M = 6.16, SD = 2.48, modern arch. M = 4.31, SD = 2.16). Architecture accompanied by baroque music was judged as more extraverted (F(1, 75) = 26.499, p = 0.000). In the interaction effect* architecture × music*, architecture was judged as more extraverted when it was accompanied by music in the same style (F(1, 75) = 10.977, p = 0.001; baroque arch./baroque music M = 7.45, SD = 1.99, baroque arch./modern music M = 4.60, SD = 2.10, modern arch./baroque music M = 4.55, SD = 2.33, modern arch./modern music M = 4.03, SD = 1.93).*
*(c) Mean value diagram for the semantic scale* unbalanced – balanced *in study II. In the interaction effect* architecture × music*, architecture was judged as more balanced when it was accompanied by music in the same style (F(1, 75) = 21.737, p = 0.000; baroque arch./baroque music M = 6.24, SD = 1.67, baroque arch./modern music M = 4.26, SD = 2.21, modern arch./baroque music M = 3.95, SD = 2.19, modern arch./modern music M = 5.29, SD = 2.94).*
*(d) Mean value diagram for the semantic scale* incoherent – coherent *in study II. In the interaction effect* architecture × music*, architecture was judged as more coherent when it was accompanied by music in the same style (F(1, 75) = 18.300, p = 0.000; baroque arch./baroque music M = 7.10, SD = 2.03, baroque arch./modern music M = 4.83, SD = 2.54, modern arch./baroque music M = 4.90, SD = 2.92, modern arch./modern music M = 6.06, SD = 2.31).*



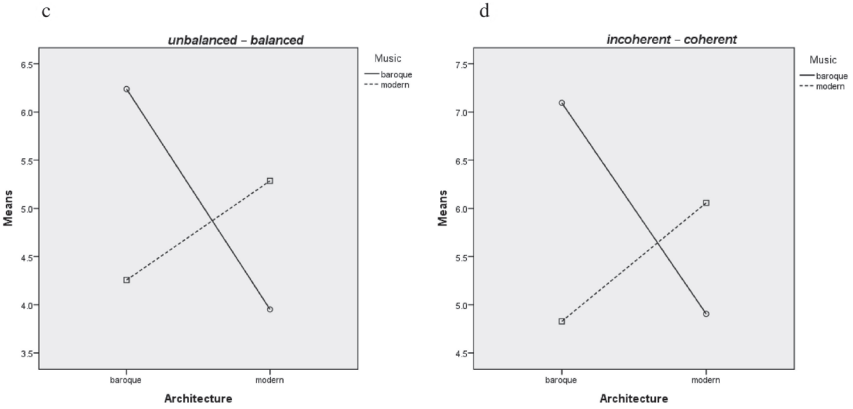



*Table 13.2: Overview of significant effects found in both studies (α = 0.05)*

| **Dependent variables** | **Independent variables (Factors)** | | |
|---|---|---|---|
| **(Rating scales)** | **Architecture** | **Music** | **Architecture × Music** |
| | **Levels: baroque/ modern** | **Levels: baroque/ modern** | **Levels: intermodally congruent/ intermodally incongruent** |
| aesthetic preference | | | |
| *introverted – extraverted* | baroque is rated more towards *extraverted* | | |
| *unbalanced – balanced* | | | architecture with music in the same style (= intermodally congruent) is rated more towards *balanced* |
| *bright – dark* | modern is rated more towards *dark* | | |
| *incoherent – coherent* | | | architecture with music in the same style (= intermodally congruent) is rated more towards *coherent* |
| *grave – agitated* | | | |
| *modest – bold* | baroque is rated more towards *bold* | | |
| *reason – feeling* | baroque is rated more towards *feeling* | | |
| *complete – incomplete* | | | |

