



*Legibility of Musical Scores
and Parallels
with Language Reading*



VIDAR ARILD STENBERG

GONVILLE & CAIUS COLLEGE

SUBMISSION DATE: JUNE 2017

THIS DISSERTATION IS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DECLARATION

DECLARATION

I, Vidar Arild Stenberg, hereby declare that:

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the word limit prescribed by the Degree Committee of the Faculty of Music.

Cambridge, June 2017

TABLE OF CONTENTS

TABLE OF CONTENTS

SUMMARY [xii]

CHAPTER 1 [1]

Reading Natural and Formal Languages

1.1. RELEVANCE OF READING RESEARCH IN
DIFFERENT DOMAINS [2]

1.1.1. HUMANS AND VISUALLY MEANINGFUL PATTERNS

1.1.2. READING PARADIGMS WITHIN EXPERIMENTAL PSYCHOLOGY

1.1.3. INTEGRATION OF RESEARCH AND INSTRUCTION

1.2. THE EVOLUTION OF READING AND WRITING OF
NATURAL LANGUAGE SCRIPTS [7]

1.2.1. ANCIENT READERS AND UNSEPARATED TEXTS

1.2.2. SEPARATED SCRIPT: A SUCCINCT OVERVIEW

1.2.3. NEW GENRES AND NEW MODES OF READING AND WRITING

1.2.4. WRITTEN CULTURE AT THE BEGINNING OF MODERN TIMES

1.2.5. TEXT DESIGN AND THE ARTS OF MEMORY

1.3. THE EVOLUTION OF READING AND WRITING OF
SCIENTIFIC SCRIPTS [20]

1.3.1. NUMBERS AND FORMULAS AS NOTATION

1.3.2. PARALLELS BETWEEN TEXTUAL AND NUMERICAL REFORMATTING

1.3.3. EVOLUTION OF THE INSTRUCTION OF ARITHMETIC THINKING

1.3.4. SEPARATED SYLLOGISMS IN NASCENT SYMBOLIC LOGIC

1.4. INTEGRATION, AUTOMATISM AND PATTERNING IN
PROFICIENT READERS [27]

- 1.4.1. THEORETICAL MODELS AND EXPERIMENTAL STUDIES ON READING STRATEGIES
- 1.4.2. APPLICATIONS IN INSTRUCTIONAL AND DEVELOPMENTAL STUDIES
- 1.4.3. CONTEXTUALISATION AND RECOGNITION OF INFORMATION
- 1.4.4. ACCURACY AND HOLISTIC PROCESSING

CHAPTER 2 [33]

*Contexts, Models, and Mappings for Reading
Tasks*

2.1. DEFINITIONS OF SIGHT-READING AND
PARALLELS WITH LANGUAGE READING [34]

2.2. SIGNIFICANCE OF SIGHT-READING WITHIN
MUSICIANSHIP [36]

- 2.2.1. PROFESSIONAL SIGNIFICANCE
- 2.2.2. MUSIC READING AS A FORM OF LITERACY

2.3. COGNITIVE CLASSIFICATION OF THE TASK [43]

2.4. EXPLANATORY, DESCRIPTIVE, AND VISIO-
MOTORIC COMPARISONS [45]

- 2.4.1. MULTIMODAL EXPLANATORY MODELS FOR READING PROCESSES
- 2.4.2. MUSIC-SPECIFIC EYE MOVEMENTS
- 2.4.3. EYE MOVEMENTS AND PERCEPTUAL SPANS DURING READING

2.5. NEUROPHYSIOLOGICAL MAPPINGS OF READING
PROCESSES [51]

- 2.5.1. NEUROPHYSIOLOGICAL MAPPINGS OF VISUAL ENCODING PROCESSES
- 2.5.2. NEUROPHYSIOLOGICAL MAPPINGS OF STRUCTURAL INTEGRATION PROCESSES

2. 6. MUSIC READING AND THE LEXICOSEMANTIC
ROUTE [60]

- 2.6.1. SEMANTICITY IN INTRA-MUSICAL RELATIONS
- 2.6.2. SEMANTICITY THROUGH REFERENTIALISM
- 2.6.3. SEMANTICITY AND THE DETECTION OF MEANINGFUL GROUPINGS WITHIN SETS

2.7. DIMENSIONS OF THE READING TASKS [66]

- 2.7.1. SEPARATION OR CONCURRENCE OF DIMENSIONS
- 2.7.2. ORIENTING ATTENTION

2.8. IMPROVEMENTS OF READABILITY USING TEXT
DESIGN [69]

- 2.8.1. THE USES OF SPACING TO IMPROVE READABILITY
- 2.8.2. EFFECTS OF SPACING ON RECALL AND REPRODUCTION
- 2.8.3. RESEARCH ON PERFORMATIVE READINGS
- 2.8.4. APPLYING VISUAL CUES TO MUSICAL SCORES
- 2.8.5. TAKING UP SLOBODA'S GAUNTLET

CHAPTER 3 [75]

*Experiment I: Establishing Possible Effects
of Score Design on Reading*

3.1. INTRODUCTION: MOTIVATIONS FOR IMPROVING
READABILITY [76]

- 3.1.1. PROFESSIONAL PRACTICES WITH MINIMUM REHEARSALS
- 3.1.2. ELEVATED STRESS LEVELS IN MUSICIANS
- 3.1.3. MUSICIANS' OCCUPATIONAL ROLE CONCERNS
- 3.1.4. AN INITIAL ATTEMPT TO IMPROVE LEGIBILITY IN MUSICAL SCORES

3.2. METHOD: COMPARISONS OF PERFORMANCES
WITH DIFFERENT DESIGNS [80]

- 3.2.1. OVERVIEW OF AIMS AND STRATEGIES
- 3.2.2. PILOT STUDY

- 3.2.3. STATEMENT OF ETHICAL APPROVAL
- 3.2.4. PARTICIPANTS
- 3.2.5. MATERIALS
- 3.2.6. ENVIRONMENT AND PROCEDURE
- 3.2.7. RECORDING OF PERFORMANCES
- 3.2.8. OUTCOME MEASURES

3.3. RESULTS: UNWANTED EFFECT OF SURFACE STRUCTURE INTERFERENCES [92]

- 3.3.1. NON-NORMAL DISTRIBUTIONS AND OUTLIERS
- 3.3.2. PERMANENCE OF MISTAKES IRRESPECTIVE OF VERSION
- 3.3.3. MEASUREMENTS OF PERFORMANCE IMPROVEMENT AND NORMALISATION OF DATA
- 3.3.4. ANALYSIS OF MARK INCREASES IN THE SECOND READING
- 3.3.5. ANALYSIS OF EFFECTS OF SEQUENCES OF PRESENTATIONS ON THE SECOND READING
- 3.3.6. PREFERENCES BY PARTICIPANTS

3.4. DISCUSSION: CONSIDERATIONS FOR FURTHER EXPERIMENTAL DESIGNS [104]

- 3.4.1. HIGH DEGREE OF VARIABILITY AND DETECTED TRENDS
- 3.4.2. PERCEPTION BY PARTICIPANTS AND CORRELATION WITH PERFORMANCE
- 3.4.3. LIMITATIONS AND FUTURE PARADIGMS

CHAPTER 4 [107]

Experiment II: Confirming Effect of Score Design on Reading Fluency

4.1. INTRODUCTION: IMPROVING EXPERIMENTAL DESIGN AND AVOIDING INTERFERENCES [108]

- 4.1.1. RE-READING OF TEXTS AND WORKING MEMORY
- 4.1.2. RATIONALE AND HYPOTHESIS FOR EXPERIMENT II

4.2. METHOD: COMPARISONS OF PERFORMANCES
WITH TWO DESIGNS AND TWO PIECES [111]

- 4.2.1. OVERVIEW
- 4.2.2. PARTICIPANTS
- 4.2.3. EXPERIMENTAL MATERIALS
- 4.2.4. EXPERIMENTAL PROCEDURE
- 4.2.5. DATA COLLECTION ENVIRONMENT

4.3. RESULTS: SIGNIFICANT EFFECT OF DESIGN ON
NUMBER OF MISTAKES, FLUENCY, AND
ASSESSMENTS [123]

- 4.3.1. PITCH DOMAIN RESULTS
- 4.3.2. RHYTHM DOMAIN RESULTS
- 4.3.3. TOTAL NUMBERS OF MISTAKES RESULTS
- 4.3.4. NORMALISED TEMPO VARIATIONS RESULTS
- 4.3.5. BLIND ASSESSMENT BY EXPERT MUSICIANS
- 4.3.6. SELF-ASSESSMENT QUESTIONNAIRES

4.4. DISCUSSION: VALIDITY AND UNIVERSALITY OF
THE MAIN MODIFICATIONS [139]

- 4.4.1. ALL ASSESSMENTS INDICATING BETTER PERFORMANCES WITH THE MODIFICATIONS
- 4.4.2. THE IMPORTANCE OF SEPARATION OF DISCOURSE UNITS
- 4.4.3. THE UNIVERSALITY OF THE EFFECTIVENESS OF SEPARATIONS

CHAPTER 5 [143]

*Experiment III: Narrowing Possible Effects
and Increasing Ecological Validity*

5.1. INTRODUCTION: THE IMPORTANCE OF PAGE
ORIENTATION AND OTHER ECOLOGICAL VALIDITY
ISSUES [144]

- 5.1.1. CRITIQUES ON THE PREVIOUS EXPERIMENT
- 5.1.2. VERTICAL *VERSUS* HORIZONTAL ARRANGEMENTS OF TEXTS

5.1.3. STRUCTURED SPACING AND SEPARATION IN RELATION TO DIFFERENT TRADITIONS

5.1.4. IMPLEMENTING SEPARATION IN CONJUNCTION WITH SUPPORTING CUES

5.2. METHOD: COMPARISONS OF PERFORMANCES
WITH TWO DESIGNS WITH SAME ORIENTATION [148]

5.2.1. OVERVIEW

5.2.2. PARTICIPANTS

5.2.3. EXPERIMENTAL MATERIALS

5.2.4. PROCEDURE

5.2.5. DATA COLLECTION AND CODING

5.3. RESULTS: INCREASED EFFECT OF DESIGN ON
NUMBER OF ERRORS AND FLUENCY [151]

5.3.1. PITCH DOMAIN RESULTS

5.3.2. RHYTHM DOMAIN RESULTS

5.3.3. TOTAL NUMBERS OF MISTAKES RESULTS

5.3.4. NORMALISED PAIRWISE VARIATION INDEXES RESULTS

5.3.5. EXPRESSIVITY AND USE OF DYNAMIC RANGES

5.3.6. ASSESSMENT QUESTIONNAIRES

5.4. DISCUSSION: ADVANTAGES AND APPLICABILITY
OF THE MODIFIED DESIGN [164]

5.4.1. ADVANTAGES OF THE MODIFIED SCORES IN ALL ASSESSMENTS

5.4.2. ACCURACY, FLUENCY, AND REHEARSAL TIMES

CHAPTER 6 [167]

*Experiment IV: Testing the Universality of
the Novel Designs*

6.1. INTRODUCTION: ENCODING, SENSORIMOTOR
TRANSFORMATIONS, AND UNIVERSAL NOTATIONS
[168]

6.1.1. MAPPINGS IN RELATION TO INSTRUMENTAL PRACTICES

6.1.2. UNIVERSAL NOTATIONS *VERSUS TABULATURA*

6.2. METHOD: COMPARISONS OF PERFORMANCES
BY VIOLINISTS [173]

- 6.2.1. PARTICIPANTS
- 6.2.2. MATERIALS
- 6.2.3. ENVIRONMENT AND PROCEDURE
- 6.2.4. COLLECTION AND CODING OF DATA

6.3. RESULTS OF EXPERIMENT WITH VIOLINISTS [175]

- 6.3.1. PITCH DOMAIN RESULTS
- 6.3.2. RHYTHM DOMAIN RESULTS
- 6.3.3. TOTAL NUMBERS OF MISTAKES RESULTS
- 6.3.4. ASSESSMENT QUESTIONNAIRES

6.4. DISCUSSION: MODIFIED SCORES AND NOTATION
FAMILIARITY [184]

- 6.4.1. FEWER MISTAKES, HIGHER RESILIENCE, AND BETTER EVALUATION WITH MODIFIED VERSIONS
- 6.4.2. OVERRIDING FORMAT FAMILIARITY
- 6.4.3. ADVANTAGES OF STRUCTURED TEXTS FOR NON-NATIVE READERS

CHAPTER 7 [189]

*Experiment V: Separating the Design
Elements of the Modified Scores*

7.1. INTRODUCTION: INTEGRATION AND
INDIVIDUALIZATION OF DESIGN COMPONENTS [190]

- 7.1.1. INTEGRATION OF DIFFERENT DESIGN NOVELTIES IN PREVIOUS EXPERIMENTS
- 7.1.2. RATIONALE FOR EXPERIMENT V

7.2. METHOD: THREE EXERCISES MEASURING
EFFECT OF INDIVIDUAL DESIGN COMPONENTS [195]

- 7.2.1. PARTICIPANTS
- 7.2.2. MATERIALS
- 7.2.3. THE THREE EXERCISES AND THEIR INDEPENDENT VARIABLES

7.2.4. PROCEDURE

7.2.5. FINAL CONSIDERATIONS ABOUT SPACING

7.3. RESULTS: SIGNIFICANT EFFECT OF SEPARATION
BUT ONLY AT SPECIFIC LEVELS [205]

7.3.1. EXERCISE 1: SEPARATION OF LARGE STRUCTURAL UNITS

7.3.2. EXERCISE 2: SEPARATION OF SUB-PHRASAL DISCOURSE UNITS

7.3.3. EXERCISE 3: SEPARATION OF GRAPHEMES

7.3.4. PREFERENCES BY PARTICIPANTS

7.4. DISCUSSION: ASSESSMENTS OF LEVELS OF
SEPARATION [218]

7.4.1. EXERCISE 1: SEPARATION ONLY AT PHRASE LEVEL

7.4.2. EXERCISE 2: SEPARATION AT MANAGEABLE INFORMATIONAL LEVEL

7.4.3. EXERCISE 3: SEPARATION AT MINIMUM UNIT LEVEL

7.4.4. CONCURRENCE OF PARTICIPANTS' PREFERENCES

CHAPTER 8 [221]

*General Discussion and Future Research
Directions*

8.1. TRENDS THROUGH THE SERIES OF
EXPERIMENTS [222]

8.1.1. READING FLUENCY AND PERFORMANCE MODULATED BY DESIGN

8.1.2. MARKED INCREASES IN READING FLUENCY WITHIN THE RHYTHM DOMAIN

8.2. FUTURE RESEARCH: NEEDED IMPROVEMENTS IN
CODING AND IN EXPANDING METHODS [226]

8.2.1. PROBLEMATIC CODING OF MUSICAL PERFORMANCE MISTAKES

8.2.2. EXTENDING PARTICIPANT SAMPLES AND USE OF DIGITAL INTERFACES

8.2.3. LIMITATIONS IN THE CHOICES OF REPERTOIRE, PARTICIPANTS AND USAGES

ACKNOWLEDGEMENTS [231]

REFERENCES [233]

APPENDICES [253]

TO CHAPTER 3 [254]

- 3.A. PIECE 1
- 3.B. PIECE 2
- 3.C. PIECE 3

TO CHAPTER 4 [264]

- 4.A. PIECE 1
- 4.B. PIECE 2
- 4.C. EXAMPLE OF CODING

TO CHAPTER 5 [270]

- 5.A. PIECE 1
- 5.B. PIECE 2

TO CHAPTER 6 [278]

- 6.A. PIECE 1
- 6.B. PIECE 2

TO CHAPTER 7 [282]

- 7.A. EXERCISE 1 —PIECE 1
- 7.B. EXERCISE 1 —PIECE 2
- 7.C. EXERCISE 2 —PIECE 1
- 7.D. EXERCISE 2 —PIECE 2
- 7.E. EXERCISE 3 —PIECE 1
- 7.F. EXERCISE 3 —PIECE 2

- 7.G. LIST OF CHORALES —FIRST PAGE
- 7.H. LISTING OF ONSETS —FIRST PAGE

LIST OF FIGURES [297]

LIST OF TABLES [305]

SUMMARY

ARILD STENBERG

*Legibility of Musical Scores and
Parallels with Language Reading*

SUMMARY

Following on the extensive literature within experimental psychology on the reading of natural language texts, I have undertaken a series of experiments on the sight-reading of musical scores that have shown that spacing of information, the structuring of the musical discourse, and the predictability of design in a score can aid its legibility in a manner similar to what has been shown in the language domain.

Cultural studies of reading —particularly the works of Saenger— point in the same direction; according to these, the change in Medieval textual scripts from scriptura continua at the beginning of the eight century to the adoption of canonical separations between words, phrases, or paragraphs (which had fully spread throughout Europe by the mid-fourteenth century) significantly decreased the cognitive load and time that had previously been needed to decode a script. Crucially, this eliminated the need for the ancient techniques of the *praelectio* (initial decoding of the text by reading it aloud) and rote repetition for its comprehension, triggering a whole new culture of private fluent reading.

Equally, the literature on music sight-reading (although lacking in systematic research based on objective measurements of legibility of texts) has proposed, based on surveys and studies of expertise, a series of cognitive models of the activity that prime, as factors that distinguish proficient readers from beginners: the integration of discursive elements into higher-order meaning units, the ability to predict upcoming information, and the awareness of the structuring of the text.

The experiments reported here compared readings using conventional scores with readings using novel scores where the suggested advantages of information separation, integration and predictability were implemented in the design. Fluency of performance was measured primarily in terms of numbers of mistakes, results showing that readers played more accurately with the novel scores. Other, more qualitative, measurements —such as spectrogram coding of tempo stability, blind expert judgment of performance quality, and participant self-assessments— all showed strong positive correlations with the

measurements of numbers of mistakes, with the novel scores producing performances that were more fluent and ranked as more trustworthy and musically satisfactory by experts and readers alike.


These results will still need to be extrapolated to many other musical practices, but they serve to open a debate on the conventions of music publishing as they stand, and are well placed to open new lines of research in score legibility and design.

Rachel was indignant with the prosperous matrons, who made her feel outside their world and motherless, and turning back, she left them abruptly. She slammed the door of her room, and pulled out her music. It was all old music —Bach and Beethoven, Mozart and Purcell— the pages yellow, the engraving rough to the finger. In three minutes she was deep in a very difficult, very classical fugue in A, and over her face came a queer remote impersonal expression of complete absorption and anxious satisfaction. Now she stumbled; now she faltered and had to play the same bar twice over; but an invisible line seemed to string the notes together, from which rose a shape, a building.

V. Woolf [1915]. *The voyage out.*

*O, learn to read what silent love hath writ:
to hear with eyes belongs to love's fine wit.*

W. Shakespeare [1609]. *The Sonnets.*



*Reading
Natural and
Formal Languages*

CHAPTER 1

CHAPTER 1

Reading Natural and Formal Languages

1.1. RELEVANCE OF READING RESEARCH IN DIFFERENT DOMAINS

1.1.1. HUMANS AND VISUALLY MEANINGFUL PATTERNS

The development of the study of reading processes has been much slower to germinate and blossom in the music domain than in the realm of natural languages, and also limited when compared with the literature regarding other formal languages (like mathematical or scientific notation). The systematic study of reading processes in music only took off in relation to the momentum gained through studies in the realm of natural languages in the late 1960s and the 1970s. Apart from a few early music psychology studies, such as those carried out on eye-movements in the 1920s [compiled in Jacobsen, 1941], or isolated efforts such as the development of a music sight-reading test (performance from a score without preparation or rehearsal) and marking system in 1942 by Watkins, skills relating to music reading only started to attract the regular interest of experimental psychologists in the 1970s [*e. g.*, Sloboda, 1974a, 1974b, 1976b, 1976c, 1977, 1978a]. It is for this reason that several hypotheses, references, and proposals in this thesis necessarily refer back to studies of natural language scripts and to studies of different formal languages.

Reading linguistic, arithmetic, or musical notations entails more than purely visual processes, and depends on other factors, common to all these domains, in the form of general comprehension processes (which determine our ability to understand statements both when reading and listening), and access to information in memory for visual symbols (which allows us to encode information from ocular fixations).

The evidence for a strong relationship between general comprehension skills and reading ability is well-documented and compelling [see, *e. g.*, in the natural languages domain, Hogaboam & Perfetti, 1978;

Perfetti, Goldman, & Hogaboam, 1979; Jackson & McClelland, 1979; Haenggi & Perfetti, 1992; in the arithmetic domain, see Ashcraft & Christ, 1995; Berg, 2008; Ashcraft & Guillaume, 2009; in the music domain, Kinsler & Carpenter, 1995], although there are processes unique to reading *per se* that influence the ability as well. Jackson and McClelland [1979] uncovered basic visual processing skills that influence reading ability independently of general comprehension processes. This reading component appears to lie within the stages of visual encoding that access representations in memory for any *meaningful* visual patterns or symbols: Jackson and McClelland [*id.*] showed the presence of a reaction-time advantage for skilled readers in matching tasks (subjects presented with pairs of items and required to respond same/different as quickly as possible); and, perhaps more relevantly, they confirmed that this advantage was *not* due to more efficient visual comparison processes or general response time, since reaction times for better readers were not faster than reaction times for poorer readers when presented with a different matching task that required a visual comparison but did not require access to memory codes (subjects presented with unfamiliar dot patterns). Thus, they inferred that the faster reaction times of skilled readers on symbol identification tasks reflected faster access to identity codes of the items in memory.

That this memory access difference between readers is multi-modal, and not specific to the processing of letters or words, had already been shown by Jackson [1980], when comparing reaction times of skilled and less-skilled readers (classified in terms of general effective reading speed) for a Picture Category matching task; the results were strikingly similar to reaction-time results for letter and word name-matching tasks, with better readers showing the same reaction-time advantage over poorer readers for making the correct response in this task as they had shown in letter and word name-match tasks (around 100 ms). Furthermore, the correlation of picture category-match reaction time with effective reading ability paralleled that for letter and word name-match reaction time. Jackson therefore proposed that individual differences in speed of access to letter and word identity codes are only but one manifestation of the effects of this underlying difference between individuals in memory access for any meaningful visual patterns.

The type of character to be read, and more particularly its familiarity to the reader, had also been shown by Jackson [1980], in a different experiment, not to affect the better-reader advantage in matching-tasks. But, importantly, it was *not* in a first test of this experiment, in which only the *physical* identity of a pair of new (made up, but consistent) characters had to be assessed, where differences in reaction times between skilled and less-skilled readers could be found; it was only when subjects learned *names* (nonsense names, in this case) for a second set of novel characters (with the same design rules as the first set), that better readers showed faster reaction times. Furthermore, the size of the better reader advantage was about the same for unfamiliar novel characters as for familiar alphabetical letters (always around 100 ms).

1.1.2. READING PARADIGMS WITHIN EXPERIMENTAL PSYCHOLOGY

A. READING RESEARCH AT THE CENTRE OF THE STAGE

Wilhelm Wundt, usually credited as the founder of experimental psychology, established his first laboratory in Leipzig in 1879, and the systematic study of the processes involved in reading can be tracked back to that laboratory, where sensation, perception and reaction time experiments in reading paradigms became some of the foremost concerns of the newly formed discipline [for a more detailed account of the seminal relevance of reading studies within experimental psychology, see Leary, 1980; Tzeng, 1981]. In those years, reading research was considered to be one of the major tools for analysing the contents of mind. In 1908 Edmund Huey carefully summarized in a scholarly fashion most of the reading research of this early period in his classic book, *The Psychology and Pedagogy of Reading*, where he states:

To completely analyse what we do when we read would almost be the acme of a psychologist's achievement, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history. [1908/1968, p. 6]

Indeed, reading research was at the centre of the stage in experimental psychology studies in the later part of the XIXth century, and several of the original findings still inform today's models of reading processes: already in 1878, Javal had discovered *saccadic* eye movements (rapid movements of the eye between fixation points), which he characterised as discontinuous, erratic and sometimes long jumps that were remarkably contrary to the reader's phenomenological experience (of reading as a fluent continuous task); in 1898, Erdmann and Dodge were able to compute the speed of saccadic movements, and provided evidence showing that no perception could take place during these rapid jumps.

But perhaps a more relevant concern in the beginnings of the systematic study of reading processes was whether the perception of individual symbols was modulated by context, and it was soon hypothesised that visual information units would be integrated into meaningful groupings: as early as in 1886 Cattell [1886a, 1886b] observed that the perceptual span (the span of visual awareness and interpretation) for letters in meaningful words was considerably greater than that for letters in random strings. This was corroborated by Pillsbury's [1897] work on the "creativity" of the apperceptive faculty: in his experiments, subjects were asked to identify distorted words (words in which a letter was blurred, replaced by another letter, or altogether missing), with the results showing that subjects typically reported seeing the stimulus word as normal, which led Pillsbury to conclude that the apperceptive process included making corrections or filling in missing elements to match a predicted or recurrent grouping of symbols. Consistently, Erdmann and Dodge [1898] also demonstrated that words could be perceived at a distance at which isolated letters could not be identified.

B. PREOCCUPATION WITH APPLICATION AND ASSESSMENT OF FINDINGS

Soon after the publication of Huey's [1908] book, though, the proliferation of research in reading suddenly came to an end and experimental psychologists' interest in mental processes gave way to the analysis and specification of the functional relationship between stimulus and response in behavioural act. In language research, the analysis of *verbal* behaviours became the focus of attention, and in the education circle, investigators became more preoccupied with a concern for application, structuring, and assessment of previous findings. In his introduction to the 1968 (*n. b.*) reprint of Huey's book, Kollers comments that "remarkably little empirical information has been added to what Huey already knew" [1908/1968, p. xiv].

C. ANALYSIS OF DEEP STRUCTURES IN LANGUAGE

The return of an interest in natural language reading research did not happen until several decades later, in the late 1960s, and was brought about by several factors. Tzeng [1981] has pointed at the renaissance of the Cartesian idea of "innateness" as crucial, and gives pre-eminence to Chomskian transformational linguistics (which follow Cartesian postulates) in the shift of researchers' attention from descriptions of surface structure toward analyses of deeper structures in natural languages [see, *e. g.*, Chomsky, 1966]. Another factor could have been the development of psychochronometric procedures (*i. e.*, reaction time experiments) to a level of sophistication such that their reliability could be established independent of stochastic or subtractive methods [see, *e. g.*, Sternberg, 1970 —somehow unveiling a newly gained pride in the field]; such procedures would prove, in any case, to be very useful in the study of natural languages for experiments of word recognition, sentence verification, or text comprehension. But possibly the most important factor was the *funding* of research related to the improvement of education, with the goal of strengthening its scientific and technological foundation [for a detailed historical account, see Venezky, 1977], a factor ultimately linked to political and economic concerns in regimes that depend on the dissemination and proliferation of information [on the crucial roles of information in what he terms Well-Being Societies, see García-Calvo, 1993, 2013]. Finally, it can also be speculated, as Tzeng [1981] suggests, that advances in computer technology at the time, and the electronic implementation of simulations of higher mental processes such as problem-solving or comprehension, could have led to comparisons with "natural cognitive behaviours" and thus generated novel questionings of the understanding of natural language reading.

1.1.3 INTEGRATION OF RESEARCH AND INSTRUCTION

Furthermore, not only have the experimental studies and theoretical models on reading processes flourished much later in the music domain, but also the integration of research and instruction has been much less promoted, or even addressed, regarding music, whereas experimental psychologists researching reading of natural languages have frequently claimed the need of such an integration [see, *e. g.*, Tzeng &

Singer, 1981]. This has been particularly the case for experimental psychologists that have been interested in solving reading problems, which have even proposed that experimental research in reading should be constrained within the framework of reading instruction [Tzeng, 1981].

Although Wundt had a certain antipathy to the whole field of education and insisted that pedagogical and scientific interests were two different matters, findings of early experimental reading research had a (perhaps unintentional) profound effect on reading instruction. The impact of these basic research findings on the improvement of education in general and reading instruction in particular is also recorded in Huey's [1908] account of those pioneering years, where we furthermore also can clearly see his personal effort in trying to relate basic research findings to pedagogy, especially in the visual domain, as he starts to look for appropriate lengths of printed lines, appropriate type sizes, *etc.* As Buchner [1909] commented in his early review of Huey's work: "probably its most striking feature is the tempered, yet progressive mixture of science and practice" [p. 149].

Singer [1981], when giving a historical overview of reading instruction in the US, accounts for a total of 14 studies in visual perception in the period 1884–1910 conducted by American researchers who had studied with Wundt in Leipzig, which would have heralded the subsequent period of application of scientific findings to instruction, characterised mainly by the transition from oral to silent (visual) reading. Singer [*id.*] also sees Huey's [1908] discussion of meaning in reading, and the derived definition of reading as *reasoning*, as aiding this emphasis on the instruction of silent reading in the following years. In fact, reading aloud (in natural languages) never regained its supremacy: today, teachers use reading aloud for the very initial instruction but switch to silent reading as soon as they can (they then employ oral reading for diagnosis, or for readings to an audience). Some authors [particularly Saenger, 1982, 1989, 1990, 1996, 1997] view this transition from spoken to silent reading as paralleled over historical time in Western societies in the transition from continuous unseparated scripts to the use of spaces between words and have thoroughly researched its evolution and dissemination, seeing this transition as a crucial aspect of the configuration of a culture.

1.2. THE EVOLUTION OF READING AND WRITING OF NATURAL LANGUAGE SCRIPTS

In his seminal study of the historical evolution of reading fluency, *Spaces Between Words*, Saenger [1997] tracks the developments from the Ancient World's custom of writing with rows of letters uninterrupted by space (*scriptura continua*) to the canonical separation of words, phrases and paragraphs that we use nowadays, and that was more or less established, in Europe, by the end of the Middle Ages. I will argue in this thesis that the benefits derived from the introduction of spacing in language scripts could have a parallel in musical scores, and therefore will start by looking in somewhat detail at the stages of the evolution described by Saenger.

1.2.1. ANCIENT READERS AND UNSEPARATED TEXTS

The uninterrupted writing of ancient *scriptura continua*, used in Europe until at least the VIIth century *A. D.*, appeared for the first time in Indo–European languages with the adaptation by the Greeks of the Phoenician alphabet by adding symbols for vowels, thus creating a writing system that had a complete set of signs for the attempt at unambiguous transcription of pronounced speech. Before this introduction of vowels to the Phoenician alphabet, all the ancient languages of the Mediterranean world —syllabic or alphabetical, Semitic or Indo–European— were written *with* word separation by either spaces, points, or both in conjunction [Wingo, 1972; Anderson, Parsons, & Nisbet, 1979]. After the introduction of vowels, word separation was no longer felt necessary to eliminate an unacceptable level of ambiguity. While the very earliest Greek inscriptions were written with separation by interpuncts (points placed at midlevel between words), Greece soon thereafter became the first ancient civilisation to employ *scriptura continua* [Turner, 1971; Saenger, 1997].

The Romans borrowed their letter forms and vowels from the Greeks, and they, too, discarded word separation as superfluous. Furthermore, Romans were reluctant to emulate certain Greek practices that may have helped Greek scribes and readers to control unseparated text. Foliation, pagination, and catchwords, all present in Greek papyri codices, were never employed by the Romans, and paragraphing was received into Latin only with hesitancy and often confined to certain genres of texts [Turner, 1977].

From the modern point of view, it seems inexplicable that two modes for facilitating lexical access, the use of vowels and the use of word separation, were not combined at an early date to form a hybrid transcription that would have greatly facilitated reading by incorporating cues for word recognition similar to those of the modern separated printed page.

But the ancient world did not possess the desire, characteristic of the modern age, to make reading easier and swifter, because the advantages that modern readers perceive as accruing from ease of reading were seldom viewed as advantages by the ancients. These include the effective retrieval of information in reference to consultation, the ability to read with minimum difficulty a great many technical, logical, and scientific texts, and the greater diffusion of literacy throughout all social *strata* of the population [Roberts & Skeat, 1983; Saenger, 1997]. We know that the reading habits of the ancient world, which were profoundly oral and rhetorical by psychological necessity as well as by taste, were focused on a limited and intensely scrutinised canon of literature. Because those who read relished the mellifluous metrical and accentual patterns of pronounced text and were not interested in the swift consultation of books, the absence of interword space in Greek and Latin was not perceived to be an impediment to effective reading, as it would be to the modern reader, who strives to read fluently [Parkes, 1993; Saenger, 1997]. Furthermore, long-term memory of texts frequently read aloud also compensated for the inherent graphic and grammatical ambiguities of the languages of late antiquity.

Also, the notion that the greater portion of the population should be autonomous and self-motivated readers was entirely foreign to the elitist literate mentality of the ancient world. For the literate, the reaction to the difficulties of lexical access arising from *scriptura continua* did not spark the desire to make script easier to decipher, but resulted instead in the *delegation* of the labour of reading and writing to skilled slaves, who acted as professional readers and scribes [Sirat, 1988; Saenger, 1997].

In reading, the eyes move over the lines in quick jumps (saccades, generally considered to be < 50 ms), separated by longer pauses (fixations, generally between 100–500 ms). Saccades are too brief for useful vision, and recognition processes operate solely on the intake during fixations, which for a given point in the text might occur only once, or on repeated occasions (regressions). The historical evolution of word separation would change the format of the page and reduce the quantity of fixations and saccades for the reader: an ancient reader of Greek or Latin texts would probably have needed more than twice the number than what a modern reader would do using a canonically separated text [Fisher, 1976; Saenger, 1997]. Readers who habitually read unseparated writing would undoubtedly improve their reading rates over time, and get to a certain degree of habituation; however, they would always need to manage the unseparated texts with more numerous ocular fixations and regressions than what is needed with a separated text [Senders, Fisher, & Monty, 1978; Downing & Leong, 1982; Solomon & Pelli, 1994].

Due to the limitations imposed by the script, the ancient reader of a text would need an initial preparation or *praelectio*, in which he would read orally —aloud or in a muffled voice—, because overt physical pronunciation aided the reader to retain phonemes of ambiguous meaning and ascription [Taylor & Taylor, 1983; Banniard, 1989; Saenger, 1997]. Oral activity helped the reader to hold in short-term memory the fraction of a word or a phrase that already had been decoded phonetically while the cognitive

task of morpheme and word recognition, necessary for understanding the sense of the initial fragment, proceeded through the decoding of a subsequent section of the text. The aural retention of inherently ambiguous fragments often was essential until a full sentence was decoded.

1.2.2. SEPARATED SCRIPT: A SUCCINCT OVERVIEW

A. INTERPUNCTUS, HEDERAE, AND SCRIPTURA CONTINUA

In the classical age Roman books and inscriptions were written with separation by medial points (*interpunctus*) placed at midlevel in the line, but these points were generally not accompanied by quantities of space any greater than that ordinarily placed between adjacent letters within a word [see examples in Wingo, 1972]. In the IInd century *A. D.*, words in inscriptions were frequently separated by an ivy-leaf-like decorative design forming a space-filling character known as *hedera*, which more closely resembled a letter of the alphabet than a point [see examples in Parkes, 1993].

While from a grammatical point of view texts using *interpunctus* or *hederae* may be separated, these signs are not susceptible to rapid visual detection, while space of sufficient quantity is readily perceived [on the physiological consequences of the use of symbols rather than space to separate words, see *e. g.* Hochberg, 1970; Fisher, 1976]. Experimental work has shown that the placing of symbols within the spaces between words, while preserving separation in a strictly grammatical sense, greatly reduces the advantages of word separation and produces ocular behaviour resembling that of unseparated text [Fisher, 1976] (most modern readers will experience difficulty in decoding the text in *Figure 1.2.*, for example).

In the Imperial period, most Roman texts discarded word separation and used *scriptura continua* (with the presence or not of *interpuncti* not substantially changing the continuity of the script, as mentioned above). At the very end of the Roman Imperial period, the vulgar Latin tended to be pronounced without distinct word endings, which itself tended to make a formulaic ordering and grouping of words more recurrent. Towards the end of the IVth century, St. Jerome (*c.* 342–420) introduced a very first attempt at including visual cues that reflected the structure of a text, possibly as a reflection of the by then conventionalised sequencing of the language: the *cola et commata* formatting in the manuscript text of the Vulgate, in which each line represented either a phrase, a clause, or a sentence.

The Vulgate became a primary model for medieval Latin. The enhanced observance of conventions of word order and word grouping increased the efficiency of reading, as has been confirmed in experimental works looking at the eye-voice span (the distance between ocular fixation and oral utterance), showing that texts using predictable word order are easier to decode than texts using convoluted orders [*e. g.*, Levin & Kaplan, 1970; Klein & Klein, 1973; Levin & Addis, 1979].

B. (RE)INTRODUCTION OF WORD SEPARATION

In the British Isles in the late VIIth century, the reintroduction (as it had been the case in pre-vocalic scripts of Antiquity) of word separation, first by Irish and then Anglo–Saxon scribes, marks a dramatic change in the relationship of the reader to the book. The first datable word-separated manuscript is Dublin, Trinity College 60 (A. I. 15), the *Book of Mulling* [Anonymous, before 692 (see *Figure 1.1.*)]. The first VIIth-century Irish grammarian who has left indication of the appeal of separated script is Virgilius Maro: his graphic distinctions between conjunctions and word terminations, or conjunctions and demonstratives are indicative of a new desire to avoid ambiguity [see Löfstedt, 2003]. In England, the earliest Anglo–Saxon author that offers direct insight into the pedagogical implications of word separation is Aldhelm (*ca.* 640–709), bishop of Sherborne, who redefined *prosodiae* (ancient Latin term for the signs complementing script that aided the reader in recognizing the words), which had been until then added by the reader, as new signs regularly provided by the *scribe* for the reader's correct distinction of the words. Some of Aldhelm's most relevant comments concern the explicit signs for word demarcation that he termed *passiones* (meaning "feelings"), a term unknown to the technical vocabulary of ancient grammarians. Aldhelm apparently preferred it to *prosodiae* because these signs removed ambiguity and enabled the reader to lend suitable expression to his voice by not mistaking word endings. The description of word demarcation signs as *passiones* seems to indicate, though, that these Insular graphic innovations were not meant merely to avert misplaced word limits or stress, but also to speed the comprehension necessary for reading with expression [see Aldhelm, *ca* 640–709/2009].

C. CONSEQUENCES ON READING HABITS

As a consequence of the reintroduction of word separation by space, even readers of modest intellectual capacity could read more swiftly, and they could understand an increasing number of inherently more difficult texts, which were furthermore increasingly more foreign to their cultural or social contexts. Word separation also allowed for an immediate oral reading of texts, which eliminated the need of the arduous process of the ancient *praelectio* (see SECTION 1.2.1.). Word separation, by altering the neurophysiological process of reading, simplified the act of reading, enabling both the medieval and the modern reader to receive silently and simultaneously the text and encoded information that facilitates both comprehension and performance. Empirical evidence has corroborated the neurophysiological importance of separation by showing that eye movements, processing times and accuracy are affected by its usage: the suppression of the boundaries between words decreased the percentage of words read correctly and increased the decision times in a word identification paradigm [Jarvella & Lundberg, 1987, with Swedish readers]; respecting word boundaries when splitting a script into lines decreased gaze durations [Li, Zhao, & Pollatsek, 2012, with Chinese readers, which is particularly salient since Chinese printing and writing systems do not customarily respect word boundaries]; removing or replacing interword spaces slowed reading times and impaired normal eye movement behaviour [McGowan *et al.*, 2014, with English

readers]. Therefore, Saenger [1997] argues that "[...] the most crucial change in the relationship of the reader to the book from antiquity to modern times was the consequence of the evolutionary process through which space was (re)introduced into text" (p. 20).

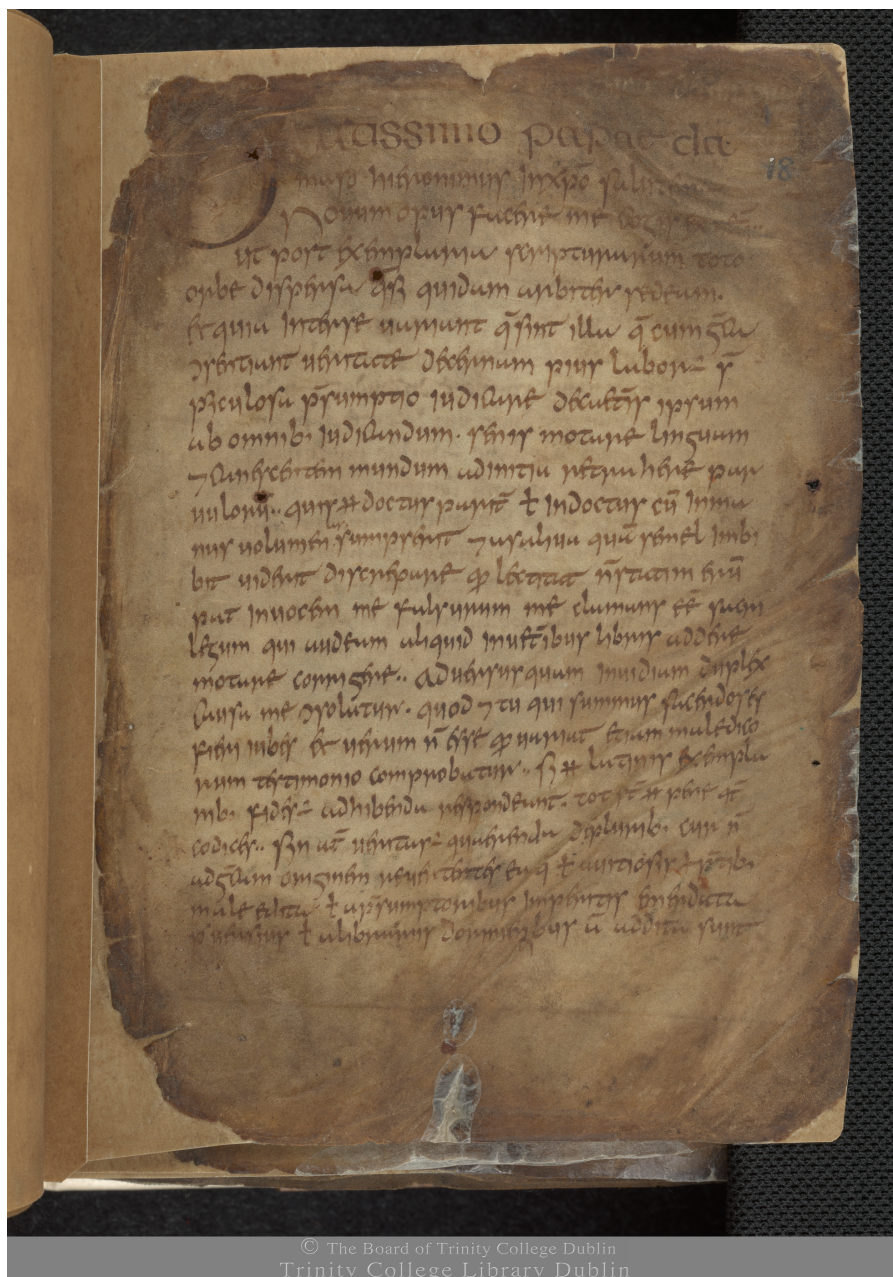


Figure 1.1.

The first datable word-separated manuscript: Dublin, Trinity College 60 (A. I. 15), the *Book of Mulling* (Anonymous, before 692). Shown here is the first page of the text (folio 1r). Saenger has argued that "[...] the most crucial change in the relationship of the reader to the book from antiquity to modern times was the consequence of the evolutionary process through which space was introduced into text".

Finally, in the period from the IXth to the XIth century, the shape of words (the distinct visual image of a word, usually called Bouma Shape after the eponymous Dutch psychologist) emerged gradually even clearer in the written page through the pervasive use of abbreviations, *prosodiae*, punctuation, terminal forms, and other related graphic innovations that enhanced word image (or item image, in the case of standardised abbreviations).

Saenger [1997] considers this whole process, triggered by a re-conception of the visual spacing of texts by the end of the VIIth century so dramatically influential on the relation between reader and text that he insists that "it is unsurpassed by any other alteration in the act of reading between the patristic age and the sixteenth century" (p. 21) (*i. e.*, above transformations in calligraphy, dissemination, or imprinting techniques, all ultimately submitted, according to his view, to the crucial introduction of spacing).

1.2.3. NEW GENRES AND NEW MODES OF READING AND WRITING

Saenger [1997] notes that new genres of books and new modes of reading and writing originated in Ireland and England already from the seventh to the ninth centuries, after separated script started to become habitual. Of particular relevance for the possible comparisons with music scripts are the following:

POCKET GOSPEL BOOKS. Ancient paintings depicting reading suggest that to resist the tunnel vision imposed by *scriptura continua*, the reader tended to distance himself physically from the page [see (parametrically unclear, albeit primary) examples in Anonymous, Vth cent. (*Figure 1.2.*); see also Metzger, 1968; Parássoglou, 1979]. The late medieval shift of attention from the *figura* (shape upon which vision fixates) of the letter to the *figura* of the word produced a compacted and smaller script that brought the reader closer to the page, creating, as Saenger [1997] has proposed, "a greater intimacy between the writer, the reader, and the book" (p. 93). It could be argued that the same shift can be observed in the change from the notation in a *codex* intended for group usage towards the private study score —although systematic motif or phrase separation (using spaces) has not been introduced in music scripts until today (2017). The plainchant musical practices in Western Europe in the XIIIth–XVIth centuries, using the square notation (*i. e.*, resolving previous stroke notation into a series of discrete squares) that facilitated singing from a *codex* by a group of singers (all singing the same line), are consubstantial with the large size of manuscripts at the time, a size that would decrease significantly for scripts intended for instrumental or court (private) music, in parallel with the development of mensural notation (more appropriate for individual parts) from the XVIth century on [see Bent *et al.*, 2001, section III].



Figure 1.2.

Folio 3 (recto) of the copy of the Vergilius Opera kept in the Biblioteca Apostolica Vaticana, Rome (Vat. Lat. 3867 —3r). The picture represents a professional reader, at good distance from the reading stand from which he would run the *praelectio* (first preparatory readings of the text) and then the rote repetitions leading to his performance of the text. Also note the use of *scriptura continua* (with *interpunctus*, in this case) in the text around the figure, a hindering to fluent reading for modern eyes accustomed to spaced texts.

VERNACULAR TEXTS. Writing in the vernacular, like using word separation, was a way of accommodating reading. The first transcriptions of a European vernacular other than Latin using the Latin alphabet occurred in the same geographical regions and at approximately the same time as the acceptance of the

separation of words by space [Van Uytvanghe, 1976, 1983]. The recurring phenomenon of less rigorous word separation for the vernacular can be explained both by the oral character of vernacular literature, which had a broader, semiliterate audience, and by the fact that reading one's native tongue required fewer graphic aids [Milner & Regnault, 1987; Kendrick, 1988]. The restrained use of *prosodiae* in vernacular texts, in which the *diastole* (a comma-shaped mark placed under the line to indicate word separation) and the hyphen occurred rarely, is yet another sign of the greater ease of reading one's own language. In a comparable manner, musical notation will need more accuracy and clear prescription when familiarity with idiomatic conventions can *not* be assumed. For instance, a historical turning-point in musical notation as Guido of Arezzo's reforms (compiled mainly in his *Micrologus*) [*ca.* 1030/1978], based on the introduction of a staff (previous notation was based on proportions and contours, but not on accurate discrete categories), changed the relationship between writing and music in the greater part of Europe in a remarkably short space of time, but it was not only the intrinsic merits of the reform that lay behind its Europe-wide success; the socio-historical context was determinant too. The dissemination of staff notation took place in the era of the crusades and the investiture struggle, and Guidonian notation belonged to the arsenal of the reforms of Pope Gregory VII (1073–85): it would facilitate liturgical reform and preserve the unity of *centralized* uses (*i. e.*, no longer based on familiar local conventions) [see Bent *et al.*, 2001, section III].

COMPILATIONS OF *AENIGMATA*. These were Latin riddles written in verse, a favourite Irish and Anglo-Saxon genre of text [see, *e. g.*, the late Merovingian collections compiled by Glorie, 1968], which arguably could reflect a conception of the written page, as Saenger [1997] puts it, "[...] as a field for the artificial mental manipulation of word images, rather than the transcription of oral speech" (p. 98). Similarly, a move from transcription of vocal inflexions in neumatic notation towards the use of notation as a reflection of the manipulations of the compositional process or even as constituent part of it can be traced in musical scripts. In fact, the change to pitch-specific notations of plainchant (as opposed to early contour-based notations), initially in the form of alphabetic notations, as early as the IXth century [compiled in the *Musica Enchiridis* group of treatises, Anonymous, IXth cent./1995] and from the XIth century on, in staff-based systems [proposed by Guido of Arezzo; modern edition in Palisca, *ca.* 1030/1978], was probably initially prompted [see Bent *et al.*, 2001] by the needs of theoretical texts, many of which contained musical examples (*i. e.*, music for mental manipulation rather than for performance), preceding use in liturgical chant books. As Bent *et al.* [2001] put it: "the alphabetization of the individual notes of the scale was thus at first a purely theoretical procedure and was intimately connected with the use of diagrams as teaching instruments" (p. 85). On the introduction of the staff, these authors [*id.*] explain: "As early as the first period of medieval music notation, *theoretical* and *pedagogical* writings often specified the exact intervallic structure of music examples they cite. For this purpose, horizontal lines and symbols were employed. These methods, however, remained confined to theoretical texts..." (p. 87; italics mine).

VISUAL COPYING. The use of visual memory in copying language scripts, not essential for the scribe who wrote to dictation or who dictated to himself, became intrinsically attractive to the scribe charged with the complication of adding marks linking comments in the margins to apposite points within the text, with the inclusion of construction notes to help the reader understand the underlying sequence, or with the marking of punctuation to delineate the grammatical structure of a text. Similarly, instrumental notation of the late Middle Ages, possibly with parallel intentions in terms of clarity of structure, tends to gradually make the division of time easier with the increased use of bar-lines. Vertical lines had been used through staves in medieval score notation, but only to divide whole sections from one another; but by the XVth and early XVIth century, some keyboard and lute sources included the visual separation of units of one or more bars, either by a space left between them or by a bar-line [Bent *et al.*, 2001, section III]. However, and importantly, the visual spacing and separation of materials in music never became a systematic or intrinsic part of notation, in the manner that Canonical Separation [to use Saenger's terminology, 1990, 1997] in language texts, as established throughout Europe by the XIIth century, configured space in almost all written or printed texts thereafter, by establishing a fixed or measurable separation of all words, which were furthermore separated *only* by space. In fact, in later musical scores, from the XVIth century on, the spacing of discourse units was abandoned (probably in relation to economic intentions, with the willingness to include as much music in a sheet as possible), and only the use of barlines persisted and was to be adopted generally in mensural notation from the XVIIIth century on.

All in all, several changes in the degree of intimacy with the text, in the attempts at accurate prescription, in the reflection of compositional processes, or in the use of aiding visual cues, can be found in the evolution of both natural language and music scripts, with the remarkable difference that spacing and separation has never been deemed important in the music domain. *Why did this never occur in musical texts?* It could be speculated that the ultimate reasons have to do with the social, cultural, and economical functions that music has played in Europe in the Modern Times. As was the case with natural language scripts in the Ancient world, music reading between the XVIIIth and XXth century was focused on a relatively limited and intensely scrutinised canon of literature, and the reading habits were profoundly rhetorical by taste, with the refined details of agogic or prosodic interpretation of the text being particularly relished, with mid- or long-term memory of a well known repertoire compensating for the ambiguities of notation, and with a class of specialised professionals in charge of the transmission and [re]interpretation of that literature.

1.2.4. WRITTEN CULTURE AT THE BEGINNING OF MODERN TIMES

Saenger [1997] goes as far as to claim that "With the general acceptance of separated script and the practices it made possible, Europe entered the modern world as we know it" (p. 256). In any case, it

seems apparent that separated and structured writing affected a wide range of cultural practices in different social *strata* at the end of the Middle Ages:

THE MANUSCRIPT BOOK. The late medieval appetite for using writing as a means of refining the subtleties of intellectual discourse had profound consequences on the mode of composing texts. The task of composing lengthy works of synthesis ultimately led to the development of the author's autograph manuscript, with composition on quires and sheets of parchment, rather than on wax tablets, meaning that the authors, without assistance, could rapidly revise and rearrange their texts while composing them. The new ease in writing ultimately enhanced the author's sense of intimacy and privacy in his work [for examples of this eased and flowing writing, see the compilation by Thomson, 1969]. Could a clearer or more flexible design of musical scripts trigger a different compositional approach or influence the creative process?

PRIVATE STUDY AND HERESY. Psychologically, silent reading emboldened the reader because it placed the source of his curiosity completely under personal control. Private, visual reading encouraged individual critical thinking and contributed ultimately to the development of scepticism and intellectual heresy. The mere possession of certain writings could be legal grounds for formal charges of heresy [*e. g.*, Hudson, 1971, describes the case of Lollard Writings in England]. Could the design of musical scores contribute in the same way to analytical, creative, or adventurous (re)interpretation of a given repertoire?

PRIVATE READING AND PERSONAL EXPRESSION. Just as separated written Latin had facilitated the birth of scholasticism, separated vernacular writing allowed for the transference of the subtleties of fully developed scholastic thought to a new lay audience. The receptivity of Europe's elite to making private judgements on matters of conscience owed much to a long evolution that began in the late VIIth century and culminated in the XVth century in the manner in which men read and wrote. This enhanced privacy represented the consummation of the development of separated writing and constituted a crucial aspect of the modern world. Could the introduction of separation and structuring of the musical discourse on the page ease the access to the information by less informed readers? And could this have consequences on the relation with —and expression of— the musical texts?

Looking at the cultural impact of separated and structured scripts in the realm of natural languages, a possible parallel could perhaps be proposed in the musical realm: it could be argued that a musical culture that would try to enhance creative adventurousness incorporating flexible ways of notating the compositional intentions, or that would allow for interpretative choices beyond the rhetorical inflexions of a well established repertoire, or promote the private reading by new audiences that do not have access to scholastic information on a given musical idiom, would benefit from questioning certain notational and information transcription conventions that are ultimately linked to very specific musical practices.

The main line of research for this thesis has been on the more immediately measurable effect of novel score designs (including separation and structuring of the materials) on sight-reading performances, but the possibility that the imbrication of notation with creation, interpretation and diffusion of musical cultures might have medium- and long-term consequences if novel and flexible approaches to notation are set forth, can not be completely discarded when looking at the impact of a similar process in the reading and writing of natural languages.