Due to the different methods, study settings, and participant pools used in the two studies, we can regard the results found by both studies as well confirmed (see Table 13.2). However, we should not disregard entirely those results that are found only in one study: the distinct patterns in both studies indicate that the two study designs were variously sensitive for specific effects. The high numbers of participants made study I especially sensitive, allowing relatively weak effects to become statistically significant. This higher sensitivity may explain why differences between the styles were found for all but one of the rating scales. The exception is the scale *grave – agitated*, which did not display significant differences in ratings in either study. The second study demonstrated a higher sensitivity for all effects where the music influenced ratings of the architecture. Thus, only the second study found two carry-over effects, where musical qualities were attributed to the architecture that was to be rated. Furthermore, the second study found six stylistic congruence effects, only two of which were manifest in the first study. It is quite possible that the online study design alleviated the influence of the music. It could be speculated that some Internet users have learnt to block out background music, or that some listened to it at a low volume.

## 3 General discussion

Four main effects for the factor *architecture* coincide in both studies: baroque architecture is judged more towards *extraverted*, *bright*, *bold*, and *feeling*, whereas modern architecture is connected with the opposing poles *introverted*, *dark*, *modest*, and *reason*. Thus, hypothesis H1 is confirmed: Semantic scales are able to capture at least some aspects of the experience of architectural style.

It can be concluded that even very broad architectural style concepts such as *modern* and *baroque*, which generalise over a wide range of buildings with different functions and architects, not to mention times and places of construction, are consistently connected with specific semantic scales. This is by no means obvious: it could be assumed that these style categories are too general to be correlated with specific perceptual qualities or semantic associations. Not all scales, however, were successfully applied to the two styles: the scale *grave – agitated* was probably unsuitable for judging the architectural styles, and for some of the other scales, only one study found an effect. On the basis of these results, we conjecture that a limited set of rating scales can be compiled that is consistently and non-randomly applied by subjects to a specific style.

The study did not investigate the nature of the connection between styles and semantic scales. It might be conjectured that some semantic scales capture certain

perceptual qualities of the two styles, allowing the participants to express aspects of their experience in viewing the architecture. An alternative explanation would be that the styles in question are recognised as belonging to a category and the respective natural-language terms (e.g. *modern* or *baroque*, for English viewers) came to mind; the participants may then have rated the styles on the basis of semantic associations (connotations) of these terms.

The second hypothesis concerned 'carry-over effects' between musical and architectural styles, proposing that qualities attributed to the music would influence the ratings of architecture asked of the participants. The results remain inconclusive: Whereas in study I, no semantic scale varied significantly with the background music, study II showed two effects: baroque music influenced the ratings of architecture towards *extraverted*, while modern music influenced the ratings towards *dark*. Further studies will be necessary to determine if some semantic scales consistently foster such carry-over effects.

The third hypothesis is confirmed by both studies: intermodal congruence or incongruence of styles influences the judgment of architecture for semantic scales such as *unbalanced – balanced* and *incoherent – coherent* and possibly for *complete – incomplete* (the second study did not confirm an effect for this last scale, but it was present in the pilot study). This points towards an important interaction effect between modes: for rating scales which are inherently applicable to intermodal relations between styles, the experience of stylistic congruence or incongruence 'imposes' itself onto aesthetic perception and judgment. This is an important result. Participants in both studies were explicitly asked to concentrate on and evaluate the style of the buildings they were shown, but they involuntarily adapted their ratings on these scales to the stylistic relation between the music and architecture they experienced. Theories of perception assuming the modularity of sensory processing, which have long dominated in psychology (cf. Fodor 1983, for a recent discussion see Deroy 2014), have recently been amended to incorporate multisensory processes (cf. Calvert et al. 2004) and these should include intermodal congruence effects as well. Similarly, research on intermodality in linguistics and semiotics should consider such effects, and focus more often on experimental studies that can help formulate and verify the principles of intermodal interaction.

## 4 Conclusion

For multimodality research, it is of central importance to understand how semiotic modes work together in multimodal texts. However, although a number of approaches for describing and annotating relations between modes have been

developed (cf. Schriver 1997; Marsh/White 2003; Martinec/Salway 2005; Baldry/Thibault 2006; Liu/O'Halloran 2009; O'Halloran/Smith 2011; Bateman 2011, 2014), the mechanisms that underlie interactions between semiotic modes are still insufficiently explained.