1.2.5. TEXT DESIGN AND THE ARTS OF MEMORY

A. ENHANCING LEGIBILITY AND ENHANCING MEMORISATION

So far my focus has been on the practical and social consequences of the enhancements of *legibility* of texts in the late Middle Ages with the introduction of interword spaces and the corollary usages of sentence and paragraph separations, as well as of pervasive punctuation and structuring signs. However, the texts to which this battery of new design devices and signs were added were frequently of a specific type, namely scriptural, to be used within the ritual practice of the Roman church (with the signs making reference to the formulaic traditions prescribed for the ritual reading of those texts). It has been proposed that musical notations were in fact initially additional signs included in these texts to serve as guides in their oral performance, and that the early function of notation could be seen as a further refinement and specialisation of punctuation signs (not so much derived lineally from specific punctuation marks, but rather with these providing a repertoire of available signs and a proximity in terms of function) [Treitler, 1989, 2007; see, however, Barrett, 1997, on musical notation having not so much a performative as a symbolic or status-imbuing function in Carolingian text compilations, and certainly not facilitating legibility or short-term memory].

In types of texts not intended for overtly ritualistic reading, but for other functions as meditation, pedagogy or compilation, the novel visual designs that accompanied the legibility revolution from the VIIIth century onwards were sometimes oriented towards different *memorisation* strategies, and would focus on facilitating the navigation through the text by underlining its relational construction (both at internal level and in terms of association with other texts or contexts), rather than its performative aspect [see examples in Carruthers, 2002, 2009; see also Carruthers, 2014, on designs instigating the social engagement or debate with a text].

In relation to these apparently diverging approaches, a pertinent question, bearing in mind my ultimate aim of revising the design of present-day common-practice musical scores, would thus be whether the facilitation of their immediate legibility could be in conflict with a broader or deeper understanding of their contents (and of their 'relational construction').

B. TWO TYPES OF ARTS OF MEMORY

Early studies of memorisation practices in the Middle Ages [*e. g.*, Carruthers, 1993] insisted on distinguishing between these two strategies in a scribe's text design: facilitating the remembrance of something through its exact reproduction, or facilitating its reconstruction or cognitive 'translation'. This seemed to be in consonance with the apparently separate usages of *memoria verborum* or *verbatim*, based on rehearsal and rote repetition, and *memoria rerum* or *sententialiter*, reconstructive memory, which could be translated into English as 'remembering the substance'.

More recently, however, a different analysis has been proposed: Carruthers's [2008] diagnostic is that most scholars of the arts of memory "have made a basic error when considering the relationship of [reconstructive] memory craft to rote learning, by thinking both to be methods for initially memorizing the basic contents of educated memory" (p. xii); what had been allegedly overlooked was the fact that when a mnemotechnic organisational strategy was used for the internalisation of a text, what was being imposed was a divisional system onto something that was *already* well known (by rote repetition or exposure).

As the *memoria rerum* would be an organisational device imposed on information apprehended by other means, simplifying the process of decoding and initially reading a text should concomitantly facilitate the ultimate memorisation by saving time and energy in the acquisition process. Furthermore, the requisites of musical practice —certainly in the more common contemporary settings— do not favour performances based on remembrance of only the major outlines of a piece or the main contents of its sections; rather, musicians are expected to add their interpretation to a fixed text that can not be substantially altered. In common present-day musical practices performers are required to adhere *verbatim* to the contents of a score —whilst at the same time overlaying cohesive and convincing choices of production (*e. g.*, in terms of prosody, agogic, or intention).

C. PRIMARY SOURCES ON *VERBATIM* MEMORY AND STRUCTURING

The recommendations of *memoria verbatim* by teachers of Antiquity and the Middle Ages seem to point in the same direction that I will be proposing with modifications to be introduced in musical scores: a recurring advice by these teachers is to divide the text structurally and to work with manageable segments rather than long streams of information.

Already in the oldest surviving Latin textbook on rhetoric, the *Rhetorica ad Herennium* [attribution to Cicero widely contested; see Anonymous, *ca* 80 BC/2014] the rules for memorisation are related to what one can take in during a single memorial *conspectus* or glance; the specific recommendation is that the person remembering should start by forming units with a definite starting-point, with their extent being governed by the need for the person remembering to see at a glance, clearly and without confusion, what the unit

further contains (all the more relevant in Latin *scriptura continua*, obviously). In his *Institutiones Oratoriae* [ca 95 AD/1970], Quintilianus also observed that he who *properly divides* the basic structure of an oration can never err when recomposing (or remembering) it again. The Roman school master of the IVth century Caius Julius Victor states in his *Ars Rhetorica* [IVth cent./1980] that the procedure for memorising depends upon the symbiotic activities of *divisio* and *compositio* (*divisio* being the procedure of chunking a text into short fragments for memorising, and *compositio* that of putting the segments together in their order). His contemporary and colleague Consultus Fortunatianus also says that the best procedure for memorising is first to divide a long piece into sections, next memorise, and then join one piece to the next in order: *Quid vel maxime memoriam adiuvat? Divisio et compositio: nam memoriam vehementer ordo servat* (What best helps memory? Division and Composition; for order most secures the memory) [see Carruthers & Ziolkowski, 2004; also Carruthers, 2002]. As a final example I will mention the advice in the *Didascalion /De Studio Legendi* (Didascalion, or The Study of Reading) by Hugh of St Victor [selection in Harkins & Van Liere, XIIth cent./2012], a work widely copied (in at least 125 separate manuscripts) and referenced up to the late middle ages and beyond. In this curricular treatise, Hugh of St Victor, an author of recognised influence at the time, extolls the dependency of all wisdom upon an organised memory, and gives the concrete advice to *colligere* (to gather together) while reading, integrating into a compendious outline elements which are presented individually in a string. Importantly, his advice is not only directed at reading as we would most usually practise it nowadays (with an informative aim, in as little as a single pass) but at users trying to retain a given text in memory. As the authors before him (he references the principles basic to mnemonics as coming from the *Rhetorica ad Herennium*, Cicero, Quintilianus, and many others), Hugh of St Victor insists on inculcating the principle of dividing a long script into a number of short, securely retained segments that can then be gathered and ordered in memory: *Memoria hominis hebes est et brevitate gaudet* (the memory of men is lazy and rejoices in brevity) [see also Carruthers, 2002].

Thus, the design strategies for facilitating *memoria verbatim* (which is what is required from musicians nowadays) would differ from those oriented towards underlining the major arches of a discourse or its metalinguistic or referential potentials, which would be helpful for a *memoria rerum* (understanding the gist of a passage or text, and being able to use it appropriately in other contexts). These *verbatim* strategies should underline the minor structural divisions in a text (those that can be grasped in a glance or *conspicuous*, perhaps), as these are the ones that are crucial for the apprehension of the segments that can then be assembled in memory.

The design demands for enhancing *verbatim* memory (particularly visual separation at word, phrase, and clause level) were therefore similar to those required for facilitating the readability of a text, and any examples of increased legibility in manuscripts from the VIIth century onwards are implicitly examples of how greater spacing and structuring explicitness were also supporting precise mnemonic approaches.

1.3. THE EVOLUTION OF READING AND WRITING OF SCIENTIFIC SCRIPTS

1.3.1. NUMBERS AND FORMULAS AS NOTATION

A. THE IMPORTANCE OF NOTATIONAL DESIGN

Historians of science have pointed out the importance of notational design in the development of science, and have proposed that critical breakthroughs in scientific knowledge have been achieved when easily decoded symbols were employed. Tzeng and Hung [1979] did put forth the hypothesis that a scientific notational system could be better or worse depending on how it may enhance or retard thinking processes. They used as example the convenience of Arabic-numeral notation compared to Roman-numeral notation for attempts at simple arithmetical processes like multiplying two not too large quantities: the adoption of the new and more versatile ideographic symbols (Arabic numerals) allowed for the grasping of the essence of the operations in terms that the phonograms of ordinary language or the complex Roman numerals could not afford. Seeing that reading different writing systems may entail different cognitive processes—which in turn impose different problems for the (beginning) reader—and in a sense thus arguing for linguistic determinism from the written language perspective, Tzeng and Hung [1981] propose that, in interaction with the processing capacities of each individual, thinking can be complicated or marred "because the graphemic symbols used to represent certain concepts happen to be clumsy and thus require a great deal of mental resource (central-processing capacity) in order to hold them in our working memory, let alone to further operate with them" (p. 238). Could it be worthwhile, thus, to assess whether conventional musical notation and publishing use the most efficient designs in terms of processing resources?

To further illustrate the importance of notational design, Tzeng and Hung [1981] give a more specific example from the history of calculus (the branch of mathematics dealing with the findings and properties of derivatives and integrals of functions), where notation was all important: calculus was invented independently by Newton and Leibniz, but the mathematical symbols used in most calculus texts today are the deliberate design of Leibniz; although undoubtedly the greatest talent and thinker of his time, Newton's notational system for calculus did not measure up to his inferential power. Tzeng and Hung [1980, 1981] argue that Newton's symbolic system failed to meet the challenge of later mathematical developments due to a lack of design clarity and a lack of neat visual distinctions, whereby certain symbols were more difficult to distinguish from each other than in Leibniz's system; similarly, in Newton's fluxionary notation (the notation of derivatives) certain symbols introduced ambiguity because their visual position in a formula could lead to mistaken functional links. But such perceptual factors, important as

they can be, may only be secondary: it is the *conceptual* advantage of Leibniz's notational system that makes him, according to Tzeng and Hung [1981], "the most successful and influential builder of symbolic notation that the science has ever had" (p. 239). More specifically, it was the flexibility of his symbols, the ease with which one could pass from one calculus process to another (*e. g.*, passing from the derivative to the differential), and the intuitional suggestions that his symbols called to mind (reaching to the very heart of geometric and mechanical applications of calculus), that led to their widespread adoption. Further, Tzeng and Hung [1981] remind us that for considerably more than a century, British mathematicians failed to perceive the superiority of Leibniz's notation, and thus, while Continental mathematicians were rapidly extending knowledge by using the infinitesimal calculus in all branches of pure and applied mathematics, in England comparatively little progress was made. It was not until the beginning of the XIXth century that there was formed, at Cambridge, a Society to introduce and spread the use of Leibniz's notation among British mathematicians.

Also Saenger [1997] argues that the fidelity of English physicists to Isaac Newton's system of notation retarded the development of physics in England, while on the Continent, Leibniz—who was also a palaeographer, Saenger reminds us—perfected a system of notation that facilitated the advancement of both physics and mathematics.

Such a deterministic view of notation can lead to a further claim: a good notational system, according to Tzeng and Hung [1981], once formed, "seems to have an independent existence and an intelligence of its own" (p. 240). In some sense, thus, a good notational system has, as Bertrand Russell puts it, "a suggestiveness which at times make it seem almost like a live teacher" [cited in Tzeng & Hung, 1981, p. 240].

B. EXPERIMENTAL STUDIES ON THE INFLUENCE OF DESIGN ON ARITHMETIC PROCESSING

Experimental studies confirm that notation can directly affect the speed with which a mathematical formula can be read and a problem solved. In an Event Related Potentials study designed to disentangle the detection of overlapping cognitive processes and thereby isolate the effect of detection of *visual* stimulus properties on arithmetic verification tests, Avancini, Soltesz, and Szucs [2015] have shown that inconsistent visual properties in stimuli (certain operands unexpectedly presented in red, in their experiment) can have a significant effect on reaction times, with the operands with inconsistent visual features processed *slower* than visually congruous stimuli (no main effect of colour was found on accuracy, though). More specifically, an unexpected red font colour elicited a larger P3b wave: this wave is usually considered to be an index of contextual integration related to the process of stimulus categorization, and is sensible to subjective expectations about stimulus probability; physically mismatching or inconsistent stimuli elicit in particular the late part of the P3b, as found by Avancini, Soltesz, & Szucs, (with the effect appearing right on and after the peak of the P3b wave: 500–580 ms). Although previous studies [Szucs &

Soltesz, 2007, 2008, 2010] had shown the late part of the P3b (between 450 and 580 ms) to be more positive to visual incongruences when the mismatch was task-relevant or conflicting with the numerical processing, Avancini, Soltesz, and Szucs [2015] showed the same effect for an use of colour that was task-irrelevant, and that did not convey any conflicting information with the *task* at hand (but conflicted with the visual consistency of the information). They concluded that visual properties of the stimuli may confound the results of many experimental procedures, and that uncontrolled physical stimulus features may strongly affect task execution in arithmetic paradigms. Although in their study the physical feature selected was colour, they advanced that "it is reasonable to assume that similar effects may be elicited by other visual characteristics found in mental arithmetic tasks" [Avancini, Soltesz, & Szucs, 2015, p. 327].

C. EXPERIMENTAL STUDIES ON STRUCTURAL RULES AND NOTATIONAL LEGIBILITY

Further, the way in which visual symbols are recognized depends not only upon their consistency within a system, but also upon the *function* that they symbolize within the system. Experimental studies have shown that knowledge of the abstract structure behind scientific symbols (*i. e.*, knowledge of the functionality of the symbols) can affect their perception in a manner analogous to language. For instance, Baron and Thurston [1973] showed that trained chemists were better at identifying orderly notated compounds than compounds in which the order of the cation and anion had been reversed (formulas for inorganic salts, acids, and bases follow strictly the "spelling rule" that the cation precedes the anion). The symbols used in both the regular and the irregular formulas were the same: formulas were created by randomly combining a very limited pool of common symbols that were correctly matched in terms of valence; simply, half of the notations reversed the structurally regular order of the combination. The reaction times of the chemists in a matching task (deciding same/not same on pairs of formulas) were significantly lower for the structurally regular condition. The authors therefore put forwards the argument that the knowledge of structural rules of a system enhance its legibility above familiarity or recognition of symbols. Notably, the authors also tested on the same task a (small) group of participants that were not at all acquainted with scientific notation of this kind, and found that they could be trained through guided practice (using certain clever patterns of distribution and repetition in the task booklets) to perform *faster* with the unconventional notation, demonstrating that at least some conventional reading usages are not motivated by cognitive or perceptual biases but are configured by practice or exposure.

It will be discussed in more detail below how the practice of automatization of different levels of coding processes has been conceptualised into theoretical and developmental models of reading, applied in instruction, and related to the detection of recurrences and patterns in information (see SECTION 1.4.). But at any rate, if music reading or musical notation processing share resources and pathways with arithmetic verification tasks, then musical notation designed for situations in which high processing speed is fundamental should probably also benefit from designs enhancing consistency and predictability.

1.3.2. PARALLELS BETWEEN TEXTUAL AND NUMERICAL REFORMATTING

According to Saenger [1997] the early medieval chronology for the emergence of new symbols for numbers, mathematical operations, and conventions of order for mathematical statements, as well as their clarification through punctuation, paralleled the emergence of intratextual space and conventions for syntactical sequence in textual manuscripts. The conventional form of graphic mathematical notation that evolved in the Middle Ages would encode the mathematical equivalents of grammatical distinctions with an efficiency equal to that of the new conventions of written Latin. Saenger sees, for example, the evolution of the reformatting of Roman numerals as following the same steps as the modifications introduced in natural languages when the cultural emphasis evolved towards the use of images directly convertible to meaning: the same gradual introduction and dissemination of clarification signs, points, spaces and redundant visual elements that generally reduced ambiguity. Saenger reasonably speculates that the same isolation of the British Isles that contributed to the ignorance of much of antique culture created a *milieu* in which scribes were open to deviation from writing conventions inherited from the Roman Empire, and that this would be reflected both in the transcription of natural languages and in the modes of computation.

In ancient Latin, numerical expressions were regularly written out as words —especially in literary texts—, and when symbols were used for numbers, they were usually letters of the alphabet, as had been the case in Greek scripts. In ancient Latin, however, only seven letters were used. Although it has been suggested [*e. g.*, Harris, 1989] that this limitation to seven symbols could have had pedagogical advantages for the early acquisition phase, from a competent reader's point of view, the Roman system necessitated prior extensive neurophysiological activity to determine how to combine the values of its discrete symbols in order to achieve an accurate comprehension of each element in a mathematical statement. This prior cognitive activity required for adding and subtracting in Roman numerals was analogous to that required for combining phonemes to achieve lexical access in the reading of unseparated Latin text.

In a study looking at the parallels between oralization in verbal and numerical exercises, Sokolov [1972] showed that readers confronting ambiguously written mathematical expressions reacted with increased oralization in the same way as readers of phonetically transcribed verbal text. Accordingly, Saenger [1997] has suggested that oralization aided the ancient reader to retain ambiguous mathematical elements in memory during the process required for manipulating them. As with written verse and prose, context probably also served as an important cue in resolving the inherent graphic ambiguity in Latin numerical statements.

When examining the reading of numbers in the early Middle Ages, and more specifically the reformatting of Roman numerals, Saenger [1997] points at two salient modifications that aimed at swift semantic processing and at clear differentiation of numerals from words in a text:

PUNCTUATION TO REDUCE AMBIGUITY. One modification introduced in the British Isles to Roman numerals in early medieval Latin codices was the division by space and/or points into decimals or decimal distinctions (*e. g.*, MCCLVI would become M. CC. L. VI —1,256). This use of punctuation of large numbers by decimal distinctions paralleled the Insular use of spacing and punctuation employed to reduce the ambiguity of Latin prose, and appeared already in English charters of the VIIIth century [see, *e. g.*, a charter of Swæfred, King of Essex, 704; a charter of Cenwulf, King of Mercia, 799 —these and more examples compiled in Hart, 1971]. This practice then migrated to the Continent in the late Xth century, and conventions of numeric punctuation would then be transferred to Arabic numbers when, in the early XIIIth century, the use of small vertical hyphens was introduced to divide Arabic numerals into hierarchical groupings of three digits.

BOUNDARIES WITH ADJACENT TEXT. Another modification introduced in medieval scripts (initially in Spain, from where it travelled to Ireland) was the use of a long terminal *j* instead of a final *i* in composite expressions of integers, as a redundant sign that clarified the boundary between numbers and adjacent verbal text. Also, Roman numerals were written with minuscule script instead of capital or uncial script (majuscule script with rounded unjoined letters) when differentiation from words was required.

Thus, by the mid XIth century, both in northern Europe and in the British Islands, the antique expression MCCLXVII was likely to be transcribed m. cc. lx. vij (1,267), suggesting a direct visual awareness of its structure and boundaries that would facilitate the rapid comprehension of the symbol. G. R. Evans [1975b] has proposed the emergence of Quadrivium Studies in XIth- and XIIth-century Schools as closely related to the need for a notational system that enhanced swift conversion of numerals to verbal expressions: Quadrivium courses involved the 'mathematical arts' of arithmetic, geometry, astronomy, and music, whereas the previous Trivium courses had focused on grammar, rhetoric, and logic.

1.3.3. EVOLUTION OF THE INSTRUCTION OF ARITHMETIC THINKING

A further example of the gradual change from oralization to visualization in late Medieval reading, in both verbal and numeral exercises, has been proposed by G. R. Evans [1975a] in his study of arithmetic textbooks produced during the late XIth and early XIIth centuries. Most of these are instructional treatises on the abacus (a device for calculating, consisting of a frame with rows of wires along which beads are slid), and Evans [*id.*] traces a new preoccupation amongst the teachers: the novel concern is that pupils should be able to *visualize* what they describe. Pupils are instructed to commit the figures to mind so that the 'eye of the mind' (*Oculus mentis*) will be able to manipulate them. Accordingly, the novel and (as Evans underlines) conspicuous use of diagrams helps to make abstractions visually accessible. The diagrams form an integral part of the work and are not later additions, as in many treatises for other subjects. But these treatises on the abacus are by no means mere sets of instructions, and the diagrams are

mostly aimed at making abstractions manageable: it is current in them to use analogies with dialectics, and Evans [*id.*] posits that it is probable that the pupils who used these texts and began to learn practical arithmetic had some grounding in elementary dialectic. The books have a strong link to Greek works on number theory and the Latin rendering of it, but the medieval writer concentrates on a small piece of a great field and gives practical instruction in its application.

1.3.4. SEPARATED SYLLOGISMS IN NASCENT SYMBOLIC LOGIC

In logic, the clearly separated and syntactically sequenced syllogisms of the late Xth-century and XIth-century Aristotelian texts were a nascent form of symbolic logic (*i. e.*, representational and abstractive rather than discursive and argumentative) [Hawkes, 1977; Treitler, 1982]. Again, the novel use of structural spacing in protoscholastic syllogisms and schematic diagrams made them an important transitional stage between the purely rhetorical logic of antiquity and the symbolic logic used by modern mathematicians and philosophers.

An Iconic Memory model of visual information processing, as postulated by Coltheart, Lea, and Thompson [1974], holds that the contents of brief alphanumeric displays are initially held in a high-capacity fast-decay visual-information store (the ‘iconic memory’). Some of these items are subsequently transferred to a more durable form of storage; the remaining non-transferred items are lost. More relevantly, in relation to the use of structural spacing, the model posits that readers (or observers, more generally) can select which items are to be transferred on the basis of physical characteristics of the items such as location, position in a sequence, shape, or other visual cues that differentiate some categories of items from others. Coltheart [1980] refined such a model and argued that iconic memory was not to be confused with visible persistence, which itself depends on mere neural persistence at the photoreceptor level and at various stages in the visual pathways; on the contrary, iconic memory would be a form of *informational* persistence, not intimately tied to processes going on in the visual system (as visible persistence is). Iconic memory would thus be post-categorical, occurring subsequent to stimulus identification. Thus, although it is defined as non-episodic (its contents can not be explicitly stated or conjured) it nonetheless carries the physical properties of a stimulus in relation to its representation in semantic or syntactic memory. In sum, it temporarily carries distinctive physical information that is relevant for the signification or the functionality of a stimulus within a system.

More recent experimental studies of visual working memory afford a similar picture of how the various features of an item are bound together into chunks. Combined features (bound together into an information unit) have been shown to produce load effects that are no greater on memory than those for the individual features themselves [Allen, Baddeley, & Hitch, 2006; the effect was particularly strong in simultaneous presentations —as happens when reading a text from a page, specially if its information is

structurally separated]. The binding of elements has moreover been found to depend on underlying organisation: recall for sentences is consistently better than for word lists (even in the presence of disruptive concurrent tasks, be they cognitive or executive) [Baddeley, Hitch, & Allen, 2009]. In a manner reminiscent of the Coltheart model, Baddely, Allen, and Hitch [2010], propose a working-memory model including an episodic buffer, a passive multidimensional store of limited capacity; this episodic buffer acts as a passive (and ephemeral) store, capable of keeping bound features and making them available for processing, but not itself responsible for the act of binding (which will be enhanced if it is emphasised by the visual display).

The novel use from the Xth century onwards of structural spacing for the nascent symbolic logic was therefore an efficient way of loading the different items in syllogisms and diagrams with physical properties that immediately helped their symbolic value and functionality within the complexities of the deductive arguments that were just then coming into existence.

1.4. INTEGRATION, AUTOMATISM AND PATTERNING IN PROFICIENT READERS

1.4.1. THEORETICAL MODELS AND EXPERIMENTAL STUDIES ON READING STRATEGIES

Since the surge of reading studies within the more general interest in transformational linguistics and analysis of deep structures in natural and formal languages in the late 1960s, theoretical models have proposed an integration of exposure to informational conventions and a capacity for merging its recurrences into iterative hierarchies. From early on, theoretical models of information processing in reading suggested that readers learn and make automatic a hierarchy of coding processes which transform input into progressively more abstract forms. In an early position paper Gough [1972] had proposed that letter identification was associated with meaning through the rapid succession of intricate steps, including transposition into abstract phonemic representation, and posited that specifying the mechanism by which letters are mapped onto entries in the mental lexicon was the fundamental problem of reading research. Following up on this intended quest, Laberge and Samuels [1974] put forwards a (highly influential and referenced) model of information processing in reading, in which visual information is transformed through a series of processing stages involving visual, phonological, and episodic memory systems until it is finally comprehended. They further proposed that the processing which occurs at each stage is assumed to be learned and made efficacious both in levels of accuracy and, most importantly, automatism—a level at which conscious attention is not necessary. Therefore, in subsequent experimental research, Samuels, Dahl, and Archwamety [1974], tested several subskills of the word recognition process, but centred their final discussion on the need for developing subskills and strategies to the point where they would become a unitary process that could be performed automatically without attention to each detail or step. In line with this argument, Samuels, Begy, and Chen [1976], in an experimental study comparing the strategies of less skilled and more skilled readers, observed that superior *word* recognition (*i. e.*, relegating letter or grapheme identification to automatism) was associated with more fluent readers, who were furthermore superior in ability to generate a target word given context and minimal visual cues. Concurrently, Terry, Samuels, and Laberge [1976] investigated the role of information in individual letters and in letter-strings in word recognition, and found that the display of words in a mirror-image transformation (such that the word would appear normal if viewed in a mirror) produced an expected increase in word recognition latency but also, significantly, an interaction with a letter degradation condition (deterioration of the quality of individual letters), suggesting that subjects had reverted to processing *component* letters. On the contrary, words displayed in normal (*i. e.*, not mirrored) presentation produced no change in latency when individual letters were degraded.

1.4.2. APPLICATIONS IN INSTRUCTIONAL AND DEVELOPMENTAL STUDIES

These theoretical frameworks and experimental studies had furthermore quite direct instructional applications, based on the contention that readers could be trained to recognize integrated groupings of basic symbols as single units, a process that would be not only faster, but also leading to better comprehension, since it would put less load on memory and attention as certain cognitive stages were automated. For instance, P. R. Dahl and Samuels [1977], developed a seven-component model for training novice readers based on developing their prediction skills: according to the authors, as information builds the reader can predict which words have a high probability of occurring; thus the reader needs to use only a minimum of visual information (of letters in a word) and is able to recognize the word more rapidly. Although their specific subdivision of the training into seven steps is not thoroughly founded and could be questioned in itself, the major drawback in trying to extrapolate such a training system to music readers is that, at least in today's (western) practices, musicians are exposed to an extremely great number of idioms and styles, and it would be difficult to imagine semantic units that could be applied transversally. On the other hand, specialised teaching of specific (historic) repertoires as it occurs in high-level conservatoires could possibly benefit from adaptations of the programme devised by Dahl and Samuels, particularly since it builds upon the assumption that aural training (which could be extended into physiological encoding in instrumental practice) should form the basis for developing visual recognition. Approaches that are heuristically or generally organised using the progression from aural to physiological to visual recognition of discourse information do of course abound in music pedagogy (*e. g.*, the 'Suzuki Method'), but attempts to systematize and exploit the automatization of the different steps in information recognition applied to a specific and cohesive idiom are not in use yet (the Suzuki Method assumes exposure to a generic mix of not well-defined styles).

Further evidence for the validity of the assumption of a correlation between reading fluency and ability to recognize meaningful (and/or recurrent) groupings of symbols was found in subsequent developmental studies indicating trends from component to holistic processing with advancing academic grade level [Samuels, Laberge, & Bremer, 1978]. More relevantly for possible comparisons with the music domain, in a reaction time categorization task, Samuels, Miller, and Eisenberg [1979], found a shift from component to holistic processing (with associated shorter latencies) for *repeated* words within a set; they also found that a word's recurrence within a set elicited a decrease in the upwards slope in latencies that is associated with increase in word length in these type of studies, and therefore suggested that with the repeated exposure to a unit of information (a word, in their specific case) there is an increase in the size of the processing unit used in recognition. It is also perhaps of relevance to note that in this study, Samuels, Miller & Eisenberg presented the words in mirror image, thereby rendering the experiment more valid for comparison with domains using less widespread or deeply internalized strings of symbols than studies using regularly printed words. Thus, in music reading, facilitating the recognition of repeated units of

information and prioritising the visual recognition of similarities at different levels of the discourse (at a comparable level to language, probably cells, motifs and short phrases, rather than structural or formal divisions) could equally ease the cognitive load of information recognition and enhance reading fluency.

1.4.3. CONTEXTUALISATION AND RECOGNITION OF INFORMATION

But not only reader experience and perception of recurrences correlate positively with the recognition of information units, also context conditions can affect the task. Patberg, Dewitz, and Samuels [1981] presented poor and inexperienced readers (students of the second and fourth grade in the USA's educational system; 7–10 years old —crucial ages in language reading development) with words of various lengths under three exposure conditions: a guiding context, miscues, and no context. They measured the word recognition latencies, assuming, as in the preceding studies in the field [Terry, Samuels, & Laberge, 1976; Samuels, Laberge, & Bremer, 1978; Samuels, Miller, & Eisenberg, 1979], that an increase in latency relative to word length would suggest a fall-back to component processing, while no increase would suggest continued holistic processing. Their findings suggested that the size of the information recognition unit is positively sensitive to aiding context and the level of development of general reading skills: the poorer readers tended to use component (letter) processing under all conditions (context, miscues, no-context), whereas the more advanced readers tended to use holistic processing regardless of how the words were presented; for medium-level readers, only the meaningful context condition made the readers shift to holistic processing. In music reading, a meaningful context could perhaps be easier to perceive if recurrent, similar, or comparable groupings of figures and symbols become visually recognisable through the design of a score.

In a language domain experiment designed to measure the facilitatory effect of similarity between cue and target words in simultaneous matching tasks, Lawry and Laberge [1981] presented readers with normal and spatially transformed targets, with the transformations including reversed order of letters. The authors found that the cue/target relationships produced markedly different effects in the reversed and normal word groups. Identical cue/target words facilitated performance in both groups, but the magnitude of the effect was significantly greater for the reversed words. When the cue and target words were highly similar the normal word group showed no sign of facilitation and even some slight degree of interference relative to when visually dissimilar target words were used (*n. b.*, in these tasks no sequential reading is performed; cue and target are presented at the same time side by side for 1.2 sec). In contrast, the reversed words showed substantial facilitation for highly similar targets relative to dissimilar targets. Thus the recognition of identity or similarities in motifs that have a relation of symmetry between them facilitates the cognitive process of matching them (obviously), and more interestingly, this facilitation is much more marked than when the motifs do not have an underlying symmetry relation. Although this is a phenomenon known and exploited by composers in infinite examples at auditory level (*e. g.*, in the last

movement of Beethoven's Hammerklavier Sonata, where a fugal exposition based on a reversed presentation of the original subject is recognisable in contour and pacing, even if the scale and specific intervals are changed), there is no literature on the visual aspect of such matching processes in music readers. It is our hypothesis that the process in music reading will be similar to the process in language, and that visual recognition of identities and similarities should ultimately enhance reading fluency. And it is a widely recurrent compositional technique in written (western) music to use a group of notes that are at a certain point played in inversion in relation to a given axis of symmetry. To use a succession of notes played backwards, for instance [a 'retrograde'; see Drabkin, 2001], is such a compositional device, dating at least back to the late middle ages (with G. de Machaut's *Ma fin est mon commencement* being a magisterial and egregious early example).

1.4.4. ACCURACY AND HOLISTIC PROCESSING

Finally, also the quest for accuracy, which is arguably more marked amongst music readers than amongst language readers —due to the intrinsically discrete character of music materials— would probably also benefit from visual cues that enhance holistic processing above component processing: at least in the language domain more accurate readers have been shown to tend towards more holistic processing than poorer readers [McCormick & Samuels, 1979, in a study of beginning readers (7–8 years old)].

If (particularly proficient) readers transform input into progressively more abstract forms an important question is at which level is this integration process carried out, or at which levels can it be effective (syllable, word, phrase, *etc.*). This is a specially relevant matter for a possible extrapolation of the models used in language reading to experimental procedures in music reading, since music's role in many cultures is based on its floating semanticity [Cross, 2005, 2009; Cross & Tolbert, 2009; Cross & Woodruff, 2009]. More specifically in the common practice of western music, it is precisely the choices in articulation and agogic rendering of a given piece that are mostly appreciated in a performance, as signifiers of a personal yet cohesive understanding of its (possible) semanticity. In this sense, if models of language reading show that perceptual integration at various and/or multiple levels facilitates processing, the translation into musical scores will be more likely to be effective, since the demarcation of information units is necessarily more ambiguous in music than in language. If in language the effective units can vary depending on task demands, it can be hypothesized that facilitation of visual integration in musical scores should be effective even without the presence of irrefutable or standardised demarcations of units of information.

There is, in fact, abundant literature in the language and in the arithmetic domains showing that the integration of information into perceptual units by readers is an iterative process (with ceilings determined by cognitive capacities), effective at different levels of the discourse. The functioning of syllables as perceptual units for information recognition, for instance, has been intensely scrutinised since the 1970s.

For example, Spoehr and Smith [1973], in experiments examining tachistoscopic reports for numbers and words as a function of the numbers of syllables in them, found that the number of syllables in the vocalization of a (two digit) number had no effect on report accuracy, which seemed thus unaffected by the duration of any implicit speech process; they further found, in apparent contradiction, that accuracy was greater for one- than for two-syllable words, but they hypothesized that the difference was due to the syllable functioning as a *single perceptual unit*. In the same way, the numbers had been processed as single perceptual units in the arithmetic system, independently of subsidiary translations into speech. It is therefore imaginable that short bindings of notes could be processed as perceptual units in music reading, independently of their vocal denomination. The use of cells or motifs (possibly closer to syllables than words, if we want to establish a simple parallel) is further at the core of most compositional processes in western concert music, and in some authors (*e. g.*, amongst many others, Bach, Webern, or Ligeti) these short groupings can be so basilar that complete pieces are constructed exclusively with the interweaving of a very limited number of them.

The findings by Spoehr and Smith [1973] seemed thus to indicate that semantic targets (units of signification within a given system) were apprehended and processed independently of their (potential) acoustic reification and their number of components (at least to a certain degree, since the possible numbers of components had a small range). More generally, they showed that reading tasks have a basically semantic nature (with semanticism understood as contextual), and that perceptual units are adapted to the mode and task at hand. In a follow-up study, Ball, Wood, and Smith [1975] further confirmed —albeit somewhat unwittingly— that complete targets (complete information units) were detected faster than partial targets or representations of the target using an extraneous mode. It has to be noted, however, that this study is marred by rather unfortunate choices of terminology. Participants searched (visually, reading) for critical words (*e. g.*, ‘salmon’) within strings of 11 words (which, unproblematically, could be meaningful sentences or random strings). The cues for search were all specified *verbally* by the experimenter at the beginning of each trial, and the participants were prompted to look in the text for either: (1) a word including a specified group of letters (*e. g.*, the letters ‘s’, ‘a’, ‘l’, and ‘m’); (2) a word including a pronounced syllable (*e. g.*, /saem/); or (3) a word belonging to a specified supraordinate category (‘fish’). Cues of type (1) were misleadingly termed ‘visual’, when a mental representation of the reconstruction of the string of the aurally given letters (‘s-a-l-m’) could hardly be considered a visual aid; it is also difficult to see how the cue of type (2) is any more ‘acoustic’ than the others, since they are all presented verbally. To confound matters even further, the authors term as ‘targets’ these three types of cues used for searching for the critical word, with type (3), for example, being a ‘semantic target’. Using this terminology, the authors posited that semantic targets [sɪz] were detected faster than visual [sɪz] or acoustic [sɪz] ones. At any rate, and more relevantly, the findings support (although, as mentioned, somehow indirectly) the substantial literature showing easier (faster, and more

accurate) information recognition for holistic processing than for component processing [*e. g.*, Samuels, Begy, & Chen, 1976; Samuels, Laberge, & Bremer, 1978; McCormick & Samuels, 1979; Patberg, Dewitz, & Samuels, 1981].

ACKNOWLEDGEMENTS

The *Book of Mulling* is reproduced with kind permission from The Board of Trinity College Dublin, through its Digital Resources and Imaging Service. The manuscript for the *Vergilius Opera* is reproduced with kind permission from the Biblioteca Apostolica Vaticana, through its Digitized Manuscripts resources.



*Contexts, Models, and
Mappings for
Reading Tasks*

CHAPTER 2

CHAPTER 2

Contexts, Models, and Mappings for Reading Tasks

2.1. DEFINITIONS OF SIGHT-READING AND PARALLELS WITH LANGUAGE READING

Most studies on music reading start by putting the task in the context of other musical activities, and frequently underline its importance within musicianship, with some authors ultimately signifying its importance as akin to being alphabetised. When these classifications and claims are made in the literature, it is usually in reference to sight-reading (reading a score for the first time, without preparation), since the conditions in which further readings of music are performed invoke processes that rely heavily on memory, be it short- or long-term memory. It could be argued that even when a score is seen for the first time, at least by experienced musicians, the reading process relies on previous knowledge of repertoire, idiom, and lexical and grammatical conventions, but this is also true for further readings of a score, and at any rate a taxonomical distinction can be established between the first reading and subsequent ones.

Having distinguished sight-reading from the use of a score in further rehearsals or performances, it can then be compared to tasks that imply similar cognitive processes of decoding and reorganizing information. For the comparison, some authors find it useful to look at the constituent sub-tasks: many of these are shared with disparate activities, and most of them, except the final production in an instrument or in the voice, are always required in prose reading. Even the final production of sound is not necessary for—or exclusive to—music reading, but only more frequently required than in the reading of natural languages.

Experimental studies of music reading reflect this similarity to prose reading, and most of the paradigms are based on designs originally used in language studies: priming, chunking, and error-detection experiments all follow methods used in prose-reading research. This does not by itself undermine the

ecological validity of these experimental design choices, taking into account the mentioned parallels between the processes triggered by music and prose reading. There is, though, a crucial difference between music- and prose-reading abilities in their distribution throughout the populations that are usually investigated: whereas prose reading is more or less universally mastered at a relatively uniform level by participants in language tests of any kind (priming, matching, error-detection, *etc.*), the differences in sight-reading skills between musicians are extremely marked, even within otherwise homogeneous groups of participants. The same problem applies, perhaps even more distinctly, when naturalistic experimental designs are used, in which usual working conditions for musicians are imitated, since the results will reflect more directly the regular use of the abilities rather than their principles or potentials.

A particular problem thus for experiments in sight-reading is the measurement of results and their distribution in analysable datasets, especially in the more naturalistic experimental designs. Two basic categories of measurement could be distinguished: those that are more or less algorithmic or automatized (which itself poses numerous problems and diverse solutions), and those based on assessments using reports or questionnaires. These two ways of measuring can furthermore be correlated for solidity of results.

Experimental results, ultimately, will reflect not only the choices of method and procedure, but also an approach to music reading as a performative task or as a tool for other goals, that translates into the significance and priority that different musical cultures give to the ability. Unfortunately all the efforts that have been put into research in the ways in which language reading and text readability can be improved are completely absent in the music literature.

2.2. SIGNIFICANCE OF SIGHT-READING WITHIN MUSICIANSHIP

2.2.1. PROFESSIONAL SIGNIFICANCE

Sight-reading is considered by many authors as an important skill for a professional musician [*e. g.*, Sloboda, 1978b, who remarked already in 1978 the phenomenal amount of sight-reading required of instrumentalists in Britain; Lehmann & Ericsson, 1996; Waters, Underwood, & Findlay, 1997; Kopiez & Lee, 2008; Lehmann & Kopiez, 2009; Zhukov, 2014; Zhukov *et al.*, 2016]. In present-day practices of production and rehearsal, given the increasing demands in efficiency and time management, sight-reading skills could perhaps be considered of particular importance —as Sloboda pointed out [1978b], many professional musicians (nowadays probably even more so) could simply not perform their jobs without a high level of reading skill. On top of this, the historically unequalled diversity of idioms and styles that a musician is confronted with nowadays means that in the case of any repertoires that are not part of core specialities, performance will be relying on reading more than on contextual or long-term knowledge. There are therefore many cases in which sight-reading ability is almost *essential* for the performer: students at any level having constraints in the time given to prepare a piece; orchestral musicians (certainly in Britain); studio musicians, or busy professional accompanists —to name just the more salient.

The importance given to music reading varies in different societies, and it might also reflect preoccupations with accuracy and efficiency. In the experiments carried out for this thesis a slight trend could be observed showing British students as more competent and confident in sight-reading exercises than students from continental Europe¹. The fact that production and rehearsal times for British orchestras and ensembles are usually much more limited than in continental Europe could be a determining factor —there are unfortunately no studies on the subject.

One of the very few studies looking at orchestras from the point of view of management and computational efficiency of their materials, carried out by Bellini, Fioravanti, and Nesi [1999], concluded by recommending the use of digitized versions of the scores, allowing storage of multiple versions of a given piece that could be adapted to different performative circumstances, eliminating the amount of repetitive work involved in hand-written amendments and at the same time maintaining visual consistency. Their stress on design consistency for the different versions could resonate with research on the presentation of printed language advocating predictability and clear structuring of the scripts,

¹ This is, however, merely an observation: due to the differences in materials, instruments (British students were violinists, whereas continental students were percussionists) and coding of mistakes, a statistical comparison can not be established. It is, nonetheless, a research avenue that could be pursued in the future.

particularly for situations in which readers will greatly benefit from accessibility [*e. g.*, in materials for children, Watts & Nisbet, 1974; in textbooks, Hartley & Burnhill, 1976b; in instructional or informational texts, Hartley, 1978, 2004; in journal abstracts, Hartley & Betts, 2007].

2.2.2. MUSIC READING AS A FORM OF LITERACY

A. DEFINITIONS OF LITERACY

Considering the importance of music reading for overall musical competence, Sloboda [1978b], lamented the little attention devoted to music reading by teachers, educationalists and psychologists (we could add music publishers/editors to the list), particularly when compared to the attention devoted to the reading process in language (see, *e. g.*, SECTION 1.1.2.). But even if, as Sloboda [*id.*] states, “the ability to read in one’s native tongue is, in most cultures, an almost essential qualification for full membership of society” (p. 3), there have been numerous attempts at social organisation that did not entail literacy (all prehistoric ones, to start with), and that probably functioned in a smoother and more efficient way than many a literate society. The debate on whether literate transmission of information is inferior or superior to the oral one lies in fact at the core of civilisation; in his dialogue with Phaedrus, Socrates [Plato, 370 BC approx./2009] remembers the reprobatory words of the Egyptian god Ammon against the impertinent invention of letters by Theuth of Naucratis:

[...] for this discovery of yours will create forgetfulness in the learner’s souls, because they will not use their memories; they will trust to the external written characters and not remember of themselves. The specific technique which you have discovered is an aid not to memory, but to reminiscence, and you give your disciples not truth, but only the semblance of truth; they will be hearers of many things and will have learned nothing; [...]; they will be tiresome company, having the show of wisdom without the truth. (p. 7)

Apart from the contraction of certain mnemonic abilities, and the superficiality of knowledge attained through reading, an even more fundamental shortcoming of written transmission, its rigidity, is mentioned sometime later in the dialogue:

I can not help feeling, Phaedrus, that writing is unfortunately like painting; for the creations of the painter have the appearance of life, and yet if you ask them a question they preserve a solemn silence. [...] You would imagine that they had intelligence, but if you want to know anything and put a question to one of them, the speaker always gives one unvarying answer... (p. 9)

Finally, he also underlines the weakness of written transmission in terms of the possibility of true access to an author's intentions (a most recurrent theme in classic academic writings on music):

[...] when they [the texts] have been once written down they are tumbled about anywhere among those who may or may not understand them, and know not to whom they should reply, to whom not; and, if they are maltreated or abused they have no parent to protect them; and they cannot protect or defend themselves. (p. 10)

All these reproaches can easily be translated to the music domain: the weakening of musical memory, the access to reminiscence and not true content, the tiresome performances apparently loyal to the discourse fixed in a score, the rhetorical repetition of effectist gestures, all are burdens of conceptions of musical literacy that emphasize rigid decoding of a score and the contextually unaware regurgitation of stale discourses.

But reading could also be a process by which the discourse is enriched—as Borges [1989] put it, with his unexpansive yet unflinching enthusiasm for literacy: “every reading implies a collaboration and almost a complicity” (p. 127). García-Calvo has insisted, from a political perspective of repulsion of vacuous social interactions, on this complicity, even friendship, with writers of the past [García-Calvo, 1986, 2002]. Borges [*id.*] seems furthermore to take a stance that could be seen as opposite to the one posited in the Platonic traditions: “A written volume, in itself, is not an aesthetical item, it's a physical object amongst others; the aesthetic experience can *only* occur when the volume is written or read.” (p. 142; italics mine). Thus, the interpretation of a script, the collaboration or complicity between author and reader, could—at least in some cases—be consubstantial to the relevant aesthetical process. To deny this as a possibility is to fall back into regarding a script or notation as a transparent medium, which is the main pitfall of the Platonic argument. In a study looking at a collaborative project between a composer and a pianist that included unconventional grids for performance, Clarke [2006] has suggested (extrapolating from that particular experiment) that:

The performer's role is *not to reproduce* either the score, or the sound that the score seems to specify, in performance. Rather, it is a question of taking the music apart each time you play it, interpret it, [...] and react to and work with whatever the notation on the page elicits. (p. 44; italics mine)

Recently, P. Dahl [2013] has proposed that music reading could be integrated in a concept of Musical Literacy, but using a definition for the term that would be very similar to the one accepted by international cooperation agencies for language literacy. His proposed definition of literacy would therefore include, on top of the ability to decode and communicate written musical notation, the ability to interpret, understand, and transform the materials.

One of the main actors in implementing and coordinating (language) literacy programmes internationally is the United Nations Educational, Scientific, and Cultural Organization (UNESCO), which has significantly expanded its definition of literacy in the last decades. Thus, from initially [1978 report, as cited in Benavot *et al.*, 2006] considering literacy as a set of skills that would be determined by the context in which they were applied, practised and situated, and that would allow an individual to function and communicate effectively in his community, the latest working definitions for their Literacy Programmes Design & Delivery take into account the importance of using literacy skills *meaningfully*. The latest report [Benavot *et al.*, 2006] tellingly stresses that literacy should also include a set of cognitive skills that are independent of context (*i. e.*, susceptible of extrapolation), and that afford individuals an *active* and broad-based learning process. Thus literacy in the latest sense can be defined as not only the capacity to decode a script and reproduce it within a given tradition, but rather the capacity to use these skills for personal development, contextual awareness and critical reflection. A musical tradition based on the rhetorical reproduction of texts within an established and limited repertoire (as was the case of European concert music in the XIXth and, in some ambits, the XXth and even the XXIst centuries) would thus not entitle its players to properly qualify as literate, according to the more recent standards of the UNESCO.

Truly literate performers should therefore not only be able to identify the notational material, but also be able to compute, understand, interpret and communicate it (in a non simply rhetorical way; *i. e.*, not just reproducing it). A definition of music literacy following the UNESCO guidelines would thus certainly take into account that reading is by no means a merely visual activity, and that it is only completed when it allows the performer to use the material in a structurally organized, informed, and contextually aware way.

B. TWO COMPONENTS OF LITERACY

Reading literacy could thus be divided into two phases of ability: decoding a script, and using it appropriately in a critically informed manner. For the first phase, today's musical practices hinder the mere legibility of scripts, since this has been shown to depend on pattern recognition and predictability of the materials (thus also indirectly on a player's expertise in a particular idiom) [*e. g.*, Waters, Townsend, & Underwood, 1998, in a study of pianists; Sloboda *et al.*, 1998, in a study of performance consistency specifically reflected in playing positions; Fine, Berry, & Rosner, 2006, studying choir singers]. In today's practices, the variety of musical languages that coexist in the development and the repertoire of a musician, and the fact that the different roles required in music production tend to be more separated and defined than in other music practices throughout history, result in less context-related information being passed on to the performer (something that eases the predictability of the materials substantially). And the second phase would neither be attainable for someone equipped only to reproduce the instructions in a score without contextual or structural understanding of the discourse. A totally valid intuitive approach (probably more effective than an analytical or dissectional one) could be used by players immersed in a tradition where creation, transmission, and reception were contemporary and based on a shared *lingua*

franca, but an uninformed ‘intuition’ will normally just fall back into convention when performance is alienated from creation (both in the sense of authorship and of interpretation of materials), as is the case in the standard repertoire of much concert music performed nowadays. A complicity or friendship with an author of the past cannot be developed in a framework of reverential formalised relations.

At any rate, it could be argued that the more reflective or creative ‘second phase’ of literacy, as we have identified it, will not begin to flourish if the script is not easily legible. But this is not such a straightforward contention as it might seem, though, since it can be of artistic interest for an author to challenge performers by *not* being clear in a score or a script for action. Estranged notations or scripts can function as a foil to the ever-present danger of performers falling back on easy habits and lazy complacency [see Clarke, 2006].

C. AN EXAMPLE OF ARTISTIC INTEREST IN ESTRANGED NOTATION

In 2007, the group I then directed, Ensemble Madrid, had to give the Spanish première of *Piano Trio* by Kevin Volans [2005]. It is a powerful and striking work in two movements, lasting a total of 25 minutes. *Figure 2.1.* shows the first bars of the second movement. Please notice the time signature, 13/16, and the tempo indication (the quaver at 240). On top of this, the pianist has to play triplets. This *seems* nonsensical, or unplayable. In fact, after a few rehearsals, and the very night before our first personal meeting with Kevin, two of the players called me in panic and told me they refused to play the piece and to attend the rehearsal with the composer the next day. This was just a few weeks before the actual concert, which had been programmed and publicised months in advance. I negotiated with them that they would at least attend the meeting with the composer, listen to what he had to say, and if after that they were still not convinced, we would have to cancel the concert and face the consequences. Not knowing any of this (although perhaps suspecting it), Kevin showed up the next day and explained in a calm, composed, and altogether charming way, what his intentions were with this notation; I could see the anxiety of the players transforming first into relief, and into utter enjoyment after a few trials.

The music is actually conceived within a fast 3/4 time signature, with a comma at the end of each bar. What the composer wanted was that the discourse should *not* flow, and that each bar should be attacked with energy, almost angrily, and as a separated utterance from the next bar, “like in some violent passages of flamenco music” (in Kevin's words). When we asked him why he had not just written the music in 3/4 he convincingly explained that with that time signature players have a tendency to almost imperceptibly shorten the last quaver and that, particularly when a pattern is repeated over this signature the bars tend to get slightly shorter and shorter with each repetition, maybe because musicians unconsciously relate music in 3/4 with a flowing and amiable discourse. By making the players *uncomfortable* with his choice of notation, Kevin attained here exactly the kind of expression he had in mind.

The image shows the first five bars of the second movement of a piano trio by Kevin Volans. The score is written for Solo Violin, Solo Cello, and Piano. The time signature is 13/16. The tempo is marked as quarter note = 240. The Solo Violin part has a dynamic marking of *ff* and includes triplets (x3) and quadruplets (x4). The Solo Cello part also has a dynamic marking of *ff* and includes triplets (x3) and quadruplets (x4). The Piano part has a dynamic marking of *ff* and includes triplets (x3) and quadruplets (x4).

Figure 2.1.

First bars of the Second Movement of *Piano Trio* (2002, rev. 2005) by Kevin Volans. Notice the 13/16 signature.

D. IMPROVING TEXTS AS A LITERACY STRATEGY

Increasing the legibility of musical scripts does therefore not necessarily mean that clarity should be the only possible editorial strategy, or that it should be applied in all circumstances. In the cases in which clarity of scripture would be welcome and beneficial it could be applied, whereas in cases where cryptic transmission might elicit a more rewarding response from performers it should obviously not.

Thus, in situations where constraints of production means, energy, concentration, or simply time, are at play, it could perhaps be useful to at least facilitate the first phase of literacy, the decoding of the scripts, and hope that performers will apply their effort to the development of expansive, critical, contextualised or somehow creative utterances. In appraising the recent publication of a guide to musical notation by Gould [2011] (her *Definitive guide to music notation*), Sir Simon Rattle puts it in apparently straightforward terms:

"[...] we need [notation] solutions and rules that will make our life easier, save rehearsal time and frustration, and will ultimately lead to better performances. What is important for a musician is to be able to spend rehearsal time on the music itself, without the hindrance of trying to decipher it" (p iv)

With the caveats mentioned above, which make the expression 'music itself' somewhat ambiguous, since in some cases it might be part of the artistic intentions of the composer to make notation an intrinsic part of the communication with the interpreter, the general quest for facilitation of reading fluency and ultimately of musical literacy is very welcome, particularly when coming from such a relevant and influential figure.

Music notation, however, cannot reach as “definitive” a state as the title of Gould’s [2011] treatise could suggest. In any living culture there will be a constant interaction of notational presuppositions, necessarily formalised and static at the moment of production, with more long-term processes or evolutions derived from musical practices and priorities. As is the case with most conventions of representation, new rendering or performance styles will be invented and esteemed, and they will eventually make the consensus on preferred representation shift across historical periods. As Blackwell [2013] has put it: “It would be naïve to assume that the conventions of today are the final and perfect product of technical evolution” (p. 8).

Therefore, and rather than focusing on the ‘definitive’ form in which notation might be conventionalised today (which furthermore, and fortunately, is not completely standardised) it seems that making musicians’ lives easier, saving rehearsal time and frustration, and ultimately leading to more suitable performances could be better attained by researching the functional relationship between representation and response in behavioural act, as was done in the language domain already after Huey’s [1908] seminal studies, and preoccupying ourselves with a concern for application, structuring, and assessment of cognitive, behavioural, and neurophysiological findings.

2.3. COGNITIVE CLASSIFICATION OF THE TASK

Sight-reading has usually been classified as a transcription task in which notational information is converted into kinaesthetical information [Sloboda, 1978b, 1984; Fine, Berry, & Rosner, 2006]. Handwritten or typed copying of a text in 'real-time', as well as reading a text aloud, are considered similar instances of tasks that are target-driven and use temporal references [Shaffer, 1978, 1981, 1982, 1984a, 1984b]. Time references seem to play a much more restrictive role in music than in language but, the teleological drive, it is argued by Shaffer [1982, 1984a], means that the internal timing in all these tasks is based on a schedule of events—the targets—not on strictly timed onsets (with the targets simply being more or less rhythmical in nature in the different tasks). According to Shaffer, thus, tempo or rhythm requirements are perhaps less determining in differentiating music and language reading. This seems to be consistent with the views offered by various studies on eye-hand and on voice-hand spans in music reading, where expertise is correlated with *flexibility* in the span and the *adaptation* of this span to the underlying structure of the material, rather than being dependent on a fixed mode or speed of visual scanning [Rayner & Pollatsek, 1997; Truitt *et al.*, 1997; Furneaux & Land, 1999].

More importantly, the cognitive processes of music and language reading share the use of hierarchical structuring [Hauser, Chomsky, & Fitch, 2002; Hauser & McDermott, 2003; Fitch, Hauser, & Chomsky, 2005; McDermott & Hauser, 2005; for the neural correlates of this hierarchical processing, see Tettamanti *et al.*, 2002a; Tettamanti *et al.*, 2002b; Tettamanti & Weniger, 2006; Tettamanti *et al.*, 2009; Enrici *et al.*, 2011], which is in fact also to be found in several other typically-human, non-linguistic functional domains such as tool manipulation, visuospatial processing, and gestural diachronism [Patel, 2003; Tettamanti *et al.*, 2009; Enrici *et al.*, 2011]. Tettamanti and Weniger [2006] have argued in a position paper that whenever the processing of perceptually discrete elements calls for structural integration, some kind of syntactic processing is taking place, which explains the activation of a supramodal hierarchical processor—Broca's Area—for a multitude of disparate linguistic, cognitive and sensorimotor tasks. Convergent experimental evidence seems to suggest that Broca's area does indeed play an essential role in the human-specific faculty of extracting hierarchical structural regularities from materials or utterances [Friederici, 2004; Koelsch *et al.*, 2004; Friederici & Brauer, 2009; specifically on the processing of visuo-spatial sequences, see Bahlmann *et al.*, 2009; see also Greenfield, 1991, on the phylogeny of these faculties].

Of all the activities that rely on such structuring processes, music reading and language reading bear apparent formal resemblances, which have been related to similarities in the perceptual processes of the two reading modes. This has led to comparisons in their explanatory and acquisitional models, in the

measurements of the psychological correlates they elicit or convoke, and in the recognition and definition of their constituents.

2.4. EXPLANATORY, DESCRIPTIVE, AND VISIO-MOTORIC COMPARISONS

2.4.1. MULTIMODAL EXPLANATORY MODELS FOR READING PROCESSES

A. HUMANS ORGANIZING HIERARCHICALLY STRUCTURED SEQUENCES

Like language, music seems to be a human universal, and the nature of these two cognitive systems arises, according to biolinguistic explorations, from a shared, species specific computational ability [Berwick, 1997; Niyogi & Berwick, 1997; Berwick *et al.*, 2011b; Berwick *et al.*, 2013; Miyagawa, Berwick, & Okanoya, 2013; Bolhuis *et al.*, 2014; Hauser *et al.*, 2014; Everaert *et al.*, 2015; an important line of research to define this cognitive capacity as uniquely human has been the comparison with birdsongs —see *e. g.*, Berwick *et al.*, 2011a; Beckers *et al.*, 2012], with the computations involving perceptually discrete elements organized into hierarchically structured sequences [on the specific similarities in structural integration in language and music, see Patel, 1998; Patel *et al.*, 1998; Patel, 2003; Patel *et al.*, 2008].

Introduced in Generative Grammar theories in the late 1980s, the Principles and Parameters approach (separating the problems of descriptive adequacy from those of explanatory propositions) opened the possibility for investigations on certain factors in language attainment that are language- or even organism-independent. It opened thus the possibility for the attempt to account for properties of language in terms of general considerations of computational efficiency [for an overview, see Chomsky, 2005, 2006, 2017; for an interdisciplinary approach to computational mechanisms see Hauser, Chomsky, & Fitch, 2002]. The current (2017) explanation of how language is recursively generated is through Merge, an operation that takes objects already constructed, and reconstructs a new object from them, in a manner considered close to optimal, relying on these general principles of computational efficiency [Chomsky, 2017]. In a recent paper on models of music structure, Katz and Pesetsky [2011] argue that all formal differences between language and music are a consequence of differences in their fundamental building blocks, and that in all other respects, language and music are identical; for them, music, like language, contains a syntactic component in which headed structures are built by an abstractive process: iterated, recursive, binary Merge [see also Pesetsky, 2009, on the unity of processes underlying phenomenological diversity]. (See further SECTION 2.5.2. for the neurophysiological mappings of these overlappings.)

B. THOUGHTS AND EXTERNALISATIONS

This abstractive process would generate a 'language of thought', with externalization (and hence communication) as a secondary process [Chomsky, 2007, 2009, 2017]. In this sense, it has been debated in the music reading literature whether (silent) music reading is or is not a form of music perception, and at which point of the cognitive process a valid understanding of the musical material by the reader is

completed (with independence from production). Music teachers, and some authors in the music psychology literature, underline the efficacy and usefulness of silent studying for performers as a help to develop mental representations of the music [particularly in the case of expert musicians, which have acquired action-effects mental representations —see Drost *et al.*, 2005a, 2005b; although some of these representations might be instrument specific —see Drost, Rieger, & Prinz, 2007], whilst others go further, asserting that music reading (at least in expert musicians), despite its atypical input modality, is a true species of music perception, in that much of what is read is analysed for musical significance *prior* to the formulation of motor commands for response [*e. g.*, with musicians being superior to non-musicians in written reports of notation —Sloboda, 1976c, 1978a; on the implication of music reading in musical knowledge and perception, see Sloboda, 1984].

C. LINKING MOTORIC PLANNING AND IMAGERIES

The involvement of motor as well as *spatial* processes in visual imagery², particularly when dynamic images are convoked, has been confirmed in (relatively) early Brain Electrical Activity maps [*e. g.*, Williams *et al.*, 1995], leaving the question of whether motor imagery is primarily motoric or perceptual in character without a simple neurophysiological answer, also due to the highly distributed nature of motor control [Annett, 1995]. Relevantly, Annett [1996], although acknowledging that the physical basis of imaginal representations of actions is best understood in terms of the mechanisms of motor control, showed that *imagery* forms an essential mediating link between the motoric (unconscious, demonstrable in behaviour) encoding and a conscious (verbalisable or visualisable) encoding [see also Sloboda, 1978c, specifically on the links between phonological and visual encoding in language]. This link was exploited in studies investigating the recall of movement patterns (gestural, not musical, although temporally sequenced), showing better rehearsal strategies and remembrance for motoric patterns that were demonstrated through imagery (visually) than for those presented via guided (but blind) motorics [Hall *et al.*, 1997]. A musical score might thus function as a map of images that link motoric encoding and the abstractive processes of ‘musical thought’, and in this sense, be a powerful tool, if properly designed, for the recall of sequenced patterns of movement.

At any rate, musicians (above a certain level of expertise) have been known to understand musical notation at abstractive level, not necessarily (or at least consciously) linked to motoric encoding. Eye-tracking findings by Goolsby [1994b], showing that music notation is processed before performance by experienced readers, or by Draï-Zerbib, Baccino, and Bigand [2012], as well as by Draï-Zerbib and Baccino [2013], specifically addressing cross-modal (*i. e.*, not performance-linked) reading competence, support these observations, showing that experienced musicians are capable of hearing what they see and *vice versa*, that they are able to integrate multimodal information, and that musical discourse can be

² A visually descriptive or figurative abstraction of an action, in this context.

perceived by expert musicians with modal independence of information. Furthermore, a behavioural and electrophysiological study by Schon and Besson [2005] confirmed that musicians were able to expect tonal auditory endings for materials presented to them based on visual information (without playing). In this study, strong interactions seemed to exist between visual and auditory codes, with visual information influencing auditory processing as early as at 100 ms. Similarly, in a neuromagnetic study by Yumoto *et al.* [2005a], an auditory-imagery-based negativity (peaking at approximately 150 ms) was linked to experienced musicians identifying mismatches between a tone sequence (in auditory input) and the score they were simultaneously reading, with the authors suggesting that it reflected an early neural process of deviance detection modulated by visual input. Lu *et al.* [2014] have further shown that long-term (western, score-based) musical training alters basic audiovisual integration at an early stage (directly after the N1 wave — 100 ms) even for non-musical audiovisual stimuli.

Early neural processing of musical notation by experienced musicians seems thus to entail an initial correspondence between visual input and *auditory* imagery, rather than an inevitable motoric encoding. In this sense the parallel with language reading would not necessarily be modulated by a bias towards performance in music reading, with both processes basically mapping a visual input to an auditory representation that might be externalised or not [see, *e. g.*, Yumoto *et al.*, 2005b, on an audiovisual mismatch during silent (language) reading; compare with the similar mismatch described above for music reading — Yumoto *et al.*, 2005a]. However, most musicians' experience is that reading is in many circumstances intimately linked to motoric encoding, and that specifically the learning of a musical piece requires the development of a strong link between sensory and motor representations. Revealingly, D'Ausilio *et al.* [2006], using transcranial magnetic stimulation of the cortical motor representation involved in musical performance, searched excitability changes in piano players during auditory presentation of a rehearsed and a non rehearsed piece, and found an increased motor excitability for the rehearsed but *not* for the non-rehearsed piece; moreover, they observed an increase of excitability over time (as the number of accumulated rehearsals increased during a 5 day training period), with very basilar cortico-spinal facilitation being encoded at the end of that period for the rehearsed music.

D. NEURAL PROCESSES INDEPENDENT OF MODE AND USE

In sum, these cognitive models and the related findings make one of the more apparent differences between music reading, as mostly performance-oriented, and prose reading, as mostly thought-oriented, a not necessarily crucial one in terms of auditory/phonological representation of the visual input, even if in practice, music, particularly recognisable or rehearsed music, will also trigger a tight cortical (or even more basilar) auditory-motor co-representation. Furthermore, at neural processing level, even though the classic view could be that action and perception (performing and reading) are two extremes of mental operations, a more integrated view has been favoured in the last three decades, on the basis of several discoveries, specially the discovery of mirror neurons in the premotor cortex (initially in monkeys), which

has been seen as a strong demonstration of sensory and motor processes sharing the same neural substrates. These mirror cells show sensory-motor properties such that observed, heard, and executed goal-directed actions can equally activate them [see D'Ausilio, 2007; Fadiga, Craighero, & D'Ausilio, 2008; D'Ausilio, 2009, for mirror mechanisms in the processing of sounds and music].

2.4.2. MUSIC-SPECIFIC EYE MOVEMENTS

Profiles of processing reported by Goolsby [1994b] accounting for eye movements used by sight-readers indicate, nonetheless, a number of differences between music perception from processing notation and perception resulting from language reading. These differences include: (1) opposite trends in the control of eye movement (*i. e.*, the better music reader fixates in blank areas of the visual stimuli and not directly on each item of the information that was presented); (2) a perceptual span that is vertical as well as horizontal, and (3) more eye movement associated with the better reader.

FIXATIONS FOLLOWING STRUCTURAL INFERENCES. The fact that (at least expert) music readers prioritise groupings and the spacings that mark them, over item-by-item fixations has been accounted for since early examinations of (mainly expert) visual perception of musical notation. In an investigation of the nature of the abstraction processes for musicians in which various types of interference in linguistic or musical mode were presented to readers, Sloboda [1976c] observed that neither concurrent letter naming nor concurrent memorization of pitches appeared to cause a decrement in the original visual task, suggesting that musicians may not use naming or pitching transformations in coding the visual input. Moreover, when measuring the effect of item positions in the likelihood of identification of spelling and notational errors, Sloboda [1976b] found that errors were least likely to be detected when they occurred in the middle of words or musical phrases, demonstrating that a high number of inferences occurred at these positions; using parallels with language reading, it was further suggested that inference in music reading usually results from structural than visual factors [see also Sloboda, 1976a, for the structuring model he originally proposed for words].

THE VERTICAL COMPONENT OF THE SPAN. That the span has a vertical component in the case of music has an obvious explanation in the configuration of the musical material, especially in the cases where the reader is scanning several staves. This would be further accentuated by the importance that the perception of *contours* allegedly has in music reading, above identification of individual notes [as proposed by Sloboda, 1978a, 1981, 1984].

THE EFFECT OF STRUCTURAL BOUNDARIES. The importance of the perception of structure in the discourse (be it musical or linguistic) has been repeatedly investigated in reading studies. Improvements in comprehension of essays were found in average readers to whom texts were presented formatted so that points between phrases had extra space added to them [Jandreau, Muncer, & Bever, 1986; Bever *et al.*,

1991; Jandreau & Bever, 1992]. Similarly, the expressive timing patterns that indicate confidence and assimilation of discourse in performances by expert musicians have been shown to be more frequently present in events notated as phrase boundaries [Palmer, 1989; see also Palmer & Van de Sande, 1993, 1995 on demarcations and planning in music performance]. The effect of clause boundaries (units of grammatical organization next below the sentence in rank) in prose reading was examined by Rayner, Kambe, and Duffy [2000]: with a target category noun referring to an antecedent, readers regressed from the category noun more frequently when it was clause final than when it was not clause final, and readers made longer initial saccades when their eyes left the category noun when this word was in clause final position [see also Hirotani, Frazier, & Rayner, 2006, on the role of intonation in these language demarcations]. These findings are concurrent with measurements of eye movements in music reading that suggest that skilled musicians look farther ahead in the notation (and then back to the point of performance), when sight reading, than non-proficient music readers do [Goolsby, 1994b, 1994a; Rayner & Pollatsek, 1997].

2.4.3. EYE MOVEMENTS AND PERCEPTUAL SPANS DURING READING

The influence of structural units on the visual processing in music reading has also been considered in studies of the eye-hand or the eye-voice span. Sloboda [1977] found that the presence of structural markers increased span, and tended to cause span to extend exactly to a phrase boundary. These results suggested for Sloboda [*id.*] a clear analogy between the cognition of music and language, in that knowledge of abstract(ive) structure would be of adamant importance in the organization of immediate visual processing of text [see also Sloboda, 1985, on how structuring with notational symbols for meter determined processing of text].

The span can be measured in two ways: as a note index (number of notes between hand/voice and eye) or as a time index (length of time between fixation and performance). Tellingly, in their measurements of the effect of skill on this, Furneaux and Land [1999] encountered that professionals showed significantly larger *note* indexes than the amateurs, whereas all subjects showed similar mean *time* indexes, with no significant differences between any of the skill levels. This means that the length of time that information is stored in the memory buffer is not strongly related to ability, but that experts can fit more information into their buffers in a given lapse of time, thereby being prepared to perform more rapidly and fluently. The exact information stored will furthermore be modulated by the comprehension of structure in the case of experts [for a general model of the knowledge base of the oculomotor system, see Land & Furneaux, 1997; see also Brochard, Dufour, & Despres, 2004, on the positive effect of music reading expertise on mental imagery integration].

Not only expertise, but also practice, can affect eye movements in music reading: measurements with an eyetracker by Goolsby [1994a] indicated that music readers used fewer but longer fixations after practising melodies that they had to read.

The tempo of performance has also been shown to influence the duration and amplitude of eye movements, albeit not in a directly proportional or correlated manner. Kinsler and Carpenter [1995] have observed that higher tempos necessarily impose constraints in the average time between saccades, which has to be reduced, whereby their mean amplitudes have to be increased [see convergent results in Waters, Underwood, & Findlay, 1997]. The time of execution of individual saccades appears, though, to be entirely unrelated to the time of execution of elements of the performance itself, and there is no tendency towards an 'entrainment' between musical tempo and the sequence of fixations and saccades. This sequence is rather determined (in good and expert readers) by target points in the discourse that are structurally relevant.

It has been argued by Madell and Hebert [2008] that measures such as the eye-hand span and the perceptual span, as evaluated in numerous studies, are relatively simple, and that more complex phenomena such as context effects have to receive more attention in eye movement studies; more specifically, these authors propose that more focus should be placed on fine-grained structural properties like interval size or tonal-harmonic expectation. Although the effects of contextual expectations have been addressed in various music reading studies [*e. g.*, Sloboda, 1974b, 1976b, 1976c, 1977; Schon & Besson, 2002, 2005; Fine, Berry, & Rosner, 2006] these did not use measurements of eye movements, and on the other hand, investigations that have been based on eye-tracking have not specifically or directly addressed the influence of perception of syntactical or grammatical constructions in the manner that it has been done in language studies [*e. g.*, McClelland & Oregan, 1981; Rayner, Kambe, & Duffy, 2000; Hirotsu, Frazier, & Rayner, 2006], but rather have been designed to account for the mere concentration of visual information, regardless of its functionality [*e. g.*, Goolsby, 1994b, 1994a], finding simply that eye movements are reduced when participants read music with high-concentrated visual information.

2.5. NEUROPHYSIOLOGICAL MAPPINGS OF READING PROCESSES

Another respect in which parallels and overlappings of music and prose reading processes are indicated by numerous authors is that of the brain mappings related to these abilities. As Jancke [2012] has observed in his overview of recent studies, the advent of modern brain imaging techniques in particular has challenged the previous dual treatment of music and language as separate faculties, since findings in the neuroimaging field show that several neural modules are similarly involved in language and music [even if the strength of the connectivity architecture can present major intersubject variability —see Wang & Liu, 2014].

Electroencephalography (EEG), functional Magnetic Resonance Imaging (fMRI) and Positron Emission Topography (PET) studies have all been used to examine the shared and distinct cortical areas involved in the processing of language and music. The earliest reported study on music reading, to this author's knowledge, was carried out by Sergent *et al.* [1992], and found (using PET) that reading musical notations and translating these notations into movement patterns on a keyboard resulted in activation of cortical areas distinct from, but adjacent to, those underlying similar verbal operations [see similarly Sergent, 1993; also Sergent, 1994, on the technical difficulties confronted in early PET studies].

In the last decade, especially since more wide-spread access to fMRI technologies became available, studies looking at the neurophysiological mapping of the processing of natural and formal languages have abounded and the literature has grown exponentially in recent years. Of particular relevance for the study of reading are the surveys of brain regions involved in the visual coding of symbols and texts, on the one hand, and the mapping of syntactic, sequencing or structuring resources, on the other.

2.5.1. NEUROPHYSIOLOGICAL MAPPINGS OF VISUAL ENCODING PROCESSES

A. LINKS BETWEEN THE ENCODING OF WORDS AND NOTATIONS

The initial step in any reading task, be it in the realm of the natural languages or in that of the formal languages, is the visual encoding of the presented information. This first step of the process presents remarkable similarities in terms of its neurophysiological mapping when music and language scripts are approached, and in both cases an area of the cortex lateral to the midportion of the fusiform gyrus is triggered (see *Figure 2.1.*, in dark-blue), an area usually labelled as the 'Visual Word Form Area' (VWFA).

The establishment of a link between the visual encoding of words and notes can be traced as far back as to the studies of the biological foundations of reading impairments by Dejerine [1892]. As many (most, in fact) of the studies of impairments after it, Dejerine relied on an individual peculiar case: in this first study,

the patient had lost his capacity to read after a lesion that had not affected speech, and was in addition a musician, with the impairment in word reading associated to a loss of the capacity to read notes; this patient's lesion was located in the left fusiform gyrus, which suggested a first hypothetical link between the encoding of words and notation.

In a recent experiment using fMRI technology on the left fusiform gyrus, Mongelli, Cohen, and Bartolomeo [2014] found that music-induced responses in the VWFA did *not* differ from the responses induced by the information category for which the area is normally selective (and named after): the visual recognition of words. The important role that this specific area plays in the initial stage of reading, in the visual identification of language items, has been noticed for at least two decades now [see, *e. g.*, Cohen *et al.*, 2000]. Its role as a prelexical representation of visually presented words has been posited [Dehaene *et al.*, 2002] and contested [Glezer, Jiang, & Riesenhuber, 2009; see more details below] and it has further been thoroughly studied as an area activated invariantly and independently of presentation factors like letter case [Dehaene *et al.*, 2001], text position in the visual field [Cohen *et al.*, 2000], the use of printed or hand-written scripts [Qiao *et al.*, 2010], and even variations in reading direction or type of character (alphabetic, syllabic or morphosyllabic) [Bolger, Perfetti, & Schneider, 2005] [for theoretical models of this saliently robust yet adaptive attuning, see Dehaene *et al.*, 2005; Dehaene & Cohen, 2011]. At any rate, this region of the left lateral occipitotemporal sulcus is thus always activated, whenever literate humans read a text. What was particularly remarkable in the experiment by Mongelli, Cohen, and Bartolomeo [2014] was that music-induced responses were compared with those induced by words and by pictures of faces and diverse (non-symbolic) objects, with similar patterns of activation in the VWFA evident only in the case of words and notes: in other words, the VWFA appears specialised for processing abstract symbolic forms, irrespective of whether these are words or music notation. Along the same lines, their fMRI study also addressed three other specific areas of the ventral cortical visual stream, each responsible for specialised decoding tasks; these category-selective regions were the Fusiform Face Area (FFA), involved in the discrimination of facial features, the Parahippocampal Place Area (PPA), activated for processing of spatial placements, and the Lateral Occipital Complex (LOC), which plays an important role in object recognition (see *Figure. 2.2*). All the categories of stimuli (written words, various types of pictures, musical notation) induced activations in these category-selective regions that were weaker than the activations induced by the preferred category (that is, *e. g.*, words for the VWFA, faces for the FFA), in all regions. The *only exception* was music notation activation of the VWFA, where, as mentioned, there were no significant main differences between word and notation stimulations. Furthermore, when looking at interactions between musical expertise and VWFA responses, the authors found that in a group of expert music readers (professional conservatoire-trained musicians), VWFA activation was even stronger for written music than for words (with as consequence slight displacements of the mappings of word activations).

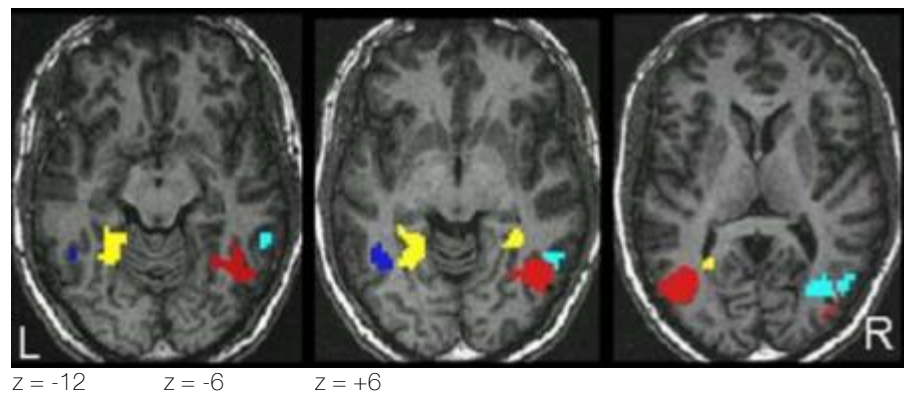
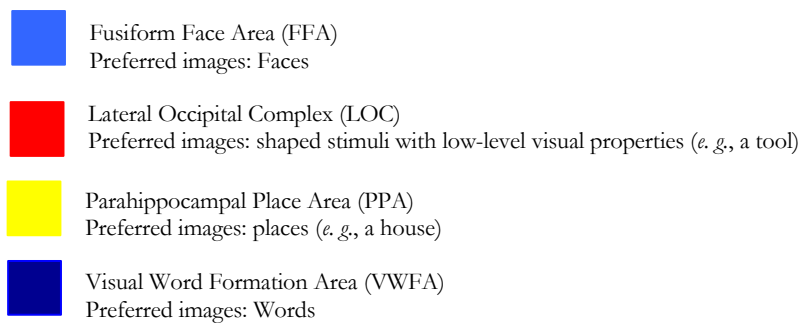


Figure 2.2.

[Adapted from Gaillard *et al.*, 2006, and Mongelli, Cohen, & Bartolomeo, 2014]. Three horizontal slices at different z coordinates (-12, -6, +6), showing various category-selective areas in the ventral visual cortex:



B. SELECTIVITY FOR THE INTEGRATION OF SYMBOLS

These results are of particular importance since the VWFA performs specific computations for reading that cannot be interpreted as mere generic visual recognition processes. Its robust attuning to text independently of visual presentation parameters has already been mentioned, but it has furthermore also shown sensitivity to orthographic regularity, with activations increasing for stimuli forming approximations to real words [Vinckier *et al.*, 2007]. Finally, and possibly most relevantly to our hypothesis and experiments, the VWFA has been shown to contain a representation of language based on neurons highly selective for *whole* words [Glezer, Jiang, & Riesenhuber, 2009]. These findings by Glezer, Jiang, and Riesenhuber supporting the existence of a neural representation coding for whole words contrast with previous models [*e. g.*, Dehaene *et al.*, 2002] which posited a sublexical representation in the VWFA, although the main issue determining the design of paradigms in this research line seems to have been more the setting of priorities than an overt subscription to a given theoretical model; the research conducted so far has had more interest (fascination, almost) in the exquisite attuning characteristics of this visual cortex area than in its role within the various steps that constitute the reading process. A clear example of this focus on the details of the processing rather than on its function within reading tasks, are

studies showing the capacity of the VWFA to distinguish between items and their mirror images, but only in the case of letters (*e. g.*, ‘p’ and ‘q’), and not in the cases of other types of images [*e. g.*, Pegado *et al.*, 2011; Dehaene *et al.*, 2010].

This lack of a wider perspective and of integration with other lines of reading research can probably be justified due to the relative novelty of the discoveries and to the unavoidable (due to obvious logistic considerations of space and mobility) use of short and fragmented materials in fMRI acquisition setups. On the contrary, and as mentioned in SECTION 1.2., experimental psychology established very early on [*e. g.*, Cattell, 1886a; Pillsbury, 1897] that humans tend to integrate visual information units into meaningful groupings and that, in reading, these groupings (*e. g.*, morphemes, words, or phrases) modulate the perceptual span and ability to a higher degree than what mere strings of individual non-related units (graphemes) do. In fact, the most recent studies on the VWFA by Glezer and colleagues [Glezer *et al.*, 2015; Glezer *et al.*, 2016] posit that the area contains a visual lexicon based on neuronal representations highly selective for individual written real *whole* words, and use this as basis for experimental paradigms; consequently, they showed for example that subjects could be trained to learn and recognize novel pseudowords, which in fMRI comparisons of neural selectivity elicited the same tight tunings that were observed for real words in the VWFA, and that differed from the broadly tuned responses elicited by totally novel (untrained) pseudowords. Whether this attuning in the VWFA would be shown for larger-than-word integrations, particularly in common phrases or usual word bindings has not yet been researched.

C. INTERACTION WITH EXECUTIVE AND TIMING CIRCUITS

It should also be noted that the identification of words in this area of the occipitotemporal cortex has been recently shown to be exclusively orthographic (*i. e.*, not phonologically or acoustically triggered) [Glezer *et al.*, 2016]; but at the same time, that reading as a whole (particularly when reading aloud) relies as well on brain circuits involving the Temporo-Parietal Cortex for grapheme-to-phoneme conversions. this means that, for an activity such as sight-reading, or lyric reciting, an executive component or circuits involved in timing and time keeping, as for instance the Cerebellum, will also necessarily be summoned [*e. g.*, Keren–Happuch *et al.*, 2014]. The VWFA plays thus a crucial but not exclusive role in reading processes.

D. MULTIMODAL ATTUNEMENT TO MEANINGFUL COMBINATIONS OF SYMBOLS

But at any rate, the fact that the VWFA, the ‘visual dictionary’ of the brain, also showed comparable activations to music notation and not to other images is of crucial importance for design proposals looking at the configuration of novel scores, since establishing parallels between increased recognisability of words and of notes could lead to the argument that design strategies or visual cues and aids implemented to increase the recognisability of words in texts could be transferred to notes (or rather, note

motifs or patterns) in musical scores. I will dedicate a section below (see 2.7.) to discuss in some detail the specific visual aids that could be transferred, oriented both to the facilitation of visual encoding processes and to the facilitation of the perception of structure (see also next section, 2.5.2).

What is remarkable is that the VWFA seems thus to be attuned and sensitive to *meaningful* combinations (or at least combinations aimed at meaning) of symbols from a given set, irrespective of the visual properties of said set; furthermore, the form-to-phonology (or form-to-production) mapping of a given set can also use several routes without affecting the level of activation of the VWFA, since Chinese characters, which are in this respect dramatically different to the functioning of Roman alphabet symbols, have also been shown to attune to the sensitivity of the VWFA [Bolger, Perfetti, & Schneider, 2005; Liu *et al.*, 2013]. This could be taken as indicating that symbolic systems irrespective of their surface qualities or ultimate mapping route are all visually encoded in the same manner through the VWFA, since it manifests activation for systems with great variations in type of character or even reading axis. It should finally also be mentioned that this consistent activation has even been shown in within-subject comparisons for individuals (native Japanese) using two different symbolic scripts: the Kanji (logographic, one character per word or phrase), and the Kana (one character per syllable) [Nakamura *et al.*, 2005].

Visual encoding seems thus to be equally efficacious for logograms and for structured, integrated strings of characters. Since models of music sight-reading [*e. g.*, Sloboda, 1976c, 1976b; Wolf, 1976] propose the presence of processes including abstract coding for groupings of items, in concordance with expertise studies [*e. g.*, Sloboda, 1978a; Palmer & Drake, 1997; Drake & Palmer, 2000], showing an effect of training on the tendency to perceive groupings or contours of notes rather than individual symbols, and equally in accordance with eye-movement studies [*e. g.*, Kinsler & Carpenter, 1995] demonstrating a tendency towards hierarchical structuring of symbols rather than towards using item-by-item fixations in proficient music readers, the implementation of visual cues in a score that facilitate the integrated encoding of symbols by separating groups of notes could be an efficient strategy in trying to enhance the fluency of sight-reading.

2.5.2. NEUROPHYSIOLOGICAL MAPPINGS OF STRUCTURAL INTEGRATION PROCESSES

The hypothesis that resources might be shared for structural integration processes and sequencing in music and language is supported by (A) experiments that reveal interactive effects between music-syntactic and linguistic-syntactic processing through the use of interference paradigms, and (B) by studies of reading impairments in different domains, which place particular interest in possible multi-domain associations or dissociations.

The findings, both in terms of interactive effects of music and language syntax, and in terms of associated impairments for both faculties, show the presence of shared neural resources, as will be detailed below.

These findings resonate with a theoretical positioning by Katz and Pesetsky [2011] positing that the differences between music and language arise from the intrinsic characteristics of their building materials (*e. g.*, discrete *vs.* gliding pitches), not from the way in which these materials are structured into discourse (*i. e.*, the way they are processed syntactically) [see Pesetsky, 2009, for a precedent of this position]. Katz and Pesetsky [*id.*] further argue that syntactic movement (Internal Merge, in Generative terminology) is found in music in the same way as in language, and can be exemplified by the Perfect Cadence (in tonal music, a specific ordering of harmonic functions).

But even Katz & Pesetsky's contention that the characteristics of the building materials of language and music are intrinsically or categorically different could be taken issue with, on the grounds of the literature that indicates a common (behavioural, cognitive, and neural) basis for processing linguistic and musical relationships. For example, in an Event Related Potentials comparison of the auditory processing of linguistic prosody and the processing of pitch intervallic relationships in music, Glushko *et al.* [2016] report a positive potentials shift (*i. e.*, a specific increase in activity) at the onset of musical phrase boundaries that strongly resembles the language Closure Positive Shift. Moreover, the language Closure Positive Shift in musicians was found to be less prominent (indicating less processing effort) than in non-musicians, suggesting more efficient processing of prosody in language as a result of musical expertise. In addition, and relevantly for reading studies, Drury *et al.* [2016] have provided evidence that mechanisms that subserve speech prosody processing —specifically referring to the components elicited in Closure Positive Shift— play an active role during silent (*i. e.*, not overtly acoustically based) reading as well³. In other words, the *lack* of cognitive/neural differentiation between music and language extends far beyond the syntactic domain [for a recent overview of the relationship between music and language accounting for these processing similitudes, see Jentschke, 2016]. However, since the study of the syntactic domain has generated some relevant debates in the last years, it will be reviewed somewhat in detail.

A. INTERACTIONS BETWEEN SYNTACTIC PROCESSINGS

Hoch, Poulin–Charronnat, and Tillmann [2011], for example, found interferences as an effect of a musical chord's tonal function (violating or not the music-syntax expectations at the end of a fragment) on the syntactic processing of language (processing of sequence endings —a decision task on the last word), with the results being considered particularly relevant to the observation of common *structuring* processes since there was no interference of music syntax with *semantic* expectations. In an overview of papers focusing on similarities between music and language in terms of structural integration and cognitive sequencing, Tillmann [2012] posits that, since music can be described as a sequence of events that are structured in various parameters, studying music processing can provide insights into how complex event sequences are

³ This interaction with production (prosody) could seem to be in contradiction with the findings in visual identification studies showing independence from phonology [Glezer *et al.*, 2015; Glezer *et al.*, 2016], but the processing mentioned here is at phrase level (openings and closures), not at early word identification level (see also SECTION 2.5.1.C.).

learned, perceived, and represented by the brain; Tillmann proposes that this would account for the recent development of research looking into shared resources between music processing and the processing of other structured stimuli. (Incidentally, Tillmann only mentions pitch and time as musical parameters for which events can be structured, but any contemporary musician would understand density, texture, or dynamics—to name just the more obvious—as parameters that can be similarly structured.)

But it is in fact the study of the *sensitivity* to structure in the human brain that has triggered some of the research debates in the neuromapping of the processing of music and related stimuli (particularly language, but also arithmetic). More specifically the relevance or precision of some of the allegedly structural manipulations implemented in these studies has been questioned. Fedorenko *et al.* [2012] have pertinently argued that, although the use of structural violations in music processing studies is an approach that has high temporal acuity and is thus well suited for analysis, it might be problematic, if not well designed, since violations sometimes recruit generic processes that are engaged by irregularities across many different domains. To which one can add that the assumed homology between rule violations in the musical and linguistic domains in some of the studies in syntactic integration (most notably, Slevc, Rosenberg, & Patel, 2009) breaks down when one looks at what was actually done: Ross [2014] covers this in detail. Along the same lines, Koelsch *et al.* [2005] had demonstrated that all of the brain regions that respond to structural violations in music also respond to other (more evident) auditory manipulations, such as unexpected timbre changes. Equally, Tillmann *et al.* [2006], in studying cognitive priming in music, argued that strong musical manipulations might explain certain activations, supposedly related to violations, by mere sensory deviance detection.

Therefore, rather than the use of unexpected elements in a sequence (the ‘violation’ approach), the re-arranging or ‘scrambling’ of elements of a given parameter has been proposed as a way of reliably measuring brain responses that differentiate structured from randomised discourses. Still, Fedorenko *et al.* [2012], using a scrambling procedure that manipulated musical structure by randomizing the pitch and/or timing of each note of a musical excerpt, find brain regions with representations of musical structure that coincide with previous neuroimaging findings [*e. g.*, Fiebach & Schubotz, 2006; Tillmann *et al.*, 2006] in pointing at the inferior frontal cortex (Broca’s Area; see details below) as highly sensitive to musical structuring. What Fedorenko *et al.* [2012] further show is that these regions are sensitive to parameter-specific (*i. e.*, pitch or time, in this case) manipulations, thus reinforcing according to these authors the view of these regions as detectors of structural violations that can be related to the predictions in a particular parameter, rather than being generic detections of unexpected salience of an event. As was the case in the study by Tillmann [2012], though, the use of restricted parameters (pitch and time only) and the use of materials (arrangements of pop entertainment from the ‘50s and ‘60s) that are highly homogeneous and predictable in all their aspects (and certainly in melodic and rhythmic terms) implies a somewhat reductionist vision of what can be termed as music, since in many other forms of musical

composition ambiguity or floating intentionalities are primed. The predictability of the materials furthermore means an increased risk of Type I statistical errors when measuring sensitivity to incongruities.

Similarly looking at sensitivities to syntax in specific structures (again, only small-scale melodic and temporal structurings were considered), and using to this end the examination of the auditory perception of song and speech, Merrill *et al.* [2012] compared utterances of both modalities (language and music) in three conditions: spoken and sung sentences; hummed speech prosody and song melody; use of only the rhythmical component of speech or music. Systematic contrasts within these conditions showed a great overlap between song and speech at all levels in the Bilateral Temporal Lobe [Alonso *et al.*, 2014, showed agreeing results with patients suffering hippocampal lesions], although they suggested a differential role of the Inferior Frontal Gyrus (IFG) and Intraparietal Sulcus (IPS) in processing song and speech: while the left IFG coded for spoken words and showed predominance over the right IFG in linguistic (prosodic) pitch processing [similarly confirmed in Alonso *et al.*, 2014], an opposite lateralization was found for pitch in music; also, the IPS showed sensitivity to discrete pitch relations in music as opposed to the gliding pitches used in speech. But these differences between language and music processing arise, as predicted by the model by Katz and Pesetsky [2011], from the idiosyncrasies of their building materials, not from the way in which the structural integration is processed.

B. ASSOCIATED AND DISSOCIATED IMPAIRMENTS IN WORD READING AND MUSIC READING

Neurophysiological mappings show that music reading impairments can be either dissociated from or associated to word reading impairments [for a review of the literature on brain damage and music reading, see Hebert & Cuddy, 2006].

On the one hand, disorders can occur limited to one domain. Lebrun *et al.* [2012] describe a case of congenital amusia that appears limited to music since the patient's audiometric as well as intellectual and language skills are normal. Moreover, in spite of an initial assessment of severe problems with both melodic and rhythmic discrimination and memory for melodies, a closer study found that, in singing, the patient made more pitch than time errors. Concurrently, an analysis of electric brain responses found abnormal mismatch negativities only to small changes in pitch.

Contrary to these disorders in fine tuning of intervals, impairments in timing, sequencing and structuring of discourse seem to be modally interlinked in many cases. For instance, temporal processing deficits have been associated with dyslexia, and specifically timing interval evaluations have been shown to have much larger differential threshold values (leading to evaluations and ordinal comparisons that are less precise) in dyslexic patients [Rousseau, Hebert, & Cuddy, 2001]. Congruently, in a case of music dyslexia,

Hebert *et al.* [2008] found a dissociation of pitch and rhythm reading abilities (with marked timing deficits) not quite the same as in control cases.

Patel [2003] has argued that the fact that syntactic comprehension problems in Broca's aphasia are not selective to language but influence music perception as well could be a further proof of these links [see Patel *et al.*, 2008 for an example of aphasic individuals showing impaired processing of musical syntactic relations]. Focusing more restrictedly on the perception of time, in a study aimed at identifying brain activations during sustained perceptual analysis of auditorially and visually presented temporal patterns (rhythms), Schubotz, Friederici, and von Cramon [2000] consistently found that the neural network supporting time perception involves the same brain areas (Broca's and adjacent) that are responsible for the temporal planning and coordination of movements. This would be very much in line with the conception of Broca's Area as a processor for *any* hierarchically organised sequential behaviours. Indeed, Schubotz and von Cramon [2001] subsequently found activations within the same neural network for subjects attending to sequences of visually presented stimuli, with independence of whether this was followed by motor reproduction or by perceptual encoding; however, the motoric and pre-motoric parts of the network were (as expected) activated to a *higher degree* before motoric reproduction (specifically, the pre-Supplementary Motor Area, the Supplementary Motor Area, the Primary Motor Cortex, and the Medial Cerebellar Cortex).

2. 6. MUSIC READING AND THE LEXICOSEMANTIC ROUTE

As has been noted above (see SECTION 2.5.1.), early visual identification of meaningful groupings of symbols follows similar pathways (with responses in the VWFA) for both music and language scripts. However, visual identification is only a first step in the reading process. According to the Dual Route theory [for a meta-analysis of studies of these two possible access routes, see Jobard, Crivello, & Tzourio–Mazoyer, 2003; also Mongelli, Cohen, & Bartolomeo, 2014, for the relations to legibility] the visual decoding is followed by two kinds of treatments. On the one hand, the graphophonological (the indirect) route applies *individual* symbol-to-sound correspondences (grapheme-to-phoneme in language; note to sound in music). On the other hand, the lexicosemantic (direct) route facilitates the association of visually integrated forms (whole words or phrases in language; ‘motifs’ or ‘gestures’ in music) to their meanings. It is only by the conjunction of both routes that final phonological or lexical output (execution, in the case of instrumentalists) is achieved. It has to be noted, though, that the concept of ‘meaning’ is problematic when applied to music, and therefore a possible lexicosemantic route that would combine with the graphophonological route (or ‘graphoexecutive’, for instrumentalists) is questionable in music reading (see Figure 2.2.) [for an overview of theories on the elusive relation of music and meaning, see Cross & Tolbert, 2009].

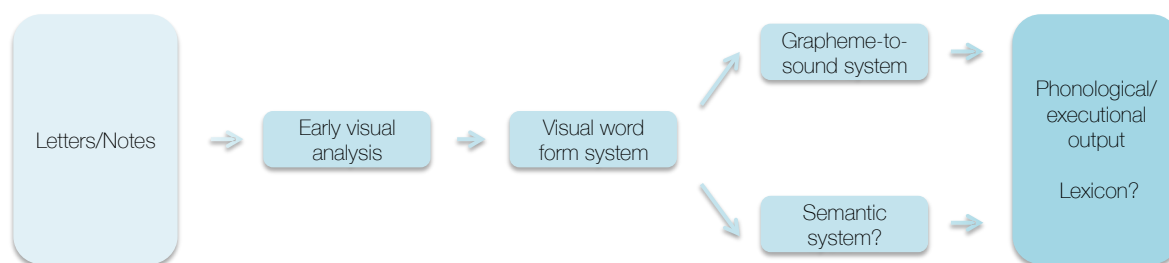


Figure 2. 3.

[Adapted from Jobard, Crivello, & Tzourio–Mazoyer, 2003; see also Mongelli, Cohen, & Bartolomeo, 2014]. Schematic model of the Dual Route Theory, both for language and music reading. The semantic system and the output lexicon do not have exactly the same characteristics in both modes.

But since the semantic route is more efficient and direct, it would be particularly useful to enhance it in reading situations —as most music sight-reading situations are— where time constraints and stressful conditions apply. Therefore, in this section I will discuss somewhat in depth the possible semanticity of (readable) music. A basic distinction that is normally made in terms of the possible meaning of musical

discourse is that between what could be termed ‘absolutist’ and ‘referentialist’ approaches [see Gabrielsson, 2009]. An absolutist approach to meaning will focus on intra-musical relations, with music considered self-contained. Referentialism implies, on the contrary, that music points to extra-musical phenomena.

2.6.1. SEMANTICITY IN INTRA-MUSICAL RELATIONS

The tradition of discussing music, or musical meaning, purely in self-referential terms goes back to, at least, the mid XIXth century. The Austrian music critic, aesthetician and historian Eduard Hanslick challenged at the time a tradition of aesthetic thought that located the meaning of music in a loosely defined ‘expression of feelings’, and his work has remained a touchstone in musical-aesthetic debates to the present day [Hanslick, 1854/1905; on Hanslick’s historical relevance, see Grey, 2001]. This vision of music as a self-contained language was equally defended by none other than Theodor Adorno, who aimed to decipher the content of music (even its historical and social content) from the interior of the musical work, with the concept of musical material being central to his thinking [see, *e. g.*, Adorno, 1968/1997, 1969/2002; on Adorno’s positioning and his highly influential role, see Paddison, 2001]. Many major composers of the XXth century subscribed to this absolute conception of music, and focused on the intra-musical relations of the discourse, particularly when serialised or algorithmic approaches to creation were adopted [to name just one extreme example of total serialism, see Messiaen, 1949, *Mode de valeurs et d’intensités* —this particular composition had a crucial impact on the generation of Boulez and Stockhausen]. Igor Stravinsky’s saying that “if music appears to express something, this is only an illusion” is frequently cited in this respect [*e. g.*, Fisk, 1997, p. 280; Gabrielsson, 2009, p. 141]. In this sense, composers would thus be in charge of ordering elements in a self-contained world: “musical tones inhabit and form a universe of their own, and with the human mind have created musical materials and reduced them to order” [I. Stravinsky, cited in Oliver, 1999, p. 213]. In a parallel with language, one could say that by taking this approach composers are creating their own artificial grammars: “The function of the creative artist consists in making laws, not in following laws ready made. He who follows such laws ceases to be a creator” [F. Busoni, cited in Oliver, 1999, p. 224].

It is thus possible, and common in XXth- and XXIst-century aesthetics, for music to not aspire to semanticity in the sense of pointing towards anything outside the formal rules of its own grammar, and certainly not pointing towards defined real concepts. This lack of connection with the real world and the concepts that natural languages present can lead to polyvalent significances of a given musical discourse when translated to the real world circumstances of different individuals. In fact, some authors see these ambiguous (subjective) meanings ascribed to music as an essential, delineating part of the social and cultural functions fulfilled by music [*e. g.*, Cross, 2005, 2009]. Furthermore, when considered as a communicative medium, music has been proposed as *optimised* for the management of uncertainty, and

precisely therefore complementing language in the human communicative toolkit [Cross & Woodruff, 2009].

However, even in absolute, non-referentialist music, an internal meaning *can* be derived from the ways in which a recognisable unit of the discourse (a cell, a motif, a fragment), relates to other units of the discourse that are interpretable as capable of bearing similar kinds of meanings. In this endogamous defining relation, the ‘sense’ —to use the Fregian term— of a given musical expression arises by virtue of its articulation *within* a certain compositional system or musical idiom. Meaning would thus be a consequence of a musical gesture’s capacity to be bounded by relationships to other gestures, within a system governed by formal rules. Using a more specific ambit of reference, Meyer [1967, 1973] had similarly suggested that music’s meaning could be observed at those moments when an established pattern has given rise to an habitual response, but the pattern has then been interrupted in some way [for a contextualisation of Meyer’s proposals and terminologies, see N. Cumming, 1991; Sparshott & Cumming, 2001]. Meyer applied this model of patterns and predictions to analyse XXth-century trends in music [1967], and also developed techniques for analysing melody based on the Gestalt ideas of pattern-completion and good continuity [1973]. In the later 1970s and early 80s Meyer turned to the study of short recurrent patterns in music, and proposed that they could function as cognitive schemata (*i. e.*, including a conception of what is common to all members of a class; therefore with predictive value).

Following Meyer’s methods (or similar detailed examinations), in non-referentialist music, through the analysis of the repertoire of specific composers or idioms, certain recurring patterns that gain meaning by their internal contextualisation in a piece can be defined. Consequently, their visual identification and encoding could be enhanced expecting a prevalence of the direct route to execution (and the creation of a ‘lexicon’). Equally, memory-based, semantic representations of these patterns would form the basis for direct lexicosemantic interpretations. Thus, the increased neuronal specificity for new (learned) words in the VWFA studied by Glezer *et al.* [2015], showing that novel shapes can be added to this visual dictionary of the brain could perhaps be exploited in music pedagogy, particularly having in mind completed repertoires of prolific authors (could thus a ‘dictionary of Bach gestures’ be compiled?).

2.6.2. SEMANTICITY THROUGH REFERENTIALISM

In music with absolute aspirations ‘reference’ —to continue with the Fregian nomenclature—, through mapping to the real world, would thus be excluded. The (relatively rare) exceptions to this (in written music) would be the specific cases where music has an overt referential or signalling component, for example in Messiaen’s compositions referencing birdsongs [*e. g.*, 1953, 1956, 1956–58, 1960]; although Griffiths [2001] has argued that Messiaen’s famous musical copying of the songs of particular species he had heard in nature, apart from including the necessary adaptations to human instruments, temperament

(tuning), and timescale, were also dependent on adjustments and aesthetical priorities, meaning that “his birds are recognizably themselves, but they are also recognizably his”. Along the same lines, Griffiths [*id.*], further posits that “one of the attractions of birdsong for Messiaen may have been that it allowed him to diminish or ignore the distinction between reality and representation”, making an argument for Messiaen’s use of birdsongs in an iconic (not symbolic) manner that would integrate the icon into his own modes of composition. Examples in the music of other composers (*e. g.*, in many a ‘pastoral’ evocation) arguably also always show a merging of the transcription of what is given and the artist’s assertion of an individual compositional technique or formal conception.

A. REFERENTIALIST MUSIC AND SEMIOTIC ANALYSIS

In a semiotic interpretation [using C. S. Peirces’s categories; see Hoopes, 1991] it could therefore be argued that musical gestures that attempt to reference the external (non-musical) world would be defined as iconic (*i. e.*, pointing by virtue of resemblance), rather than indexical (*i. e.*, based on necessary connections) or symbolic (based on established convention) [*n. b.*, we are talking here of the curious attempts at reference to reality through music; on the other hand, the relation between note and sound is clearly of symbolic nature; see in this respect Nattiez, 1990]. The fact that aesthetical preferences or compositional needs will necessarily distort the resemblance to reality in an artistic production even limits the applicability of the term ‘iconic’, particularly from the functional point of view; the sign, in the best cases, becomes a part of the composition, and loses or blurs, as Griffiths proposed, its representational function. Cross and Tolbert [2009] survey the incidence of this ambiguity related to musical discourse (*n. b.*, not to musical notation), and conclude that it “limits the explicitness with which the bases for its meanings can be articulated in semiotic terms” [p. 25].

B. MUSIC AND EMOTIONAL ASSOCIATIONS

The adscription of meaning to music, apart from through imprecise or questionable semiotic avenues, could perhaps be progressed through the study of its undeniable emotional impact. In this respect, a first important distinction has to be made, between perception and induction of emotions; we may perceive an emotion, or we may feel an emotion in response to the music [Juslin, 2009; Gabrielsson, 2009]. This distinction is often made in modern research, for two reasons: the underlying mechanisms may be different depending on the process involved, and the types of emotions may be different in each case [but see, on the blurred borders between these alternatives, Kallinen & Ravaja, 2006; also P. Evans & Schubert, 2008; E. Schubert, 2013].

Since the *perception* of emotions takes an analytical perspective, ‘meaning’ in this case is a (mostly academic) exercise in dimensioning within a framework of stylistic or historical conventions —rather than the expression of a connection of musical discourse to external references (which would be an adscription of

meaning *sensu stricto*). The following paragraphs will therefore review literature on the *induction* of emotions.

FOCUS ON REPERTOIRE. The abundant literature looking at the induction of subjective emotional (and related psychophysiological) responses to music has largely focused on repertoire or at any rate on complete pieces, and not on the syntactic components or subcomponents of the discourse (contrary to the error-detection or priming literature, which has by large narrowed its focus to specific grammatical links). Therefore, in terms of music reading and its possible dependency on access to a ‘lexical’ output and, more specifically, the possible association of a given visual pattern in a score to a certain (emotional) meaning, the literature on the psychophysiology of induction of musical affect can generally not offer much guidance at this stage (2017).

RESPONSES TO ACOUSTIC CHARACTERISTICS. On the other hand, results in this literature [*e. g.*, Egermann *et al.*, 2015] confirm that changes in the subjective dimension of arousal involve a basic, universal response to low-level acoustical characteristics of music (or sound, rather) [see further Irrgang & Egermann, 2016, on how these basic relations are equally reflected in expressive movements]; it could therefore be useful to cue visually in a score the scaling of these sonic characteristics (loudness, attack, range, frequency), since they seem to elicit universal responses in subjects (although research so far has been conducted on listeners, not on readers, and not on specific syntactic elements, as mentioned) —this is one of the many research avenues that could be explored in novel score designs, if interdisciplinary integration with studies of music and emotional responses was advanced.

DIMENSIONS OF RESPONSE AND THEIR UNIVERSALITY. Studies in the literature looking at the emotional meaning of music normally present to participants a rating interface with limited dimensions, in order to avoid completely subjective descriptions of meaning. In fact these interfaces in most cases focus only on two dimensions: arousal (with scaling from ‘relaxing’ to ‘energising’) and valence (scaling from ‘negative’ to ‘positive’). Contrary to what has been found for the subjective dimension of arousal (eliciting universally convergent responses), the dimension of valence might be mediated by cultural learning [see Egermann *et al.*, 2015], although recent experiments [Irrgang & Egermann, 2016] have found that both the qualities of arousal and valence could be predicted by acceleration data of participants, when assessing their expressive movements stimulated by music (and therefore present patterns mediated by basilar, universal responses). In any case, the adscription of meaning to musical discourse through the scaling of its valence ratings has equally, as in the case of arousal or other less studied dimensions [for instance, Joy, Transcendence, Wonder, Power, or Tenderness, as proposed by the Geneva Emotional Music Scale, GEMS; see, *e. g.*, Labbe & Grandjean, 2014; Pearce & Halpern, 2015] has been done through the study of reactions to whole pieces or extended fragments, which renders it of very limited value for the implementation of local visual cues that could facilitate a lexicosemantic route in music reading.

2.6.3. SEMANTICITY AND THE DETECTION OF MEANINGFUL GROUPINGS WITHIN SETS

But even if the referential function of music is, as seen, at least problematic, the reading of music is nonetheless the reading of a regulated and formalised system of symbols, that clearly elicits activations in the brain in an area dedicated to the processing of visual forms that bear a meaning within a set, even if it is not a meaning ultimately referring to reality. In this respect, it is worth noting that the preferential responses for another formal symbolic system, namely the number system, are equally located in the inferior temporal gyrus, close to (but distinct from) the VWFA [this adjacent area has been labelled Visual Number Formation Area, VNFA; see Shum *et al.*, 2013b, 2013a; Merkley, Wilkey, & Matejko, 2016]. A similarly distinguishable area in the visual cortex having preferential responses to music notation has not been reported so far (2017), but musical notation has, as mentioned, been proven to distinctly activate the contiguous Visual *Word* Formation Area.

2.7. DIMENSIONS OF THE READING TASKS

2.7.1. SEPARATION OR CONCURRENCE OF DIMENSIONS

Whether hierarchical integrations in perception, and more specifically in reading, are performed by focusing attention to an identified dimension or concurrently for several dimensions has been a debate present in the language and music literatures, perhaps more saliently so in the latter, given that in music different dimensions (particularly rhythm and pitch) are quantifiable by definition.

The starting point for the focus on the different dimensions that underlie lexical and grammatical integration is the assumption that selectively attending to one-dimensional rather than multi-dimensional stimulus features does improve behavioural measures of performance, as does orienting attention to a particular value of a defined dimension. This is any reader's experience, and has been shown by both behavioural and neuroimaging measurements. For instance, Coull [2004], in a study summarising behavioural and neuroanatomical correlates of temporal aspects of attention, found that focusing on temporal rather than non-temporal stimulus features improved performance in attention tasks, as did focusing on particular moments in time. Further, these effects were accompanied by specific increases in activity of functionally specialised, and anatomically discrete, brain regions (with the Frontal Operculum playing a pivotal role). But whether voluntary or conscious focus reflects a subdivision of processing in everyday, functional reading (where one-dimensional or discrete focus is not regularly consciously applied) could be debatable.

Studies looking at measurements regarding only one dimension can show results that point in an opposite direction to those found in studies that do not separate dimensions. For instance, Kinsler and Carpenter [1995] measured eye movements in subjects whilst they read and performed lines of music consisting of rhythmic information only, and found a tendency to fixate individual salient details of the notation such as notes and barlines rather than the spaces in between. Apart from the concerns about ecological validity that could be raised by such materials, all lines of study of reading seem to concur on the importance of processes of integration, pattern recognition and prediction, and do present the progressing through materials unit-by-unit using (relatively long) fixations on details as a feature of less skilled readers [*e. g.*, Goolsby, 1994b, 1994a].