In practice, multimodal analyses often approach discourse either by concentrating on the separate semiotic modes, or by assuming that meaning is produced by all modes together, licensing researchers to draw on each mode at will when explaining the overall meaning. However, apart from mode-specific contributions and holistically produced meanings, specific interactions between semiotic modes, where the presence of one mode has a definable influence on the formal, semantic, and/or stylistic properties of another mode, play a significant role in the creation of multimodal meaning. It is thus important to understand which types of interactions can be assumed (cf. Oviatt 1999) and to integrate them into a general model of text/discourse analysis that adequately represents all participating modes and textual levels.

The methods presented in this paper allow the measurement and comparison of meanings connected with semiotic artefacts that belong to different semiotic and/or perceptual modes. It was demonstrated how intermodal influences can be detected as rating differences on semantic scales, and how statistical methods are used to verify that these differences are caused by aspects of the other mode, or relations between the modes such as congruence or incongruence, and not by confounding factors. Other combinations of modes, and other aspects apart from stylistic qualities, can be investigated with the methods detailed in this paper; we therefore propose that the used methods can be regarded as an experimental approach to multimodality.

## References

Baldry, Anthony/Thibault, Paul (2006): *Multimodal Transcription and Text Analysis*. London: Equinox.

Bateman, John A. (2011): "The Decomposability of Semiotic Modes". In: O'Halloran, Kay L./Smith, Bradley A. (eds.) (2011): *Multimodal Studies. Exploring Issues and Domains*. London: Routledge, pp. 17–38.

Bateman, John A. (2014): *Text and Image. A Critical Introduction to the Visual-Verbal Divide.* New York: Routledge.

Bateman, John A./Schmidt, Karl-Heinrich (2012): *Multimodal Film Analysis. How Films Mean*. London: Routledge.

Bateman, John A./Wildfeuer, Janina (2014): "A Multimodal Discourse Theory of Visual Narrative". *Journal of Pragmatics* 74, pp. 180–208.

Buhrmester, Michael/Kwang, Tracy/Gosling, Samuel D. (2011): "Amazon's Mechanical Turk: A New Source of Inexpensive, yet High-Quality, Data?" *Perspectives on Psychological Science* 6(1), pp. 3–5.

Calvert, Gemma A./Spence, Charles/Stein, Barry E. (2004): *The Handbook of Multisensory Processes.* Cambridge: MIT Press.

Chapados, Catherine/Levitin, Daniel J. (2008): "Cross-Modal Interactions in the Experience of Musical Performances. Physiological Correlates". *Cognition* 108, pp. 639–651.

Davidson, Jane W. (1993): "Visual Perception of Performance Manner in the Movements of Solo Musicians". *Psychology of Music* 21, pp. 103–113.

Deroy, Ophelia (2014): "Modularity of Perception". In: Matthen, Mohan (ed.): *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199600472.013.028.

Fodor, Jerry A. (1983): *Modularity of Mind. An Essay on Faculty Psychology*. Cambridge: MIT Press.

Fricke, Ellen (2012): *Grammatik multimodal. Wie Wörter und Gesten zusammenwirken.* Berlin/New York: de Gruyter.

Hess-Lüttich, Ernest W.B./Wenz, Karin (eds.) (2006): *Stile des Intermedialen. Zur Semiotik des Übergangs*. (= *Kodikas/Code. Ars Semeiotica* 29, 1–3). Tübingen: Narr.

Kress, Gunther/van Leeuwen, Theo (2001): *Multimodal Discourse. The Modes and Media of Contemporary Communication.* London: Arnold.

Liu, Yu/O'Halloran, Kay L. (2009): "Intersemiotic Texture. Analysing Cohesive Devices between Language and Images". *Social Semiotics* 19(4), pp. 367–388.

Machin, David (ed.) (2014): *Visual Communication*. Handbooks of Communication Science 4. Berlin: de Gruyter.