Music, due to the intrinsic scalability of its main dimensions, provides a useful domain in which to study how the different attributes of multidimensional stimuli are processed both separately and in combination, but as Waters and Underwood [1999] have suggested, results do not support either strong independence or interactive models of processing. In a reaction-times and electrophysiological study, Schon and Besson

[2002] have suggested that pitch and duration could be processed independently, but the fact that the procedure of their experiments was based on participants making a match/mismatch judgement on a specific dimension (either pitch or duration) which they were required to analyse does necessarily exclude the possibility of supramodal integration; their finding that congruency of targets in the irrelevant dimension did not have a significant effect on the measured Event Related Potentials could therefore just be showing that music readers are capable of using selective attention if needed or prompted to, not that this is *always* (or most frequently) the case for a musician when interacting with a score.

2.7.2. ORIENTING ATTENTION

At any rate, and most importantly, it has also been shown in research comparing internally (*i. e.*, guided by mental representations) and externally (*i. e.*, guided by extrapersonal targets) directed attention that these two crucial aspects of cognition are subserved by extensively overlapping networks [see, *e. g.*, Nobre *et al.*, 2004, on spatial attention]. It could therefore be argued that, even if focusing attention on one dimension of the musical score is not the most usual practice in sight-reading, it could be possible to guide the attention of the player to salient aspects of the discourse without necessarily conflicting with more habitual reading strategies. There could thus be scope for guided adaptive distribution of attention without major friction with personal representations or expectations of spatial layout. That would ultimately mean that it should be possible to direct attention to structurally relevant elements of the musical discourse, and perhaps improve the readability of musical texts by using design strategies that orient reading (even underlining certain dimensions over others) without disrupting the musician's previous mental representations.

Furthermore, Lepsien and Nobre [2006] have shown that attentional orienting can bias information processing not only in the perceptual domain but in the working memory domain as well, optimizing goal-directed behaviour: by using spatial orienting cues that appeared *after* perceptual events (thus enclosing a unit of information) the cost of retrieving items from within working memory was diminished, enhancing performance to a similar degree as spatial 'precues' appearing before perceptual events. Concurrently, brain imaging data showed a high degree of overlap between brain areas and dynamics involved in spatial orienting in the working memory domain compared to the perceptual domain. These findings led Lepsien & Nobre [*id.*] to suggest that the neural representations of objects in working memory can be directly modulated by visual spacing cues, and may include early roles both in guiding spatial shifts within mnemonic contexts and, most relevantly for us, in the selection of memorized targets amidst distracting stimuli.

If visual spacing can act as a 'retro-cue' that helps retrieve items from working memory, it could perhaps be usefully implemented in the design of musical texts to enhance the detection and performance of cells,

motifs, gestures, or phrases in the discourse, by diminishing the cognitive load of retrieval. Even though the findings by Nobre *et al.* [2004] and by Lepsien and Nobre [2006] were not strictly in reading modes but more generally applicable to spatial orientation with multi-dimensional objects, some influential designers of (two dimensional) diagrams —what a score arguably is— conceive the interaction with them in terms of ‘visiting’ or even ‘living in’ these diagrammatic places, and propose that a diagram can ultimately be evaluated in terms of the ‘user experience’ that it affords regarding convenience, comfort or usability, as any other multidimensional object [Blackwell, 2001, 2002, 2008]

2.8. IMPROVEMENTS OF READABILITY USING TEXT DESIGN

One perspective that has —to this author's knowledge— never been taken in music reading studies, and that has been covered by plentiful literature in the language-reading domain, is the analysis of how the graphical design of a text affects its readability. This became a particularly prominent object of scrutiny with the generalisation of desktop-publishing in the 1980s, and has become even more relevant with the profusion of screen-based devices that are used for reading nowadays. Already in the 1980s text design researchers were advocating for an assessment of print-based research in order to work on the new problems that were arising in the design of electronic text (computer-processed and/or screen-based text). In music reading research, though, considerations on the design of a score, on its layout, spacing, or typography have never been *systematically* addressed —in contrast, throughout the XXth century and up to today, there have been numerous examples of personalised notations and score designs, always subservient to the aesthetic or compositional needs of individual composers.

2.8.1. THE USES OF SPACING TO IMPROVE READABILITY

Looking at printed (language) text, it has been observed [*e. g.*, Hartley & Burnhill, 1976a; Hartley, 2004; Hartley & Betts, 2007] that crucial design decisions can be made taking into account the use of white space to help display the underlying structure of the text [see, *e. g.*, on the phenomenal growth of structured scientific abstracts in the last decade, Hartley, 2014]. In relation to this, some designers [*e. g.*, Schriver, 1997; Hartley, 1999] have considered (particularly for less able readers) the use of unjustified text (text that uses consistent or fixed measurements of white spacing between signs), with the beginning and end points for each line determined by syntactic considerations related to the underlying hierarchical structure of the text rather than by fixed margins.

The use of structured space to improve (language) readability has been researched extensively, and has repeatedly and consistently been shown to be helpful for readers, particularly average and poor readers. Furthermore, it has been shown that, provided that the space is not inserted randomly but following the structure of the discourse and respecting its syntactic and formal boundaries, any form of reasonable spacing will improve the readability of a text. Jandreau, Muncer, and Bever [1986] demonstrated improvement in reading performance for poor readers when presented with texts where the major phrase⁴ structure boundaries were marked with increased spacing, and equally when presented with texts where spacing corresponded to both the major and the minor phrase structure boundaries. Moreover, the space

⁴ In this and other parsing or chunking studies, 'phrase' refers to a *small* group of words standing together as a conceptual unit. It is shorter than a Clause, which traditionally is defined as containing a subject and a predicate.

was added algorithmically, without a deep functional analysis of the levels of grammatical embedding or hierarchy. Equally using a computationally heuristic analysis in order to assign extra spaces between word groups, Bever *et al.* [1991] similarly found that readability of texts was improved, and that the spacing added with a simple algorithmic parsing was more effective than complex spacing reflecting a phrase-structure analysis assigning each interword space a size proportional to the depth of the phrase structure at that point. In line with the hypothesis that it is the structuring of the text (even if using a simplified or crude parsing) what facilitates the reading and not merely the increased spacing between items, Bever *et al.* [*id.*] introduced a baseline condition assigning a constant amount of space between each word on a line, which proved detrimental for readability.

2.8.2. EFFECTS OF SPACING ON RECALL AND REPRODUCTION

If structured spacing can affect the reading of texts and the spacing can be added in different manners as it does not necessarily benefit from an extremely detailed correlation with grammatical complexities, the decisions on how to structure a text leave some degree of choice for other parameters to be included, which will depend on function and intentionality (something possibly appealing to composers and music educators). This is of particular relevance since, within the constraint of maintaining integration of design and (basic) discourse structure, the final format of a text can, however, have an important impact on the end user in terms of recall or performance. Hartley [1993] tested recalls by readers with texts in either traditional format or a ‘chunked’ format, where phrases were segmented and presented in separate lines, with additional indentation to show sub-groupings (the texts described genealogical lines). It was found that ‘chunking’ did not significantly affect the amount of recall (both formats were well-spaced and followed the structure of the text). However, the format of presentation affected the format of the recall in a highly significant way: *all* of the participants in the ‘chunked’ condition recalled their texts in a chunked format, and virtually all of the participants with the traditional text wrote out their passages in the traditional way.

Music sight-reading can be considered as a transcription task, where information in one form (notational) is converted to another (performative or kinaesthetic), like reading aloud or copying a text [Fine, Berry, & Rosner, 2006; see also Shaffer, 1978, for a comparison of music performance and typing skills; Shaffer, 1982, for a general theoretical framework for multi-modal skilled performances]. The results in Hartley’s [1993] memory transcription task are therefore pertinent, although in music-reading there is furthermore a time constraint that was not present in the memory task. Sight-reading might be closer to tasks including the presence of an underlying pulse or at the least some level of temporal restriction. Mahalski [1995] timed a group of university students as they read and copied course notes (native language, but unrelated to their own course) with different layouts from an overhead projector screen, and found that copying was faster and errors fewer when a paragraph of unfamiliar notes was laid out to emphasize the

meaningful structure of the text. In line with the findings mentioned above [Jandreau, Muncer, & Bever, 1986; Bever *et al.*, 1991; Hartley, 1993], for texts that had already been structured into meaningful parts with a certain cohesive design and spacing, presentation with different options similarly using structured spacing but varying in surface did not further reduce copying time or errors.

It is also perhaps worth noting that all studies mentioned so far used exclusively typographic and spatial cueing, but no special marks as dividers, guidelines or visual tags [see Gilreath, 1993, for a proposed taxonomy of these graphic cues]. Systematic care was taken with line breaks, phrase grouping, and structured presentation, but no extra symbols were required. Similarly, Jandreau and Bever [1992] carried out reading comprehension tests with average (college) readers, varying the spaces between phrases in the essays to be surveyed: they found that texts formatted so that points between phrases had fractional extra space added to them were comprehended better than normally formatted text; they suggested that poor to average readers specifically lack perceptual strategies for *grouping* word sequences into phrases, and that the extra spacing facilitated fluent reading and comprehension precisely at this level.

2.8.3. RESEARCH ON PERFORMATIVE READINGS

Even if structuring and grouping of information can be advantageous for language readers, music reading most frequently entails conversion to performance, rather than comprehension or copying tasks (which come up in analysis or score study). But the literature on performative (*i. e.*, oral) language reading is convergent with the reviewed studies showing visually spaced parsing as a facilitator for general fluency, memory tasks or comprehension. LeVasseur *et al.* [2006] showed that fluency in oral reading was facilitated by text formatted to preserve phrasal boundaries. In addition, this structured formatting resulted in significantly fewer false starts as the beginning of lines following a return sweep. Importantly—specially for comparison with music reading—the structurally separated text was also marked higher (by two independent experts) in a fluency rating scale giving the greatest weight to proper phrasing in reading aloud: reading in stylistically organized phrasal groups was ranked higher than reading in a listwise, word by word manner. The authors included thus not only speed and accuracy as dimensions of fluency but also proper expression, giving prominence to correct phrasing, emphasis and intonation. The increase in fluency found in readings of the syntactically cued texts entailed thus appropriate phrasing or prosody as well as facile word recognition. A follow-up study by LeVasseur, Macaruso, and Shankweiler [2008], found that also for *repeated* readings fluent prosody was most strongly facilitated by visually cued texts, printed with added spaces between phrases and with ends of lines at clause boundaries. Repeated reading of texts is perhaps the most universally recurrent strategy for any musicians rehearsing a piece, and LeVasseur, Macaruso, & Shankweiler showed that repeated readings with language texts also resulted in gains in fluency, with this gain, crucially, being observed for the repeated readings of meaningful texts, but not in repeated readings of random lists of difficult words.

Prosodic structuring of a text may be an important skill not only for expressive oral reading, but also, as Beggs and Howarth [1985] have proposed, a skill related to the progress of young readers towards fluent, adult reading, since the planning of oral utterances, containing the lexical items visible on the page within an appropriate prosodic envelope, will be linked to interiorization and assimilation of the text. Therefore, Beggs & Howarth, [*id.*] used in their reading paradigms texts that were enhanced not only through spacing, by adjusting line breaks to coincide with pauses, but also by graphically representing the stressed words in each phrase (typed in boldface). In experiments with children at what these authors considered a critical point in learning to read (8–11 years old), participants comprehended text better when the prosodic stress was made visible on the text.

2.8.4. APPLYING VISUAL CUES TO MUSICAL SCORES

In texts designed to be visually informative, many other graphic cues have been commonly used, preserving or not orthographic regularity, and including or not extra mark cueing; the literature covering the options is vast and detailed [see, *e. g.*, Hartley, Burnhill, & Fraser, 1974; Burnhill *et al.*, 1975; Hartley & Burnhill, 1976b, 1977; Hartley, 1978; Bringhurst, 1992; Gilreath, 1993; Hartley, 1999, 2004; Hartley & Betts, 2007].

However, in my approach to the visual representation of musical discourse, I have decided to *not* to include any extra marks or symbolic cues in scores that are usually already quite information-dense (at least in comparison to language scripts). I will base my modifications of musical scores on the extensive literature (examples of which have been reviewed in this section) showing that the use of structured grouping, separation and spacing of materials can be of help for readers, increasing accuracy, limiting hesitations and favouring stylistically cohesive expression.

I have decided to limit the research to the manipulation of spacing, layout, and engraving rules, rather than innovating with the creation of symbols or specific diagrammatic signs, for several reasons, all of which could be bundled around the concept of ‘economy’: (1) the principles governing present day scores have been developed and elaborated within whole fields of study through centuries [as Blackwell, 2013, has pointed out, principles of this category should not be underestimated, let alone disdained]; (2) learning to read music is an arduous and time consuming process that, unlike language reading (at least in contemporary cultures), generally requires personal effort and commitment from the learner (meaning that music readers could be less open to diagrammatic novelties than users of other systems); (3) (also building on the previous point) using a conventional graphic vocabulary of lines and symbols that are understood in our culture to represent music will mean that these elements will not draw attention to themselves, and therefore highlight the structuring function of the spacing and layout [see the pioneering and highly effective work on pictorial representation by I. Sutherland, cited in Blackwell, 2013]; (4) there exists

already, as mentioned above, an extensive, systematic, and rigorous historical (see specifically SECTION 1.2.2.) and psychological literature (see SECTIONS 1.4.1. and 2.4.1.) analysing the use of structured spacing and layout for the facilitation of legibility in natural and (certain) formal languages, with no alternative (or simply any literature) on the issue in the music domain; and (5) (in concordance with the psychological literature) from the point of view of designers, the framing of regions of information has been considered a powerful graphic resource—at the same level as marks and symbols—which can *efficiently* help the identification of shared membership and the segregation or nesting of multiple elements within a discourse (simply by separation with white space, possibly supplemented by alignment grids) [see Bertin, 1967; MacEachren, 1995; Blackwell & Engelhardt, 2002; Blackwell, 2013].

2.8.5. TAKING UP SLOBODA'S GAUNTLET

Some forty years ago now, Sloboda [1978b], in a comprehensive paper on the psychology of music reading, reviewed historical development of research, qualitative and quantitative approaches to its evaluation, coding strategies including measurement of reading span or memory tasks, and pedagogical approaches to the transmission of the skill. He finished with a speculative section where he discussed the possibility of improving musical text, and started the section thus:

In all that has been said so far we have assumed a perfect, immutable score which it is the reader's job to understand and execute. Many of the difficulties involved in reading, however, may be attributable to faults in the score itself, and not to the reader. (p. 15)

In the complete absence of research into these matters at the time, Sloboda [*id.*] proposed a general principle that could guide those interested in the legibility of their musical materials: “to adopt any practice which makes it easier for the reader to perceive the structure of the music” (p. 16).

Excluding the numerous novel or alternative notational systems that many XXth-century composers have devised in order to specify the requirements for their own aesthetics, or the numerous recent attempts at representing music through digital (non stave based) interfaces [*e. g.*, Greussay *et al.*, 1980; Blum *et al.*, 1983; Roads *et al.*, 1986; Battier, 2015], to this author's knowledge the situation remains exactly the same today: in spite of the overwhelmingly numerous literature on legibility of (natural and some formal) language scripts, there has never been an attempt at applying any of the tested principles, paradigms, procedures or designs onto musical scores.

The following chapters are an account of what happened when we took up Sloboda's gauntlet.

ACKNOWLEDGEMENTS

Figure 2.2., showing various category-selective areas in the ventral visual cortex, is an adaptation from Gaillard *et al.* [2006], and Mongelli, Cohen, and Bartolomeo [2014]. *Figure 2.3.*, showing the schematic model of the Dual Route Theory, is an adaptation from Jobard, Crivello, and Tzourio–Mazoyer [2003].



*Experiment I:
Establishing Possible Effects
of Score Design on Reading*

CHAPTER 3

Experiment I: Establishing Possible Effects of Score Design on Reading

3.1. INTRODUCTION: MOTIVATIONS FOR IMPROVING READABILITY

3.1.1. PROFESSIONAL PRACTICES WITH MINIMUM REHEARSALS

In a recent debate on the training of professional musicians, responding to a position paper by Friend [2011] on the quality of education in British conservatoires, Tilch [2011] strongly objected to a proposed structuring of conservatoire orchestra rehearsals, on the grounds that it included too long and meticulous preparation of pieces. Tilch spoke as a member of the Philharmonia Orchestra in London of 17 years to that point and, being of German descent and having been educated in German institutions, mentioned how she was still in awe, at the rehearsals of the Philharmonia, at the sight-reading skills of her British-trained colleagues. The objection against an academic planning of long and meticulous preparation of repertoire for aspiring *professional* musicians was thus that it did “pay no heed whatsoever to the realities of orchestral life and working pace, especially in British [professional] orchestras, where time and money are key and rehearsals are kept to an *absolute minimum*” (p. 7; italics mine). According to Tilch, the academic rehearsal schedules proposed by Friend were overlooking the (nowadays) crucial ability to sight-read, if not overtly stifling it.

The debate could be seen as a sign of a change in priorities and usages in professional orchestras through the last decades, particularly in British orchestras. Friend is one of the most outstanding English-born violinists, a highly influential pedagogue, and was Concertmaster for the London Philharmonic and the BBC Symphony in the 1960s and 1970s. His careful and dedicated approach to music making is unquestionable, but Tilch’s position is also defensible as realistic and empathetic to the worries and needs of aspiring professionals.

3.1.2. ELEVATED STRESS LEVELS IN MUSICIANS

A SELF-PERCEPTION AND REPORTS

In this scenario, today's musicians (specially orchestra musicians) are frequently confronted with the conflicting stressors of public exposure/criticism on the one hand and rapid repertoire assimilation on the other. Psychophysiological studies among performing musicians have found elevated stress levels with negative health effects [for a systematic review of studies, see Vervainioti & Alexopoulos, 2015, who found that public exposure, unknown repertoires, and criticism were recurrently reported as stressors in the literature]. Moreover, the self-reported frequency and severity of performance-related musculoskeletal pain disorders have been strongly related to music performance anxiety [see Kenny & Ackermann, 2015, for a cross-sectional survey of Australian professional orchestral musicians], which in its turn is significantly affected by repertoire assimilation.

Performance related musculoskeletal disorders have in fact been recognized as a common phenomenon amongst musicians, specially professional orchestral musicians. Moreover, musculoskeletal disorders create a significant artistic and financial burden to orchestras and performing bodies [Chan & Ackermann, 2014]. At the same time, Chan and Ackermann [*id.*] found that musicians attributed excessively high or sudden increases in repertoire load a major role in these disorders. A cycle of friction between management and performers can thus be triggered by the incorporation of new repertoires to be rehearsed (and thus to be read). The influence of increased rehearsal load on musculoskeletal disorders might be modulated by perception issues amongst musicians, but in any case the majority of professional orchestral musicians report these sort of disorders at a certain point in their careers [84% in a survey by Ackermann *et al.*, 2014]. Besides, with respect to psychosocial screening, Ackermann *et al.* [*id.*] found that 32% of surveyed musicians returned a positive depression screen, and a 22% for post-traumatic stress disorder.

B PHYSIOLOGICAL MARKERS

At any rate, self-report measurements of emotional and experiential states have been shown to sometimes correlate with changes in levels of non-volitive physiological markers associated with specific psychological states. There is, for instance, substantial evidence for the effects of stress on cortisol levels. Increased cortisol levels have been related to a wide range of stressful performative experiences and conditions [*e. g.*, maternal separation for infants, Larson, Gunnar, & Hertzgaard, 1991; or attention tests with rewarding/punitive conditions, Muller, Budde, & Netter, 1992]. In a study of professional choir singers, Beck *et al.* [2000] found that cortisol concentrations (as measured in pre- and post-performance saliva samples) increased significantly (37%) during performance.

C ANXIETY AND SUCCESSFUL PERFORMANCES

However, as Beck *et al.* [2000] acknowledge, performing music is a complex of experiences, and some aspects of it may compensate or even overcome others. The authors also measured levels of secretory immunoglobulin A (S-IgA); its proportion to the whole number of proteins in a saliva sample is considered a reliable indicator of immune system responses (with healthy enhancements of the immune system reflected in higher levels of S-IgA). Counteracting in a way what had been observed with the cortisol levels, immunoglobulin levels also increased significantly (240%) with performance. The authors proposed that the immune system functioning is likely to be affected both as a response to anxiety (inhibition; increased cortisol levels) and as a function of the pleasure of the singing experience (activation; increased immunoglobulin levels). Importantly, the singers involved rated the analysed performance as highly satisfying, and the authors ventured that for the majority of singers the positive emotions overcame the anxiety during the course of the performance.

Anxiety is, and will be, in many cases, part of a performative experience. It could be argued that it is how this anxiety is evaluated and counterbalanced with other factors what could distinguish successful from frustrating performances. Brotons [1994] examined the physiological, psychological, and behavioural components of performance anxiety and performance quality in music students, and found that jury conditions triggered significant increases in heart rate and State Anxiety scale scores when compared to non-jury conditions, but no significant changes in performance quality ratings (of recordings) by judges, or in ratings based on behaviour analysis of videotaped students' performances.

3.1.3. MUSICIANS' OCCUPATIONAL ROLE CONCERNS

But in fact, the personality traits related to evaluative processes of many (professional) musicians are associated with increased levels of anxiety and occupational stress. In a study of highly accomplished singers, Kenny, Davis, and Oates [2004] found that music performance anxiety was *not* related to occupational ranking or to issues related to the physical environment (dust, humidity) or working conditions (travel, rehearsal rooms), but rather to *trait* levels of anxiety (a predisposition towards experiencing and reporting anxiety), associated with high personal commitment and strain in the work environment. Notably, these singers concurrently showed higher occupational role concerns than normative samples [see furthermore Yoshie, Kudo, & Ohtsuki, 2008 on the correlation of trait anxiety to heightened perceptual sensitivity to physiological and behavioural changes].

Put simply, musicians tend to care a lot about what they are doing [Liston, Frost, & Mohr, 2003, in a study with music students, found 'catastrophising' to be the main trait predictor of musical performance anxiety]. Reading music, in this sense, cannot be justly compared with language reading, which only

exceptionally (*e. g.*, for an actor, or someone giving a speech) is related to occupational stress, perfectionism, aspiration, and performance anxiety.

3.1.4. AN INITIAL ATTEMPT TO IMPROVE LEGIBILITY IN MUSICAL SCORES

Legibility of musical texts could be considered therefore particularly relevant, due to its performative implications. Scripts in many other professional or highly committed domains can be imagined to have a crucial importance on major issues as, for instance, security or personal health. But in very few ambits is a text specifically intended for reading under stressful conditions, including strict pacing and time constraints, by individuals with personality traits commonly including the combined predisposition to perfectionism and to develop psychosocial disorders.

The main aim of the initial experiment was then to generally determine if the visual aspect of a score can affect its performance, and more specifically its legibility. If legibility can at all be affected by the visual cues in a score, a whole new line of research could open, in which different elements of the design of a score (integrated or separately) should be tested for their effect on accuracy, expression, or memory.

It has to be taken into account that at this stage the design of the paradigm, the design of the coding strategy and the possible analysis were partly heuristic, since there was no specific experimental research to build upon.

3.2. METHOD: COMPARISONS OF PERFORMANCES WITH DIFFERENT DESIGNS

3.2.1. OVERVIEW OF AIMS AND STRATEGIES

As the aim of this and subsequent experiments would be to evaluate the impact of score design on *legibility*, I decided to focus on the reading of materials that would not form part of the core repertoire of the participating instrumentalists, in order to avoid, in as far as possible, the involvement of long- or short-term memory of a body of pieces regularly performed. At the same time (although this might seem a contradictory aim), I wanted to use a set of pieces that would be part of common practice, that could not be criticized for being selected from an obscure author's catalogue in order to match the agenda of the experiments and hypotheses. The selected pieces would furthermore need to form a homogeneous corpus, and needed to be manageable (in terms of length and difficulty) in sight-reading exercises. This led me to select pieces from the collection of four-part chorales by J. S. Bach, and to show them not to keyboard players or singers (who would in most cases recognise the materials immediately), but to percussion players reading on mallet instruments (marimba, vibraphone or xylophone).

This had the advantages of using an unquestionably central and seminal corpus of pieces in modern western traditions, and at the same time of working with performers that are usually noted for not being the most proficient in reading or thoroughly familiar with this repertoire. There were, moreover, a few secondary advantages:

1. The attacks on mallet instruments are relatively sharp and with a short decay, which meant the coding of mistakes and hesitations would be somewhat clearer than with other instruments (this proved very beneficial in subsequent experiments, when filtering the recordings through spectrogram onset detectors; see SECTIONS 4.2.5.B. and 5.2.5.).
2. As the performances would not only be audio- but also video-recorded, this meant that in certain passages that were difficult to code (*e. g.*, in determining whether a delay on the attack of a note was due to expressive or articulatory intentions, or due to a hesitation), the kinaesthetical component of performance would be much easier to interpret (compared to performances on other instruments less dependent on body position and movement). See SECTION 3.2.8. on coding protocols.
3. Having been a percussion degree undergraduate myself at the *Real Conservatorio Superior de Música* in Madrid, I was aware of the idiosyncrasies and limitations involved in adapting materials for these instruments. Reading on a mallet instrument is broadly equivalent to reading on a keyboard with only fingers 1–2 in each hand. This is furthermore complicated

by the fact that distances between keys are unstandardized (varying depending on instrument brand or make), irregular (changing within instrument depending on octave), and large (involving up to shoulder, torso, and whole body displacements). I was aware that it was likely that the pieces by Bach would have to be adapted for two- or maximum three-voice settings, and that a few small adjustments would be needed in order to avoid unidiomatic positions. In order to have a conventional version to compare to, these pieces would have to be adapted from well-known keyboard editions rather than from cantatas or primary sources (see SECTION 3.2.5.A. for details on selection and adaptation of materials).

4. There would be a (small, but highly relevant) network of colleagues, presently percussion teachers at various prestigious conservatoires in Europe (Amsterdam, Barcelona, Rome, Saragossa) that would afford access to reliable participants, furthermore with access to information about their competences and initial sight-reading abilities (which would be very helpful in counterbalancing procedures). See SECTION 3.2.4. on recruitment of participants.

The feasibility and pragmatics of this setup were tested in a pilot study, after which I proceeded to the first experiment proper.

3.2.2. PILOT STUDY

A pilot study was implemented at the *Accademia Nazionale di Santa Cecilia* in Rome, with 4 students (all male) of their Percussion Department, within the *Alta Formazione* (Master; MMus) courses. Each participant read four Pieces (two-voice arrangements of Bach chorales 005, 006, 015, and 031 from the collection by K. Schubert, 1765/1990) in two different Versions (Conventional, following Schubert, and Modified, with our design novelties). A brief presentation of the general aim of the study to all the students of the class of *Maestro* E. Giachino was followed by individual video-recorded sessions of four students selected subsequently. An adequate individual tempo was gauged for each participant on the basis of a short enquiry on number of study years and priorities (although all were oriented towards orchestral —non-pitched— percussion), and a short trial of materials (one phrase) similar in difficulty to the Pieces to be read in the test proper. *Figure 3.1.* presents the timeline.



Figure 3.1.

Schematic timeline for the PILOT EXPERIMENT: After a short presentation to the class, participants volunteered to participate; the pre-test used a short fragment, to decide upon an individualised tempo; in the test, four Pieces were read, two times, alternating Conventional (C) and Modified (M) Versions of the score. Order of Pieces and Versions were counterbalanced between subjects—shown here is the sequence for participant 1.

Approximate durations: presentation 5 min; individual sessions 20 min.

It was hypothesized that contextualisation (reading in a new key signature and how this relates to the previous one) might be a significant factor in sight-reading, particularly for not very proficient readers. In order to be able to factor out in statistical analysis the effect of the alternation of the Pieces (which were for the rest of very similar difficulty), these were ordered with an attempt to avoid the creation of reinforced tonal gravitations between them, and at the same time to try to have a regular tonal distance between the ending of one Piece and the beginning of the next. The key signatures and lengths were as follows: Piece 1, G Major, 8 phrases (digressions to d minor and a minor); Piece 2, F Major, 4 phrases (digression to C Major); Piece 3, d minor, 6 phrases (with strong presence of a minor); Piece 4, a minor, 5 phrases (digressions to C Major and G Major). Thus, of the 24 possible orderings of the four Pieces—and also taking into account that I would have a limited pool of participants—I included only 2-4-3-1 and 4-2-1-3 (avoiding, *e. g.*, the links 2-3 or 3-4 as too tonally close, or the link 1-2 as being potentially disorienting for beginning readers, specially due to the length of the consolidation of G Major in Piece 1). These two orderings of Pieces added to the two possible orderings of Versions created four possible sequences of performance.

Participants were given 30 seconds to look through the Piece, after which a metronome would start clicking at the decided tempo for two bars (7 beats since all Pieces started on an anacrusis). The metronome was kept on for the first phrase, until the first fermata. Participants then continued reading, trying to maintain the initial tempo.

Although the participants were highly competent as orchestral percussionists (the main aim of these prestigious Alta Formazione courses) their level of proficiency as sight-readers with mallet instruments was somewhat lower than expected. Thus, even though tempi were set relatively slow, the number of mistakes was rather high, with all participants deviating substantially from the presented score (adding and

eliminating fragments, changing figures, stopping the discourse, *etc.*). Therefore, a detailed coding of the readings was not deemed useful or even feasible. However, participants showed genuine willingness to co-operate, interest in the project, and all four of them manifested preference for the Modified Versions over the Conventional Versions when questioned at the end of the session. Finally, and perhaps most relevantly, it was observed that participants tended to perform more or less at the same level of sight-reading proficiency with both Versions (in terms of deviations from the score), but that there was a clear tendency towards tempo stability and confidence (even with wrong pitches and mistakes) with the novel Versions.

The possible effect of the arrangement of tonalities and lengths could not be statistically tested, although it was observed that these beginning readers did tend to either ignore the key signature or continue with the mental framework acquired in a previous Piece. For the first experiment, after the Pilot study, it was decided to reduce the number of Pieces and to standardise their length (in terms of phrases and confirming cadences). Since participants were expected to be better readers in the First Experiment, there would be no time allocated for visual inspection of the Piece before starting the metronome. Therefore, it was decided to henceforth use a cover sheet that would leave the time and key signature visible before starting to play, in case there would be a transfer effect as the one observed in the Pilot study.

All in all, the exploratory pilot was useful to understand that there would be a certain interest from instrumentalists in the subject, and also in order to understand various logistical constraints, such as the angle of recording or the placement of microphones, the management of participants —particularly the importance of having someone inside the academic institution in charge of room bookings, instruments, and scheduling—, and specially, the preparation of the tests (the importance of a more thorough pre-testing session) for the performances to be statistically analysable.

3.2.3. STATEMENT OF ETHICAL APPROVAL

After completing the exploratory pilot, the project for the experiment proper was presented to the Faculty of Music's Research Ethics Committee, which considered it using agreed institutional procedures and approved it.

The main experimental procedures were described to the participants in advance so that they were informed about what to expect. Participation (although organised by the teachers) was voluntary. Written consent was obtained for participation in the test and for use of the recordings (even if they would not be recognisable in the recordings, which focused on their hands —this was also explained beforehand). Participants were informed that they could withdraw from the research at any time (none of them did). For the questionnaires, participants were given the option of omitting questions they did not want to answer. Participants were debriefed at the end of the study —although the main hypothesis, the expected

better performances with the modified Scores, was never revealed (in order not to affect subsequent participants, who might be informed by their colleagues).

3.2.4. PARTICIPANTS

For the experiment proper, implemented in the *Conservatorium van Amsterdam*, the organization, scheduling, and recruiting of participants was left to the teachers of their Percussion Department (with the only indication that a high number of participants was preferred). This proved an excellent strategy in terms of having committed, punctual and disciplined participants. In fact, one of the logistic findings of the series of experiments that started with this co-operation was that the organisation and scheduling that a conservatoire teacher can ask from/impose on his students is unparalleled when trying to organise a similar study with free-lance participants (even if generously paid). Future research in music reading (or performance, in general) could benefit enormously—at least in terms of implementation or participants' dedication—from the co-operation with Conservatoires and Music Colleges.

Fourteen participants (4 female) from all levels of the Department (from students in the preparatory course to semi-professional Masters students) were organised to participate in two 30-minute sessions each, in two consecutive days (one session each day). The mean age was 21.29 years (*SD* 2.95), mean number of years studying 10.86 (*SD* 2.91), and mean years of professional experience 1.71 (*SD* 1.94): a relatively homogeneous group, except in terms of professional experience (See Table 3.1.)

Table 3.1.

*Data and Descriptive
Statistics for the Students*

Part.	Gen.	Age	Study	Prof.	Metr.
1	M	22	14	5	62
2	M	19	9	2	64
5	F	26	10	3	64
6	M	20	11	0	52
7	M	20	8	1	60
8	M	20	11	1	66
9	F	28	15	6	64
10	M	19	14	0	64
11	M	18	10	0	48
12	M	22	4	1	56
14	F	23	14	2	70
Mean		21.29	10.86	1.71	61.00
SD		2.95	2.91	1.94	6.50

Study = number of years of formal musical education

Prof. = years of professional experience—if any

Metr. = speed (in beats per minute) at which they performed

3.2.5. MATERIALS

A. SCORES

The Pieces to be read were two-voice transcriptions for Vibraphone of Bach chorales from the K. Schubert [1990] collection: nrs 075, 230, and 336. The Soprano part was kept in the original octave, in the right hand stave; Alto and Tenor voices were cleared; the Bass was transcribed an octave higher in the lower stave (using a G-clef, which is not unusual in Vibraphone scores). In a few cases, and when strictly needed, slight amendments were introduced in the Bass line, mainly in order to keep it within the range of the instrument. For examples, see APPENDICES 3.A., 3.B., and 3.C.

The Conventional Version maintained the page orientation (landscape), layout and spacing of the Schubert edition for Breitkopf & Härtel. The distances between notes, the sizes of bars, the placement of barlines, the disposition in systems¹, the distances between these, all follow the original version.

The Modified 1 Version included the following changes:

1. The page orientation changed from landscape to portrait, in order to allow for the layout of musical phrases on a vertical axis, with one phrase per line.
2. Each line of the musical text contains thus only one phrase, and nothing more. An upbeat for the next phrase, at the end of a bar, is therefore included in the next line, leaving the final bar of a line, and the first bar of the next line, incomplete.
3. The distance between notes is used as a digital cue, with a constant measure (throughout the piece) between notes of the same rhythmic value; furthermore, the distances are proportional to the rhythmical values of the notes (*e. g.*, the distances between quavers being exactly the half of the distances between crotchets).
4. Only left-side justification is used for the lines of the text —there is no attempt to fill up each line. (This is in fact an unavoidable consequence of applying design changes 2 and 3.)
5. The beaming of quavers was systematised, prioritising the integration of four graphemes (quavers) rather than two. In spite of being overtly intended for instrumental practice, the beamings in the B. & H. edition appear to follow an underlying textual conception of the material, since the beaming of quavers is not at all systematic, and can only respond to implicit factors. In the Modified Version, runs of four quavers are used whenever possible, and only avoided in the weak/strong alternations of beat (from the 4th beat of one bar to the 1st beat of the next, and from the 2nd to the 3rd beat within one bar).

¹ A 'system' is a line of text in music engraving; it can contain one or more staves —*e. g.*, in a string quartet score, a system (a line of text) would include four staves.

6. The appearance of barlines was modified to dashed thin lines, in order to interfere less with the rhythmic patterns within a segment, particularly as the rhythmic values of the notes are now signified systematically and precisely by the distances (the spacing) between them. Similarly, final barlines (and end of section barlines) were simplified into solid single lines, in order to keep distances and spacings constant.
7. Although the size of the staves and distance between them is the same as in the Normal Version, the distance between systems was increased (from 14.6 to 22.5 points) so that it would be visually distinguishable from the inter-stave one. The distance was incremented even further when marking a major structural division (from 14.6 to 30 points). This was the case in Piece 3 (chorale 336; see APPENDIX 3.C.) between the second and third system: apart, obviously, from the ending, it is only at the closure of the second phrase (thus second system in the Modified Versions) that Bach includes confirmation of the main tonality with a perfect cadence. This is a usual compositional device in many chorales, and habitually it is then from the third phrase on (in short chorales) that the piece starts meandering into adjacent tonal frameworks, only to return firmly to the main key in the final phrase, as is the case in Piece 3 (in fact, here, with a final phrase that almost replicates the 2nd).

The Modified 2 Version included the following additional change:

8. Sub-divisions of each phrase are marked with short white gaps in the staves (the gap also slightly increasing the distance between the notes it separates). No white gaps disturb a run of short notes (quavers or semiquavers), nor do they interfere with the original intentions of the author (his word-setting). The B. & H. edition by K. Schubert, following the original Breitkopf edition by C. Ph. E. Bach, is intended chiefly for amateur performance on a keyboard —‘[Für] den Liebhabern der Orgel und des Claviers’ in C. Ph. E. Bach’s words in his preface [as cited in Bach, 1784–87/1990, p. 4]—, therefore including no texts. But the original compositional intention is vocal, and adapting the design systematically to the text that ultimately underlies the instrumental version was deemed a valid algorithm for deciding these sub-divisions.

Scores were created using Sibelius 7.1.0. on a MacBook Pro running OSX 10.7.5. The Normal Versions were created using the default settings of the program, then modified through the standard dialog windows in order to imitate the B. & H. edition as closely as possible. To create the modified Scores, some unusual paths had to be taken in order to avoid the program’s (many) limitations:

1. Each system was in fact just one long bar. The hidden time signature was 51/16. The shortest figures in these pieces are semiquavers, and the longest phrase (phrase 1 in chorale

075; see APPENDIX 3.A., *Figure 3.A.3.*) is 12 beats long (48 semiquavers) and includes three subdivisions (three gaps of a semiquaver each).

2. Engraving rules and spacing rules were set to force the distances between equal notes to be constant, and to exactly double the value of their first subdivision, eliminating all ‘minimum spaces’ and ‘extra spaces’.
3. Having switched off the uncomfortable ‘magnetic layout’ feature for all items, any elements that were not notes were then graphically superimposed to the staves.
4. Barlines were just vertical lines superimposed to the systems at constant distances from the following note.
5. Equally, the gaps were just white rectangles superimposed to the systems at constant distances from the following note. A hidden semiquaver silence was left unused under each rectangle.
6. At the end of most systems, one or more wide white rectangles were superimposed over the semiquavers (out of the 51 available) that had not been used in that line.

These solutions would obviously prove limited in other circumstances (the documents created in this manner maintain a precarious equilibrium that makes them very difficult to export to other formats). But ultimately, one of the goals of these experiments could be to create either a plug-in for Sibelius 7 or to cooperate in the development of an alternative application.

Scores were printed out in A3 format (usual for percussionists, as the music stand is normally separated from the player by a large instrument). They were printed in high-grammage paper (200 g/m²), in order to prevent the sheets from bending or wearing out with use. The Conventional Version was presented in landscape orientation (following the Edition Breitkopf original), and the two Modified Versions in portrait orientation. All three Versions used staves of exactly the same size (14 mm, measured as the vertical distance between the first and the fifth line). The size of notes and all other musical symbols were also exactly the same in the three Versions. The Music Fonts used were all from the ‘Opus Standard’ family—the default used by the Sibelius programme—and the same for the three Versions.

B. INSTRUMENT

A good quality Yamaha vibraphone (standard size) was provided by the conservatoire—an instrument that all participants were familiar with. The choice of mallets was left to the participants, both in terms of hardness (although medium-hard was explicitly recommended) and number (two or four), since what was of interest was the difference in performance between Versions for a given performer, not the particular achievements or style of that performer. It is further not unusual for percussionists nowadays to play with personal mallets (in the same manner that a string player will use a personal bow).

3.2.6. ENVIRONMENT AND PROCEDURE

On the evening before the experiment all students of the department were gathered at the rehearsal hall for a presentation of the project (the research hypothesis not being disclosed). One student played an example for the rest of the participants, so they would understand the procedure.

On both subsequent days, readings were held in a smaller percussion studio at the Conservatorium van Amsterdam—a room and environment that was highly familiar for all participants. There were no incidents with the scheduling, and no interruptions of the sessions.

Each session lasted 30 minutes. It started with a questionnaire in order to compile participants' data, and in order to have an initial idea of his/her possible proficiency in sight-reading, which was thought to be correlated with years of study and years of professional experience.

Participants then took two attempts at reading one phrase in the same style of the music to be performed in the test (this phrase was taken from chorales 249 and 185 of the Schubert collection, alternating between participants). This was also used as reminder of the procedure of the test, and in order to set an appropriate tempo at which to test each participant. *Figure 3.2.* shows a schematic timeline.



Figure 3.2.

Schematic timeline for EXPERIMENT I: on the evening before the tests, the project is presented to the participants (without revealing the research hypothesis), and an example of procedure is played through; on the next day, the participant takes a Pre-Test with a short fragment of music in order to decide upon performance tempo; this is followed by the recorded test (which will be used for coding); on the following day, the test is repeated (also used for coding), and the participant is questioned on preferences. Order of Pieces and Versions were counterbalanced between subjects.

Approximate durations : day 1, 10 min; day 2, 30 min.; day 3.; 25 min.

3.2.7. RECORDINGS OF PERFORMANCES

Room 808 at the Percussion Department was used for all recordings, during two days —this studio was soundproof, and there were no disturbances. Sessions were recorded with a JVC Card-based digital video camera standing on a tripod pointing downwards, filming the keys of the marimba and the movements of the mallets. Participants were assured that their identities would remain anonymous (only their hands and mallets were inside the recording frame).

The recorded performances were sequenced as shown in Table 3. 2. Since the number of participants was limited, there was no attempt at full counterbalancing, but two possible sequences of the nine readings

that each participant would perform were distributed evenly amongst the subjects (6 with sequence 1 / 5 with sequence 2).

Table 3.2.

Sequences of Pieces and Versions Performed

Sequ. FIRST			SECOND		THIRD	
Part.	Piece	Versions	Piece	Versions	Piece	Versions
1	2	Co - M1 - M2	3	Co - M2 - M1	1	M1 - Co - M2
2	2	Co - M1 - M2	3	Co - M2 - M1	1	M1 - Co - M2
5	1	M1 - M2 - Co	3	M2 - Co - M1	2	M2 - M1 - Co
6	2	Co - M1 - M2	1	Co - M2 - M1	3	M1 - Co - M2
7	2	M1 - M2 - Co	3	M2 - Co - M1	1	M2 - M1 - Co
8	2	M1 - M2 - Co	3	M2 - Co - M1	1	M2 - M1 - Co
9	3	Co - M1 - M2	1	Co - M2 - M1	2	M1 - Co - M2
10	1	Co - M1 - M2	2	Co - M2 - M1	3	M1 - Co - M2
11	2	M1 - M2 - Co	1	M2 - Co - M1	3	M2 - M1 - Co
12	3	Co - M1 - M2	2	Co - M2 - M1	1	M1 - Co - M2
14	1	M1 - M2 - Co	2	M2 - Co - M1	3	M2 - M1 - Co

Co = Conventional

M1 = Modified only with phrasal structuring

M2 = Modified with phrasal and sub-phrasal structuring

A total of 198 clips were recorded (11 participants * 3 readings each * 3 Pieces * 2 sessions). Recordings were transferred from the digital cards to a MacBook Pro laptop, and then converted into ‘.mov’ files (video files), which were used for the marking of reading mistakes.

3.2.8. OUTCOME MEASURES

NUMBERS OF MISTAKES. All of the 198 clips were marked for mistakes in the Pitch and the Rhythm domain, which were then added up, to produce a Total number of mistakes for each reading. Results will be discussed for each of these three sub-sections. In both the Pitch and Rhythm domains, one error was counted for each deviated, wrong, eliminated or added item.

In the Pitch Domain, a Deviation could be a hit on the edge of the key, producing a noticeably bad quality of sound, or a hit in the gap between the intended key and the adjacent one, whereas a Wrong item would be produced by playing an altogether different pitch to the one written.

In the Rhythm Domain, a Deviation would be marked when there was a noticeable tempo alteration but the proportions between note values were not changed, whereas a Wrong marking would be used when an altogether different value was played.

The major, most noticeable errors, produced by eliminating a note altogether or adding notes that were not written, counted thus as two in the Total of errors, since they were disruptive both in the Pitch and in the Rhythm Domain.

EVALUATIONS BY PARTICIPANTS. At the end of the test participants were asked to report which Version they felt was easier to read .

3.3. RESULTS: UNWANTED EFFECT OF SURFACE STRUCTURE INTERFERENCES

3.3.1. NON-NORMAL DISTRIBUTIONS AND OUTLIERS

In spite of our use of personalised tempi for the sight-readings, one of the main problems encountered in this first experiment was the high variability of results between participants: whilst some of them found the exercise too easy, others barely coped with it. In fact, out of the 14 participants, the data from 3 participants had to be discarded since they did not complete the exercise properly or in a way that afforded viable data (they hesitated and/or stopped too much, and did not finish some of the Readings). Amongst the 11 remaining participants there was still a high degree of variability in terms of numbers of mistakes made, as can be seen in *Figure 3.3.a.*, where the error bars represent 95% Confidence Intervals, and even more clearly in *Figure 3.3.b.*, where the error bars represent 1 Multiplier of the Standard Deviation.

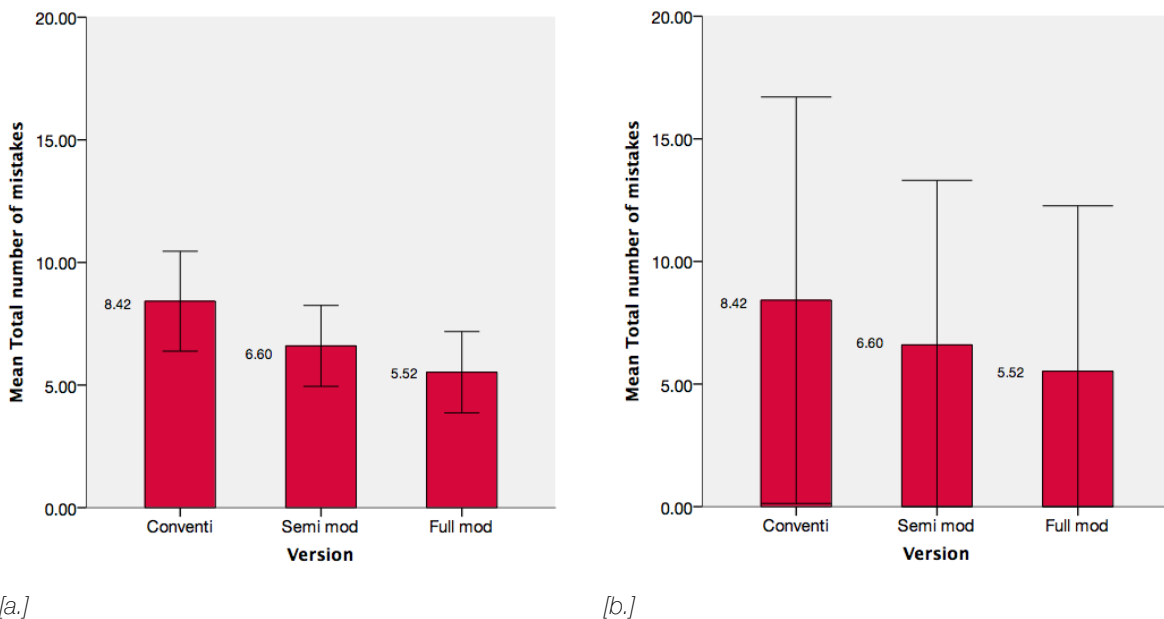


Figure 3.3.

Mean Totals of mistakes in the sight-reading performances as a factor of the Version used: Conventional, Semi-Modified, or Fully-Modified. No statistically significant differences (for all pairwise comparison p s $>.05$). Notice high variability in numbers of mistakes: in (a) Error Bars represent Confidence Intervals (95%); in (b) Error Bars represent Standard Deviation (1 Multiplier).

As can be seen in *Figure 3.4.*, the *third* Reading of each Piece was particularly problematic in terms of distribution of the numbers of mistakes, with several outliers present (with $N = 11$); this was particularly

the case in the second day of readings (Thu 16 Feb). This was due to the fact that by the Third Reading (especially on the second day —the Pieces were the same on both days) most participants had a good grasp of the music and performed almost without mistakes (in fact, there were several ceiling-effect cases, with no mistakes whatsoever), which results in the performances by the few participants that had not yet mastered the Pieces appearing as outliers.

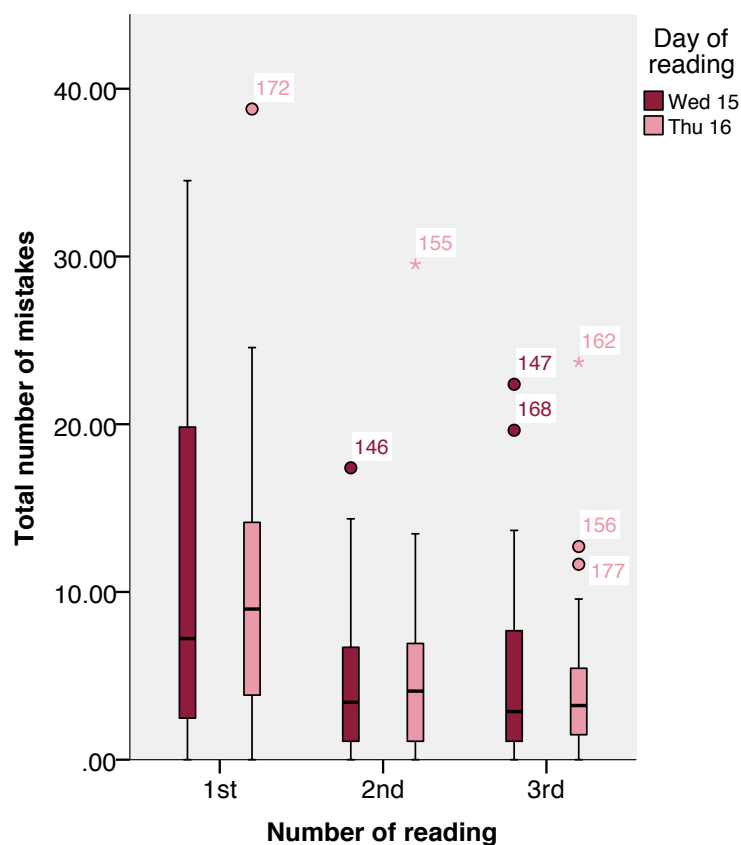


Figure 3.4.

Boxplots of the distribution for each Reading. The number of the Reading (1st, 2nd, or 3rd Reading of a Piece) affects the distribution of the Total number of Errors and the appearance of outliers. Most outliers appear on the 3rd Reading, on the second day.

- = outliers
- * = extreme outliers ($> 3x$ boxed values)

Furthermore, the distribution of datapoints was not normal in relation to any of the factors (number of Reading, day of reading, or Version) that could have had direct bearing on the number of errors, as can be seen in Table 3.3., showing the results of a Shapiro-Wilk test of normality, where the possibilities of normal distribution are highly unlikely in all cases [all p s $< .001$]. Correspondingly, attempts to transform these

variables to conform to normal distribution using \log_e and \log_{10} functions proved unsuccessful [with both \log_e and \log_{10} transformations all p s still $< .01$].

Table 3.3.

*Tests of Normality for the
Datapoints of the Numbers of
Total Errors*

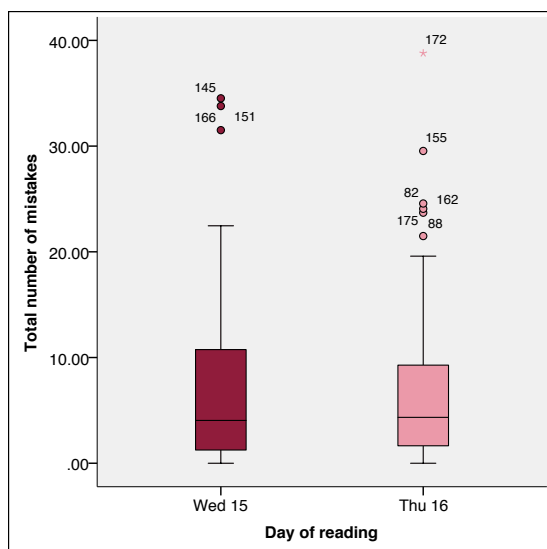
FACTORS	SHAPIRO-WILK		
	Stat.	df.	Sig.
Version			
Conventional	.829	66	.000 ***
Semi modif.	.831	66	.000 ***
Full modif.	.768	66	.000 ***
Number of read.			
1st	.900	66	.000 ***
2nd	.806	66	.000 ***
3rd	.803	66	.000 ***
Day of read.			
Wed 15	.809	99	.000 ***
Thu 16	.809	99	.000 ***

*** Significant at $p < .001$. Normality cannot be assumed.

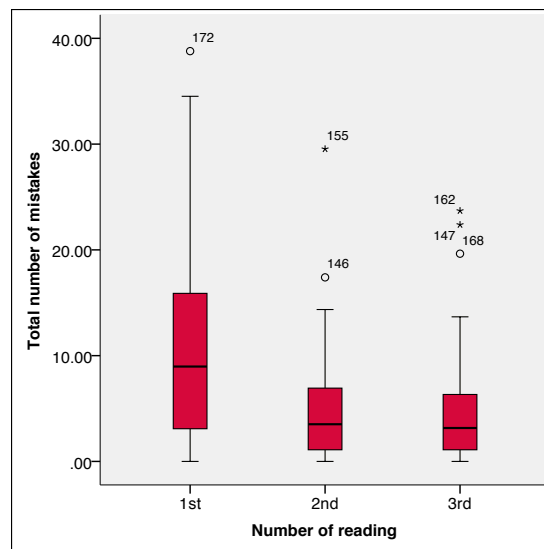
This can also be observed in the related boxplots shown in *Figure. 3.5.a., 3.5.b., and 3.5.c.*, where it is further worth noting that the expected trends observed in the number of Readings and in the day of reading are replicated in the comparison of Versions. In the boxplots of the number of Readings, the quartiles become more compressed as the players become more familiar with the Pieces (observe the changes from the first to the 2nd Reading, and from the second to the third); similarly, the quartiles are more compressed for the second day of reading, as would be expected. As this core of quartiles reflects the general trend of participants towards fewer mistakes and more stable or predictable performances, the number of outliers (performances that deviate from the core trend) increases markedly. What is relevant to our analysis is that the same changes can be observed when comparing the Versions, with changes observed from Conventional to Semi-Modified, and from Semi-Modified to Fully-Modified.

Since these evolutions were observed in relation to the use of the different Versions, it was decided to analyse further the effect of Version on the developments of performance fluency, hypothesising that the Modified Versions would elicit larger improvements from the 1st to the 2nd and from the 2nd to the 3rd

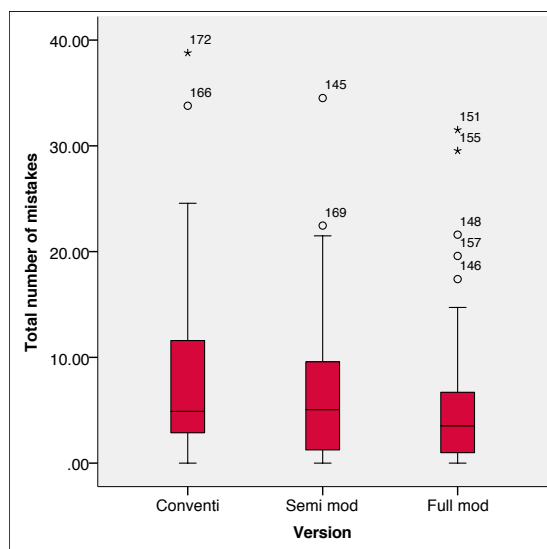
Readings. The effect of Version on the initial (1st) Reading could thus not be analysed due to the large deviation from normal distribution observed (Table 3.3.).



[a]



[b]



[c]

Figure 3.5.

Boxplots of datapoints for the factors affecting the number of mistakes. The same trends can be seen as participants progress [a.] from day 1 to day 2, and [b.] from 1st to 2nd to 3rd Reading, and when comparing this to [c.] the effect of the use of a Conventional, Semi-Modified or Fully-Modified Version.

- = outliers
- * = extreme outliers (> 3x boxed values)

Finally, and before proceeding to the analysis of the evolution of the performances and the effect of Version, I will include some considerations on the persistence of encoding errors in the psychological and pedagogical literature, which led to the decision of measuring *differences* with prior performances rather than absolute values when analysing the patterns of fluency evolution from one Reading to the next.

3.3.2. PERMANENCE OF MISTAKES IRRESPECTIVE OF VERSION

It was observed when coding the performances that there was a ‘carry-on’ or ‘mistake permanence’ effect when players read a Piece for the second or third time. This distortion whereby a particular note, cell or passage that is problematic in a first Reading will tend to elicit problems in subsequent Readings is something that is common in any musician’s experience, but that was not taken into account in the design of this initial experiment: we followed an intuitive or direct paradigm design of having a performer read the *same* piece of music in different Versions, but this meant that the possible effect that a given Version would have on the number of mistakes would be strongly modulated by how correct the *previous* performance had been. This could also explain the abundance of outliers in the 2nd and 3rd Reading and, at least partially, the generally non-normal distribution of the data. In order to take account of this ‘mistake permanence’ effect, it was decided to measure not only the number of mistakes per performance but, crucially, the *decrease* in number of mistakes from the previous Reading.

The persistence of encoding errors has not been studied systematically in the sight-reading literature, but in other psychological processes its effect has been thoroughly examined. Again, the literature in the language domain is vast, and the phenomenon is described and acknowledged in numerous paradigms and language cognition modes. The debate centres more on whether error reoccurrences are due to error *learning* or to the fact that some items tend to pose intrinsic difficulties.

Warriner and Humphreys [2008], for example, have cleverly tested this error persistence in tip-of-the-tongue (TOT) states, which allowed them a certain control of mistake probabilities, and of the timing of the erroneous state (timing of the delays until the correct answer was supplied); these authors argue that a longer delay amounts to greater implicit *learning* of the erroneous state, and their results showed that TOTs were almost twice as likely to reoccur on words that had previously elicited a TOT and had been followed by a long (imposed) delay than those that had been followed by a short delay.

Howe [1970, 1972], on his side, designed studies to measure the effect of repeated presentation on attempts at written reproduction of prose passages, and found that each attempt was very closely related to the previous recall: the contents of a given recall attempt, even when incorrect, were more likely to recur in succeeding recall trials than were nonrecalled items, despite repeated presentations of the correct material; recalls by participants showed thus high stability of initial meaningful retentions. In music, this will be somewhat complicated by the fact that meaningfulness is less well-defined, and can include more subjective components. It has been proposed by Tulving [1966], in fact, that increasing recall or confidence over successive practice trials in learning is a consequence of *subjective* organization and of the development of higher-order memory units (what Tulving terms ‘S’ —Subjective— units). Although Tulving’s experiments were done in the language domain and using free-recall paradigms (*i. e.*, the recall of items could be done in any order), his conclusions might resonate with many musicians’ experiences:

“rehearsal is effective in producing increments in recall only if it permits the subject to organize the material into appropriate S units.” [1966, p. 197]. Thus, according to Tulving’s model, in conditions where repetition of the material does not result in the development of higher-order S units, or where it results in the development of inappropriate units, as happens in the misreadings and mistakes analysed in our experiment, it may even retard the mastery of the material.

More specifically researching the effect of error persistence on repetitive reading, Amlund, Kardash, and Kulhavy [1986], equally found that persistence of initial encoding errors was remarkably stable (both in time and content). After completing one to three readings of a text, all participants were asked to complete free and cued recall measures three times: test and immediate re-test in a first session, and another delayed test a week later. Free recalls were scored for intrusions: each unit that was semantically inconsistent with the original passage as well as each incorrect version of an original unit was scored as an intrusion. Their analysis showed the probability that intrusions would persist on delayed test (after a week) to be higher for subjects in the two- and three-readings groups than for subjects in the one-read group. Furthermore, in the delayed tests, subjects in the two- and three-readings groups had a significantly greater probability of repeating the *same* intrusions they had reproduced on the immediate re-test than subjects in the one-read group. Similarly, on cued recalls, errors were significantly more likely to be repeated on delayed recall, once they had been reproduced on immediate re-test. Thus, errors overall were extremely resistant to long-term correction, and the authors concluded that efforts to improve accuracy should be implemented either prior to (if at all possible, by familiarisation) or during first contact with materials.

3.3.3. MEASUREMENTS OF PERFORMANCE IMPROVEMENT AND NORMALISATION OF DATA

The account of numbers of mistakes, due to the distortions introduced by the persistence of initial encoding errors, was therefore substituted by the measurement of the improvements in performance from one Reading to the next. Due to the numerous outliers in the data derived from various ceiling effects in the third Reading, the focus was set on the improvements in the 2nd Readings. Furthermore, the presence of ceiling effects in the third Reading, apart from creating a distribution with outliers, also meant that it was not possible to measure how marked or relevant a decrease in number of errors from the previous performance had been.

In order to signify the improvements in performance from the first to the 2nd Reading, firstly a mark (M) out of a 100 point scale was assigned to each performance, with the following formula:

$$M = 100 - \left[\frac{\text{Mistakes}}{\text{Events}} * 100 \right]$$

where Mistakes is the number of mistakes in a given mode (*i. e.*, pitch, rhythm, or totals), and Events is the

total number of events (*i. e.*, the total number of onsets that should be elicited by that score). The highest possible mark is thus 100, for a performance without mistakes; the lowest possible mark (theoretically) could be 0, but none of the non-discarded participants performed with so many mistakes: the lowest mark in the pitch domain was 56.70; the lowest in the rhythm domain was 81.25; the lowest accounting for total numbers of mistakes was 70.22. Table 3.4. shows the minimum and maximum marks for the 1st Readings (irrespective of Version; note also the large variance), and the minimum and maximum improvements (sometimes retrogressions; note also the more contained variance) elicited by each Version.

Table 3.4.

*Descriptive Statistics for the
First and Second Readings*

Mode	MARKS				
	FIRST READ. (absolute / any vers.)		Vers.	SECO. READ. (improvement)	
	Mini.	Maxi.			Mini.
PITCH					
	68.55	99.50	Con.	-0.99	24.23
	56.70	98.45	M. 1	1.55	23.71
	61.89	98.90	M. 2	0.55	29.70
RHYTHM					
	88.75	100	Con.	-7.50	3.75
	83.75	100	M. 1	-1.25	12.50
	81.25	100	M. 2	-2.50	15.00
TOTALS					
	78.65	99.76	Con.	-4.01	13.99
	70.22	99.23	M. 1	0.15	18.11
	72.07	99.46	M. 2	-0.01	22.35

Con. = Conventional

M. 1 = Modified only with phrasal structuring

M. 2 = Modified with phrasal and sub-phrasal structuring

Once the values of the increases (in some cases decreases) in markings from the first to the 2nd Readings were established, they could be logarithmically transformed in order to apply a parametric analysis —the data without transformation was still not normally distributed. In order for the logarithmic transformation to be correctly computed, since some cases had shown a negative development from the 1st to the 2nd Reading (and logarithms of negative numbers are not possible), a fixed amount had to be added to all the measurements of differences between 2nd and 1st Reading, so that all data would be of positive sign. Table 3.2. shows the tests of normality results for the original and \log_e transformed variables (with the fixed amounts added).

Table 3.5.

Tests of Normality for Original and Transformed Data of Mark Increases in the Second Reading

		SHAPIRO-WILK							
		ORIGINAL			Min.	Computation	TRANSFORMED		
MARK INCREASES		Stat.	df.	Sig.			Stat.	df.	Sig.
PITCH									
	Incr. Conve.	.834	11	.026 *	-.99	Log _e (incr.+1)	.863	11	.062
	Incr. Mod. 1	.837	11	.029 *	1.55	Log _e (incr.+1)	.939	11	.504
	Incr. Mod. 2	.797	11	.009 **	.55	Log _e (incr.+1)	.963	11	.809
RHYTHM									
	Incr. Conve.	.846	11	.038 *	-7.50	Log _e (incr.+8)	.547	11	.000***
	Incr. Mod. 1	.979	11	.962	-1.25	Log _e (incr.+8)	.972	11	.910
	Incr. Mod. 2	.917	11	.297	-2.50	Log _e (incr.+8)	.961	11	.786
TOTALS									
	Incr. Conve.	.914	11	.273	-4.01	Log _e (incr.+5)	.921	11	.328
	Incr. Mod. 1	.916	11	.284	.15	Log _e (incr.+5)	.969	11	.875
	Incr. Mod. 2	.806	11	.011 *	-.01	Log _e (incr.+5)	.902	11	.196

Conve. = Conventional Version; Mod. 1 = Semi-Modified Version; Mod. 2 = Fully-Modified Version.

* Significant at $p < .050$ - normality cannot be assumed.

** Significant at $p < .010$ - normality cannot be assumed.

*** Significant at $p < .001$ - normality cannot be assumed.

As can be seen in Table 3.5., parametric analysis could thus be performed for the totals and pitch data for marking increases in the 2nd Reading; for the rhythm data, only non-parametric tests would be implemented.

3.3.4. ANALYSIS OF MARK INCREASES IN THE SECOND READING

A. PITCH MARKINGS INCREASES

Results followed the hypothesis and reflected better performances (better increases of marks) in the 2nd Reading with the Modified Versions than with the Conventional Versions. A Paired-Samples t -Test showed that there was a significantly higher increase in pitch marks (*i. e.*, decrease in number of pitch mistakes) when the Semi-Modified Version was used in a 2nd Reading than when the Conventional Version was used [$t(10) = -2.146$, $p = .028$]; further, there was no significant difference in performance between the Semi-Modified Version and the Fully-Modified one [$t(10) = .903$, $p = .388$], although the

difference in performance between the Fully-Modified Version and the Conventional Version did not reach statistical significance [$t(10) = -1.623, p = .068$].

Results were checked with the use of non-parametric tests on the original, non transformed data: although a Friedman Ranking Test did not show a significant main effect of the Version played in the 2nd Reading on the pitch mistakes [$\chi^2(2,11) = 5.091, p = .078$], a Wilcoxon Signed Ranks Test confirmed in pairwise comparison that the increase in mark had been higher for the Semi-Modified Version than for the Conventional Version [$Z = -2.134, p = .016$], that there were no significant divergences between the Semi-Modified and Fully-Modified Versions [$Z = .267, p = .790$], and that the difference in performance between the Fully-Modified Version and the Conventional Version was not significant [$Z = -.889, p = .187$].

B. RHYTHM MARKINGS INCREASES

Due to the lack of normal distribution of the rhythm data, only non-parametric tests were implemented (on the original, non transformed data for the 2nd Reading): a Friedman Ranking Test showed a significant main effect of the Version played in the 2nd Reading on the amount of rhythm mistakes [$\chi^2(2,11) = 7.220, p = .013$]; in pairwise comparisons, a Wilcoxon Signed Ranks Test showed that the increase in mark (*i. e.*, the decrease in number of mistakes) was higher for the Semi-Modified Version than for the Conventional Version [$Z = -2.701, p = .003$], that there were no significant differences in rhythm performance between the Semi-Modified and Fully-Modified Versions [$Z = .153, p = .878$], and that the increase in mark (decrease in mistakes) was higher for the Fully-Modified Version than for the Conventional Version [$Z = -1.785, p = .037$].

C. TOTAL MARKINGS INCREASES

The results for the analysis of Total numbers of mistakes were similar to those obtained for the pitch and rhythm mistakes: a Paired-Samples *t*-Test showed that there was a significantly higher increase in total marks (*i. e.*, decrease in number of total mistakes) when the Semi-Modified Version was used in a 2nd Reading than when the Conventional Version was used [$t(10) = -2.619, p = .013$]; there was no significant difference in performance between the Semi-Modified Version and the Fully-Modified one [$t(10) = .845, p = .418$], although the difference in performance between the Fully-Modified Version and the Conventional Version did not attain statistical significance [$t(10) = -1.527, p = .079$].

Since data transformations had been computed, these results were checked with the use of non-parametric tests on the original, non transformed data: a Friedman Ranking Test confirmed a main effect of the Version played in the 2nd Reading on the total number of mistakes [$\chi^2(2,11) = 6.727, p = .035$]; further, a Wilcoxon Signed Ranks Test confirmed that the increases in marks had been higher for the Semi-Modified Version than for the Conventional Version [$Z = -2.667, p = .004$], that there were no significant

divergences between the Semi-Modified and Fully-Modified Versions [$Z = .178, p = .859$], and that the difference in performance between the Fully-Modified Version and the Conventional Version would not amount to statistical significance [$Z = -1.334, p = .091$].

3.3.5. ANALYSIS OF EFFECTS OF SEQUENCES OF PRESENTATIONS ON THE SECOND READING

Due to the limited number of participants and the subsequent impossibility for full counterbalancing (see Table 3.2.), and due to the non-normal distribution of mistakes (possibly related to the discussed permanence-of-mistakes effect), the data for the two days of performance (Wed 15 Feb and Thu 16 Feb), and for the Readings of the three Pieces (chorales 075, 230 and 336) had to be integrated for each participant into one general more normalised measurement per Version (A, Conventional; B, Semi-Modified; C, Fully-Modified) and mode (Total mistakes, Pitch mistakes, and Rhythm mistakes), thus obtaining 9 manageable datapoints per participant and number of Reading (1st, 2nd and 3rd Reading, although ultimately, as explained above, only Reading 2 was analysable).

For this reason, only the effect of the general Sequence of Presentation per participant (the order of presentation of the Versions for the nine Readings; *e. g.*, ABC-ACB-BAC; see Table 3.2.) rather than the specific sequences for each Piece (chorales 075, 230 and 336) could be analysed as a Between-Subject Factor. Since ultimately the order of the Versions would interact with the ordering of the Pieces, only two sequences were used: (1) ABC-ACB-BAC, and (2) BCA-CAB-CBA, and the Version Sequence Factor had therefore only two possible values. In a similar manner, only the effect of Piece Sequence (and not the effect of each Piece separately) could be analysed as a Between-Subjects Factor —although all the six possible orders of presentation of the Pieces were used, and the Piece Sequence Factor had thus six values.

A. EFFECT OF SEQUENCES ON PITCH MARKINGS INCREASES

The sequence of presentation of the Versions and the sequence of presentation of the Pieces did not have a significant effect on the pitch markings of the 2nd Reading. In a General Linear Model (using the log_e normalised data for the 2nd Reading) with ‘Sequence of Version Presentation’ and ‘Sequence of Piece presentation’ as Between-Subjects Factors, Multivariate Tests showed no main effect of Version Sequence [$F(2,1) = .941, p = .242$], nor of Piece Sequence [$F(10,4) = 1.756, p = .160$]; equally, there was no effect of the interaction Version Sequence x Piece Sequence [$F(4,4) = .924, p = .557$].

B. EFFECT OF SEQUENCES ON RHYTHM MARKINGS INCREASES

A General Linear Model with Multivariate Tests could not be implemented, due to non-normal distribution of the rhythm data (see Table 3.5.).

C. EFFECT OF SEQUENCES ON TOTAL MARKINGS INCREASES

Similarly to what had been found in the Pitch domain, the sequence of presentation of the Versions and the sequence of presentation of the Pieces did not have a significant effect on the Total markings of the 2nd Reading. In a General Linear Model (using the \log_e normalised data for the 2nd Reading) with ‘Sequence of Version Presentation’ and ‘Sequence of Piece presentation’ as Between-Subjects Factors, Multivariate Tests showed no main effect of Version Sequence [$F(2,1) = 167.844, p = .154$], nor of Piece Sequence [$F(10,4) = .520, p = .817$]; equally, there was no effect of the interaction Version Sequence x Piece Sequence [$F(4,4) = 1.012, p = .496$].

3.3.6. PREFERENCES BY PARTICIPANTS

Participants generally preferred the Modified Versions over the Conventional Version. More specifically, 8 out of 11 participants stated a preference for the Modified 1 Version (the partially Modified Version — see APPENDIX 3.A.2.), and 2 out of 11 preferred the Modified 2 Version (the Fully-Modified Version, including gaps between sub-phrase units — see APPENDIX 3.A.3.). However, and saliently, these Version preferences did not always match the Versions with which they made the biggest improvement from the first to the 2nd Reading (see Table 3.6., leftmost columns): 5 out of 11 participants performed best (improved their marks the most) with the Modified 1 Version, whereas 6 out of 11 performed best with the Modified 2 Version. A Pearson analysis of the correlation between the preferred Version and the best Performance confirmed its non-significance [$r = -.194, N = 11, p = .568$].

If the preferences and better performances were to be coded simply as a binary option (Conventional /Modified) without taking into account the differences between the two Modified Versions, the overwhelming majority (with only one exception in Part. 1) would favour the Modified Versions (see Table 3.6., rightmost columns).

Table 3.6.

Participants' Preferences and Best Performances

Part.	THREE OPT.		TWO OPT.	
	Pref.	Best	Pref.	Best
1	Conv	Mod 2	Conv	Modi
2	Mod 1	Mod 1	Modi	Modi
5	Mod 1	Mod 2	Modi	Modi
6	Mod 1	Mod 2	Modi	Modi
7	Mod 2	Mod 1	Modi	Modi
8	Mod 2	Mod 2	Modi	Modi
9	Mod 1	Mod 1	Modi	Modi
10	Mod 1	Mod 1	Modi	Modi
11	Mod 1	Mod 1	Modi	Modi
12	Mod 1	Mod 2	Modi	Modi
14	Mod 1	Mod 2	Modi	Modi

Conv. = Conventional Version

Mod 1 = Modified Version including phrasal divisions

Mod 2 = Modified Version also including sub-phrasal divisions

3.4. DISCUSSION AND CONSIDERATIONS FOR FURTHER EXPERIMENTAL DESIGNS

3.4.1. HIGH DEGREE OF VARIABILITY AND DETECTED TRENDS

The performances by eleven percussion students showed a high degree of variability in a series of sight-reading exercises. This had been predicted to a degree, since I had made the assumption, following personal and academic experience, that sight-reading abilities vary markedly between musicians, even within a relatively homogeneous group (see Table 3.1.; specifically, note the small deviations from the means for age; the main factor differentiating subjects seems to be the amount of professional experience²).

This variability, however, was larger than expected (see *Figure 3.3.*, particularly *Figure 3.3.b.* —note that the standard deviations are larger than the values of the estimated means). The evolution of this variability as the test progressed could be characterized by two main trends: on the one hand, the students that coped well with the exercises and started showing ceiling effects (readings with no or very few mistakes) after being exposed to the materials several times (specially in the third readings in both sessions, and more markedly in the second session). This substantially lowered the means of mistakes as the tests progressed, and made the less proficient students appear as outliers in an analyses using boxplots of the data distribution (see *Figure 3.4.* and *Figure 3.5.*). Since the number of participants was already limited, it was decided to maintain all subjects that had completed the test following the protocol in the dataset. This further meant that any results obtained with general linear models had to be contrasted with non-parametric analysis (*i. e.* not assuming normal distribution in the datasets).

A salient feature was the observation that the same patterns of gradual separation between participants tending towards ceilings of performance (performances without mistakes) and participants struggling with the tests, which were apparent when comparing readings (2nd and 1st, or 3rd and 2nd; see *Figure 3.5.b.*) or when comparing sessions (2nd and 1st day, see *Figure 3.5.a.*), could be signalled independently when comparing performances with different versions (see *Figure 3.5.c.*), showing a tendency towards more compact boxplots and more presence of outliers in the modified versions, with the tendency being particularly marked for the fully-modified version. It appeared thus that performances with these versions had some of the traits (at least the distribution of the dataset) of performances informed by repeated exposure and contextualisation (*i. e.*, ‘rehearsed’ performances).

² This was measured as the number of years since they had started having paid public performances. It is usual for conservatoire students - specially at high-level competitive conservatoires - to engage in professional activities before finishing their degree.

The specific analysis of the advantages for performance presented by the modified versions was complicated by the (tentative) design of this first experiment: it was unfortunately not taken into account that the possible improvements in legibility offered by the modified designs would interact (significantly) with the intrusions of mistakes from the previous readings, and that therefore the absolute numbers of mistakes in the 2nd and 3rd readings would be a less reliable indicator of performance fluency than the relative *increases* in fluency from the previous readings. Since there were furthermore several ceiling effects by the third reading, only analysis of the increases in fluency markings for the 2nd reading were implemented.

Analysis of the mark increases (reflecting increased fluency) in the 2nd reading showed that the Semi-modified versions elicited significantly larger improvements from the 1st readings than the conventional versions, and that there was no statistical difference between the behaviour of the semi-modified and the fully-modified versions. The improvements in reference to the 1st readings when using the fully-modified versions were also larger than the improvements when using the conventional versions, although this difference did not reach statistical significance. Since the measurements were established as a comparison with the 1st reading, it can not be precisely quantified to which extent the lack of statistical significance in this latter case was determined by intrinsic qualities of the design of the fully-modified version or by the quality of the previous performance: in Table 3.4, it can be seen that the maximum increases evoked by the fully-modified versions are slightly larger than the maximum increases drawn out by the semi-modified versions (the minimums are not substantially different) but, at the same time, the performances that they are coming from had a higher range (higher minimums, similar maximums) than the performances that preceded the semi-modified versions.

3.4.2. PERCEPTION BY PARTICIPANTS AND CORRELATION WITH PERFORMANCE

Notably, even though more participants (6 participants out of 11) had their best mark increases with the fully-modified version than with the semi-modified version (5 out of 11; no participants performed best with the conventional versions), only two stated a preference for the fully-modified version above the others. This is somehow an encouraging result for the fully-modified design, though, since it seems to preclude placebo effects. The presence of gaps in the stave might be initially disconcerting or even unappealing, but results show that reading fluency is increased nonetheless.

3.4.3. LIMITATIONS AND FUTURE PARADIGMS

Finally, attention should also be drawn to the lack of effect of the sequences of presentation on the facilitating effect that the modified versions had on the reading process. Neither the sequence of pieces nor the sequence of versions nor their combined order had a significant effect on the results. It has to be

noted, however, that the order effects could not be thoroughly analysed, due to the impossibility of covering full counterbalancing with the number of participants available, due to the use of only two different sequences of versions (1, ABC-ACB-BAC; 2, BCA-CAB-CBA, see Table 3.2.), and due to the analysis only focusing on the 2nd reading of each piece.

All in all, the experiment showed that the number of mistakes in sight reading situations can be modulated by the design of the score, which acted as a green light for further and more detailed explorations of these effects. This first experiment also established the understanding of methodological and logistical issues that would have to be refined in future paradigms. More specifically, an extended pre-test, that would act as a training session for the participants could help in avoiding losing valuable data from participants that do not perform according to the protocol, as was the case here. In general, the aim should be to try to increase (valid) participation numbers and to simplify the design of the test proper, so that complete counterbalancing and effect of order analysis can be implemented, and statistical analyses of the effect of novel designs is not significantly affected by uncontrolled variability.

ACKNOWLEDGEMENTS

I am grateful to the participants, all of whom engaged kindly and selflessly in the experiment. I am equally grateful to *Maestro* Viktor Oskam (who sadly has left us since, due to unsolvable health issues —*Requiescat in Pace*) and *Maestro* Gustavo Gimeno, who organised the seamless and tremendously useful sessions with their students at the *Conservatorium van Amsterdam*. For the Pilot Study, I am very grateful to *Maestro* Edoardo Giachino and his students at the *Accademia Nazionale di Santa Cecilia* in Rome.

*Experiment II:
Confirming Effect of Score Design
on Reading Fluency*

CHAPTER 4

Experiment II: Confirming Effect of Score Design on Reading Fluency

4.1. INTRODUCTION: IMPROVING EXPERIMENTAL DESIGN AND AVOIDING INTERFERENCES

4.1.1. RE-READING OF TEXTS AND WORKING MEMORY

In a study examining the re-processing of expository text, Haenggi and Perfetti [1992] found that working memory (short-term memory concerned with immediate conscious perceptual and linguistic processing) played the major role for comprehending text-implicit information (*i. e.*, information on how the text is constructed, not information on contextualisation or references from the text —this was termed script-implicit), whereas prior knowledge on a subject was relatively more important for the retrieval of explicit information and script-implicit information. Assuming that music reading (and re-reading) shares processes and cognitive strategies with language reading and assuming, in very general terms, that most music readers do not have thorough musicological knowledge of the scores that they are surveying, it could be inferred that the re-reading of a musical script will be similarly modulated by working memory storage of the implicit structure of the piece (more than by their general knowledge of the idiom or repertoire). In fact, it was observed in the previous experiment (see SECTION 3.3.2.) that musicians without extensive knowledge of the vocabulary or stylistic conventions that they were performing, when re-reading the scores, tended towards a permanence of mistakes from one reading to the subsequent (irrespective of, or interacting with, the version of the text presented). Based on Haenggi & Perfetti's findings, it is likely that they were relying on working memory rather than on prior knowledge in the re-readings of the music (particularly in 2nd and 3rd readings).

Bowers [1993], in a longitudinal study of poor and average readers which were followed for two years in their fluency progress, found, after controlling for word-recognition skill (the most significant factor in reading fluency), that both phonological awareness (measured in phoneme deletion tasks) and digit-naming speed were consistently related to speed and errors on initial reading of a text and on its re-reading. Especially after practice (in 2nd, 3rd, and 4th readings of a text) symbol-naming speed contributed significant additional variance to reading speed prediction (on top of the main effect of word-recognition), and the author proposed that the role of processes indexed by symbol-recognition speed should be included in models of fluent and accurate reading. Moreover, Bowers [*id.*] proposed that the role of symbol-processing is more relevant in iterations of reading tasks than phonemic sensitivity. The longitudinal approach in Bowers's [*id.*] study gives a reassuring weight to his model, as well as to his complementary proposal of shared resources for natural and formal language symbol processing [see concurrent findings in Bowers, 1995; Young & Bowers, 1995].

Both the studies by Haenggi and Perfetti [1992] and by Bowers [1993] draw a picture of reading abilities in relation to repeated trials where the short-term engraving of implicit structures can play a dominant role above explicit information or sensitivity to produced utterances.

Bowers's model of shared resources and pathways for language reading fluency and symbol recognitions could be reified in music reading —iterated readings of the symbols in a musical score could be using shared cognitive strategies with re-readings of language. If the *re-reading* of a musical piece, in the same manner as a language text, is modulated by short-term text-implicit information more than by the explicit appearance of the script, we were introducing, in the previous experiment, very unwanted variability in the measurement of the effect of the visual design of a score on the number of mistakes (which were in fact generally carried over from one reading to the next). A much more appropriate experimental design would be to include different texts (ideally of similar difficulty), each with a different design, and observe the effect of graphic design on the first and subsequent readings of each text. In this way, possible misreading or mistakes carried through in short term memory would not be interfering with the possible fluency facilitation afforded by each design, as happened in the previous experimental design, were each piece was presented in three different versions.

Furthermore, Young, Bowers, and MacKinnon [1996] in an examination of Assisted Repeated Reading methods (ARR; pedagogical methods for poor or disabled readers including guided repetitions of oral reading processes, mostly focusing on the prosodic modelling of utterances by an instructor) found that, on top of the benefits provided by prosodic modelling, substantial additional gains in re-reading were produced by conditions that included the practice of intact text. If in the present experiment the levels of significance of a difference in performance (in numbers of mistakes, mainly), between the conventional and the modified version of the score is kept in 2nd readings, that would mean an additional gain in

fluency with the version that performs best (it is hypothesized that it will be the modified version), in consonance with the findings by Yong, Bowers, & MacKinnon [*id.*]. At any rate, LeVasseur, Macaruso, and Shankweiler [2008] have already shown that reading with fluency and cohesive prosody was strongly facilitated by repeated readings of space-cued text (spaces and layout marking the phrase and clause boundaries) providing visible support for the underlying structure.

4.1.2. RATIONALE AND HYPOTHESIS FOR EXPERIMENT II

It was decided to continue working with percussionists in order to build upon the experience gained in the Pilot Study and in Experiment I. It was expected that this should facilitate the running of the experiment and, ultimately, yield results that had an analysable distribution. Another reason to continue with these instrumentalists was our quest to evaluate sight-reading performances that would not tap into long-term memory in the form of repertoire knowledge or recognition of the piece. For ecological validity considerations, we did not want to use music that was composed *ex-professo* for the reading task, but rather use short pre-existing pieces from the common repertoire. In this sense, it was hoped that percussionists would be less prone to recognising the (rather canonical) chorales that we were proposing for the sight-reading than would instrumentalists (pianists, organists) or singers who habitually perform the pieces. In both previous studies, none of the participants identified the exact pieces they were playing, although a few of them recognised the style as that of J. S. Bach's chorales.

For obvious methodological reasons, the ceiling effects encountered in some of the readings in the previous experiment were problematic for statistic analysis, particularly on the second day and last readings. This meant that the chosen metronome marks were probably too low. It was therefore decided for this second experiment to adopt more uncomfortable tempi for participants (*i. e.* higher metronome markings). With the same intention of reducing the risk of ceiling effects, a third reading and a second day of recordings were not included in the experimental procedure.

To further test the hypothesized facilitatory effect of the modified version on legibility and ultimately on performance, three outcome variables were added: 1) an algorithmic spectrogram analysis of tempo stability; 2) blind evaluation of the performances by expert musician, and 3) a slightly expanded assessment of participants' preferences, now including comments on their stated choice.

4.2. METHOD: COMPARISONS OF PERFORMANCES WITH TWO DESIGNS AND TWO PIECES

4.2.1. OVERVIEW

The project for the experiment was considered by the Faculty of Music's Research Ethics Committee and received approval. In this experiment participants were individually tested on sight-reading performance using conventional and modified scores (one Modified Version only). The difficulty of the test was set to match the individual abilities of each participant by using different *tempi* (metronome marks; higher for more proficient readers). Participants filled in a questionnaire and went through a pre-test, which were used to establish these individual *tempi*. Counterbalancing for the different conditions (which Pieces and Versions were read in which order) was done by assigning participants to four different groups, for which fixed sequences of readings were set.

4.2.2. PARTICIPANTS

Sixteen Percussion students (4 female) at the Classical Music Department of the *Conservatorium van Amsterdam*, from all six study years, in the Probationary, Bachelor and Master levels, took part in the study. The data for one of them had to be discarded since he did not complete the task appropriately. Data was collected (see Table 4.1.) on their Age, Years of Music Studies, Degree, Year into Degree, Years of Professional Experience, Years of Piano training, Metronome Tempo at which they did the readings, and Group to which they were ascribed.

Participants were recruited through their teacher, *Maestro* G. Gimeno (they were all students in his class, in different years of the programme); *M^e* Gimeno also organised the schedules and ensured the commitment of the students. They were not paid for participation, but they all manifested interest in the project and clear willingness to co-operate. They were informed of the general purpose of the study (a study of sight-reading abilities in relation to performance materials), but were at all moments kept naive (even after finishing their participation) as to the research hypothesis, that the modified scores would elicit better performances.

Participants were divided into four groups, each reading a different sequence (see Table 4.1.) of Pieces (Piece 1 in G minor; Piece 2 in B minor) and Versions (Conventional or Modified). Each participant did four readings (two for each Piece), in the pre-established order.

Table 4.1.

Sequences of Pieces and Versions Read by Participants in Each Group

Group	1st reading	2nd reading	3rd reading	4th reading
I	Pie. 1 C. Ver.	Pie. 1 C. Ver.	Pie. 2 M. Ver.	Pie. 2 M. Ver.
II	Pie. 1 M. Ver.	Pie. 1 M. Ver.	Pie. 2 C. Ver.	Pie. 2 C. Ver.
III	Pie. 2 C. Ver.	Pie. 2 C. Ver.	Pie. 1 M. Ver.	Pie. 1 M. Ver.
IV	Pie. 2 M. Ver.	Pie. 2 M. Ver.	Pie. 1 C. Ver.	Pie. 1 C. Ver.

Pie. 1 = Piece 1 in g minor

Pie. 2 = Piece 2 in b minor

C. Ver. = Conventional Version

M. Ver. = Modified Version

Groups were organised considering the predicted level of performance of the participants, trying to ensure a homogeneous level of performance across groups. Having the First Part of the Questionnaire (see APPENDIX 4.A.) filled-in in advance, it could be gathered to a certain extent which participants would have a higher competence in sight-reading: from observation in previous experiments, and from trends gathered in the literature on sight-reading [for an early overview of approaches to music-reading focusing mainly on expertise, see Sloboda, 1984; for examples of models assessing experience and cognitive abilities as factors for sight-reading expertise, see J. I. Lee, 2003; Kopiez *et al.*, 2005; Kopiez & Lee, 2006; Kopiez & Lee, 2008; Lehmann & Kopiez, 2009; for evaluations methods of sight-reading expertise, see Zhukov, 2006, 2014] there seemed to be a positive correlation between reading skills and other experience-related variables. For music students in this experiment these were: age, number of years of formal musical education, years into the degree, and years of professional experience (see Table 4.2.).

In the particular case of percussion students having to perform sight-reading with pitch-content in two staves, another variable to be considered when predicting their abilities is whether or not they also play the piano and the extent to which they have been formally trained in it. This information was also requested in the pre-test questionnaire and accounted for when distributing participants: one of the four most competent pianists was placed in each group (Table 4.2.).

All participants reported normal or corrected to normal vision. All participants reported normal hearing, without any impairments. Only one participant (nr 2) reported reading difficulties, more specifically a Non-Verbal Learning Disability, but he had received treatment, and was further not an outlier in any of the measurements (he was the most enthusiastic about the novel scores in the post-test questionnaire, however).

Table 4.2.

Data and Descriptive Statistics for the 15 Percussion Students

Part.	Age	Study	Degr.	Year	Prof.	Piano	Metr.	Group
1	17	7	Proba.	1	0	1	60	I
2	22	5	BMus	2	2	1	70	I
3	19	7	Proba.	1	0	10	53	I
4	20	8	BMus	5	2	4	74	I
5	21	11	BMus	4	0	0	66	II
6	19	10	BMus	2	0	0	40	II
7	28	12	MMus	6	4	12	78	II
8	22	14	BMus.	3	3	0	50	II
9	20	15	BMus	4	0	3	80	III
10	21	11	BMus	4	2	6	88	III
11	23	10	BMus	5	0	5	80	III
12	19	10	BMus	3	2	3	74	III
13	23	6	BMus	4	2	2	52	IV
14	23	9	MMus	6	6	4	74	IV
15	24	15	BMus	5	5	19	92	IV
Min.	17	5	Proba.	1	0	0	40	
Max.	28	15	MMus	6	6	19	92	
Mean	21.40	10.00	-	3.67	1.87	4.67	68.73	
SD	2.67	3.12	-	1.63	1.96	5.31	14.94	

Study = number of years of formal musical education

Degr. = degree that they were studying for at the time

Year = years into that degree

Prof. = years of professional experience —if any

Piano = number of years playing the piano —if any

4.2.3. EXPERIMENTAL MATERIALS

All Pieces that were sight-read both in the pre-test and the test were adaptations of Bach chorales, chosen from the collection of 371 Four-Part Chorales for Keyboard Instrument edited by C. Ph. Emmanuel Bach in 1784–87 (the more complete collection after the success of his initial 1765–69 anthology of his father’s works). The modern re-edition used as reference was by Breitkopf & Härtel, edited by K. Schubert in Wiesbaden in 1990 [Bach, 1784–87/1990]. Incidentally, the 1784–87 edition was published by Johann Gottlob Immanuel Breitkopf, ancestor of the family that still runs Breitkopf & Härtel nowadays. Albeit anecdotal, this is perhaps in line with the ‘artisanal’ approach that is characteristic of (and defended by) most of the major music publishing houses in Europe. This means that their editors are highly competent and fantastically knowledgeable professionals, but the transmission and implementation of publishing conventions is unquestioned and based simply on following a given tradition. In this sense, the head publisher of one of London’s most prestigious music publishing houses was uneasy about the materials of our experiments declaring that the Conventional Versions ‘didn’t look right on the page’, without being

able to offer an analytical or constructive alternative [name withheld; personal communication, Fri 10 Jul 2014].

Schubert, in his preface to the modern Breitkopf & Härtel edition [Bach, 1784–87/1990] does actually acknowledge that there is an adherence to tradition and respect for the original sources (and their design, although this is not explicit) in modern music publishing: “The Breitkopf [1784–87] edition is the major source for Bach’s harmonized chorales, the source of which almost all later collections and editions were based” (p. *ii*). Most editions of the Bach 371 chorales follow in fact the same general design and presentation: A4 size (A4+, 230 x 297 mm, in the specific case of the Breitkopf & Härtel modern edition), in landscape format, with 2–3 chorales per page, and a 1.5–2 cm indentation for the first line of each chorale. There are no major differences in general design (*i. e.* paper size, orientation, layout, and spacing) between the different editions: for example, the version by *Editorial Boileau*, in Barcelona, [year unstated], the most commonly used in Hispanic countries, follows the Breitkopf & Härtel edition in terms of ordering and distribution of the chorales on the pages, simply changing the titles to a (somewhat strident) Spanish translation; the musical fonts and the quality of printing are different, but the final result does not change substantially the spacing, layout or readability of the scores. By comparing our Modified Versions to the modern edition by Breitkopf & Härtel we are thus using an example that is not particularly unreadable or badly laid-out, but rather an industry standard. If anything, the Breitkopf & Härtel is perhaps the commercial edition with clearest fonts and engraving rules, which are furthermore very close to the engraving rules applied by default in the software used for the experimental materials (the AVID ‘Sibelius’ package, also an industry standard, used amongst others by Schott and Faber & Faber in London). If our experiments yield better performances with the modified scores, it can not be fairly claimed that this was due to the fact that the Conventional Versions ‘didn’t look right on the page’.

Chorales chosen as the basis were nr 031, in A minor (Piece 1), and nr 285, also in A minor (Piece 2). Piece 1 was presented transposed to G minor (see APPENDIX 4.A.) and Piece 2 transposed to B minor (see APPENDIX 4.B.). One of the difficulties when sight-reading music is the identification and assimilation of the tonal framework that contains the music, therefore the Pieces had signatures that were not too close in the circle of fifths —and yet had the same number of accidentals so that one Piece could not be considered more difficult than the other in this respect.

Both chorales chosen are in fact ultimately based on the same melody, the anonymous *Wo Gott der Herr nicht bei uns hält*, first compiled by Joseph Klug in his *Geistliche Lieder* collection in 1529. The mastery of Bach with this material allows him to create two chorales that have the same basic structure (5 phrases in C signature all starting with upbeats, with strong cadences at the end of phrase 2 and 5) and directional contour, and yet have a different harmonic and modulatory plan in each Piece, as well as different lines for the individual voices in each of them (all of which matched perfectly our needs).

For the pre-test chorales 006 in F Major (Piece 0) and chorale 224 in F Major (Piece 00) were used as foundation. They were both presented transposed into C Major. This transposition intended to prevent participants predicting the materials of the test proper by being in a Major key, different to the keys of both Piece 1 (G minor) and Piece 2 (B minor) to be read afterwards, and equally distant from them in the circle of fifths (thus not favouring the prediction of, or predisposition towards, any of them). These chorales are also slightly shorter (eight bars in place of ten) than the ones used for Pieces 1 & 2; the intention was to avoid boredom or tiredness effect on the participants.

The adaptations consisted mainly in eliminating the Alto and Tenor parts, and changing octaves in the Bass line for it to fit into a low G-clef line. A system with G-clef lines for both hands was considered more feasible for a slightly heterogeneous (in terms of reading abilities) group of percussionists than a system with the more usual combination of G-clef and F-clef, since some of them could have little or no piano training.

The Pieces were further slightly tweaked, within the idiom, so as to make them contain an equal number of notes (an equal number of onsets, 105 in each Piece; see APPENDICES 4.A. and 4.B.), paying particular attention to the shorter notes, quavers and semi-quavers (generally serving melodic functions as auxiliary notes), of which the numbers were adapted to be similar in both Pieces (more specifically, 48 quavers in Piece 1 and 51 quavers in Piece 2; 4 semi-quavers in Piece 1, and 2 semi-quavers in Piece 2). Ideally, the difficulty of both Pieces should be as similar as possible, so that ‘Piece’ could then be factored out when comparing the mean levels of performance for the two Versions. In statistical analysis it is of course possible to implement a model using Piece as an Offset Variable, but obtaining results without an effect of Piece is more desirable, for two reasons: 1) the quantification of the impact of the Piece is always going to approximate (even using a logarithmic scaling as is usually done in this cases); and 2) the presentation of the results will not have such a clear and directly apprehensible quality as with absolute measures.

The Conventional Version followed the Breitkopf & Härtel edition as closely as possible, allowing for the aforementioned adaptations; the Modified Version included the same changes as described for the Modified 2 Version in the previous experiment (see SECTION 3.2.5.A.; see *Figure 4.1.* for a short example; see APPENDICES 4.A. and 4.B. for full Pieces).

Scores were created with the programme Sibelius (v. 7.1.0.), following the procedure described in Experiment 1 (see SECTION 3.2.5.A.)

(Ach lieben Christen, seid getrost)

[a.] First system of Piece 1 in Conventional Version

[b.] Conversion of the two first phrases to Modified Version

Figure 4.1.

The first system of Piece 1 in Conventional Version and its conversion to the Modified Version. The integration of short notes and the implicit underlying text are used to chunk the material. The caesura signs signify the major divisions of the line into separate phrases to be set in separate systems in the Modified Version; the ticks signify the sub-divisions of the phrases, to be marked with white spaces in the Modified Version. Also note the differences in note spacing, beaming, and barline appearance between the Versions.

Reading materials were printed in A3 paper, in landscape format for the Conventional Versions and in portrait for Modified Versions, with staves of exactly the same size in both cases. It is usual for percussionists to require scores bigger than standard, since the music stand is far away from the player.

A cover sheet was used, which made only the key signature, time signature and first two beats of a Piece visible before the reading test began.

4.2.4. EXPERIMENTAL PROCEDURE

Individual sessions with the participants lasted approximately 30 minutes: 5 minutes questionnaire; 10 minutes pre-test; 10 minutes test; 5 minutes post-test questionnaire.

Each individual session started with the First Part of a questionnaire (see APPENDIX 4.A.) to help establish a general level of reading competence for the participant.

A pre-test was then used with a view to familiarize the participant with the procedure and materials, and to judge his/her sight-reading abilities more in detail, since this would determine an individual tempo at which he/she would perform the test.

The pre-test included two complete readings of two short Pieces, allowing for the trial of four different tempos. The aim was to find a tempo for each participant that would be at the required threshold of his/her abilities. This had to be pondered carefully, particularly since each Piece was to be read two times: from previous experiments, I knew that too slow a tempo would prompt a ceiling effect, specially in the second reading, which is not ideal for analysis (the differences, if any, would not be statistically significant); and too fast a tempo can block the participant completely, not allowing him/her to complete the task, or can lead to performances with such a high number of errors that data shows violations of normality that can not easily be corrected/transformed.

The pre-test also served to introduce participants to the modified scores. The Pieces in the pre-test were played only using these Versions. The purpose of the modified scores was explained with the following protocol:

“In these scores we have introduced a series of graphical modifications departing from conventional layout and spacing rules, and we are wondering if this could enhance the readability of the music. These new engraving rules follow the underlying structure of the piece, marking its sections, phrases and subdivisions. Thus, you will see that there is more spacing between sections, each system includes only one phrase, and the subdivisions of a phrase are marked with white gaps (whilst pointing at the relevant elements of the score)”.

After the pre-test, before starting the recorded test, each participant was instructed with the following protocol:

“You will be presented with a series of musical pieces that you should read at the best of your abilities. In the pre-test I looked for a tempo at which I think you will be out of your comfort zone, so do not worry if you make mistakes —for research purposes these mistakes are useful. But at any rate, do not stop and go back to where you made a mistake. Imagine that this is a live situation, or imagine a recording

session for a movie where every second costs a lot of money and only fragments of your performance will be used. As in a live situation, your aim is also to try to keep the initial tempo throughout.”

Both explanatory texts were repeated in cases of doubt, especially when language barriers were detected.

No scrutiny period was allowed for between the unveiling of the score (by taking away the covering sheet) and the start of the metronome. At the start of the metronome, participants counted three beats and commenced playing, with a late start being considered as a mistake. The metronome was left on for the duration of the first phrase; participants were instructed to keep the same tempo until the end of the Piece, stopping to rest only at the fermatas at the end of each phrase. This procedure had also been rehearsed at the pre-test.

It was observed in the previous experiment that including three readings for each Piece did in fact complicate the analysis, rather than yield more analysable data (see SECTION 3.3.1., on the problems with the third reading, particularly the presence of ceiling effects). The strategy for this experiment was to have participants read each Piece only two times, and to use more demanding (faster) *tempi*, in order to avoid ceiling effects.

After the four readings (two for each Piece), part II of the questionnaire was completed, including a short interview on which Version the participants had preferred and the reasons for that (Figure 4.2. shows an outline of the whole procedure).



Figure 4.2.

Schematic timeline for EXPERIMENT II: after completing a questionnaire on experience and musical education, the participant takes a pre-test in order to get familiarised with the setup and, most importantly, in order for the experimenter to gauge the metronome time at which the test will yield analysable results; after this comes the test proper, where the participant reads two different Pieces (with two readings for each Piece); finally, the participant is questioned on preferences and comments on the novel design.

Approximate duration: 30 min.

4.2.5. DATA COLLECTION ENVIRONMENT

A. RECORDINGS

Participants played on an Adams 5-octave marimba, a good model of concert quality, albeit in a slightly worn state. Some cases of sympathetic rattling occurred when specific notes in the lower register were played loudly, but participants insisted that this was a common phenomenon that any percussionist would be used to, and that it would not affect performance. At any rate, the set-up was the same for all participants and all readings.

Conditions were otherwise uniform for all participants. All participants used two medium-hard mallets (of their own choice). The same studio (Room 808 at the Percussion Department) was used for all recordings—the studio was soundproof, and there were no disturbances. Sessions were recorded in the same manner as in Experiment 1 (see SECTION 3.2.7.).

A total of 60 clips were recorded (15 participants * 4 readings each), which were used for the marking of reading mistakes (See previous experiment, SECTION 3.2.7. for details of data transfer). These files were at a later stage converted to ‘.sv’ (Sonic Visualiser) files, for an analysis of their spectrograms, and also into ‘.m4a’ audio files, that were sent out to external referees.

B. SPECTROGRAM ANALYSIS

All performances were run through Sonic Visualiser (Release 2.3.), a free software application for viewing and exploring audio data for the purposes of music analysis and annotation, developed at the Centre for Digital Music of the Queen Mary University in London. The program can represent the spectra of audio events in visual form, with customisable levels of detail (see an example in *Figure 4.3*). This kind of analysis was a further reason to use percussionists in the study, as the attacks are very clear and distinguishable from the development and decay of the sound, which facilitates the algorithmic identification and quantification of onsets. This would have been similarly the case for any other instruments with a clear attack and short decay, (*e. g.*, piano or guitar; depending very much however on the instrumental writing), but for the reasons explained above (see SECTION 4.1.1. on the rationale) percussionist were deemed particularly useful for this experimental design.

These spectrograms were processed through an automatized algorithm using a ‘Note Onset Detector’ within the ‘Transform’ functions of the program. The settings were adjusted in several trials, in order to create an algorithm that would detect all note onsets, but not any other noises (see Table 4.3.).

After the Note Onset Detector had marked all the notes of a recording, the subdivisions of each beat were eliminated, leaving a graphic with approximately 40 marks per recording: 8 marks per phrase * 5 phrases per Piece (both Pieces had exactly the same structure); sometimes there were deviations from this

number, if a participant eliminated or added a complete beat (see, *e. g.*, *Figure 4.3.*, after beat 3.2.). Each of these markings by the Note Onset Detector Plugin were quantified in milliseconds. This would allow for a parametric comparison of beat durations.

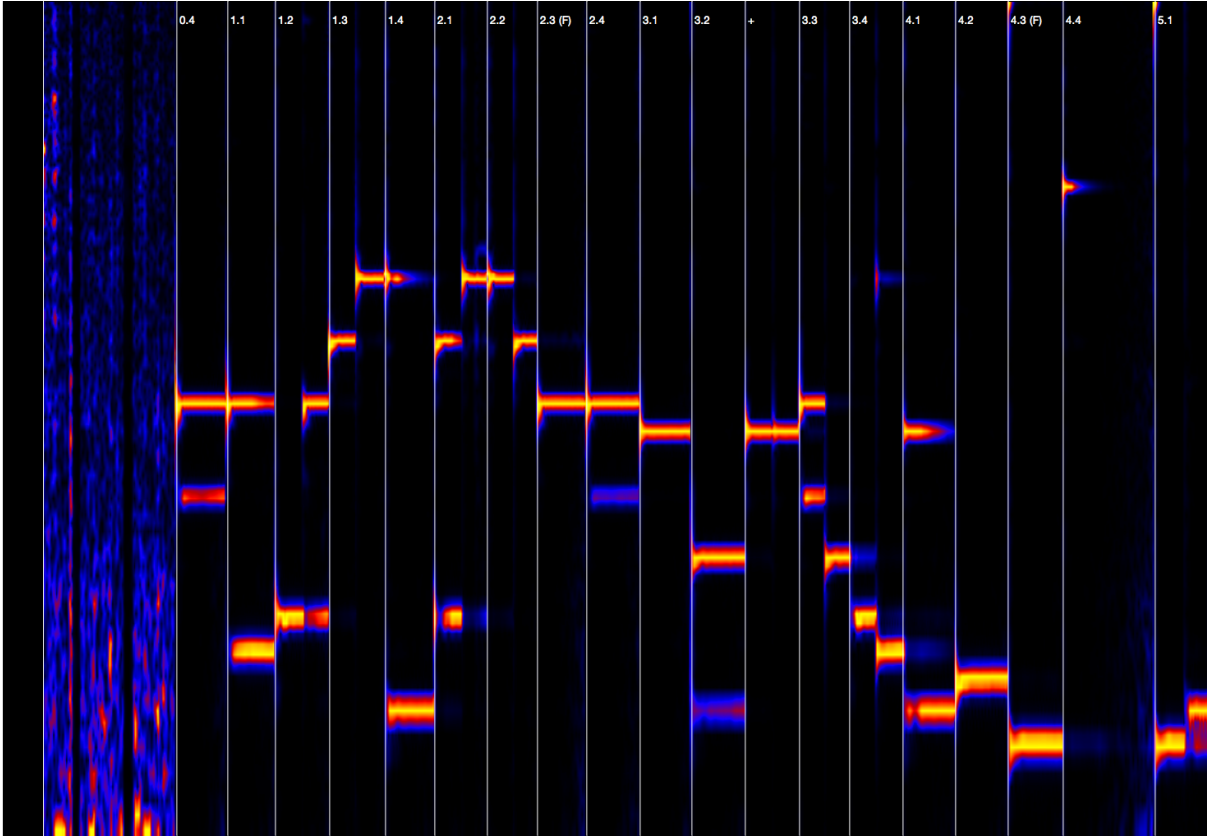


Figure 4.3.

Visualisation of the two first phrases and beginning of third phrase of a performance by participant 1. The vertical lines represent beats and the horizontal lines the notes played. This spectrogram immediately confirms four mistakes that had been detected in the coding of the video file: (1) the player does not pause (as required and instructed) at the end of phrase 1, as can be seen in beat 2.3 (bar 2, 3rd part), which is of the same width as the adjacent pulses; (2) the same happens in beat 4.3, where no fermata (pause) is included; (3) an extra pulse is included after beat 3.2 (marked with +); and (4) the third phrase starts with a hesitation since the attack is very weak in 4.4., and the lapse until the next note (beat 5.1) is much bigger than between the rest of pulses. Beats are quantified in milliseconds (not shown here), making detailed parametrical analysis of tempo stability possible.

Table 4.3.

*Note Onset Detector Settings
for the Sonic Visualiser
Programme*

PLUGIN PARAMETERS	
Program	--
Onset Detection Function Type	Spectral Difference
Onset Detector Sensitivity	10.0000 % [sic]
PROCESSING	
Window size	2048
Window increment	1024
Window shape	Hann

C. EVALUATION BY EXPERTS

The recordings were sent in audio format (no video) to two expert judges: Maestro G. Gimeno, teacher at the *Conservatorium van Amsterdam*, and Maestro L. Ferrandiz, teacher at the *Escuela Superior de Música de Cataluña*, in Barcelona. They were asked to decide blindly on 15 pairs of recordings (one pair per participant) which of the two performances would be marked higher in a sight-reading exercise, for example at an orchestral audition (where applicants play behind a curtain). The judges were not informed of the use of two different Versions of the scores in the pairs of recordings (although they knew the general aim of the research). The pairs included only second readings, and were presented in random order (see Table 4.4. for details).