Marsh, Emily E./White, Marilyn D. (2003): "A Taxonomy of Relationships between Images and Text". *Journal of Documentation* 59(6), pp. 647–672.

Martinec, Radan/Salway, Andrew (2005): "A System for Image-Text Relations in New (and Old) Media". *Visual Communication* 4(3), pp. 339–374.

Mastandrea, Stefano/Bartoli, Gabriella/Carrus, Giuseppe (2011): "The Automatic Aesthetic Evaluation of Different Art and Architectural Styles". *Psychology of Aesthetics, Creativity, and the Arts* 5(2), pp. 126–134.

McGurk, Harry/MacDonald, John (1976): "Hearing Lips and Seeing Voices". *Nature* 264, pp. 746–748.

Molholm, Sophie/Ritter, Walter/Javitt, Daniel C./Foxe, John J. (2004): "Multisensory Visual-Auditory Object Recognition in Humans. A High-Density Electrical Mapping Study". *Cerebral Cortex* 14, pp. 452–465.

O'Halloran, Kay L./Smith, Bradley A. (eds.) (2011): *Multimodal Studies. Exploring Issues and Domains*. London: Routledge.

Osgood, Charles E./Suci, George J./Tannenbaum, Percy H. (1957): *The Measurement of Meaning*. Urbana: University of Illinois Press.

Oviatt, Sharon L. (1999): "Ten Myths of Multimodal Interaction". *Communications of the ACM* 42 (11), pp. 74–81.

Paolacci, Gabriele/Chandler, Jesse/Ipeirotis, Panagiotis G. (2010): "Running Experiments on Amazon Mechanical Turk". *Judgment and Decision Making* 5, pp. 411–419.

Schriver, Karen A. (1997): *Dynamics in Document Design: Creating Texts for Readers.* New York: Wiley.

Siefkes, Martin (2011): "Style: A New Semiotic View on an Old Problem". *Kodikas/Code. Ars Semeiotica* 34 (1–2), pp. 15–25.

Siefkes, Martin (2012): *Stil als Zeichenprozess. Wie Variation bei Verhalten, Artefakten und Texten Information erzeugt*. Würzburg: Königshausen & Neumann.

Siefkes, Martin (2013): "Stil und Gesellschaft. Plädoyer für eine Allgemeine Stilistik". Research report, University IUAV of Venice. Online: http://siefkes.de/dokumente/Siefkes_Stil_und_Gesellschaft.pdf [last accessed: 01 July 2015].

Siefkes, Martin (forthcoming): "How Semiotic Modes Work Together in Multimodal Texts: Defining and Representing Intermodal Interactions". 10plus1 – Living Linguistics 1, 2015.

Siefkes, Martin/Schöps, Doris (2013 [2014]): *Neue Methoden der Diskursanalyse. Zeitschrift für Semiotik* 34(1–2).

Spence, Charles (2007): "Audiovisual Multisensory Integration". *Acoustical Science and Technology* 28, pp. 61–70.

Vatakis, Argiro/Spence, Charles (2008): "Evaluating the Influence of the Unity Assumption on the Temporal Perception of Realistic Audiovisual Stimuli". *Acta Psychologica* 127, pp. 12–23.

van Atteveldt, Nienke/Formisano, Elia/Goebel, Rainer/Blomert, Leo (2004): "Integration of Letters and Speech Sounds in the Human Brain". *Neuron* 43, pp. 271–282.

Vines, Bradley W./Krumhansl, Carol L./Wanderley, Marcelo M./Levitin, Daniel J. (2006): "Cross-Modal Interactions in the Perception of Musical Performance". *Cognition* 101(1), pp. 80–103.

Wildfeuer, Janina (2012): "Intersemiosis in Film: Towards a New Organization of Semiotic Resources in Multimodal Filmic Text". *Multimodal Communication* 1(3), pp. 276–304.

Wildfeuer, Janina (2014): *Film Discourse Interpretation. Towards a New Paradigm for Multimodal Film Analysis.* London/New York: Routledge.

## List of Figures



## List of Tables