It was decided to send second readings only since the differences in these were not so obvious, and were the ones that more likely could had been misjudged in the Comparison of Errors markings. Also, it was decided that the second readings (more fluent in general) would be more appropriate for the consideration of these two highly regarded *Maestri*.

Table 4.4.

*List of Pairs of Recordings
Sent to Referees for Judging*

Pair	(Part.)	Piece 1	Piece 2
1	(7)	M. Ver.	C. Ver.
2	(14)	C. Ver.	M. Ver.
3	(4)	C. Ver.	M. Ver.
4	(6)	M. Ver.	C. Ver.
5	(12)	M. Ver.	C. Ver.
6	(2)	C. Ver.	M. Ver.
7	(3)	C. Ver.	M. Ver.
8	(8)	M. Ver.	C. Ver.
9	(5)	M. Ver.	C. Ver.
10	(10)	M. Ver.	C. Ver.
11	(9)	M. Ver.	C. Ver.
12	(11)	M. Ver.	C. Ver.
13	(13)	C. Ver.	M. Ver.
14	(15)	C. Ver.	M. Ver.
15	(1)	C. Ver.	M. Ver.

C. Ver. = Conventional Version

M. Ver. = Modified Version

All recordings of second readings.

D. EVALUATION BY PARTICIPANTS

At the end of the test participants were asked to report which Version they felt was easier to read. No further gradations were asked, thus only allowing for three possibilities: (1) preferred Conventional; (2) thought there was no difference, and (3) preferred Modified. After this, they were asked to briefly comment on the reasons for their preference.

4.3. RESULTS: SIGNIFICANT EFFECT OF DESIGN ON NUMBER OF MISTAKES, FLUENCY, AND ASSESSMENTS

Results will be discussed in terms of coded mistakes (in the pitch and rhythm domain, then added up), of adherence to tempo, of assessment by experts, and of assessment by participants themselves.

CODING OF MISTAKES. The coding was done in the way described for Experiment 1 (see SECTION 3.2.8.). All of the 60 clips (15 participants * 2 Pieces * 2 Readings) were marked for mistakes in the Pitch and the Rhythm domain, which were then added up, to produce a Total number of mistakes for each Reading. Results are discussed for each of these three sub-sections, to show that there was no trade-off in numbers of mistakes between domains, with all three accounts of mistakes (Pitch, Rhythm, and Total numbers) showing the same trends (*i. e.*, fewer mistakes in performances using the Modified Versions). It is also useful to analyse the results in three categories to observe that, even though all three show the same trends, the differences are particularly marked in the Rhythm Domain.

STATISTIC TESTS AND GRAPHS. All tests were implemented with the IBM SPSS package (Version 20). For the purpose of rapid visualisation of trends, clustered bar charts are included for the comparisons of Pitch, Rhythm, and Total Numbers of mistakes, as well as for the comparisons of tempo stability (nPVI values) in the performances. Additional clustered bar charts are presented at the end of sections where there was an effect of specific three-way interactions (including order effects) that seemed remarkable .

4.3.1. PITCH DOMAIN RESULTS

Since the data for the previous experiment showed problematic violations of normality, two different approaches were taken for the analysis of this experiment in order to ensure the avoidance of analysis errors (Type I or Type II): (1) counting on a more thorough pre-test and settings of individualised *tempi*, assume normal distribution and proceed to comparison of means using Paired Samples *t*-Tests and General Linear Models; (2) using a more sophisticated assumption on the distribution of data (a Poisson Probability Distribution), implement a *Generalized* Linear Model. Results for both types of analysis showed clear convergence, with both indicating that the Modified Versions produced significantly fewer mistakes than the Conventional Versions.

A. COMPARISONS OF MEANS

NORMALITY TESTS Possibly due to the extended pre-test, and the more careful selection of individual performance *tempi*, adjusted to elicit an analysable number of errors from each participant without having

him/her overwhelmed by the difficulty of the task, the data showed no violation of normality in this experiment, as indicated by a Shapiro-Wilk test (see Table 4.5.).

Table 4.5.

*Numbers of Pitch Mistakes:
Shapiro-Wilk Tests of
Normality*

Variable	Sta.	df	Sig.
C Ver. 1st Reading	.93	15	.284
M Ver. 1st Reading	.94	15	.425
C Ver. 2nd Reading	.93	15	.308
M Ver. 2nd Reading	.90	15	.132
C Ver. 1+2 Readings	.95	15	.610
M Ver. 1+2 Readings	.93	15	.349

C. Ver = Conventional Versions

M. Ver = Modified Versions

All p s > .05; normality can be assumed.

EFFECT OF PIECE A General Linear Model with Piece (Piece 1, chorale 031 in g minor; Piece 2, chorale 285 in b minor) and Reading (1st Reading, 2nd Reading) as within-subjects factors revealed that the effect of Piece was not significant [$F(1,14) = .64, p = .437$], and a clear main effect of Reading, as expected (fewer mistakes will normally always be made in subsequent Readings than in 1st Readings) [$F(1,14) = 18.15, p = .001$]. The interaction between Reading and Piece was also non-significant [$F(1,14) = .75, p = .326$].

PAIRED SAMPLES A t -Test showed that the number of pitch mistakes was significantly lower for performances using Modified Versions than for performances using Conventional Versions, when comparing the total number of pitch mistakes made in both Readings (1st + 2nd Reading) [$t(14) = 3.07, p = .004$], when comparing only 1st Readings [$t(14) = 2.33, p = .017$], and when comparing only 2nd Readings [$t(14) = 2.87, p = .006$].

ORDER EFFECTS Finally, in order to assess whether the order in which the Versions were presented to participants (reading a Conventional Version first or reading a Modified Version first) had an effect on the number of pitch mistakes, a General Linear Model with Reading (1st, 2nd) and Version (Conventional, Modified) as within-subject factors, and with Order of Presentation (Conventional or Modified first) as a between-subjects factor was implemented, revealing a main effect of Reading and Version only [$F(1,13) = 17.07, p = .001$; $F(1,13) = 9.65, p = .008$, respectively], and that the interactions among Reading and Order of Presentation and among Version and Order of Presentation were both non-significant [$F(1,13) = .09, p$

= .765; $F(1,13) = 2.35$, $p = .149$, respectively]. The three-way interaction among Reading, Version, and Order of Presentation was also revealed to be non-significant [$F(1,13) = .00$, $p = .968$].

B. ANALYSIS WITH A GENERALIZED LINEAR MODEL

The numbers of mistakes coded in the analysis of the performances were in fact distributed as ‘counts’, that is, following a Poisson Loglinear Distribution, which assumes no negative numbers, no fractions, and a limited range for the positive integers. Assuming this distribution implies assumption of non-normality of the data, but allows for a nuanced and well-adjusted analysis of the mistakes as they had been coded.

MAIN EFFECTS Analysis using a Generalized Linear Model with Version (Conventional / Modified) and Reading (1st / 2nd) as within-subject factors and Order of Presentation (Conventional then Modified / Modified then Conventional) as a between-subjects factor, assuming Poisson Probability Distribution and a Log Link Function, showed a highly significant effect of Version [$\chi^2(1, N = 15) = 24.80$, $p < .001$], as well as (predictably) of Reading [$\chi^2(1, N = 15) = 52.04$, $p < .001$]. The order in which the Versions were presented to the participants was shown to be non-significant [$\chi^2(1, N = 15) = .05$, $p = .820$].

INTERACTIONS The effect of Version by Reading was also non-significant [$\chi^2(1, N = 15) = 1.45$, $p = .227$]. Means of number of mistakes in the Pitch Domain were lower for the Modified Versions than for the Conventional Versions in both Readings (see *Figure 4.4*). Pairwise comparisons of Estimated Marginal Means using the Sidak correction showed that the 1st Readings of the Modified Versions elicited significantly fewer pitch mistakes than the 1st Readings of the Conventional Versions [$M = 25.39$, $SE \pm 1.30$, *vs.* $M = 31.20$, $SE \pm 1.45$, respectively; $p = .006$]. Similarly, the 2nd Readings of the Modified Versions elicited significantly fewer pitch mistakes than the 2nd Readings of the Conventional Versions [$M = 16.02$, $SE \pm 1.03$, *vs.* $M = 22.47$, $SE \pm 1.23$, respectively; $p < .001$]. Also worth remarking, the *1st Readings* of the Modified Versions elicited mean numbers of pitch mistakes not significantly dissimilar to the numbers elicited by the *2nd Readings* of the Conventional Versions [$M = 25.39$, $SE \pm 1.30$, *vs.* $M = 22.47$, $SE \pm 1.23$, respectively; $p = .103$] (See *Figure 4.4*).

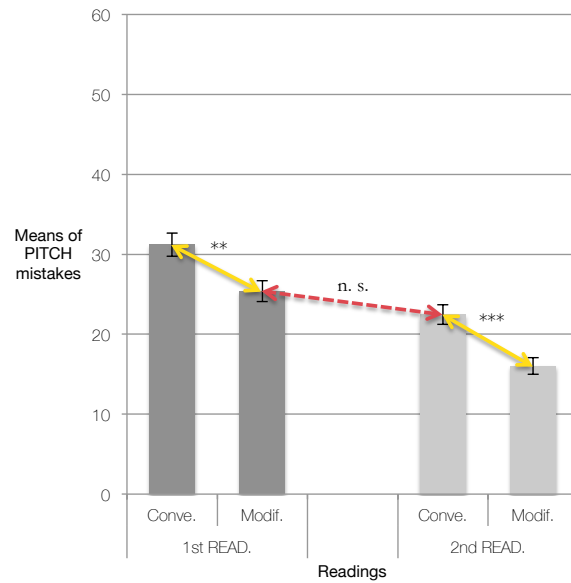


Figure 4.4.

Pairwise Comparisons of Estimated Means of PITCH Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st and 2nd Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

** $p < .010$

*** $p < .001$

ORDER EFFECTS The model showed the interaction between Version and Order of Presentation to be generally significant [$\chi^2(1, N = 15) = 4.17, p = .041$]. However, pairwise comparisons using the Sidak correction showed that the order of presentation was non-significant when specifically analysing Conventional Versions Readings [$M = 28.18, SE \pm 1.33, vs. M = 24.88, SE \pm 1.34$, for tests respectively starting with a Conventional or a Modified Version; $p = .158$], as was the case with Modified Versions Readings [$M = 19.19, SE \pm 1.10, vs. M = 21.19, SE \pm 1.25$, for tests respectively starting with a Conventional or a Modified Version; $p = .230$].

Finally, the model showed the three-way interaction between Version, Reading, and Order of Presentation to be non-significant [$\chi^2(2, N = 15) = .63, p = .727$]. All pairwise comparisons of estimated means were non-significant [all $ps > .100$].

Analysis assuming a Poisson Probability Distribution of the data showed thus the same trends and tendencies as the initial analysis using Paired Samples t -Test, but the picture delineated here is even more clear cut and boldly supportive of the research hypothesis than the comparison of means with normality assumption. In order to ensure that this more precisely adapted but less commonly practised analysis will be safe to implement in further experiments, it was equally tested on the Pitch and Total results.

4.3.2. RHYTHM DOMAIN RESULTS

A. COMPARISONS OF MEANS

NORMALITY TESTS The data for the numbers of mistakes in the rhythm domain showed no violation of normality, as indicated by a Shapiro-Wilk test [for all variables, $p > .05$].

EFFECT OF PIECE A General Linear Model with Piece (Piece 1, in g minor; Piece 2, in b minor) and Reading (1st, 2nd) as within-subjects factors revealed a non significant effect of Piece [$F(1,14) = .22, p = .644$], and a main effect of Reading, [$F(1,14) = 8.61, p = .011$]. The interaction between Piece and Reading was also non-significant [$F(1,14) = .34, p = .532$].

PAIRED SAMPLES A t -Test confirmed that the number of rhythm mistakes was significantly lower for performances using Modified Versions than for performances using Conventional Versions, when comparing the total number of rhythm mistakes made in both Readings (1st + 2nd Reading) [$t(14) = 5.21, p < .001$], when comparing only 1st Readings [$t(14) = 3.85, p = .001$], and when comparing only 2nd Readings [$t(14) = 3.79, p = .001$].

Also worth noticing is the fact that there was no significant difference [$t(14) = .13, p = .780$] between the total amounts of rhythm mistakes for both Readings (sums of 1st and 2nd Readings) using the Modified Versions and the rhythm mistakes made in only the 1st Readings of the Conventional Versions. Perhaps clearer in meaning or implications is the fact that the comparison of rhythm mistakes between *1st Readings* of Modified Versions and *2nd Readings* of the Conventional Version showed no significant difference [$t(14) = .24, p = .679$]. (If at all, in these comparisons, the Conventional Versions elicited slightly higher numbers of mistakes).

B. ANALYSIS WITH A GENERALIZED LINEAR MODEL

In general, and as was the case with the analysis of pitch mistakes, this analysis assuming a Poisson Probability Distribution of the data showed the same trends and tendencies as the analysis using Paired Samples t -Test, but with an even more clear-cut difference between significant and non-significant effects.

MAIN EFFECTS Analysis of numbers of rhythm mistakes using a Generalized Linear Model assuming Poisson Probability Distribution showed a highly significant effect of Version (Conventional / Modified) [$\chi^2(1, N$

= 15) = 64.01, $p < .001$], and a highly significant effect of Reading (1st / 2nd) [$\chi^2(1, N = 15) = 47.40, p < .001$]. The order in which the Versions were presented to the participants (Conventional then Modified / Modified then Conventional) was shown to be non-significant [$\chi^2(1, N = 15) = .81, p = .365$].

INTERACTIONS The effect of Version was not significantly modulated by Reading [$\chi^2(1, N = 15) = 1.45, p = .227$]. Indeed, means of number of mistakes in the Rhythm Domain were similarly lower for the Modified Versions than for the Conventional Versions in both Readings (see *Figure 4.5*). Pairwise comparisons of Estimated Marginal Means using the Sidak correction showed that the 1st Readings of the Modified Versions elicited fewer rhythm mistakes than the 1st Readings of the Conventional Versions, and that the difference was highly significant [$M = 14.63, SE \pm .98, vs. M = 23.61, SE \pm 1.27, respectively; p < .001$]. Similarly, the 2nd Readings of the Modified Versions elicited fewer rhythm mistakes than the 2nd Readings of the Conventional Versions, with a highly significant difference [$M = 8.27, SE \pm .74, vs. M = 15.83, SE \pm 1.02, respectively; p < .001$]. Also worth remarking, the *1st Readings* of the Modified Versions elicited means of rhythm mistakes not significantly dissimilar to the numbers elicited by the *2nd Readings* of the Conventional Versions [$M = 14.63, SE \pm .98, vs. M = 15.83, SE \pm 1.02, respectively; p = .402$] (See *Figure 4.5*).

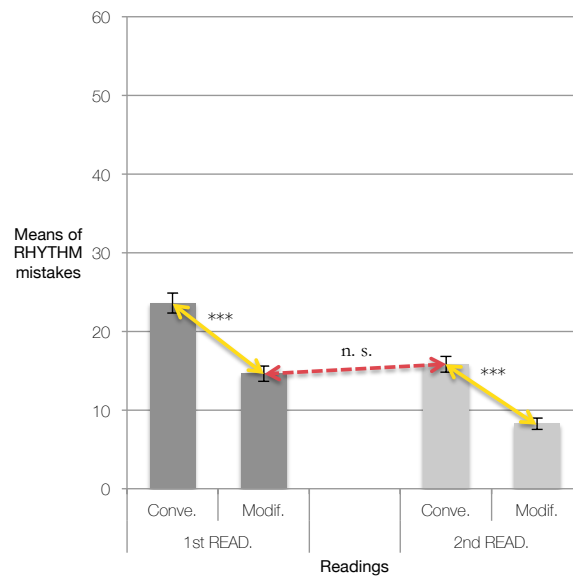


Figure 4.5.

Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st and 2nd Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

*** $p < .001$

ORDER EFFECTS The model showed the interaction between Version and Order of Presentation to be non-significant [$\chi^2(1, N = 15) = .85, p = .356$]. Pairwise comparisons using the Sidak correction confirmed that the order of presentation was non-significant both for the Conventional Versions Readings [$M = 20.62, SE \pm 1.16, vs. M = 18.13, SE \pm 1.14$, for tests respectively starting with a Conventional or a Modified Version; $p = .236$], and for the Modified Versions Readings [$M = 11.00, SE \pm .83, vs. M = 11.01, SE \pm .91$, for tests respectively starting with a Conventional or a Modified Version; $p = .991$].

However, looking more in detail into the three-way interaction between Reading, Version, and Order of Presentation, this combined source of variability showed a significant effect [$\chi^2(2, N = 15) = 7.36, p = .025$]. Pairwise comparisons showed that the Conventional Version was less resilient to context than the Modified and, specifically in the 1st Reading, there was a significant difference between performances immediately at the start of the test and performances thereafter [$M = 27.87, SE \pm 1.86, vs. M = 20.00, SE \pm 1.69$, for the Conventional–Modified order, and for the Modified–Conventional order, respectively; $p = .011$]; all other mean differences were non-significant [all $ps > .100$] (See Figure 4.6.).

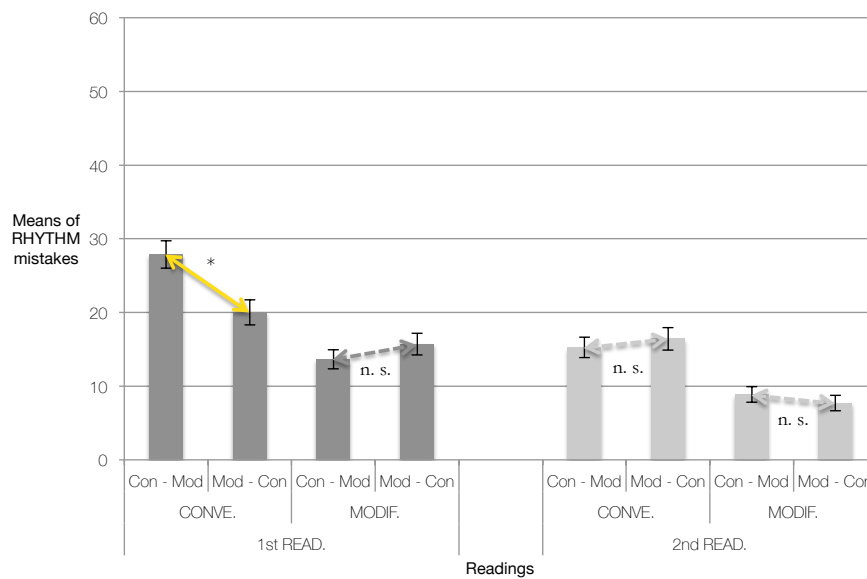


Figure 4.6.

Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.), clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then conventional). From left to right, comparisons of mistakes made in the 1st and 2nd Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

* $p < .050$

4.3.3. TOTAL NUMBERS OF MISTAKES RESULTS

A. COMPARISONS OF MEANS

The data for the totals of mistakes (the sum of pitch and rhythm mistakes) showed no violation of normality, as indicated by a Shapiro-Wilk test [for all variables, $p > .050$].

All Means of total numbers of mistakes were lower for the Modified Versions than for the Conventional Versions, and their Standard Deviations more contained. A Paired Samples t -Test confirmed that total numbers of mistakes were significantly lower for performances using Modified Versions than for performances using Conventional Versions, when comparing the total number of mistakes made in both Readings (1st + 2nd Reading) [$t(14) = 4.30, p < .001$], when comparing only 1st Readings [$t(14) = 3.29, p = .002$], and when comparing only 2nd Readings [$t(14) = 3.66, p = .001$].

Also worth noticing is the fact that the comparison of total number of mistakes between *1st Readings* of Modified Versions and *2nd Readings* of Conventional Versions showed no significant difference [$t(14) = .02, p = .891$].

A General Linear Model with Reading (1st, 2nd) and Piece (Piece 1, in g minor; Piece 2, in b minor) as within-subjects factors revealed a main effect of Reading, [$F(1,14) = 15.94, p = .001$], and no main effect of Piece [$F(1,14) = .02, p = .885$]. The interaction between Reading and Piece was also non-significant [$F(1,14) = .13, p = .764$].

A General Linear Model with Reading (1st, 2nd) and Version (Conventional, Modified) as within-subject factors, and with Order of Presentation (Conventional or Modified first) as a between-subjects factor revealed a main effect of Reading and Version only [$F(1,13) = 14.68, p = .002$; $F(1,13) = 19.03, p = .001$, respectively]; the interactions between Reading and Order of Presentation and between Version and Order of Presentation were both non-significant [$F(1,13) = .02, p = .875$; $F(1,13) = 2.12, p = .168$, respectively]; the three-way interaction among Reading, Version, and Order of Presentation was also non-significant [$F(1,13) = 3.21, p = .096$].

B. ANALYSIS WITH A GENERALIZED LINEAR MODEL

As was the case with the analysis of pitch and rhythm mistakes separately, this analysis of total numbers of mistakes assuming a Poisson Probability Distribution of the data showed the same trends and tendencies as the analysis using Paired Samples t -Test, but with an even more clear-cut difference between significant and non-significant effects.

MAIN EFFECTS Analysis of the total numbers of mistakes assuming Poisson Probability Distribution showed a highly significant effect of Version (Conventional / Modified) [$\chi^2(1, N = 15) = 80.30, p < .001$], and a

highly significant effect of Reading (1st / 2nd) [$\chi^2(1, N = 15) = 99.14, p < .001$]. The order in which the Versions were presented to the participants did not have a significant main effect [$\chi^2(1, N = 15) = .52, p = .469$].

INTERACTIONS The effect of Version by Reading was non-significant [$\chi^2(1, N = 15) = 1.45, p = .227$]. Means of Total number of mistakes were lower for the Modified Versions than for the Conventional Versions in both Readings, without significant difference between Readings (see *Figure 4.7*). Pairwise comparisons of Estimated Marginal Means using the Sidak correction showed that the 1st Readings of the Modified Versions elicited lower Total Numbers of mistakes than the 1st Readings of the Conventional Versions, and that the difference was highly significant [$M = 40.02, SE \pm 1.63, vs. M = 54.89, SE \pm 1.93, respectively; p < .001$]. Similarly, the 2nd Readings of the Modified Versions elicited lower Total Numbers of mistakes than the 2nd Readings of the Conventional Versions, again with a highly significant difference [$M = 24.31, SE \pm 1.27, vs. M = 38.34, SE \pm 1.60, respectively; p < .001$]. Also worth remarking, the *1st Readings* of the Modified Versions elicited means of Total Numbers of mistakes not significantly dissimilar to the numbers elicited by the *2nd Readings* of the Conventional Versions [$M = 40.02, SE \pm 1.63, vs. M = 38.34, SE \pm 1.60, respectively; p = .463$] (See *Figure 4.7*).

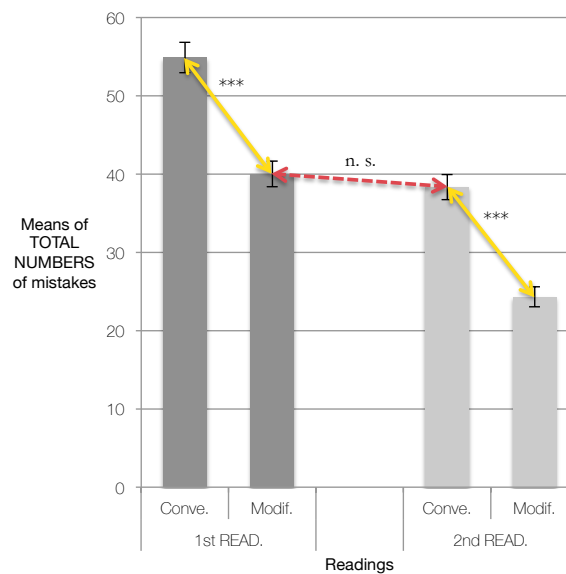


Figure 4.7.

Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st and 2nd Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

*** $p < .001$

ORDER EFFECTS The effect of Version was significantly affected by the interaction with Order of Presentation [$\chi^2(1, N = 15) = 5.05, p = .025$]. Pairwise comparisons using the Sidak correction showed that the order of presentation was in fact significant for the Conventional Versions Readings [$M = 48.90, SE \pm 1.77, vs. M = 43.03, SE \pm 1.76$, for tests respectively starting with a Conventional or a Modified Version; $p = .037$], but not for the Modified Versions Readings [$M = 30.18, SE \pm 1.38, vs. M = 32.23, SE \pm 1.55$, for tests respectively starting with a Conventional or a Modified Version; $p = .325$].

The three-way interaction between Reading, Version, and Order of Presentation did not show a generally significant effect [$\chi^2(2, N = 15) = 4.41, p = .109$]. However, pairwise comparisons showed that the Conventional Version was less resilient to context than the Modified, and in the 1st Reading there was a significant difference between performance immediately at the start of the test and performance thereafter [$M = 61.13, SE \pm 2.76, vs. M = 49.29, SE \pm 2.65$, for the Conventional–Modified order, and for the Modified–Conventional order, respectively; $p = .009$]; all other mean differences were non-significant [all p s $> .100$] (See Fig. 4.8).

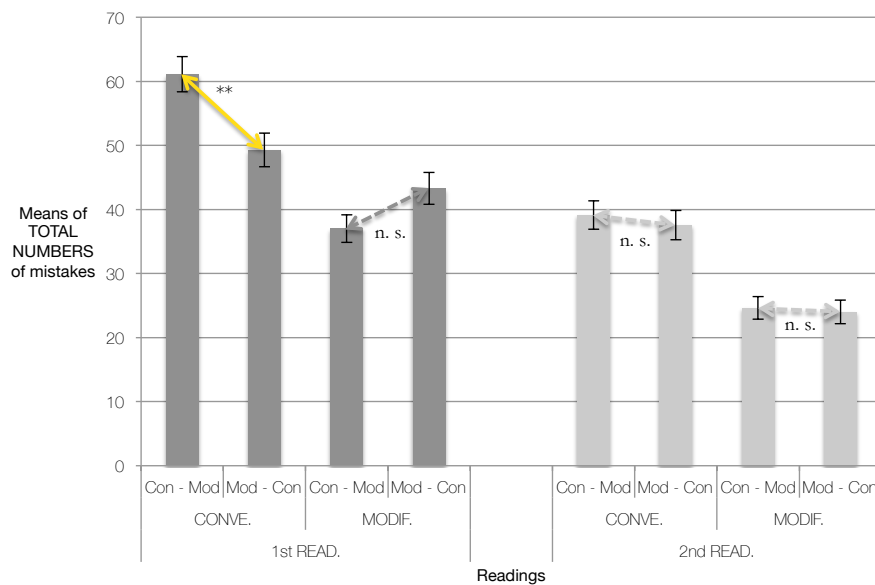


Figure 4.8.

Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.), clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then Conventional). From left to right, comparisons of mistakes made in the 1st and 2nd Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

** $p < .010$

4.3.4. NORMALISED TEMPO VARIATIONS RESULTS

A. SPECTROGRAM ANALYSIS

As mentioned, a Spectrogram Analysis was carried out with the Sonic Visualiser application using a Transformation algorithm with tailored settings (see Table 4.3.) for the detection of onsets in this particular set of recordings (see SECTION 4.2.5.B. for coding details).

Participants were accompanied by a metronome throughout the first phrase; the metronome was switched off at the first *fermata* (pause sign in the score). Participants were requested (see SECTION 4.2.5. on procedure and protocols) to try to complete the rest of the exercise at the same tempo, as is the case in any reading exercise implemented, for instance, in conservatoire examinations or orchestral auditions.

The comparisons of beat durations or tempo stability could not be done simply by using a threshold of permitted deviation from the original tempo though, since a deviation at a certain point of the Piece could mean that all the rest of the beats from that point on would be deemed incorrect. And it was indeed observed that the strategy followed by many participants in the more complicated passages was to slightly lower the tempo, and then keep the new (slower) tempo for the rest of the Piece, or at least until the next *fermata*.

Therefore it was decided to look for *sudden* variations of tempo that were not kept consistently throughout a passage, and that are usually the sign of a hesitation or a noticeable rhythmic mistake by the reader. Using a formula originally developed to measure duration variability in language [see Low, Grabe, & Nolan, 2000, for an initial attempt at quantification; Grabe & Low, 2000, for the formalised expression of variability; see a recent application in Tan & Low, 2014], and soon afterwards employed in studies of comparative rhythmic variability in Music and Language [pioneers were Patel & Daniele, 2003a; see also an examination of possible historical influences on nPVI in Patel & Daniele, 2003b], a Pairwise Variability Index (PVI) was assigned to each *pair* of beats within a phrase:

$$PVI = [100 * (b_k - b_{k+1})] / [(b_k + b_{k+1}) / 2] \quad [\text{Formula 4.1.}]$$

Where b_k is the length in milliseconds of any beat in a phrase except the last and penultimate, and b_{k+1} is the length in milliseconds of the immediately subsequent beat.

In Formula 4.1. PVI values outside the -18 – +18 range corresponded with rhythmic mistakes or hesitations that were audible in the recordings; the indexes were in fact used to double-check or confirm some of the markings of Deviations (the less clearer cases) in the compiled lists of Rhythm Domain mistakes.

Finally, a normalised Index (nPVI) could be assigned to each performance, by calculating the Average of the PVI's in that performance; nPVI values were then used to do a comparative study of the rhythmic stability of the Readings using the modified scores and those using conventional scores.

B. NPVI COMPARISONS

The data showed a slight violation of normality, as indicated in a Shapiro-Wilk test (see Table 4.6.a). This was easily corrected with a logarithmic transformation (base e) of the data (see Table 4.6.b).

Table 4.6.

[a.]				[b.]			
<i>nPVI Values: 1st Tests of Normality</i>				<i>nPVI Values: 2nd Tests of Normality After log Transformation</i>			
Variable	Sta.	df	Sig.	Variable	Sta.	df	Sig.
C Ver. 1st Reading	.83	15	.011*	C Ver. 1st Reading (Log)	.91	15	.187
M Ver. 1st Reading	.91	15	.159	M Ver. 1st Reading (Log)	.93	15	.304
C Ver. 2nd Reading	.91	15	.187	C Ver. 2nd Reading (Log)	.93	15	.286
M Ver. 2nd Reading	.93	15	.320	M Ver. 2nd Reading (Log)	.98	15	.990
C Ver. 1+2 Readings	.95	15	.554	C Ver. 1+2 Readings (Log)	.97	15	.868
M Ver. 1+2 Readings	.95	15	.570	M Ver. 1+2 Readings (Log)	.97	15	.963

C. Ver = Conventional Versions
M. Ver = Modified Versions

* $p < .05$ Normality can not be assumed

C. Ver = Conventional Versions
M. Ver = Modified Versions

All $ps > .05$ Normality can be assumed

A Paired Samples t -Test (of the normalised data) showed that the nPVI values were significantly lower (lower meaning less tempo variability, more stability) for performances using Modified Versions than for performances using Conventional Versions; this was the case when comparing the average nPVI values for both Readings (mean of 1st and 2nd Readings) [$t(14) = 3.54, p = .001$], when comparing only 1st Readings [$t(14) = 3.44, p = .002$], and when comparing only 2nd Readings [$t(14) = 2.23, p = .021$].

Further, a General Linear Model with Reading (1st, 2nd) and Piece (Piece 1, in g minor; Piece 2, in b minor) as within-subjects factors revealed a main effect of Reading [$F(1,14) = 8.22, p = .012$], and no main effect of Piece [$F(1,14) = 2.01, p = .178$]. The interaction between Reading and Piece was also non-significant [$F(1,14) = .86, p = .215$].

Finally, a General Linear Model with Reading (1st, 2nd) and Version (Conventional, Modified) as within-subject factors, and with Order of Presentation (Conventional or Modified first) as a between-subjects factor revealed a main effect of Reading and Version only [$F(1,13) = 14.02, p = .002$; $F(1,13) = 12.92, p = .003$, respectively]; the interactions among Reading and Order of Presentation and among Version and Order of Presentation were both non-significant [$F(1,13) = 4.10, p = .064$; $F(1,13) = 2.38, p = .146$,

respectively]; the three-way interaction among Reading, Version, and Order of Presentation, however, was significant [$F(1,13) = 11.70, p = .005$], as was the case with the (somewhat related) rhythm mistakes data (pairwise comparisons offered a very similar picture to what was found for the rhythm data; see SECTION 4.3.2., Figure 4.6.).

4.3.5. BLIND ASSESSMENT BY EXPERT MUSICIANS

Two expert musicians were asked to use five possible markings (see Table 4.7.) for fifteen pairs of audio recordings (see details in SECTION 4.2.6.; Table 4.4.), assessing which of the two performances in each pair would be considered 'better' in an orchestral audition or examination situation. They were not informed that these pairs included performances using two different Versions (Conventional / Modified) of the scores.

Equivalent measurements (*i. e.*, in five grades) were created by categorising the coded Total Numbers of Mistakes of the performances into five levels. The equivalence between expert preferences and gradings derived from numbers of mistakes is shown in Table 4.7.

Table 4.7.

Assessment by Experts: Relations of Ranking Levels with Numbers of Mistakes

Markings by experts				
[1] C Better	[2] C Slightly	[3] Equal	[4] M Slightly	[5] M Better
Difference in Total numbers of mistakes in 2nd Readings (C – M)				
[1] < -5	[2] -5 – -3	[3] -2 – +2	[4] +3 – +5	[5] > +5

C = Conventional
M = Modified

In the grading following the numbers of mistakes, a positive number signifies a better performance with the Modified Version, and a negative number a better performance with the Conventional Version; a value around zero (-2 -- +2) signifies equal performances.

Table 4.8. shows the markings by both experts alongside the markings derived from the coded Total Numbers of Mistakes. As can be seen in the table, experts preferred the performances with the Modified Versions above the performances with the Conventional scores, for most participants (9 out of 15 in the case of *M^o Ferrandiz*, and 12 out of 15 in the case of *M^o Gimeno*). This ratings were furthermore in agreement with the results obtained with coding of mistakes (see statistical details below). The only case of clear discrepancy is Participant 10, where the experts felt that the performance of the Piece with Conventional design was clearly better than the performance of the Piece using modified design, whereas

the marking derived from the number of mistakes would signify a slightly better performance of the Piece with modified design.

Table 4.8.

*Assessment by Experts:
Comparison of Expert
Markings and Coded
Mistakes*

Part.	(Pair)	Mark <i>M</i> ^o . F.	Mark <i>M</i> ^o . G.	Mist.
1	(15)	3	4	4
2	(6)	5	5	5
3	(7)	5	5	5
4	(3)	2	3	2
5	(9)	5	5	5
6	(4)	5	5	5
7	(1)	5	5	5
8	(8)	5	5	5
9	(5)	4	5	5
10	(10)	1	1	4
11	(12)	5	4	3
12	(5)	3	5	5
13	(13)	5	5	5
14	(2)	3	5	5
15	(14)	3	3	3

Pair = Random order in which the recordings were sent to the experts

M. F. = *Maestro* Ferrandiz, teacher in Barcelona

M. G. = *Maestro* Gimeno, teacher in Amsterdam

Mist. = Markings derived from the coded numbers of mistakes

Experts marked blindly, based on audio recordings, without identification of participant or Version used.

A Spearman's Rank Order correlation was run to determine the relationship between the three markings. There was a strong, positive correlation between the markings by *M*^o. Ferrandiz and the coded mistakes, which was statistically significant [$r(13) = .56, p = .015$], and between the markings by the two experts, *M*^o. Ferrandiz and *M*^o. Gimeno, which was similarly significant [$r(13) = .66, p = .003$]; there was a very strong, positive correlation between the markings by *M*^o. Gimeno and the coded mistakes, which was statistically highly significant [$r(13) = .95, p < .001$].

4.3.6. SELF-ASSESSMENT QUESTIONNAIRES

At the end of the test participants were asked to report which Version they felt was easier to read, and to comment briefly on their preference (see Table 4.9). As was the case with the assessment by experts, there

was a general preference for the Modified Versions (10 out of 15 participants preferred the Modified Version; 4 were indecisive; only 1 overtly preferred the Conventional).

In a manner similar to the procedure used with the markings by experts, the preferences by participants were coded into three levels, and compared to markings derived from the Total numbers of mistakes coded for the performances, categorised into the same levels (see Table 4.10.).

Table 4.11 shows the markings by participants alongside the markings derived from the coded Total Numbers of Mistakes (in this case the totals for the sum of both Readings , 1st and 2nd). There are no cases of clear divergence.

A Spearman's Rank Order analysis showed a strong, positive correlation between the evaluations by the participants and the Totals of Mistakes coded, which was statistically significant [$r(13) = .63, p = .005$].

Table 4.9.

Evaluation by Participants: Preferred Versions and Comments

Par.	Ver.	Comments (translated from Dutch)
1	C/M	M better for a first reading, easier to find information; C better for the second reading, better for phrasing
2	M	White gaps help to look ahead for next idea; in cases of mistake it is easier to find the way back
3	M?	Easier to chunk the information; doubted if it would work with more complex materials
4	C	Was more used to it; structure and phrasing clearer in M, though
5	M	Structure is clearer; felt more at ease; felt uncomfortable presented with C
6	C/M	M gives the feeling of having more time to look ahead; C gives better overview
7	M	C was too dense and confusing; felt more at ease and could breathe better with M
8	C/M	Did not feel there was a difference
9	M	C felt confusing; it is handy that the information is pre-chunked in M
10	M	It is clearer; felt more at ease; easier to see the phrases; easier to pick up playing after a mistake.
11	M	Easier to pick up playing after getting lost; feels more comfortable
12	M	Liked the spacing between systems, and the white gaps; if a mistake is made it is easier to pick up playing.
13	M	Felt more at ease; the structure was clearer
14	M	Liked the chunking of the material, the distance between the systems and the white gaps; C felt crowded.
15	C/M	Felt at ease and relaxed with M; reading was bit faster with C

C = Conventional better; M = Modified better C/M = Equal, saw advantages for both.

Table 4.10.

*Evaluation by Participants:
Ordinal Ranking and
Relation with Coded Mistakes*

Evaluations		
[1] C Better	[2] Equal	[3] M Better
Difference in Total mistakes (C – M)		
[1] < -5	[2] -5 – +5	[3] > +5

C = Conventional
M = Modified

In the difference of total numbers of mistakes, a positive number signifies a better performance with the Modified Version, and a negative number a better performance with the Conventional Version.

Table 4.11.

*Evaluation by Participants:
Comparison of Markings*

Part.	Eval.	Mist.
1	2	3
2	3	3
3	3	3
4	1	1
5	3	3
6	2	3
7	3	3
8	2	3
9	3	3
10	3	3
11	3	3
12	3	3
13	3	3
14	3	3
15	3	2

Eval. = Markings derived from participants' preferences
Mist. = Markings derived from the coded numbers of mistakes

4.4. DISCUSSION: VALIDITY AND UNIVERSALITY OF THE MAIN MODIFICATIONS

4.4.1. ALL ASSESSMENTS INDICATING BETTER PERFORMANCES WITH THE MODIFICATIONS

Fifteen Percussionists were presented with two pieces of very similar level of difficulty, with each of the pieces following a different set of layout, spacing and design rules (one conventional, and one proposed by us). The scores using our proposed modified design elicited significantly fewer mistakes in sight-reading performances, so much so that a second reading with the conventional score was equivalent to a first reading with the modified score. The differences between the performances were also significant in terms of stability of tempo (avoidance of hesitations and brusque changes of speed), with the modified scores producing performances with smaller values for indexes measuring the variability between pairs of events (pairs of note onsets). Finally, both blind assessments by experts and assessments by the musicians themselves gave preference to the performances with the modified scores. These assessments showed furthermore a very strong positive correlation with the measurements in numbers of mistakes and tempo stability.

4.4.2. THE IMPORTANCE OF SEPARATION OF DISCOURSE UNITS

Regarding the numbers of mistakes, a recurring comment in the participants' assessments was that the separations of discourse units signified with white space in the novel scores seemed effective in order to halt the cumulative effect that making a mistake within a strictly timed (stressful) framework can sometimes trigger. The comments by the musicians (see Table 4.9., particularly participants 2, 10, 11, and 12) suggest that the white gaps acted as a 'cognitive firewall' that would both prevent a knock-on effect of mistakes and help re-gather the line of discourse after a deviation. This could have been the product of slower tempi or of a strategy of slightly lowering the tempo of performance at the gaps, in order to gather the information ahead. But the contrary was the case, as the significant differences in nPVI values indicated; performances with the modified scores were also more fluent (fewer hesitations and stops) and adherent to tempo (see SECTION 4.3.4.).

These results seem to replicate findings in the language domain indicating that interword spacing serves as a very effective segmentation cue for eye-movement guidance and therefore reading fluency (fewer mistakes, and more stability). Relevantly, this seems to be the case even in languages that are not normally spaced. Sainio *et al.* [2007] have shown how interword spacing facilitates both word identification and gaze focus for Japanese native readers using Hiragana scripts (syllabic, non ideographic, although not normally spaced). Building upon these results, and taking into account that universality has been a core debate in the fields of linguistics and psycholinguistics for many years [see, *e. g.*, Chomsky, 1966, 2006,

2012], Liversedge *et al.* [2016] have recently argued for a universality of representation and process in *reading*, after examining behaviour in languages that differed fundamentally in visual presentation and orthographic ruling (more specifically Chinese, English, and Finnish). If certain visual and recognition processes in language reading are universal, and universally supported—independently of custom—by structured spacing, it could perhaps be the case that these advantages of visual cueing could be extrapolated to the reading of formal languages as well—an example of which could be music sight-reading.

4.4.3. THE UNIVERSALITY OF THE EFFECTIVENESS OF SEPARATIONS

It could be argued that the benefits of adding spacing to scripts is only relevant for readers that are accustomed to the conventions of a given language and that it might be less facilitating for learners of a second language or for situations where the inclusion of novelties in design might act as distractors for readers for whom comprehension is not as automatized, or perhaps supported by multi-modal usages (as is the case of a native speaker, thoroughly supported by mere exposure). But in a study with second-language readers using Pinyin (romanized Chinese) texts, Bassetti [2009] found that interword spacing of texts facilitated sentence-picture verification tasks in native English speakers. Similarly, Bassetti and Lu [2016], using a multiple-choice gap-filling task, found that English readers of Chinese as a second language read faster with interword spacing than without.

It has to be noted, though, that in the studies by Bassetti [2009] and by Bassetti and Lu [2016], tasks using Hanzi ideograms in the texts were not facilitated by adding interword spacing (nor were they hindered). The latter was the also the case with Kanji (ideographic) texts for Japanese readers in the study by Sainio *et al.* [2007] (see above). These authors argued that spacing information was redundant in Kanji scripts, since the visually salient characters would serve as effective segmentation cues themselves.

Given the results in the present experiment, it seems that music reading includes processes that are closer to the strategies used by readers of systems that represent phonemes—and are word-spaced—, as romanized Pinyin Chinese, than to those used for systems that represent integrated morphemes, as the Chinese Hanzi character scripts do. In fact, one of the more defining characteristics of music in opposition to natural languages would be its structural and thorough reliance on *discrete* sounds as a basilar part of its construction of discourse. A system relying on the use of conventionally defined discrete events that are then generatively combined will benefit in its scripts from visual cues similar to those used in phoneme-based language scripts. These cues seem to facilitate integration into meaningful units of the infinite combinatorial possibilities, and help the reader to recognize the recurrent formations of a given idiom, or more simply the recurrent motifs in a given fragment. On the contrary, in ideographic systems the recognition of sub-morpheme units is not of crucial importance at a common usage level, and word

hierarchy is less relevant than morphemic information, as was shown by Bassetti and Masterson [2012] in a study with native Chinese readers: removing morphemic information had strong negative effects on experienced and inexperienced readers, and subsequent adding of interword spacing had no facilitative effects.

The universality of representation and process in reading proposed by Liversedge *et al.* [2016] could thus be limited to notational systems representing items product of generative processes that have as basis a restricted set of conventionally defined and discrete elements. Western musical notation as we nowadays use it in common practice would mostly fit into this category of systems. However, musical notation can take, and has taken, many other forms: from early (at least mid-IVth millennium *A. C.*) Sumerian, Babylonian and Assyrian pictographic symbols, to the ideographic characters used as solmization syllables in China (probably dating back to the IInd century *A. C.*) —designating not absolute but movable points in a scale; or from the stylized contour shapes of medieval western neumes to the open suggestive notations used by the European avant-garde from the late 1950s onwards [for an overview see Bent *et al.*, 2001, particularly section II; for an example of successful of avant-garde open notation, see Toop, 2001, section 2.].

The results obtained here with common-practice present-day notation could therefore be more specifically related to numerous studies with western (phoneme-based) language scripts demonstrating improvements in reading fluency as a result of the formatting of the text by adding spacings that reflect the structuring of the discourse. Jandreau, Muncer, and Bever [1986] showed that readability could be improved by reflecting in the space size between words the structure boundaries of English phrases, and that this spacing strategy was effective at different levels of detail of structuring boundaries. These somewhat ambiguous (albeit telling) findings were addressed in a follow-up study, where Bever *et al.* [1991] tried to elucidate which level of parsing was the most efficient for the spacing of printed text, and found that assigning extra spaces to word groups corresponding to major phrases resulted in the best comprehension, above using an even spaced algorithm, using space sizes derived from prosodic analysis, or using spacing proportional to the syntactic depth of the phrase structure at a certain juncture. It has to be noted, though, that these are parsing strategies for adding *extra* space between words —interword spacing is in all cases assumed as standard. In our modification of the design of musical scores we therefore implemented extra spacing at the major phrase level, and spacing at sub-phrase level (figuratively equivalent to words), which is not standard in music scripts.

ACKNOWLEDGEMENTS

I am grateful to the participants, all of whom engaged kindly and selflessly in the experiment. I am equally grateful to *Maestro* Viktor Oskam (*Requiescat in Pace*) and *Maestro* Gustavo Gimeno for their invaluable help in the organisation of the tests and recording sessions at the *Conservatorium van Amsterdam*.

Sonic Visualiser is a free software developed at the Queen Mary University of London for viewing and exploring audio data for music analysis and annotation.

*Experiment III: Narrowing
Possible Effects and
Increasing Ecological Validity*

CHAPTER 5

Experiment III: Narrowing Possible Effects and Increasing Ecological Validity

5.1. INTRODUCTION: THE IMPORTANCE OF PAGE ORIENTATION AND OTHER ECOLOGICAL VALIDITY ISSUES

5.1.1. CRITIQUES ON THE PREVIOUS EXPERIMENT

Following up on the results obtained in the previous experiment with readings of two pieces of similar difficulty but different graphic design (see SECTION 4.4.), where the modified scores drew out better performances in terms of mistakes, stability, and assessment (both by experts and participants), the subsequent study investigated: whether those results could have been product of confounding factors, and not of the changes in visual structuring implemented through spacing and layout cues; whether the novelties could be substituted by conventional signs, and whether the results applied only to specific circumstances. More specifically, five critiques to the previous study were addressed: (1) the bias of participants to evaluate more positively the modified scores than the conventional ones could have been product of novelty or simple politeness; (2) a possible conflation of the effect of novel design with the change of orientation in the page; (3) a possible trade-off of stability and precision against expressivity or interpretative freedom with the proposed design modifications; (4) the parsing of the discourse implemented through the inclusion of separation cues (white gaps) could be easily (and possibly more efficiently) attained with conventional music articulation symbols (specifically with slurs); and (5) even if sight-reading was truly facilitated, it does not follow automatically that musicians would feel comfortable with the novel design for a rehearsed performance. In response to these, adapted materials, coding systems, and experimental procedures were implemented (see details below —SECTION 5.2.).

5.1.2. VERTICAL VERSUS HORIZONTAL ARRANGEMENTS OF TEXTS

The most prominent and insistent critique was that previous results showing a highly significant difference in numbers of mistakes (to the degree of making a first trial with the novel scores equivalent to a second reading with the conventional scores) could have been product of a simple change in the orientation of the page in the novel scores, which would have made them look better proportionated or aesthetically appealing, thereby facilitating concentration or motivation in the musicians. This difference in aesthetic appeal could have been marked further by an excessive lower margin in the horizontal layout of the conventional version. Even if this were the case, if more aesthetically pleasing scores could help in the stressful, highly demanding, and competitive domain of professional music making in present-day circumstances, this could perhaps be a valid approach. It is, nonetheless, also a very valid point that attention gathered merely through novelty can trigger a cycle of innovative dealings which could eventually become difficult to reign in or even become counter-productive in the long run. Text familiarity has been shown to modulate eye movements in reading, particularly as it triggers higher-level processes which modulate eye-head movement coordination [C. Lee, 1999; Seo & Lee, 2002]. It has also been shown to significantly affect the *format* of recall and reproduction in memory exercises (although not the level or amount of recall) [Hartley, 1993]. Alterations of texts purely based on design novelty would therefore not benefit from the advantages in eye movement coordination and in predictability of reproduction that familiar setups afford. Again, texts to be used in highly demanding conditions of accuracy and efficiency, as musical scores are, should be treated with caution in terms of modifications and alterations of customary design patterns.

Nonetheless, the alleged advantage in legibility for the modified scores simply by being in vertical arrangement is in itself also questionable, since the literature (on language reading) comparing horizontally and vertically arranged texts points towards *horizontally* arranged texts as facilitating faster reading [Seo & Lee, 2002, quantified the advantage as a 24% faster], given equal conditions to a vertical arrangement. Seo and Lee [*id.*] have indicated that this is primarily due to larger gaze amplitude for horizontal reading, and thus to smaller numbers of saccades or fixations. Further, a higher velocity of gaze saccades for given amplitudes in horizontal than in vertical reading might also contribute to the difference in reading speed. This higher saccade velocity of horizontal saccades is at least partly due, Seo and Lee [*id.*] posit, to the physiology of the oculomotor system (simply better adapted for horizontal than vertical rapid rotation in humans); the authors observed higher velocity for given amplitudes of horizontal saccades in a control experiment devoid of lexical load, in which a sequentially stepping laser target was tracked.

5.1.3. STRUCTURED SPACING AND SEPARATION IN RELATION TO DIFFERENT TRADITIONS

At the same time, there is an abundant corpus of literature (for language reading) showing that the more important factors for legibility, above general page orientation or general layout, are the separation and

spacing of the script following the underlying structure of the discourse, with separation at interword level being particularly crucial. This literature, relevantly for the comparison with music scripts, has recently addressed the implementation of interword separation in languages that historically do not include this visual cue (*e. g.*, Chinese or Japanese Hiragana), on top of the more classic studies showing severe disruptions in reading when the cue is eliminated in languages that are usually spaced (*e. g.*, English).

A. HISTORICALLY UNSPACED LANGUAGES

Hsu and Huang [2000] found that including word spacing in Chinese texts (ideographic texts, not normally spaced) enhanced reading performance; they also found that ambiguity of spacing at sentence structure level severely degraded reading accuracy. Inhoff and Wu [2005] similarly found that ambiguity due to the absence of distinct interword spaces led readers to fixate longer and spend more time viewing critical character sequences (suggesting legal words with various possible interpretations); moreover, in light of the results, these authors argued against the strictly serial spacing of characters in Chinese (ideographic) texts. Concurrently, Shieh, Hsu, and Liu [2005], found that readers of Chinese as a second language (Taiwanese undergraduate students) using a text format with interword spacing achieved an increased reading efficiency, and ventured to use a faster reading speed, than when using other presentation modes; further, the interaction between presentation mode and text difficulty was non significant in their experiments, with the modified (spaced) presentation increasing fluency equally for high- and low-difficulty materials. As mentioned above (see SECTION 4.4.) Sainio *et al.* [2007] found parallel results, more specifically in terms of facilitated word identification and gaze guidance in Japanese Hiragana (syllabic) scripts.

B. READERS ACCUSTOMED TO SPACING

In a study of Thai-English bilinguals and English monolinguals, Winskel, Radach, and Luksaneeyanawin [2009] found that interword spaces had a selective effect on reading in Thai, as it facilitated word recognition, but did not affect eye guidance or segmentation tasks; in English texts, removal of spaces severely disrupted reading, with the effect, notably, being more accentuated in the bilinguals (very few musicians could claim to be ‘monolingual’ users of historical idioms in today’s musical practices). Moreover, Winskel, Radach and Luksaneeyanawin [*id.*] observed that in the English texts the initial gaze landing positions were significantly nearer the beginning of the embedded target words of the exercise when reading unspaced rather than spaced text. This implies narrower gaze amplitudes, which are, as Seo and Lee [2002] argued (see above), a major reason for slower reading. In experiments with English readers, Slattery and Rayner [2013] confirmed that interword spacing has a large influence on word segmentation and is important for saccade target selection, but also found an interplay with intraword spacing (spacing between letters within words), which influenced a text’s readability. Investigating further this interaction of interletter spacing with interword spacing, Slattery, Yates, and Angele [2016] have also

confirmed, based on measurement of reading rates and target word gaze durations, that the required space between words depends on the amount of space between letters, and more specifically on the default spacing for a given font.

C. HISTORICAL PRACTICES IN LANGUAGE SPACING

These latter experimental findings are syntonic with historical practices dating at least as far back as the introduction of the movable type; since then printed books provide a minimum of 1.5 units of space (the unit measured as the internal gap of the lower-case ‘n’) and usually 2 units of space, as interword separation, which, according to Saenger [1997], enables the identification of word boundaries by the parafovea, whilst smaller interword spacing would help the reader to define word boundaries only within foveal vision (the portion of the visual field where images of letters are perceived in full detail) [Saenger also cites the seminal work of McConkie & Rayner, 1975, and Rayner & McConkie, 1976, in this respect]. In any case, what Saenger [*id.*] considers fundamental in this interaction is the unambiguous distinction between interword and interletter space in order for the reading of separated writings to benefit from the economy of mental effort that modern reading habits represent (in comparison to reading of ancient *scriptura continua*) [see also Saenger, 1997, Ch. 2. for even older examples of fusions and intraword ligatures used by late medieval scribes to facilitate distinction of word shapes].

5.1.4. IMPLEMENTING SEPARATION IN CONJUNCTION WITH SUPPORTING CUES

In spite of the intention of giving priority to space between discourse units at the motif level (*i. e.*, equivalent to word level) in the design of the materials for this experiment, another factor also implemented in previous experiments (the so-called ‘proportional notation’) was also maintained on the basis of the historical practices discussed above.

In the design of the modified scores, the use of regular and systematic units of space at inter-note level could therefore enhance the separating function of the white gaps that visually structure the musical discourse at the sub-phrase level. It was hypothesized that, in a manner analogous to the spacing interaction in language scripts [see McConkie & Rayner, 1975; Rayner & McConkie, 1976], the standardised and sufficient spacing of graphemes will facilitate the acquisition of integrated pattern length information (an information proven to primarily influence efficient saccade lengths, and subsequently reading speed). It was also hypothesized that the predictability of the proportional notation (with each note value being assigned a fixed number of spacing units) would help the unambiguous distinction between the spacing of the Sub-Phrase Units (we could cautiously call them ‘Motifs’, using the term in a lax way) and the spacing of the notes.

5.2. METHOD: COMPARISONS OF PERFORMANCES WITH TWO DESIGNS WITH SAME ORIENTATION

5.2.1. OVERVIEW

The project for the experiment was considered by the Faculty of Music's Research Ethics Committee and received approval. The method followed in general lines the model of Experiment II (see SECTION 4.2.). Some relevant changes were introduced, though, all in order to address the critiques mentioned in the Introduction (see SECTION 5.1.). Namely, these were: (1) the evaluation of the different scores by non-performers based on visual inspection, and the comparison with the preferences by performing participants; (2) changing the orientation of the conventional scores, so that both Versions are presented in portrait orientation; (3) measurement of dynamic ranges, as an assessment of expressivity and confidence; (4) the inclusion of phrasing slurs in the Conventional Versions following the same parsing of the discourse suggested in the Modified Versions with the white gaps; and (5) the addition of a rehearsed performance, including study time, following the usual sight-reading tasks.

5.2.2. PARTICIPANTS

PERFORMING PARTICIPANTS. Thirteen Percussion students (4 female) at the Classical Music Department of the *Conservatorium van Amsterdam* took part. Students were of varied academic level, from the Probationary Course, the Bachelors Degree, and the Masters Degree. Three participants did not complete the test as required (one opted out, overwhelmed, and two others did not seem to understand completely the instruction, which led them to stop during some of the readings and start sections anew); their data was therefore not used in the analysis.

BROWSING PARTICIPANTS. Concurrently to gathering performing participants' preferences, another group of 49 students at the conservatoire were asked to visually inspect the materials (without performing them) and state which Version they would prefer in three hypothetical scenarios: (1) having to sight-read the scores; (2) having to retrieve specific information from them (*e. g.* articulation or dynamic markings); and (3) having to memorise the Pieces.

5.2.3. EXPERIMENTAL MATERIALS

The chorales used for the tests were nrs. 005 and 033 in the Richter Collection (See APPENDICES 5.A. and 5.B.). The materials in this experiment were all presented in portrait orientation, included the novelty of

dynamic markings, and were produced with different levels of difficulty. The Conventional Versions also included phrasing slurs.

CHANGES IN ORIENTATION. With the aim of eliminating the possibly confounding factor of page orientation both the Conventional and the Modified Version were presented in portrait setting (in A3 sheets), with the staves printed starting from the upper margin (exactly 4cm in both cases; see examples in APPENDICES 5.A. and 5.B.). The purported advantage for the modified scores simply by being in vertical arrangement or 'looking better on the page' should thereby be cancelled out.

ASSESSMENT OF DYNAMIC RANGES. In order to assess a possible trade-off of stability and precision against expressivity or flexibility of movements, abundant dynamic markings were included in the text. The dynamics were implemented following an analysis of the grammar (harmonic progressions, cadences and modulations) of each Piece, with the usual emphasis on deviations from the tonic and (suggested) modulations. This would further serve as a test of additional information retrieval, on top of pitch and rhythm encoding, as already required in previous experiments. For this same purpose, articulation marks (*staccati*) were also included.

DIFFICULTY TIERS. Also, and in order to better adjust the difficulty of the tests to the abilities of each participant (in trying to avoid the problematic ceiling effects), and since three performances were to be coded (with the third one furthermore being a rehearsed one), a system of difficulty tiers was adopted. The scores were produced in three different levels of difficulty for each of the Pieces. This was done by including one, two, or three voices in the presented materials; this would add variation in the levels of difficulty, on top of the adjustments of tempo to ability already used in previous experiments.

CONVENTIONAL SLURRING. The Conventional Versions included phrasing slurs that indicated exactly the same chunking or parsing of the materials as suggested with the separation of basic discourse units by white space in the Modified Versions.

5.2.4. PROCEDURE

At the end of the sight-reading exercises, participants were given four minutes of private study time for each Piece (individually, in a separate room). *Figure 5.1.* shows a schematic timeline for the full experimental procedure.

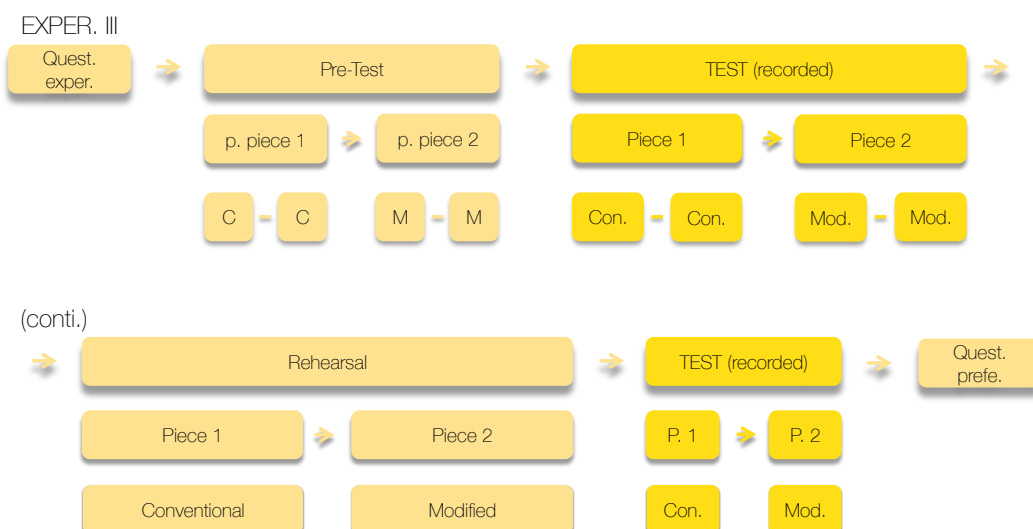


Figure 5.1.

Schematic timeline for EXPERIMENT III: the participant is questioned on experience related predictors of reading ability, and then performs a Pre-Test, in order to get acquainted with the procedure and the novel designs through performance; this is followed by the Test proper (recorded to be coded); at the end of this the participant is assigned private rehearsal time to study the Pieces, and then performs both Pieces a final time; at the end, the participant is questioned on preference (Conventional or Modified Version).

Approximate duration: 45 min.

5.2.5. DATA COLLECTION AND CODING

NUMBERS OF MISTAKES. In a manner similar to what had been done in Experiment II (see SECTION 4.3.), 66 clips (10 valid participants * 2 Pieces * 3 Readings, in this case) were coded for mistakes in the Pitch Domain and for mistakes in the Rhythm Domain, which were subsequently added up to produce a Total Number of Mistakes (see example of coding in APPENDIX 4.C.).

TEMPO STABILITY. Tempo variability indexes (nPVI) were, as in previous experiment (see SECTION 4.3.4.) coded algorithmically by an Onset Detector application in the Sonic Visualiser programme (Queen Mary University of London; release 2.2)

DYNAMIC RANGES. Loudness values were similarly assigned by the Sonic Visualiser Programme using a scale adapted for the dataset of the experiment.

PREFERENCES. Both the performing participants and the participants that only browsed through the scores coded their preferences from 1 (clearly the Conventional) to 5 (clearly the Modified) for three scenarios: sight-reading; retrieval of information and memorization.

5.3. RESULTS: INCREASED EFFECT OF DESIGN ON NUMBER OF ERRORS AND FLUENCY

As it was shown in the previous experiment (see SECTIONS 4.3.1., 4.3.2., 4.3.3., and 4.3.4.), the analysis using a Generalized Linear Model assuming a Poisson Distribution with a Loglinear Link Function offered a picture that was very similar to the one yielded by comparisons using *t*-Tests, but with sharper, better defined contrasts. It was therefore decided to implement only the Generalized Model for the data in the present experiment, particularly since the experimental design included a third Reading, which was expected to produce rather subtle differences in performance (this third Reading was furthermore a rehearsed one). Equally important is the fact that, even if less of an obvious or usually referenced type of model, the chosen analysis adjusts more correctly to the nature of the data provided by the coding of mistakes, which is not a scaled response, but rather a response in the form of a count of cases (nPVI values, even if scaled, did not show a Gaussian distribution, and equally required an extended — generalized— model for analysis)

Thus, analyses implementing Generalized Linear Models, assuming Log (*i. e.*, non-linear) Link Function, were run for the Pitch, Rhythm and Total Numbers of mistakes, as well as for the nPVI values. In the case of the numbers of mistakes, the probability assumed was Poisson, as the dependent variable consisted of counts expressed in positive integers. In the case of nPVI values, the probability assumed was Gamma, as the variable was scaled, but could not be considered normally distributed. A first Model in each section included Piece (Piece 1, in C min / Piece 2, in F# min —see APPENDICES 5A & 5B) and Reading (1st, 2nd or 3rd Reading) as Within-Subject effects, and Order of Pieces (Piece 1 followed by Piece 2 / Piece 2 followed by Piece 1) as Between Subject effect. Since in all four cases (Pitch, Rhythm, Total Numbers of Mistakes, and nPVI values) Piece had no significant effect (except for a marginal effect on the pitch mistakes; see details below, SECTION 5.3.1.), a second model was then implemented, not including the small variability due to Piece differences.

The data-sets could be considered longitudinal in this case as they comprised repeated observations (in two Readings and a rehearsed performance) of an outcome (different number of mistakes or stability values). The objective of analysis was to describe the marginal expectation (*i. e.*, the sums of counts of events, or the sums of means of indexes) in the outcome variable as a function of the chosen factors (with focus on Version) *whilst* accounting for the correlation among the repeated observations for a given subject. Because repeated observations are made on each subject, correlation is anticipated among a subject's measurements. It must be accounted for to obtain a correct statistical analysis. This was particularly the case since marked variability was observed in reading proficiency, the number of

participants was somewhat limited, and in this experimental design an interval between performances (a study time) was included. Zeger and Liang [1986] have proposed an approach to such analyses using Generalized Estimating Equations, which should offer consistent and reliable variance estimates [see similarly Liang & Zeger, 1986; see Zeger, Liang, & Albert, 1988, for a discussion of subject-specific models in which heterogeneity in regression parameters is explicitly modelled—as was the case here; see further examples in Zeger & Liang, 1991; Liang, Zeger, & Qaqish, 1992; Zeger & Liang, 1992; Liang & Zeger, 1995]

This second Model included therefore Participant as a Subject Effect and assumed an Independent working correlation matrix structure. For analysis, it included Version (Conventional / Modified) and Reading (1st, 2nd or 3rd Reading) as Within-Subject effects, and Order of Versions (starting test with a Conventional or Modified Version) as Between Subject effect. Tests of Model effects determined the main effect of each source (Version, Reading, and Order of Versions), whereas Pairwise Comparisons of Estimated Marginal Means using the Sidak Correction showed more specific differences between the Versions when taking into account the interactions with Reading and Order of Presentation.

5.3.1. PITCH DOMAIN RESULTS

The initial Model analysing the effect of Piece showed a marginal effect for this categorical variable [$\chi^2(1, N = 10) = 3.19, p = .074$]. Nonetheless, it was decided not to include the variability derived from Piece in a further model, since in the cases of Rhythm and (more importantly) Total Numbers of mistakes the effect was non-significant; furthermore, due to the limited quantity of participants ($N = 10$, having lost three) an excessive amount of predictors would lead to reaching the maximum number of step-halvings—the validity of the model fit would thus be uncertain.

MAIN EFFECTS. The finally implemented Model showed Version as a significant source of variation [$\chi^2(1, N = 10) = 6.88, p = .009$], and Reading as a highly significant source [$\chi^2(2, N = 10) = 85.82, p < .001$]; Order of Versions, on the contrary, was non-significant [$\chi^2(1, N = 10) = .04, p = .829$].

INTERACTIONS. Pairwise comparisons showed: that the 1st Readings of the Modified Versions elicited significantly fewer pitch mistakes than the 1st Readings of the Conventional Versions [$M = 25.72, SE \pm 4.38$, *vs.* $M = 35.44, SE \pm 4.65$, respectively; $p = .028$]; that the 2nd Readings of the Modified Versions elicited significantly fewer pitch mistakes than the 2nd Readings of the Conventional Versions [$M = 19.21, SE \pm 4.12$, *vs.* $M = 26.46, SE \pm 3.59$, respectively; $p = .006$]; and that the 3rd (rehearsed) Readings of the Modified Versions also elicited significantly fewer pitch mistakes than the 3rd (rehearsed) Readings of the Conventional Versions [$M = 12.61, SE \pm 2.63$, *vs.* $M = 17.37, SE \pm 2.94$, respectively; $p = .028$]. More saliently, the *1st Readings* of the Modified Versions elicited means of pitch mistakes not dissimilar to the

numbers elicited by the *2nd Readings* of the Conventional Versions [$M = 25.72$, $SE \pm 4.38$, *vs.* $M = 26.46$, $SE \pm 3.59$, respectively; $p = .850$]; similarly, the *2nd Readings* of the Modified Versions elicited means of pitch mistakes not dissimilar to the numbers elicited by the *3rd* (rehearsed!) *Readings* of the Conventional Versions [$M = 19.21$, $SE \pm 4.12$, *vs.* $M = 17.38$, $SE \pm 2.94$, respectively; $p = .850$] (see *Figure 5.2*).

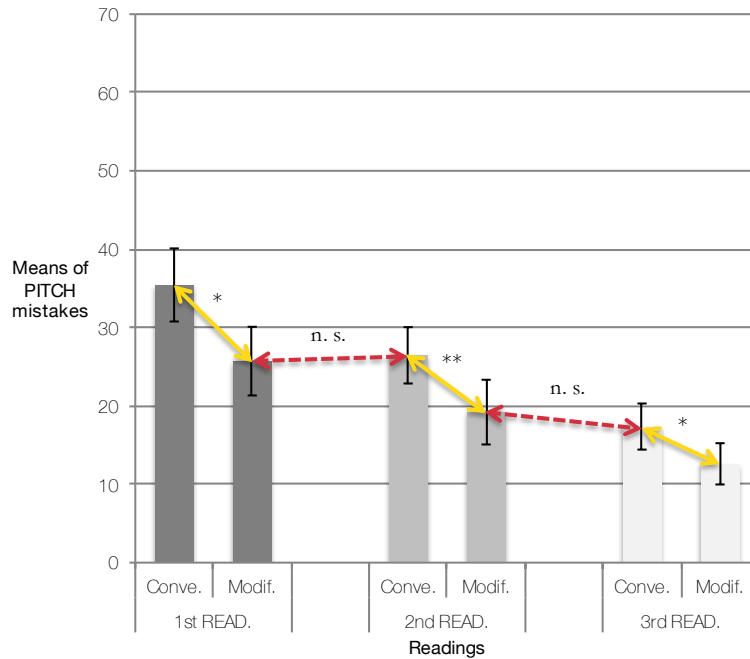


Figure 5.2.

Pairwise Comparisons of Estimated Means of PITCH Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

* $p < .050$

** $p < .010$

ORDER EFFECTS. Pairwise comparisons of the Estimated Marginal Means of Version * Order of Presentation showed that the Order of Presentation was non-significant both for the Conventional Versions Readings [$M = 26.17$, $SE \pm 5.76$ *vs.* $M = 24.55$, $SE \pm 4.46$ for Tests respectively starting with a Conventional or a Modified Version; $p = .970$], and for the Modified Versions Readings [$M = 19.00$, $SE \pm 4.15$, *vs.* $M = 17.82$, $SE \pm 4.72$, for Tests respectively starting with a Conventional or a Modified Version; $p = .970$]. Further, analysis of the three-way interactions between Version, Reading, and Order of Presentation found no significance for any of the pairwise comparisons [all $ps > .999$].

5.3.2. RHYTHM DOMAIN RESULTS

The initial model showed a non-significant effect for the Piece factor [$\chi^2(1, N = 10) = 0.11, p = .915$].

MAIN EFFECTS. The implemented Model showed Version as a highly significant source of variation [$\chi^2(1, N = 10) = 30.11, p < .001$], and Reading as a significant source [$\chi^2(2, N = 10) = 11.44, p = .003$]; Order of Presentation was non-significant [$\chi^2(1, N = 10) = 2.19, p = .139$].

INTERACTIONS. Pairwise comparisons showed: that the 1st Readings of the Modified Versions elicited a much smaller number of Rhythm mistakes than the 1st Readings of the Conventional Versions, with the difference being highly significant [$M = 10.93, SE \pm 2.10, vs. M = 21.32, SE \pm 3.35$, respectively; $p < .001$]; that the difference between 2nd Readings was similarly highly significant [with Modified scores $M = 7.75, SE \pm 1.46, vs.$ with Conventional scores $M = 15.11, SE \pm 1.89$; $p < .001$]; as was the difference between 3rd Readings [with Modified scores $M = 4.19, SE \pm 1.37, vs.$ with Conventional scores $M = 8.16, SE \pm 1.94$; $p < .001$]. Saliiently, the 1st Readings of the Modified Versions elicited means of Rhythm mistakes not dissimilar to the numbers elicited by the 2nd Readings of the Conventional Versions [$M = 10.93, SE \pm 2.10, vs. M = 15.11, SE \pm 1.86$, respectively; $p = .130$ (*n. b.*, if anything, the Modified Versions elicited fewer mistakes, albeit not significantly)]; similarly, the 2nd Readings of the Modified Versions elicited means of Rhythm mistakes not dissimilar to the numbers elicited by the 3rd (rehearsed) Readings of the Conventional Versions [$M = 7.75, SE \pm 1.46, vs. M = 8.16, SE \pm 1.94$, respectively; $p = .731$ (*n. b.*, if anything, the Modified Versions elicited fewer mistakes, albeit not significantly)]; and remarkably, the 1st Readings of the Modified Versions elicited means of Rhythm mistakes not significantly dissimilar to those elicited by the 3rd (rehearsed!) Readings of the Conventional Versions [$M = 10.93, SE \pm 2.10 vs. M = 8.16, SE \pm 1.94$, respectively; $p = .382$] (see Figure 5.3.).

ORDER EFFECTS. Pairwise comparisons of the Estimated Marginal Means of Version * Order of Presentation showed that the Order of Presentation was non-significant both for the Conventional Versions Readings [$M = 16.93, SE \pm 3.57, vs. M = 11.26, SE \pm 1.86$, for Tests respectively starting with a Conventional or a Modified Version; $p = .426$], and for the Modified Versions Readings [$M = 8.68, SE \pm 2.47, vs. M = 5.77, SE \pm 1.24$ for Tests respectively starting with a Conventional or a Modified Version; $p = .426$]. Analysis of the three-way interactions between Version, Reading, and Order of Presentation found no significance for the comparisons [all $ps > .100$].

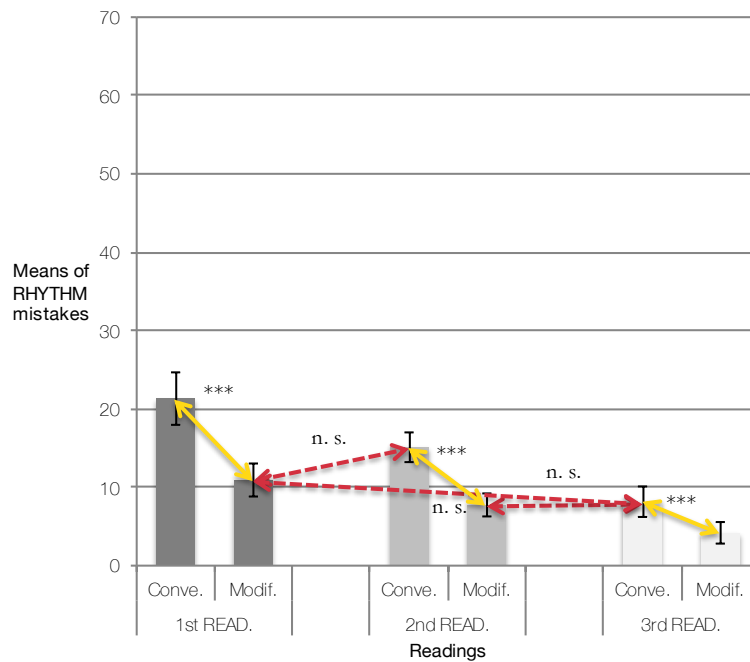


Figure 5.3.

Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

*** $p < .001$

5.3.3. TOTAL NUMBERS OF MISTAKES RESULTS

The initial Model showed a non-significant effect for Piece [$\chi^2(1, N = 10) = 2.19, p = .139$].

MAIN EFFECTS. The implemented Model showed Version as a highly significant source of variation [$\chi^2(1, N = 10) = 13.65, p < .001$], as well as Reading [$\chi^2(2, N = 10) = 32.44, p < .001$]; Order of Presentation was non-significant [$\chi^2(1, N = 10) = .50, p = .479$].

INTERACTIONS. Pairwise comparisons showed: that the 1st Readings of the Modified Versions elicited a smaller number of Total numbers of mistakes than the 1st Readings of the Conventional Versions, with the difference being significant [$M = 36.90, SE \pm 5.36, vs. M = 56.79, SE \pm 6.67$, respectively; $p = .001$]; that the difference between 2nd Readings was highly significant [$M = 27.08, SE \pm 4.91$ (Modified) $vs. M = 41.67, SE \pm 4.75$ (Conventional); $p < .001$]; and that the difference between 3rd Readings was significant [$M =$

16.73, $SE \pm 3.80$ (Modified) *vs.* $M = 25.75$, $SE \pm 4.54$ (Conventional); $p = .001$]. Saliently, the *1st Readings* of the Modified Versions elicited means of Total Numbers of mistakes not dissimilar to those elicited by the *2nd Readings* of the Conventional Versions [$M = 36.90$, $SE \pm 5.36$, *vs.* $M = 41.67$, $SE \pm 4.75$, respectively; $p = .290$]; similarly, the *2nd Readings* of the Modified Versions elicited means of Total Numbers of mistakes not dissimilar to those elicited by the *3rd* (rehearsed) Readings of the Conventional Versions [$M = 27.08$, $SE \pm 4.91$, *vs.* $M = 25.75$, $SE \pm 4.54$, respectively; $p = .750$] (see *Figure 5.4*).

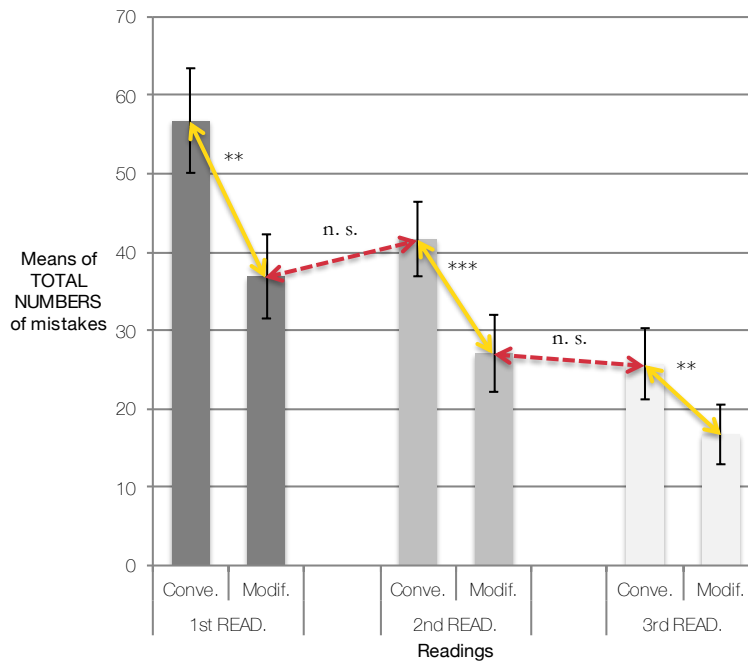


Figure 5.4.

Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

** $p < .010$

*** $p < .001$

ORDER EFFECTS. Pairwise comparisons of the Estimated Marginal Means of Version * Order of Presentation showed that the Order of Presentation was non-significant both for the Conventional Versions Readings [$M = 43.02$, $SE \pm 9.16$, *vs.* $M = 36.01$, $SE \pm 4.60$, for Tests respectively starting with a Conventional or a Modified Version; $p = .741$], and for the Modified Versions Readings [$M = 27.95$, $SE \pm 6.45$, *vs.* $M = 23.40$, $SE \pm 4.69$, for Tests respectively starting with a Conventional or a Modified Version; $p = .741$]. Analysis of

the three-way interaction Version * Order of Presentation * Reading found no significance for the comparisons [all p s > .999].

5.3.4. NORMALISED PAIRWISE VARIATION INDEXES RESULTS

The initial Model showed a non-significant effect for Piece [$\chi^2(1, N = 10) = .149, p = .700$].

MAIN EFFECTS. The data consisted of positive scale values that did however present a skewness that clearly deviated from assumptions of normality (see histogram in *Figure 5.5*). The model implemented, assuming a Gamma probability distribution, showed Version as a highly significant source of variation [$\chi^2(1, N = 10) = 26.52, p < .001$]; Reading, remarkably, had no significant effect [$\chi^2(2, N = 10) = 2.82, p = .243$]; Order of Presentation was only marginally significant [$\chi^2(1, N = 10) = 2.99, p = .083$].

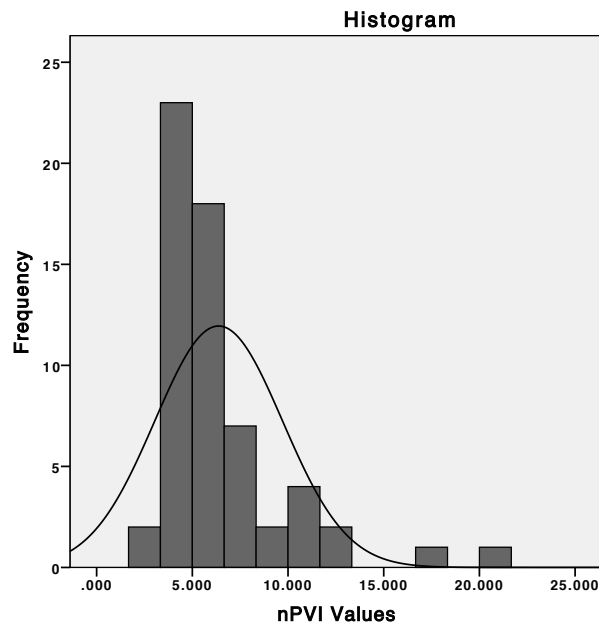


Figure 5.5.

Histogram of the nPVI values, including a normal curve. A marked skewness can be observed.

INTERACTIONS. Pairwise comparisons showed: that the 1st Readings of the Modified Versions generally elicited smaller Total numbers of mistakes than the 1st Readings of the Conventional Versions, with the difference being marginally significant [$M = 6.45, SE \pm .97, vs. M = 8.69, SE \pm 1.75, respectively; p = .065$]; that the difference between 2nd Readings was highly significant [$M = 5.11, SE \pm .31$ (Modified) *vs.* $M = 6.89, SE \pm .59$ (Conventional); $p < .001$]; and that the difference between 3rd Readings was also highly

significant [$M = 4.99$, $SE \pm .38$ (Modified) *vs.* $M = 6.72$, $SE \pm .38$ (Conventional); $p < .001$]. Saliiently, the *1st Readings* of the Modified Versions elicited indexes of stability not dissimilar to those elicited by the *2nd Readings* of the Conventional Versions [$M = 6.45$, $SE \pm .97$, *vs.* $M = 6.89$, $SE \pm .59$, respectively; $p = .926$]; furthermore, the *1st Readings* of the Modified Versions elicited indexes not dissimilar to those elicited by the *3rd* (rehearsed) Readings of the Conventional Versions [$M = 6.45$, $SE \pm .97$, *vs.* $M = 6.72$, $SE \pm .38$, respectively; $p = .987$]; along the same lines, the *2nd Readings* with the Modified Versions produced performances that were indexed as more fluent than the *3rd* (rehearsed) Readings with the Conventional Versions, with the difference being moreover highly significant [$M = 5.11$, $SE \pm .31$, *vs.* $M = 6.72$, $SE \pm .38$, respectively; $p > .001$] (see *Figure 5.6*).

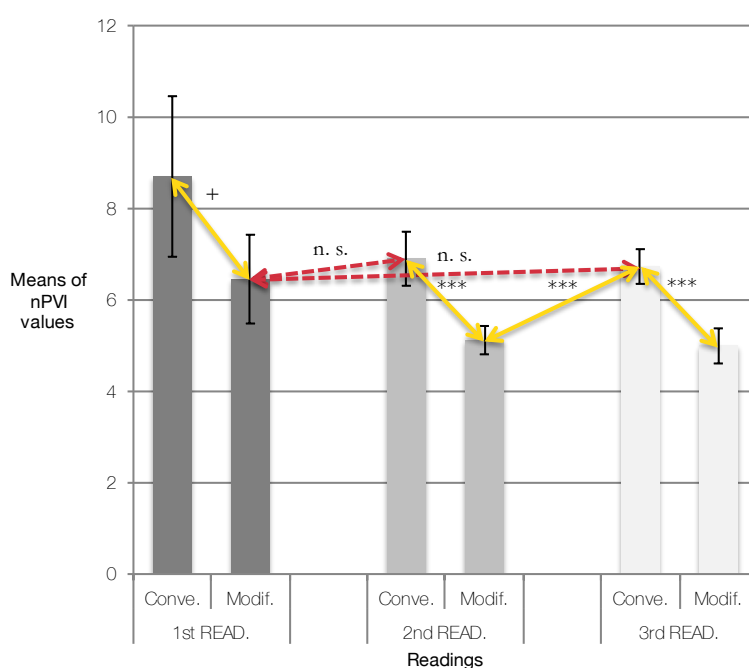


Figure 5.6.

Pairwise Comparisons of Estimated Means of nPVI VALUES for performances using Conventional Versions (Conve.) or Modified Versions (Modif.). Smaller values signify more stable performances. From left to right, Paired Differences for the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

+ $p < .100$

*** $p < .001$

ORDER EFFECTS. Pairwise comparisons of the Estimated Marginal Means of Version * Order of Presentation showed that the Order of Presentation was non-significant both for the Conventional Versions Readings [$M = 8.41$, $SE \pm 1.11$, *vs.* $M = 6.49$, $SE \pm .73$, for Tests respectively starting with a Conventional or a

Modified Version; $p = .254$], and for the Modified Versions Readings [$M = 6.23$, $SE \pm .64$, *vs.* $M = 4.81$, $SE \pm .40$, for Tests respectively starting with a Conventional or a Modified Version; $p = .254$]. Analysis of the three-way interaction Version * Order of Presentation * Reading found no significance for any of the comparisons [all p s > .100].

5.3.5. EXPRESSIVITY AND USE OF DYNAMIC RANGES

In the case of the dynamic range dataset, the variability was observed to be rather low, with most participants taking a careful approach to the interpretation of the dynamics indications. An analysis of frequencies confirmed a close to normal distribution (see Histogram in *Figure 5.7.*).

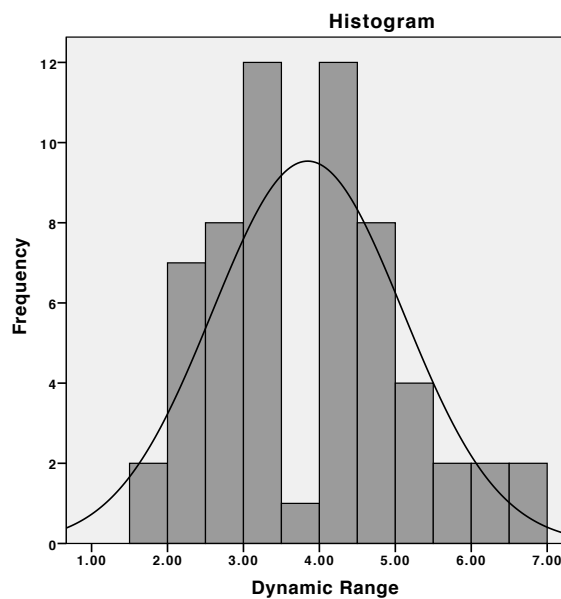


Figure 5.7.

Histogram of the Dynamic Range values, including a normal curve. The dataset ($N = 10$) seems to follow a Normal probability distribution.

Further, a Generalized Linear Model using Piece, Reading, and Order of Pieces as categorical variables found no main effect for any of these factors [all p s > .100] nor for any of their interactions [all p s > .100]. This meant that a simple Paired Samples Test (t -Test) could be implemented, without using a Subject variable nor an Offset Variable, and assuming a normal distribution —Normality was additionally confirmed with a Shapiro-Wilk test [$p > .100$ for all variables]

The Paired Samples Test showed that performances with the Modified Versions used wider dynamic ranges than the performances with the Conventional Versions, although the differences were not highly significant. Specifically, the difference in the 1st Readings was marginally significant [$M = 4.08$, $SE \pm .40$

(Modified) *vs.* $M = 3.51$, $SE \pm .45$ (Conventional); $p = .060$]; the difference in the 2nd Readings was similarly marginally significant [$M = 4.13$, $SE \pm .34$ (Modified) *vs.* $M = 3.61$, $SE \pm .45$ (Conventional); $p = .086$]; the difference in the 3rd Readings did not reach significance [$M = 4.06$, $SE \pm .42$ (Modified) *vs.* $M = 3.67$, $SE \pm .33$ (Conventional); $p = .191$] (see *Figure 5.7*). Also, there was no significant evolution for the Versions throughout the Readings: the differences between *Third* and *1st Readings* of the Conventional Versions were not significant [$M = 3.67$, $SE \pm .33$, *vs.* $M = 3.51$, $SE \pm .45$, respectively; $p = .792$]; similarly, the differences between *Third* and *1st Readings* of the Modified Versions were not significant [$M = 4.06$, $SE \pm .42$, *vs.* $M = 4.08$, $SE \pm .40$, respectively; $p = .953$] (see *Figure 5.8*).

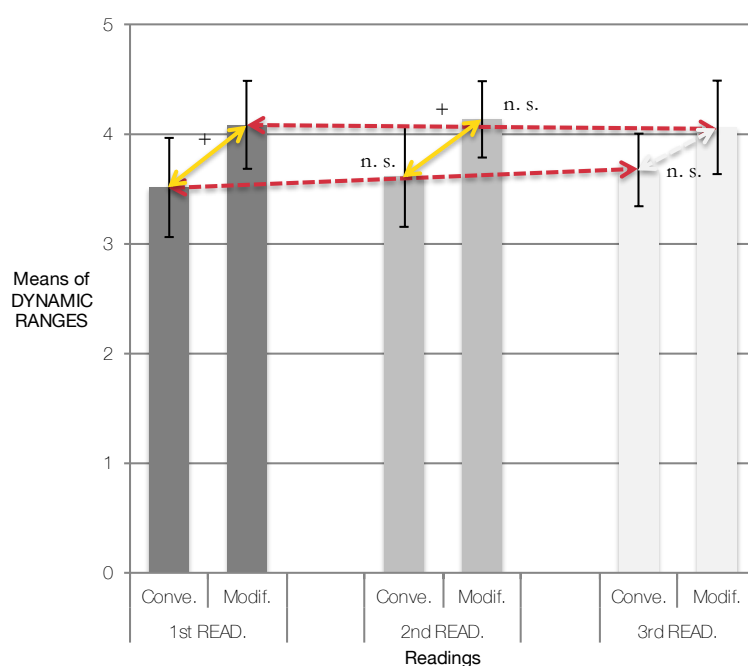


Figure 5.8.

Paired Samples Test of DYNAMIC RANGES for performances using Conventional Versions (Conve.) or Modified Versions (Modif.). Larger values signify a wider dynamic range in the performances. From left to right, Paired Differences for the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings. The vertical axis represents a Loudness Scale algorithmically implemented by the ‘Sonic Visualiser’ program.

Error bars represent Standard Error.

Significance (2-tailed):

n. s. non-significant

+ $p < .100$ (marginal)

Finally, and taking into account that this experimental design included a total of 6 performances by participant (three Readings per Version, with a study time between Reading 2 and Reading 3, see *Figure 5.1*), I analysed the effect of the position in the sequence of six Readings on the dynamic range, hypothesising that there would be a familiarity (or ‘warming up’) effect on the willingness to risk using

louder (less controlled) attacks as the experiment progressed. As can be seen in the overlaid scatterplots in *Figure 5.9.*, whilst the minimum loudness values were similar for both Versions, there was a tendency in the performances using Modified Versions to use a wider range of maximum dynamics (reflected on the wider dispersion in the figure), that were generally louder than the ones used when performing with the Conventional Versions (reflected on a higher interpolation line), and that tended to grow for the last performance after rehearsal (see increase for position 6).

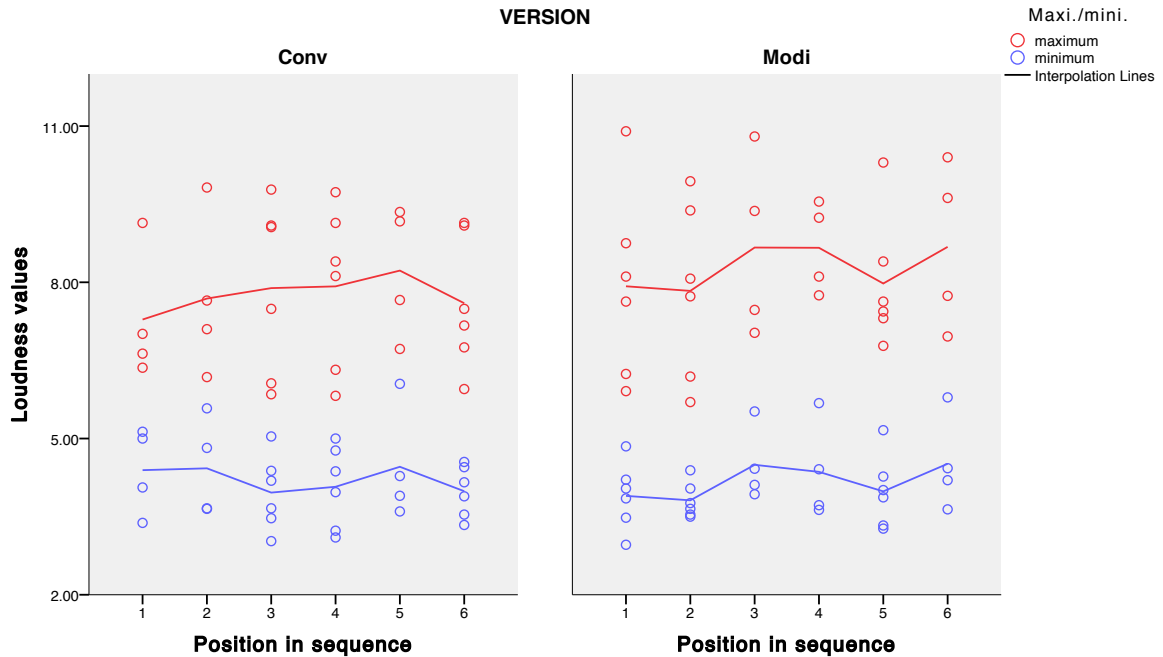


Figure 5.9.

Overlay Scatterplot of MINIMUM AND MAXIMUM LOUDNESS for performances using Conventional Versions (Conv.) or Modified Versions (Modi.). Loudness values were assigned by the Sonic Visualiser Programme (Queen Mary University of London; release 2.2) using a scale adapted for the dataset of the experiment. The horizontal axis signifies the six possible positions in the sequence of performances (Readings) of a given player. Notice the slightly wider spread and higher line of maximum levels for the Modified Version.

5.3.6. ASSESSMENT QUESTIONNAIRES

When including both the opinions of students that had participated actively in the experiment and of students that were presented with the scores merely to browse (silently) through them, a clear preference for the Conventional or the Modified Versions of the scores was not found, in terms of these facilitating sight-reading performances; students did, however, show a certain preference for the Modified Versions if having to retrieve specific information (*e. g.*, dynamics markings) from the scores; finally, the preference for the Modified Versions was marked for the scenario of a possible memorization of the Pieces (see Histograms in *Figure 5.10.*).

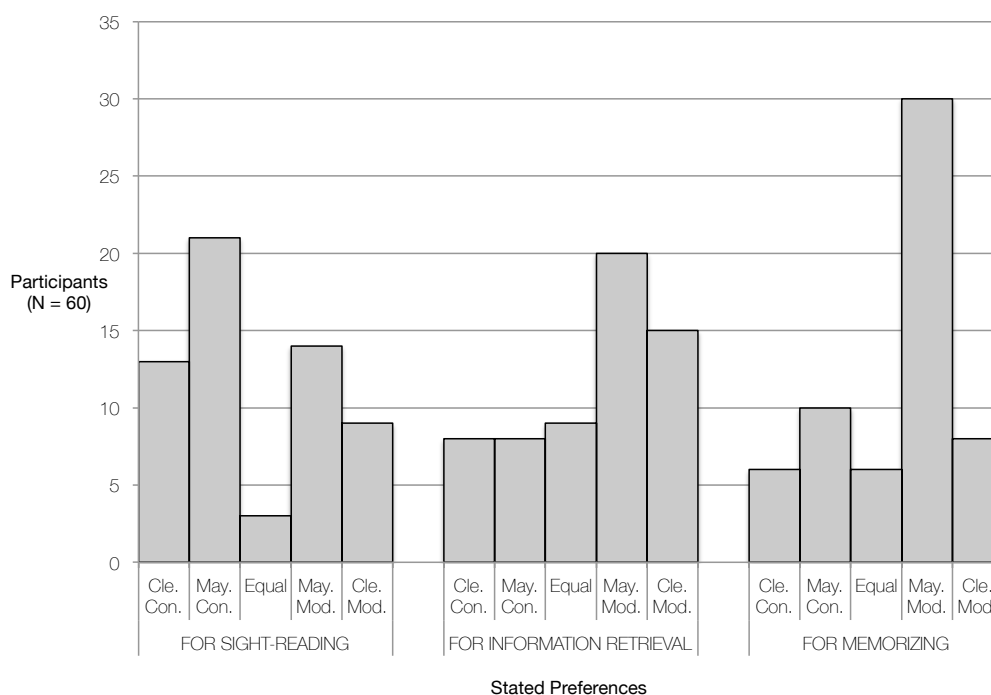


Figure 5.10.

Histograms of STATED PREFERENCES for scores using Conventional design (Con.) or Modified design (Mod.), measured in a five-point Likert scale (from ‘Clearly the Conventional one’ to ‘Clearly the Modified one’). 60 Participants (of which 11 played the scores and 49 only looked at the scores) were asked on their preferences in three possible scenarios: having to sight-read the Piece; having to retrieve information (on articulation and dynamics) from it; and having to memorise the music.

However, an exploration of the dataset using a Shapiro-Wilk test of Normality showed that the responses were non-normally distributed, particularly in the case of participants that had only looked at the scores (see Table 5.1.).

An analysis through a Generalised Linear Model, assuming a Gamma Probability Distribution (*i. e.*, with only positive values, skewed towards higher values within the range) and a Log link function (*i. e.*, assuming non-normal distribution), showed that the participants that had played through the scores stated a higher preference for the Modified scores than the participants that had only looked at them, with this difference being statistically significant in the case of having to sight-read the scores [$M = 2.53$, $SE \pm 1.18$ (looked) *vs.* $M = 3.73$, $SE \pm 0.57$ (played); $p = .024$], but with the difference not reaching statistical significance in the supposed case of having to retrieve information from the scores (*e. g.*, on articulation or dynamic markings) [$M = 3.35$, $SE \pm 0.22$ (looked) *vs.* $M = 3.82$, $SE \pm 0.53$ (played); $p = .206$], nor in the supposed case of having to memorise the scores [$M = 3.39$, $SE \pm 0.20$ (looked) *vs.* $M = 3.45$, $SE \pm 0.43$ (played); $p = .890$] (see Figure 5.11.).

Table 5.1.

*Questionnaire responses:
Shapiro-Wilk Tests of
Normality*

CASE	Look/ Play	Sta.	df	Sig.
Sight-reading	Looked	.847	49	.000
	Played	.815	10	.015
Information retrieval	Looked	.852	49	.000
	Played	.853	10	.047
Memorising	Looked	.810	49	.000
	Played	.876	10	.093

All p s < .100; normality can *not* be assumed.

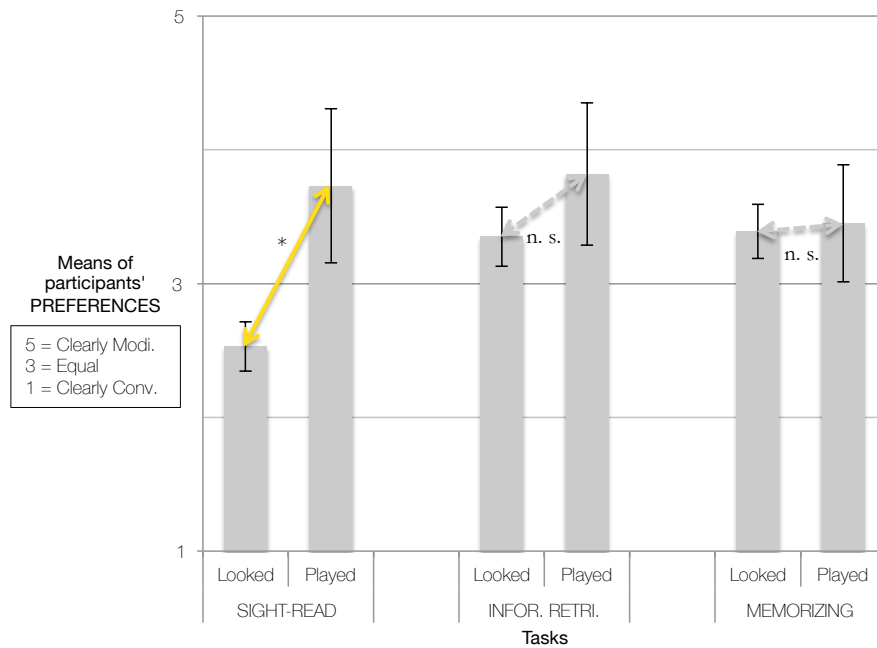


Figure 5.11.

Pairwise comparisons of Estimated Marginal Means of STATED PREFERENCES by participants for scores using Conventional design (Conv.) or Modified design (Modi.), depending on whether they only LOOKED at the scores or PLAYED them. From left to right, Paired Differences in preference in case of having to sight-read the score, having to retrieve information from it, or having to memorise it.

Error bars represent Standard Error ($n_{\text{looked}} = 49$; $n_{\text{played}} = 10$).

Significance using the Sidak correction:

n. s. non-significant

* $p < .050$

5.4. DISCUSSION: ADVANTAGES AND APPLICABILITY OF THE MODIFIED DESIGN

5.4.1. ADVANTAGES OF THE MODIFIED SCORES IN ALL ASSESSMENTS

After changing our materials, coding systems, and experimental procedure in order to deal with the aforementioned critiques, results again showed highly significant differences between performances using the conventional and the modified scores (in some aspects even more marked), with the latter evoking fewer mistakes, better tempo stability, wider dynamic ranges, more favourable assessment by participants, and better performances after rehearsal.

The apparently forceful results in the present experiment seem to confirm that the more fluent readings and better performances with the modified scores are not due to a generic advantage of vertical arrangement over horizontal arrangement: apart from the language literature (there are no studies in the music domain addressing this issue) pointing in the opposite direction (with general advantages for horizontal arrangements), both versions were now presented in portrait orientation, and the differences in accuracy and fluency seem to generate mainly from the structured spacing of the materials.

5.4.2. ACCURACY, FLUENCY, AND REHEARSAL TIMES

A. EFFICIENT MANAGEMENT OF REHEARSAL TIMES

In the present experiment three readings of each piece instead of two (as in the previous experiment) were used in the procedure. Moreover the third reading included a rehearsal/study time for each participant to prepare the third performance of each piece (four minutes per piece). The total numbers of mistakes in the first, second and third reading were smaller for performances using the modified scores than for performances using the conventional scores. Saliiently, the numbers of mistakes in the third (rehearsed) readings of the conventional versions were statistically at the same level as the numbers of mistakes in the second (unrehearsed) readings of the modified versions; in the same manner, the number of mistakes in the second reading of the conventional scores were similar to the numbers in the first readings of the modified versions (see *Figure. 5.4.*). This effect was even more significant when only rhythm domain results were taken into account, with the number of mistakes in the third (rehearsed) readings of the conventional scores being very similar (not statistically distinguishable) to the numbers of mistakes in the first (strictly the only sight-) reading of the modified versions (see *Figure. 5.3.*). Other considerations aside, it could be argued that the modified scores do afford a significantly more efficient management of practice and rehearsal time, being able to reduce it to a half —or even a third, depending on the performance priorities.

B. TIME MANAGEMENT BY EXPERT MUSICIANS

Studies on time use in instrumental rehearsals have shown that more advanced or experienced musicians can be characterised for a generally more efficient management of time and focus in practicing a piece. More specifically, Goolsby [1996], in comparing the rehearsal methods of school orchestra conductors with different levels of expertise, found that the more experienced ones talked the least during rehearsals, got the ensembles on-task the quickest, and divided rehearsal time more equally between musical sections. In a replication study seeking to determine characteristics that would define successful or outstanding Band Directors, Goolsby [1999] found, in convergence with the previous study, that novices used more rehearsal time overall, and spent more time in verbal instruction; the expert directors spent a greater percentage of the rehearsals performing than novices did. The picture that emerges from these studies is that of expert conductors being particularly time conscious, efficient and focused in rehearsals, and having a preference towards using rehearsal time playing, rather than deciphering, analysing or commenting on the score (these useful and valuable activities are consequently carried out in non-rehearsal time).

C. PERCEPTION AND RATING OF TIME MANAGEMENT

This concern with effective time management and focused pacing of rehearsal is not unilateral, though: not only are expert conductors more preoccupied with prioritising on-task behaviour and timing the rehearsal at an efficient pace, but they are also *perceived* as more competent when doing so. As shown in an analysis of ratings of video-recorded rehearsals carried out by Yarbrough and Henley [1999], musicians (all participants rating the recordings were university-level music students) give conductors highest ratings when there is a low percentage of performer off-task behaviour, many activity changes, a high percentage of performance time, and a rapid pacing in the rehearsal.

Also the interaction between conductor and performers can show similar trends towards effectiveness and prioritisation of playing time over instructional time. Worthy [2003] compared the behaviour of an expert conductor when rehearsing the same literature with a school band and a university band, and found that rehearsal with the senior students was characterized by less frequent interruptions and talking episodes. In a subsequent study observing expert wind conductors rehearsing with high-profile university bands (Intercollegiate Honour Bands, in American terminology), Worthy [2006] similarly and predictably found that rehearsals were characterized by fast pace and short episodes of conductor talking.

D. APPLICABILITY OF THE NOVEL SCORE DESIGN


Expert musicians (both players and conductors) emerge thus as interested in —and appreciative of— effectiveness of rehearsals and focused performative strategies. Furthermore, all aforementioned studies were carried out with student orchestras, and it seems reasonable to extrapolate that in professional circles

the concerns with efficacy and good rehearsal management will be even more acute (particularly in British orchestras, which work on the tightest schedules in the circuit).

In this sense, if the novelties in script layout, spacing, and information separation implemented in our modified scores seem to produce results that are equivalent to the ones elicited with conventional scores, but reducing practice or rehearsal time to less than the half, the implementation of these novel score design practices could perhaps be a welcome relief for the busy, stressed and many times overworked musicians of the era of efficiency. In the next experiments, the applicability of the design novelties on scores for different instruments must be assessed, as well as a more detailed analysis of which components of the novel design are eliciting the facilitation of legibility.

ACKNOWLEDGEMENTS

I am grateful to the participants, all of whom engaged kindly and selflessly in the experiment. I am equally grateful to *Maestro* Arnold Marinissen for his invaluable help in the organisation of the tests and recording sessions at the *Conservatorium van Amsterdam*. The students Elia Benhamou (guest student at the Centre for Music and Science) and Lauren Fink (MPhil in Music Studies at the Centre for Music and Science) were very helpful with the management of participants and implementation of the questionnaires.



*Experiment IV:
Testing the Universality
of the Novel Designs*

CHAPTER 6

Experiment IV: Testing the Universality of the Novel Designs

6.1. INTRODUCTION: ENCODING, SENSORIMOTOR TRANSFORMATIONS, AND UNIVERSAL NOTATIONS

6.1.1. MAPPINGS IN RELATION TO INSTRUMENTAL PRACTICES

In what is possibly—at least to this author’s knowledge—the first tomographic investigation of the neural networks underlying musical sight-reading, Sergent *et al.* [1992] made a distinction that has been held in the theoretical models describing the skills involved in the task [*e. g.*, Kopiez & Lee, 2006, 2008], as well as in the descriptions of expertise or degrees of achievement in the task [*e. g.*, J. I. Lee, 2003; Kopiez *et al.*, 2005]: namely, the distinction between the encoding, identification and processing of symbols, on the one hand, and the sensorimotor transformation, physiological preparation and sequencing of production on the other. Whilst the purely abstractive part of the process would be common to all instrumental and vocal practices, or more in general, to any attempt at reading and understanding a score, the organization of motor sequences required for performance will be dependent on the specificities of each practice.

In their pioneering study, Sergent *et al.* [1992] observed that two main areas were recruited when subjects had to conjunctly read and play, in addition to the areas implicated when solely reading (without performing) or solely playing (musical scales, without scores). One involved the Superior Parietal Lobule (Area 7; in both hemispheres), an area strategically placed to mediate the sensorimotor transformations for visually guided skilled actions and positions; this was consistent with neurophysiological, electrophysiological, and behavioural evidence [see Goodale & Milner, 1992, for an overview; also Taira *et al.*, 1990] suggesting that the neural substrates that play a role in the perceptual *identification* of objects are distinct from the regions mediating visually guided *actions* directed at such objects. The other area of activation specific to conjunctly reading and playing involved the left inferior frontal gyrus (Area 44),

immediately above Broca's area, which plays a critical role in organizing motor sequencing (*e. g.*, in speech production); the activation of this dorsal region of Area 44 in sight-reading performances would reflect, according to Sergent *et al.* [*id.*], a similar role in the organization of the motor sequences inherent in musical performance, and more specifically in keyboard performance, in their study.

If the visual guidance of movements [hand movements in the examples by Sergent *et al.*, 1992 and Taira *et al.*, 1990], especially in matching the pattern of movement with the spatial characteristics of the object to be manipulated or interacted with, recruits neural resources that are distinct from the visual encoding and structuring of symbols when reading, it could be the case that the effect of the novel designs on sight-reading performances in the previous experiments (Experiments I, II and III) was related to the idiosyncrasies of percussion players and/or to the motoric and physiologic requirements inherent to mallet instrument performance. Therefore, the present experiment used a similar design to Experiment III, where percussion players performed using conventional and novel scores, but this time using violinists, whose playing is quite removed from the practices of the previous instrumentalists: (1) the music is now monophonic, in one staff, and no simultaneous encoding of lines is required (it could be the case that the clearer spacing in the novel design simply enhanced this integrating aspect of performance, and not the readability of separate notes or lines); (2) similarly, the sense of pulse is now mostly relying on the right hand (it could be the case that the clarity in the novel design had simply enhanced the inference of a common pulse, and not the co-ordination of information from different lines); (3) the attack of notes is less marked and generally plays a lesser role in the idiosyncrasy of performance (contrary to its prominent role in percussive instruments, which perhaps benefitted from the clarity in spacing for their more marked performances); and (4) most importantly, the positioning and motoric sequencing is noticeably different between percussionists and violinists, with the latter requiring simply much less amplitude of movements and kinetic energy to perform, particularly in terms of sideway movements (it could be the case that enhanced clarity in the novel scores simply facilitates the linear mapping of vertical distances in the staff to horizontal displacements over a mallet instrument).

6.1.2. UNIVERSAL NOTATIONS VERSUS TABULATURA

It seems important to note at this point that one of our aims when designing the novel scores was to maintain one of the most valued and efficient aspects of conventional (western) notation: its universality and equal applicability to all instruments (and voices) used in common practice. In the history of western notation, in certain (mostly early) periods, some instrumental scores have used alphabetical and numerical notations that were specific to the requirements, technique and repertoire of such instruments, and not (easily) translatable to instrumental practices removed from the original [for a historical overview, see Bent *et al.*, 2001; Rastall, 2008]. The most important type of notation within these instrument-specific approaches is the *Tabulatura* (from the Latin *tabula* —table or diagram; henceforth 'tablature', in English),

a distinctive instrumental notation initially introduced when the vast majority of western notated music was purely vocal.

In the case of keyboard music notation, the earliest known sources [arguably initiating in *c.* 1275; see Bent *et al.*, 2001, for the debate] are nearly all in tablature, the system using letters or other symbols not found in ordinary staff notation, and which tends to specify the physical *action* required to produce the music from a given instrument, rather than an abstract representation of the music itself (although some systems like the German organ tablatures of the Late Middle Ages use a hybrid with elements of staff notation). Up to the early XVIth century most surviving keyboard sources are still notated in tablature, and the transition to mensural notation (the standardized universal abstractive notation that we roughly still use nowadays) had different timelines according to regional preferences or traditions: whereas in Italy there are printed keyboard sources using mensural notation throughout already in the late XVth-century Faenza Codex [Anonymous, XVth cent.], a somewhat convoluted mixed system using alphabetical notation supplemented by a uniform presence of mensural rhythm signs for the high/est voice/s survived in Germany into the XVIIIth century —tellingly mostly in manuscripts written by composers (including J. S. Bach) for their own use [Bent *et al.*, 2001].

Being action-based and instrument-specific, tablature scripts would generally be considered more approachable for performers without knowledge of musical theory or (abstractive) notation: Ottaviano Petrucci, credited as the first significant publisher of polyphonic music, produced from 1503 on an international repertory appealing to professional musicians (with his method of printing creating works of famed elegance), shifting from 1507 to the distribution of volumes of lute tablature, offering a repertory that departed from his previous outputs in that it addressed as much dilettantes as professionals [Boorman, 2001]. The principles on which Petrucci's tablatures rest could explain the success and pervasiveness of Italian lute tablatures for decades if not centuries to come: contrary to the limitations of keyboard tablatures in iconic terms, in this case there is a direct correlation of the diagram with the features of the instrument, with six lines of a 'staff' representing the six courses (double or paired strings) of the lute; numerals placed on these lines indicate the fret to be stopped on the relevant course, and rhythm signs (derived from mensural note shapes but lacking the note heads) placed above the 'staff' indicate the duration of the notes. This efficient notational system was similarly used in Spain for the vihuela repertoire (although in this case the sequence of strings in the diagram was reversed), and a conceptually similar system (again, with some changes in the surface of the diagrams) was used in printed French lute tablatures, which revealingly have as one of their earliest examples a publication giving instructions for beginners, Attaignant's *Très briefve et familière introduction* [see Bent *et al.*, 2001]. French lute tablature, after spreading throughout northern Europe, persisted especially in Germany, where it continued to be printed at least until the 1770s.

Guitar music from around 1550 is also notated in the successful Italian or French lute tablature, and as in tablatures for other plucked string instruments, the number of staff-lines in the diagrams vary according to the number of strings, affording an iconic and direct representation of the board on which the fretting hand (left hand in the vast majority of cases) has to operate. Tablatures for guitar remained in use until late XVIIIth century, when they yielded to ordinary mensural notation on a single staff. Taking further the iconic representation of fretting action on a diagram, tablatures of a new simplified type were introduced in the XXth century for use in popular guitar music and later in pop entertainment, with the diagram providing a schematic picture of the fingerboard, and dots representing the positions of the fingers. This tablature chord notation, like the abbreviated representation of a chord by a symbol with a capital letter (*e. g.*, the notes C, E, G, B \flat being represented by C7) lacks any indication of rhythm, which is to be supplied by the performer from knowledge of idiom or style.

In general, a notational system like the tablature is direct and efficient within the constraints of a closed repertoire and relies on idiomatic turns that are part of the performer's long-term memory knowledge, frequently provided by extended exposure. Its success in practices with highly conventionalised, formalised, and/or predictable discourses (as the Spanish vihuela repertoire, which reached the height of its development as private music-making in the court of Charles V) can also be accounted as a series of limitations, for instance in:

UNPREDICTABLE DISCOURSES. It will not be apt for music with an unpredictable grammar or the introduction of experimental repertoire. In Charles V's court, one of the preferred roles for the vihuela was to present intabulations of polyphonic compositions of well-known and highly esteemed Flemish composers (*e. g.*, Josquin des Prez, prominently). Similarly, in secular vihuela music, the *Fantasías*, a popular form of virtuoso display, were based on more or less technical exercising of adornments and *glossa* over the structure of a very familiar melody.

CROSS-NOTATIONAL COMMUNICATION. It will not function well if communication with performers using different notational conventions is required. Therefore, it is very rarely used in heterogeneous or large ensembles.

UNIVERSAL APPLICABILITY. Since the notational system derives from the playing technique of a particular instrument, it reaches maximum effectiveness in its own field, but conversely lacks in universal applicability.

On the contrary, the modifications in design applied to common practice repertoire proposed in my modified scores aim at facilitating the encounter with foreign languages (as used by non-native readers), the entrainment to a pulse (at least in the form of tempo stability), and the extraction of the structuring or compositional processes from the script (with subsequent facilitation of multi-modal usage). The system aspires thus at universality, and the experiment with violinists used therefore, in as far as possible, a

method, coding procedure and statistical analysis greatly resembling the previous experiments with percussionists (particularly Experiment III).

6.2. METHOD: COMPARISONS OF PERFORMANCES BY VIOLINISTS

6.2.1. PARTICIPANTS

The experimental procedure was approved by the Faculty of Music's Research Ethics Committee. Sixteen violinists (8 female) took part, all students at the University of Cambridge (most of them at the Faculty of Music). All students followed the procedure correctly, and no data had to be discarded.

6.2.2. MATERIALS

I composed the Pieces for the test after the manner of J. S. Bach's instrumental solo writing: two short Pieces (10 C bars each; one in g minor and one in b minor) with suggestions of polyphony, and structuring through cadences (see APPENDICES 6.A. and 6.B.). The melodic contents were created upon harmonic formulas encountered frequently in Bach's chorales.

6.2.3. ENVIRONMENT AND PROCEDURE

At the beginning of the test, a set of instructions were presented verbally to each participant. They were told that the test aimed to determine whether or not the modified notation improved sight-reading accuracy, explained the format of the pre-test and told to perform the Pieces as if it were a live situation (*i. e.* not to stop or go back if they made a mistake). During this, the participants were shown an example of the modified notation next to traditional notation. The participants were also invited to ask any questions they had about the procedure. Following this, a pre-test was conducted to determine the optimum speed for the test. In the pre-test the participants read off music in the modified notation to help them to familiarize themselves with it before the recorded test. The metronome volume was adjusted at a comfortable level for each participant. Once the participants were ready, the metronome was started and the participants came in on the fourth beat of the metronome (since the music started on an upbeat). The metronome was stopped at the end of the first phrase, but the participants were asked to try and keep to the same tempo throughout the performance. This process was then repeated twice more, at faster speeds each time in order to determine the threshold at which each participant was making substantial mistakes but could keep up with the tempo assigned. If there were still no mistakes made in the third playing (at the fastest speed), then the recorded test took place at an even faster tempo. Otherwise, we used the participants' final tempo for their recorded test.

The recorded test was similar to the pre-test in format except that it was recorded, the tempo was now fixed and there were six performances (three each of two different Pieces) rather than only three. Before

the test began the researcher again read out another set of instructions to the participant to explain the next part of the experiment. The pieces of music were covered up by the researcher between the recordings to prevent the violinists becoming too familiar with the music between the three performances (the cover sheet only left the time and key signatures visible). *Figure 6.1.* shows the schematic timeline of the procedure.

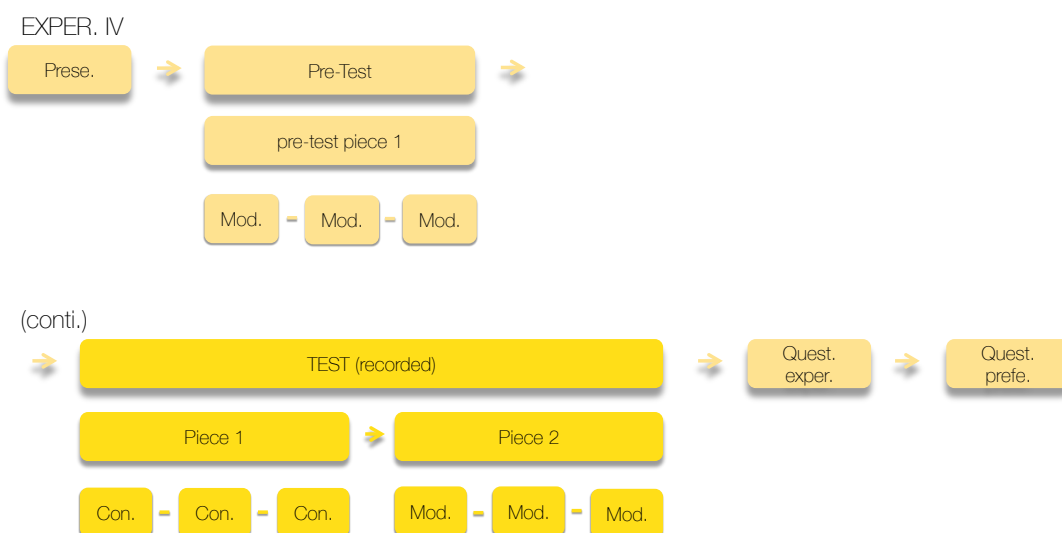


Figure 6.1.

Schematic timeline for EXPERIMENT IV: the project and novel designs are presented to the participant (without the research hypothesis nor the exact scores to be tested revealed); the participant then performs a Pre-Test, in order to get acquainted with the procedure and the novel designs through performance; this is followed by the Test proper (recorded to be coded); at the end of this the participant is questioned on musical education and experience; finally, the participant is questioned on preferences (Conventional or Modified Versions).

Approximate duration: 30 min.

6.2.4. COLLECTION AND CODING OF DATA

Recording of participants, collection of data and coding took similar forms to the procedures used in the last experiment with percussionists (Experiment III, see SECTION 5.2.). One notable exception was that the coding of mistakes was done by the students, due to limitations of time and academic function of the project, with markings of pitch errors limited to wrong notes —*i. e.*, not counting the numerous mistakes in the form of hesitations, bad attack or timbre, or, crucially, notes out of tune. The other (minor) difference was that the preferences were stated in a 3 level Likert scale instead of a 5 level one.

6.3. RESULTS OF EXPERIMENT WITH VIOLINISTS

A rather reassuring feature of this experiment was that, although the materials for performance were composed (or more precisely arranged) and edited by myself, the rating of performances (the coding of errors) was entirely carried out by the students, Maya Amin-Smith and Guy Edmund-Jones. Furthermore, they carried out their two ratings independently, and proceeded then to implement an *a posteriori* Inter-Rater reliability test which, based on a random cross-section of the recordings, showed a very strong positive correlation between their two marks [$r(16) = .82, p < .001$]

Only one of the participants (part. 7) could have been considered an outlier, since he performed with clearly more mistakes than the rest of participants. However, statistical tests implemented with and without his data yielded results that were not different in terms of significance of effects, so he was kept in the analysis. Further, and contrary to what had been the case with the percussionists (where 3 participants did not complete the test correctly), this participant did formally complete the test as all other colleagues, albeit with a noticeably higher number of mistakes.

As in the experiments with percussionists, all recordings were coded for mistakes in the Pitch Domain and for mistakes in the Rhythm Domain, which were subsequently added up to produce a Total Number of Mistakes. Results are discussed for each of these three sub-sections. Tests were implemented with the IBM SPSS Package Version 21.

Analysis of the numbers of mistakes was carried out using Generalized Linear Models with the numbers of mistakes as dependent variables, and assuming (see SECTIONS 4.3. and 5.3.) a Poisson Probability Distribution and a Log Link Function.

6.3.1. PITCH DOMAIN RESULTS

An initial model was run with Version (Conventional or Modified), Reading (1st, 2nd or 3rd Reading), and Piece (Piece 1 - g minor; Piece 2 - b minor) as categorical factors. The model found a highly significant effect of Reading [$\chi^2(2, N = 16) = 48.06, p < .001$] and, unfortunately (in spite of our precautions) of Piece [$\chi^2(1, N = 16) = 51.02, p < .001$], with Piece 1 (in g minor) eliciting significantly fewer pitch mistakes than Piece 2 (b minor) [$M = 2.54, SE \pm .37, vs. M = 5.67, SE \pm .81$, respectively]; the effect of Version appeared in this context to be non-significant [$\chi^2(1, N = 16) = .14, p = .706$].

A second Model was then implemented, using Piece as an Offset Variable. Piece becomes thereby a 'structural' predictor, and its coefficient is not estimated by the model. This is an especially useful

procedure in Poisson regression models, where each case may have different levels of exposure to the event of interest. In our experiment, each participant was exposed to Pieces of different levels of difficulty, which meant that the participant would be more or less prone to make mistakes. The effect of other factors can then be evaluated, taking into account the ‘intrinsic’ difficulty of each Piece. More specifically, the numbers of pitch mistakes per Piece were added up (summing the mistakes in the 1st, 2nd, and 3rd Readings), and the means thereof calculated: $M = 7.64$ ($SD = 6.70$) for Piece 1, and $M = 17.00$ ($SD = 14.71$) for Piece 2, confirming that indeed the second Piece was generally more difficult, even if there was a high variability in the number of mistakes as a function of the participants’ abilities. Since a Log Link function was assumed for the model, these quantities were used as logarithmically transformed values (Ln).

MAIN EFFECTS. The model, with Version, Reading, and Order of Presentation (Conventional first or Modified first) as categorical factors, and using Piece as Offset Variable, showed a marginal significance for Version [$\chi^2(1, N = 16) = 2.97, p = .085$], high significance for Reading [$\chi^2(2, N = 16) = 52.69, p < .001$], and high significance for Order of Presentation [$\chi^2(1, N = 16) = 38.94, p < .001$].

INTERACTIONS. The interaction of Version by Reading was only marginally significant [$\chi^2(2, N = 16) = 4.64, p = .098$], with the differences between Versions not being significant in the first two Readings, and reaching marginal significance in the third (see *Figure 6.2*). Similarly, the interaction between Version, Reading, and Order of Presentation reached only marginal significance [$\chi^2(1, N = 16) = 2.97, p = .085$], with the patterns for both orders being somewhat similar (see details below, SECTION 6.3.1.c., and *Figure 6.4*, for Totals of Mistakes, which presented the same trends).

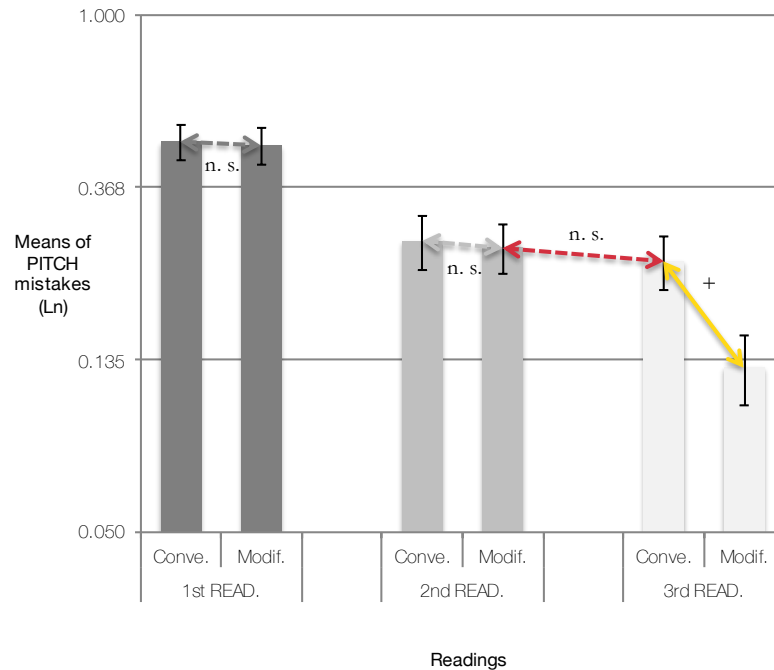


Figure 6.2.

Pairwise Comparisons of Estimated Means of PITCH Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings. The vertical axis represents logarithmically transformed values (Ln), having used the Ln of the difficulty of each Piece as an offset variable.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

+ $p < .100$

6.3.2. RHYTHM DOMAIN RESULTS

As was the case with the pitch domain results, a model with Version (Conventional or Modified), Reading (1st, 2nd or 3rd Reading), and Piece (Piece 1 - g minor; Piece 2 - b minor) as categorical factors found a highly significant effect of Reading [$\chi^2(2, N = 16) = 18.68, p < .001$] and a significant effect of Piece [$\chi^2(1, N = 16) = 4.53, p = .033$], with Piece 1 eliciting in this case more rhythm mistakes than Piece 2 [$M = 2.56, SE \pm .28$, *vs.* $M = 1.92, SE \pm .27$, respectively]. The effect of Version was nonetheless highly significant [$\chi^2(1, N = 16) = 15.83, p < .001$].

A second Model was implemented, using Piece as an Offset Variable, and with Version, Reading, and Order of Presentation (Conventional first or Modified first) as categorical factors.

MAIN EFFECTS. The model showed a high significance for Version [$\chi^2(1, N = 16) = 16.17, p < .001$], high significance for Reading [$\chi^2(2, N = 16) = 17.52, p < .001$], and high significance for Order of Presentation [$\chi^2(1, N = 16) = 15.64, p < .001$].

INTERACTIONS. The interaction effect of Version by Reading was non-significant [$\chi^2(2, N = 16) = 3.67, p = .159$], with the differences between Versions showing the same patterns in the three Readings, with fewer mistakes elicited by the Modified Versions (see *Figure 6.3.*). The interaction between Version, Reading, and Order of Presentation showed no significance [$\chi^2(1, N = 16) = 1.42, p = .922$], with the patterns for both orders being quite similar (see details below, SECTION 6.3.1.c., and *Figure 6.4.*, for Totals of Mistakes, which showed the same trends).

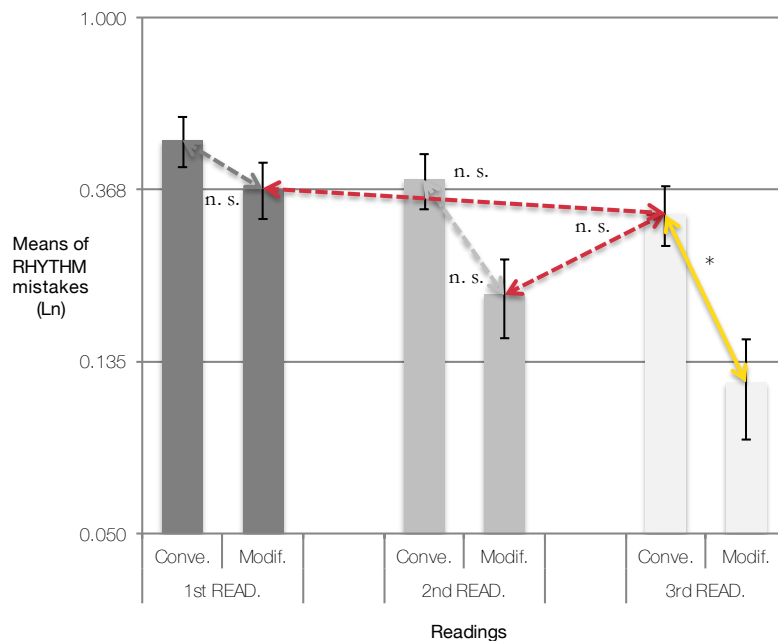


Figure 6.3.

Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings. The vertical axis represents logarithmically transformed values (Ln), having used the Ln of the difficulty of each Piece as an offset variable.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

* $p < .050$

6.3.3. TOTAL NUMBERS OF MISTAKES RESULTS

Similarly to what had been found in the pitch domain and the rhythm domain results, a model with Version, Reading, and Piece as categorical factors found a highly significant effect of Reading [$\chi^2(2, N = 16) = 66.41, p < .001$] and a highly significant effect of Piece [$\chi^2(1, N = 16) = 19.17, p < .001$], with Piece 1 eliciting smaller total amounts of mistakes than Piece 2 [$M = 5.10, SE \pm .53, vs. M = 7.58, SE \pm 1.01$, respectively]. The effect of Version was nonetheless significant [$\chi^2(1, N = 16) = 11.46, p = .001$].

A second Model was implemented, using Piece as an Offset Variable, and with Version, Reading, and Order of Presentation as categorical factors.

MAIN EFFECTS. The Model showed a highly significant effect of Version [$\chi^2(1, N = 16) = 15.80, p < .001$], Reading [$\chi^2(2, N = 16) = 68.95, p < .001$], and Order of Presentation [$\chi^2(1, N = 16) = 52.94, p < .001$].

INTERACTIONS. The interaction effect of Version by Reading was significant [$\chi^2(2, N = 16) = 8.55, p = .014$], with the differences between Versions being much more marked in the third Reading (see *Figure 6.4*). The interaction between Version, Reading, and Order of Presentation was non-significant [$\chi^2(1, N = 16) = 5.31, p = .379$], with the patterns for both Versions being quite similar in all Readings: tests starting with a Modified Version drew out higher total numbers of mistakes than tests starting with a Conventional Version; however, in the case of conventional scores this difference was significant (highly significant in the 2nd Reading), whereas the modified score was more resilient to order and the differences were not significant (except for a marginal significance in the 1st Reading —see *Figure 6.5*).

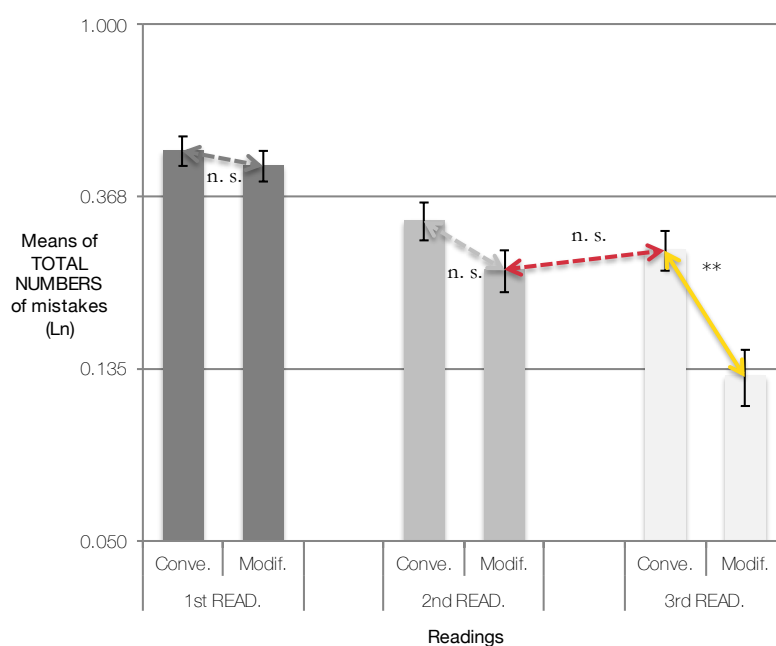


Figure 6.4.

Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). From left to right, comparisons of mistakes made in the 1st Readings, 2nd Readings and 3rd (rehearsed) Readings. The vertical axis represents logarithmically transformed values (Ln), having used the Ln of the difficulty of each Piece as an offset variable.

Error bars represent Standard Error.

Significance using the Sidak correction:

n. s. non-significant

** $p < .010$

ORDER EFFECTS Although the general interaction of Version and Order of presentation was non-significant [$\chi^2(1, N = 16) = 2.19, p = .139$] (as both Versions were affected by order in the same manner —see Figure 6.5.), pairwise comparisons using the Sidak correction showed that the Order of presentation was significant when looking at each Version separately. The effect was highly significant for the Readings of Conventional Versions [$M = .23, SE \pm .02, vs. M = .52, SE \pm .03$, for tests respectively starting with a Conventional or a Modified Version; $p < .001$], and also significant for the Readings of Modified Versions [$M = .18, SE \pm .02, vs. M = .31, SE \pm .02$, for tests respectively starting with a Conventional or a Modified Version; $p = .001$]

The three-way interaction between Version, Reading, and Order of Presentation was also non-significant in general terms [$\chi^2(4, N = 16) = 3.69, p = .449$] (as all Readings of both Versions were affected in the same manner —see Figure 6.5.). Pairwise comparisons using the Sidak correction showed, however, that the Order of presentation had a significant effect when looking at the Readings of each Version

separately. The effect was particularly marked in the Readings of the Conventional Versions, and reached significance in the 1st Readings [$M = .35, SE \pm .04, vs. M = .66, SE \pm .06$, for tests respectively starting with a Conventional or a Modified Version; $p = .010$], high significance in the 2nd Readings [$M = .18, SE \pm .03, vs. M = .55, SE \pm .06$, for tests respectively starting with a Conventional or a Modified Version; $p < .001$], and also significance in the 3rd Readings [$M = .18, SE \pm .03, vs. M = .39, SE \pm .05$, for tests respectively starting with a Conventional or a Modified Version; $p = .042$].

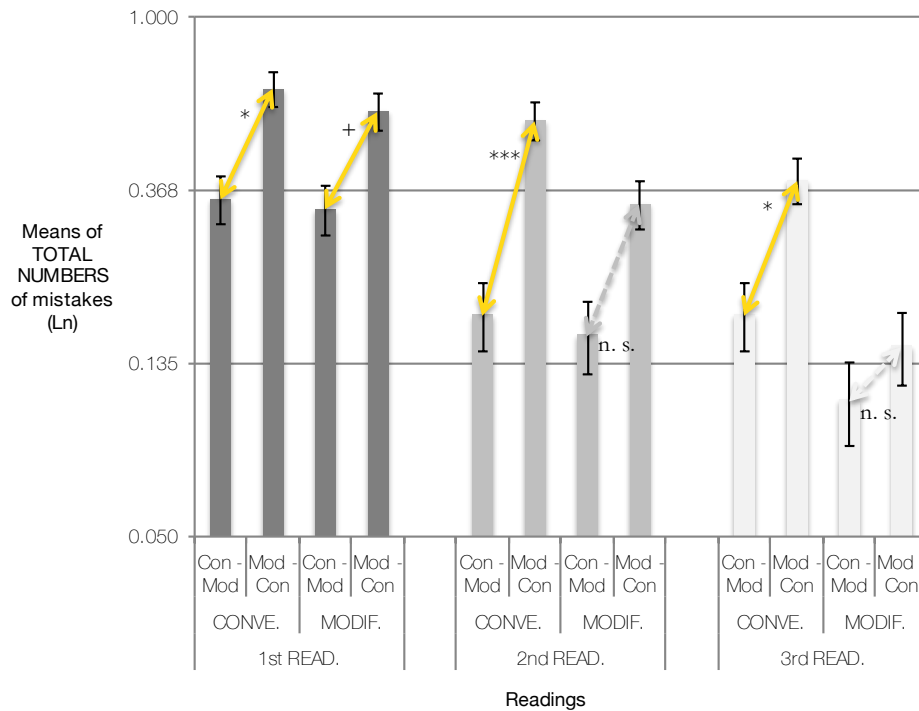


Figure 6.5.

Pairwise comparisons of Total Numbers of mistakes for Readings using Conventional (Conve.) or Modified (Modif.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then Conventional). The vertical axis represents logarithmically transformed values (Ln).

Error bars represent Standard Error.

Significance using the Sidak adjustment for multiple comparisons:

- n. s. non-significant
- + $p < .100$ (marginal)
- * $p < .050$
- *** $p < .001$

6.3.4. ASSESSMENT QUESTIONNAIRES

Following the protocol established at the beginning of the session, two participants decided not to answer questions on preferences. Of the fourteen that answered, a slight majority (8 participants) preferred the

Modified Version; two of them saw advantages for both Versions, and only four stated a preference for the Conventional Version —and all four mentioned that this preference was due to ‘just being used to it’.

The scatterplots in *Figure 6.6.* show the variability of preference (Conventional =1, Equal =2, or Modified = 3) in relation to three covariates: age of participant, number of years studying the violin, and tempo at which the Pieces were read. The students formed a rather homogeneous group in terms of age: all were between 18 and 21 years of age, with only one outlier at 24 (results without the outlier were very similar) and the analysis shows a flat response throughout the range, with a general slight preference for the Modified Version [$y = 2.29 + (1.23E-17) * x$]. The number of years studying the instrument showed more variability (from 10 to 21 years), and also a positive correlation with preference for the Modified Version [$y = 1.2 + 0.08 * x$]. Finally, the covariate most directly related to the reading proficiency of the participants, the tempo of performance at the test (between 78 bpm and 108 bpm)¹ showed a strong positive correlation with the preference for the Modified Version [$y = 3.69 + 0.06 * x$].

¹ The tempo was adjusted to match individual sight-reading skills. See pre-test procedure, SECTION 6.2.3.

Table 6.1.

Evaluation by Participants: Preferred Versions and Comments

Par.	Ver.	Comment why
1	M	Rhythm was more clearly readable, although space in music has a temporal element so confusing in that respect
2	M	I found the modified piece of music easier to play (also suspect that it [the piece] could actually be easier/less difficult?)
3	C	Because I'm used to it. Having said that, I didn't find them very different —I could feel my eyes skipping forward in the same way.
4	E	Indifferent —extra space was good for clarity but also added confusion because different to normal
5	C	Used to it, found the spacing in modified a bit confusing but I may have gotten used to it, who knows?
6	C	Probably just through experience —I've been reading normal notation for 15 years so departing from that is unusual. Would be interesting to see how it would develop with experience of the system.
7	C	Probably just because it's what I'm used to. I found it difficult to keep counting in time when the notes were split up.
8	M	Rhythm clearer, although it takes a little longer to read bars 'coming up' due to spaces.
9	M	Having the separation provided structure that you don't have time to formulate for yourself when you're sight-reading.
10	M	Easier to see each phrase/group of notes so not overwhelmed by whole extract with lots of notes
11	M	More space to read! Could separate it out a lot more easily.
12	M	There was more space so it was easier to read + process.
13	E	I found both easy to read except in the modified one I occasionally lost the sense of time signature
14	M	More space, broken up in chunks to understand phrase better —less daunting

C = Conventional better; M = Modified better; E = Equal, saw advantages for both.

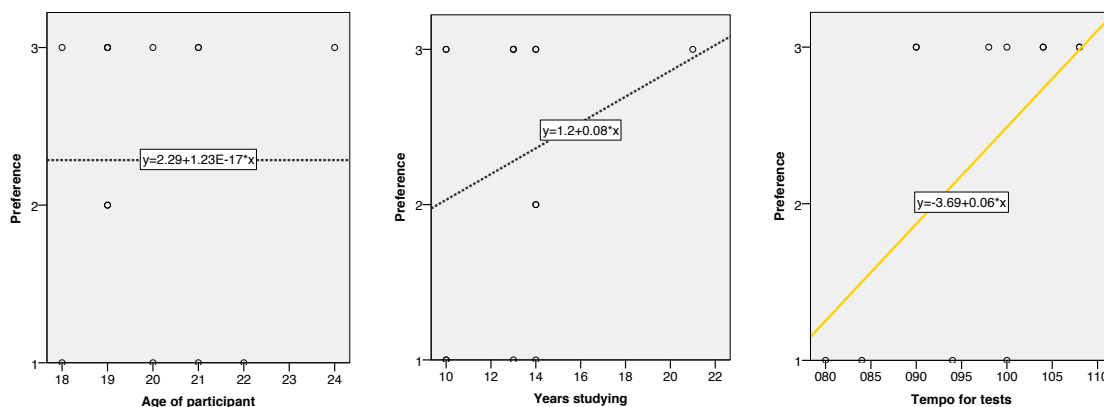


Figure 6.6.

Simple scatterplots showing the mapping of the participants' preferences (1= Conventional; 2 = Equal; 3 = Modified) to different covariates. From left to right, the relation to the Age of the Participants (from 18 to 24 years), to the number of Years Studying the violin (from 10 to 22 years), and to the Tempo at which they played in the Tests (from MM 078 to MM 108).

6.4. MODIFIED SCORES AND NOTATION FAMILIARITY

6.4.1. FEWER MISTAKES, HIGHER RESILIENCE, AND BETTER EVALUATION WITH MODIFIED VERSIONS

Sixteen violinists—all students at the University of Cambridge, in various faculties— participated in a music reading experiment, with a design very similar to previous experiments with percussionists (see particularly SECTIONS 5.2. and 6.2. for comparison of methods). Participants read short pieces of music (unknown to all participants, written expressly for the experiment) presented either in conventional format following the default settings of an industry standard music editing programme ('Sibelius' 7.1), or in a modified version that included visual cues following a structured spacing of the discourse, hypothesized to enhance readability. Measurement of reading mistakes showed smaller numbers for performances with the modified scores, with differences with the performances using conventional scores growing after each reading, and becoming statistically significant in the third reading (participants read each piece three times). The differences were particularly marked in mistakes in the rhythm domain (as had also been found in the experiments with percussionists), with the third (studied) performances with the conventional scores eliciting numbers of mistakes that were not (even marginally) significantly different from the numbers of mistakes incurred in the first (sight-) reading of the modified versions.

On top of this, analysis of the effect of order of presentation on the performances revealed the modified versions to be more resilient to context than the conventional versions (see SECTION 6.3.3. and *Figure 6.5.*), with participants performing with significantly more mistakes when using a conventional score and having to read it immediately after a modified one.

Students evaluated the modified scores in a generally positive manner, although the variability in the evaluation was shown to be significantly affected by the expertise of the participants, with the more proficient readers evaluating the modifications more positively. In fact, the only participants that evaluated the conventional scores as clearer could only adduce one—not to be underestimated—argument: that they were more *used* to it (see Table 6.1., participants 3, 5, 6, and 7). In a study of sight-reading in pianists, Sloboda *et al.* [1998] put forward the notion that certain technical and idiomatic choices made by musicians (specifically in their study: fingering decisions) are dependent on overlearned, rule-governed response sequences triggered by *familiar* visual patterns within a score.

6.4.2. OVERRIDING FORMAT FAMILIARITY

This contention could seem to resonate with studies (with language scripts) showing that format familiarity exerts a controlling influence on the coordination of certain physiological movements (eye and

head) which reading relies upon [C. Lee, 1999; Seo & Lee, 2002]. However, studies in selective impairment of musical faculties have shown that certain music reading abilities *can* be dissociated from language reading strategies, and therefore text familiarity could be acting upon musicians through different, less constraining avenues. Cappelletti *et al.* [2000] studied the case of a professional musician with very localized brain lesions after an episode of haemorrhagic encephalitis, which impaired her ability to read musical notes on the staff whilst still being able to remember and play both familiar and new melodies. The fact that musical abilities, including musical learning, can rely efficiently on presentations independent of visual input could mean that text format familiarity *could* also be overridden as a crucial factor for musical performance. This resourcefulness or ductility of musical learners could correspond with the increased benefits of the modified score in the second and particularly the third reading in our experiment (when they had had a few extra private study minutes to survey and familiarise themselves with the scores). Music reading might be less dependent on the familiarity of text format than language reading, since musicians rely probably more than language readers on acoustic and motoric resources than on conceptualising or naming systems, which could be more relevant in language reading².

In fact, music reading can, in certain cases, be completely independent of naming systems. Bevan *et al.* [2003] studied a patient with selective impairment in naming musical notes that nonetheless preserved the capacity for *instrumental* reading of notes. Music reading can thus rely on motoric encoding and instrumental conventions to a degree that could not be feasible in language reading. It is therefore likely that, above general format familiarity, it is idiomatic, motivic, or contour familiarities that could play a more relevant role in music reading —particularly, as is mostly the case, and contrary to what holds for the language domain, if the reading is performance oriented. Presentation formats that enhance the recognition of and familiarisation with these relevant internal elements of the discourse, independently of the general familiarity of the layout, could therefore be a facilitating tool for music readers. Sloboda *et al.* [1998] have proposed that responses to score materials rely upon the recognition of familiar visual *patterns* within the score.

But the familiarity of the format of presentation can also be overridden by language readers, who arguably rely more consistently on visual presentation than musicians, if the modifications introduced afford sufficient cognitive advantages to compensate for the detrimental effect of unfamiliarity. Studies with Asian (conventionally unspaced) languages are particularly relevant in this respect, since in them the (extremely recent and tentative) introduction of structured spacing is —as in our scores— acting against presentation conventions. These conventions are thoroughly established in certain Asian languages (for millennia, in some cases), and have only in the last years/decades been challenged by reading researchers. An essentially ideographic system of writing existed in China, for example, probably by early in the IInd

² Musicians, in this sense, could be more prone to rely on working memory than on general prior knowledge (see SECTION 4.1.1.)

century *A. C.*, with each character of the script representing a single monosyllabic concept [see Bent *et al.*, 2001, for details, and for chronologies of other Asian languages]. However, in experiments monitoring eye movements in native Chinese readers, Bai *et al.* [2008] found that texts modified to include spaces between words yielded reading times similar to those elicited by conventional unspaced texts. The authors' wording of the matter is also of relevance to our present discussion: "[...] sentences with *unfamiliar* word spaced format were as easy to read as visually *familiar* unspaced text." (p. 1278; italics mine). It is also worth noting that it was *structured* spacing which yielded these results, since sentences that included uniformly added extra character spacing or that included randomly assigned spacing produced longer reading times. Based on these results, and supported by similar findings in an experiment using other graphic separation devices (highlighting) as structuring tools, Bai *et al.* [2008] go as far as to postulate that "[these] experiments *clearly* indicate that words, and not individual characters, are the unit of primary importance in Chinese reading" (p. 1278; italics mine).


6.4.3. ADVANTAGES OF STRUCTURED TEXTS FOR NON-NATIVE READERS

Familiarity with graphic spacing cues could thus be of secondary importance even for language readers, particularly if the familiarisation with the presentation format is of a more generic type, and not confounded with familiarisation with recurring lexical or visual patterns within the discourse. Shen *et al.* [2012] examined the effect of word spacing for four groups of non-native readers of Chinese (all learners of Chinese as second language). Their conditions were the same as in the study by Bai *et al.* [2008]: unspaced text, word-spaced text, character-spaced text, and random ('nonword') spaced text. Eye movement measures showed least disruption to reading for word-spaced text, followed by character-spaced text, then conventional unspaced text, and finally nonword-spaced text, which yielded most disruption. The results are similar to those obtained by Bai *et al.* [*id.*] in that uniform merely sequential spacing and non structurally relevant spacing did yield worse results than spacing at word level. However, the results here also show *better* performances with the spaced texts than with the conventional texts, whereas for native speakers their levels were similar. Shen *et al.* [2012] venture that the demarcation of discourse units through spacing reduces non-native readers' uncertainty about (or unfamiliarity with) the characters that constitute such a unit, thereby speeding identification and, in turn, reading. But most importantly, the effect of introducing word spacing was uninfluenced by native language. This is of particular relevance for our discussion since the readers were native speakers of four languages with saliently different formats of presentation: English, Japanese, Korean, and Thai. This selection was particularly fortunate, since English and Korean are spaced languages, whereas (ideographic) Japanese and Thai are unspaced. Within this fundamental distinction, there are further differences between the languages: although English and Korean are both alphabetic, the set of symbols and orthographic rulings that they use are markedly different; and even though both Japanese and Thai are unspaced, Japanese is

character based whereas Thai is alphabetic. In sum, it seemed that the general familiarity with unspaced text as a native user thereof did not hinder the facilitating effect on reading fluency and speed yielded by the spacing of discourse units at a structured and relevant level, even in a normally unspaced language. In this sense, not only musicians, but language readers as well might be capable of overriding the destabilising effect of an unusual format of presentation and benefit from the structuring, guiding and legibility facilitating effects of novel spacing rules.

ACKNOWLEDGEMENTS

The experiment with violinists was implemented by two of my students at the 'Perception and Performance' Course in the Faculty of Music Tripos: Maya Amin–Smith and Guy Edmund–Jones. They took care of the recordings and the coding of the performances; the experimental design and materials were my contribution; the statistical models, their discussion, and all the figures used in this chapter are also my own. The students used different, simpler, statistical models (implementing more commonly used distribution assumptions and functions). Although their general conclusion (that the novel scores elicited better and more stable performances) was the same as mine, the statistical significance of their analysis was weaker, and they did not include a detailed evaluation of order effects and interactions. The student M. Amin–Smith submitted her work on this experiment as a project for the Part II course; although she did an excellent work in recruiting participants and carefully implementing the experiment and compiling data, her contribution can not be understood to constitute a 'substantial part' of this PhD dissertation. I am also grateful to Prof. Sarah Hawkins for the supervision and guidance of this particular project.



*Experiment V: Separating
the Design Elements
in the Modified Scores*

CHAPTER 7

Experiment V: Separating the Design Elements of the Modified Scores

7.1. INTRODUCTION: INTEGRATION AND INDIVIDUALIZATION OF DESIGN COMPONENTS

Based on the results from previous experiments at the *Conservatorium van Amsterdam* and the University of Cambridge, the next step was to plan an experiment directed at clarifying *which* of the components of the modified designs used in those experiments were more influential on the (in many cases highly) significant results found. The main feature of this new experiment would be that it should be divided into several exercises, each of them measuring the effect of one design novelty only. The differences in performance for the various versions of scores to be used were expected to be subtler to those previously found (since the possible cumulative effect of the integration of different novelties would be lost). As a consequence, I prioritised increasing the number of participants and also polishing the design of the materials so as to reduce the effect of any unwanted variability. In this sense, the use of different tiers of difficulty (exercises with 1, 2, 3 or 4 voices, using the same basic materials) was hypothesized to be a helpful tool towards the inclusion of a wider range of players. The materials were to be based on J. S. Bach's chorales, following the previous experiments, but the choice of specific fragments, the transcription process, and the standardisation of the difficulties of the pieces would have to work towards the minimisation of unwanted variability.

7.1.1. INTEGRATION OF DIFFERENT DESIGN NOVELTIES IN PREVIOUS EXPERIMENTS

In previous experiments, modified versions of musical scores showed a significant decrease in number of errors when participants sight-read certain musical pieces. The modified designs included changes in page

orientation, layout and spacing, all product of an integrated attempt to improve the readability of the scores. The two last experiments conducted (Experiments III and IV) reduced the number of modifications included in the novel scores by using the same orientation (portrait) for both the conventional and modified versions of the scores, and by including articulation marks in the conventional scores that could compensate for the possible advantages that the readers could be gaining from the more segmented presentation in the novel scores. The results of these experiments showed similar levels of significance for the differences in performance as the previous ones. To clarify to which extent each of the remaining design modifications did affect the results, a new experiment was designed, in which each main modification was individually tested against a conventional score.

Four experiments had been conducted so far at the *Conservatorium van Amsterdam* (Experiments I, II, & III) and at the University of Cambridge (Experiment IV). The design and method of the four were similar in some aspects, mainly in featuring a modified design for the musical scores to be read, in which *several* changes in different parameters of the layout and spacing of the materials were included. This was done to be sure that there were indeed differences in performance quality between the readings of the conventional and the modified scores. Differences in performance for short and idiomatically homogeneous pieces were expected to be subtle, and it was therefore decided in these experiments to look at the effect of a modified design (a whole set of rules including changes in several parameters) as a single factor that would—or not—influence performance. Another reason for changing several parameters at the same time in the modified scores was the fact that most design decisions at a certain level of the structure of a score (or in any other object, for that matter) are interdependent with decisions taken at different levels of the structure. The novel designs chosen included therefore major and minor changes, some of them more apparent than others. The four more prominent changes were:

1. Phrasing divisions. Each line of the musical text contained only one phrase, and nothing more.
2. Sub-divisions. Smaller units within each phrase were marked with short white gaps in the staves.
3. Proportional notation. Distance between notes is used as a digital cue, with a constant measure between notes of the same rhythmical value.
4. System distances. The distances between staves that did not form a unit (the notation unit formed by the right and left hand staves) were incremented. They were incremented even further when marking a major structural division in the text.

Other minor changes, that were a consequence (in some cases unavoidable) of the previous ones, included:

5. Justification. Only left side justification was used, with no attempt to fill up each line of the text. This is actually an unavoidable consequence of the combination of conditions 1 and 3 mentioned above.
6. Indentation. Since one of the aims of the use of condition 3 was to make the material more predictable in terms of location of the signs, the customary indentation of the first line of text was avoided, as it would displace the whole of the line.
7. Alignment. Further, in order to ensure the regularity and alignment of the materials — factors which were supposed to increase predictability and therefore readability— small adjustments were included in each line (*e. g.*, hidden time signatures), so that the notes of a given line would start always at exactly the same distance from the left margin of the page. This is also a consequence of proportional notation and, even if not a necessary or unavoidable one, it does add to the purpose of proportionality.

7.1.2. RATIONALE FOR EXPERIMENT V

A. MAIN AIM

The main aim of the present experiment was to look at the effect that each of the components of the novel design (each of the novel conditions listed) would have *on its own* on the performance of a short piece, when compared with a performance using a conventional score.

B. METHODOLOGICAL CONSIDERATIONS

Of the conditions listed above (SECTION 7.1.1.), only the four main components (phrasing divisions, sub-divisions, proportional notation, system spacing) of the novel design should be tested. Testing more than four design novelties would result in an unmanageable experiment, both logistically and from the methodological point of view. Since system separation has the least consequences onto other levels of structure (at least for short fragments), and can therefore be considered a relatively independent component, it was decided to only test phrasing divisions, sub-divisions, and proportional notation, all of which might have had interaction effects in previous experiments.

The conditions of justification and forced alignment are a consequence of the previous (main) ones, and can easily be included in the design without being considered a separate factor (see materials in the APPENDICES)¹; more importantly, these components of the design were always used in previous experiments, and the aim of the planned experiment was to reproduce the materials of those experiments but using only one condition (and its related consequences) per exercise.

¹ For example, in 7.A.2., when including only one phrase per line, and not wanting to introduce a different spacing between notes in relation to the Conventional Version (7.A.1.), the right hand margin must necessarily be unjustified.

C. DURATION OF THE EXPERIMENT

The experiment was divided into three short exercises, each measuring the effect of one design novelty only. The exercises were short, since sessions of more than 60 minutes per student could not be expected to be logistically possible. In Experiment III sessions of 45 minutes were required, and several students that had committed in Experiment II for a 30-minute session faltered in their interest, and excused themselves adducing other engagements. Also, priority should be given to including as many participants as possible, and with the limited time available (due to my personal travel limitations and to the limitations in available studio time, for instance), very long sessions with each participant would go against the objectives.

D. LOGISTICAL CONSIDERATIONS

Since the differences in performance were expected to be smaller in this experiment, an increase in the number of participants was attempted, to be able to achieve statistical significance. This could be carried out by changing the academic framing of the experiment, and by widening the range of participants.

Along these lines, it was decided to incorporate students from other conservatoires. The percussion departments of the conservatoires in Barcelona and Saragossa had shown willingness to participate. Extending the study to other conservatoires where the level of sight-reading proficiency might be different to the one already measured in Amsterdam should not be an unsolvable problem if the materials could be adapted to match.

E. USE OF TIERS

All previous experiments were designed as sight-reading tests. One of the keys to the completion of these, and to the harvesting of useful data, was the use of conditions that generated the right amount of mistakes: not too many to cancel out a performer, and enough for statistically significant differences to be detected (highly significant, in fact, in Experiment II & Experiment III). This was achieved by using individualised tempi for each participant, so that the scores could be played by music students of different levels of reading proficiency. In spite of this, several ceiling effects appeared in Experiment I and Experiment II. In the case of Experiment I they were frequent enough to affect the significance of the statistical tests.

A relevant methodological finding in Experiment III was the use of two tiers of participants, which enabled the possibility of having the more competent participants confronted with materials (conceptually similar, but more difficult in terms of texture) that would elicit performance errors even from the most experienced of them. This was achieved by using the same pieces with two voices for the less proficient readers and with three voices for the more experienced ones (see examples in APPENDICES 5.A .and 5.B.).

The use of tiers was in fact expanded in this experiment, but the circumstances dictated that in the end all participants used scores for tier 2 (two voices). The sight-reading level of the participants was homogeneous, and not very advanced —therefore the scores prepared with 3 and 4 voices were never used.

7.2. METHOD: THREE EXERCISES MEASURING EFFECT OF INDIVIDUAL DESIGN COMPONENTS

7.2.1. PARTICIPANTS

Twenty-one percussion students (two female), from all degree years (from Probationary to Masters), recruited from the *Escuela Superior de Música de Aragón*, in Saragossa, and from the *Escuela Superior de Música de Cataluña*, in Barcelona. All participants understood and followed the protocols, and no data had to be discarded.

7.2.2. MATERIALS

A. CHOOSING THE PIECES

All previous experiments had used transcriptions of chorales by J. S. Bach. The chorales are fairly regular in their extension, level of difficulty, idiomatic resources and instrumental range. This, added to the very consistent harmonic grammar underlying their discourse, makes them ideal for the design of these experiments, in which several Pieces with a similar level of difficulty, length and range have to be used. In the experiment designed here, as many as 12 short similar fragments were needed: each of three conditions (modifying phrase spacing, modifying sub-phrase spacing, modifying character spacing) needs two different Pieces (of very similar difficulty, length and extension), and an equal number of Pieces is needed for the pre-test.

IMPORTANCE OF THE PRE-TEST. The evolution of the pre-test, from Experiment I where it was carried out with one-phrase fragments just chosen randomly from available materials, to Experiment III where it was almost identical to the test, just slightly shorter, had been one of the factors contributing to the fine-tuning of the individual tempi and, ultimately, to the avoidance of floor and ceiling effects. In Experiment V, where the differences in performance between Versions would be more subtle, the adequate adjustment of tempi and different materials (since different tiers were initially contemplated) to the performers' capacities was deemed particularly important.

HOMOGENEITY AND MUSICAL DENSITY. In order to produce scores that would achieve the same effect in this experiment, the same source —J. S. Bach's chorales— was used as basis for the materials. This source has three features that can make the scores particularly apt for a sight-reading test: (1) it has numerous examples of pieces that have a similar texture, harmonic rhythm, range and voice leading; (2) the underlying grammar is relatively dense and regular, and in the space of a few bars complete harmonic

gestures can be detected at both phrase and sub-phrase level; (3) the pieces have a discursive nature that almost precludes repetitions *within a stanza* (one round of text; the music will usually be repeated two or more times over different *stanze*). Other *corpora* of relatively homogeneous pieces, or where at least several pieces with similar features could be found, like for example the collection of Sonatas by Scarlatti or the *Lieder* by Schubert, would require a longer selection from a piece for so many complex and rich harmonic cycles to be completed, and they include much more motivic or melodic repetitions, which are less suitable for a sight-reading test.

SEARCH FOR FRAGMENTS. J. S. Bach's chorales have been transmitted mainly in collections. The first collection was assembled by C. Ph. E. Bach and published in Berlin in 1765–69 by F. W. Birnstiel, and included 200 chorales in two volumes; after the success of this edition, C. Ph. E. Bach prepared a second one in 1784–87 in co-operation with the publisher J. G. I. Breitkopf in Leipzig, which included 371 chorales in four volumes. This formed the basis for most collections or re-editions to follow, including the most recent ones. All collections stemming from these sources (which are the most extensively used ones, like the K. Schubert 1990 edition for Breitkopf) tend to be in landscape format, and do not include the text of the chorales. For the purpose of our next experiment, though, the 1912 B. F. Richter edition for Breitkopf was deemed more useful, since it is one of the few to be presented in portrait, includes the text, and can be considered ecologically valid due to the backing of a major publisher like Breitkopf. This edition, which stems from the XIXth-century research of the Bachgesellschaft, includes 389 chorales, with and without *obbligato* instruments. The portrait format makes it valid for a comparison with the modified scores without having to change the page orientation, and the text helps to decide on the sub-divisions of the phrases without interruptions of the original word-setting.

CONSISTENCY AND VALIDITY. Apart from the important criterion of selecting consistent materials that can be factored out in a statistical analysis that will be looking at small differences in Estimated Marginal Means between performances, perhaps even more important is the fact that the criteria followed for the choice of fragments match the ones followed in all previous experiments: fragments based on 4-voice originals; in 'C' time signature; with phrasing based on anacrusis; with a regular number of beats per phrase; and giving the (more experienced) participant the possibility of detecting a teleological intention in the discourse with the presence of clear cadences at the endings.

CRITERIA FOR EXPERIMENT V. For this particular experiment, fragments that were similar or almost equal according to certain criteria were chosen. To this end, a listing of the chorales of the Richter edition (see document 'List of Chorales Ed. Richter', APPENDIX 7.G.) was created, in which the 389 chorales (214 if repetitions of melodies are not counted) were surveyed with the following criteria: 4-voices (1 = 4 voices; 0 = 5 or more voices); Time signature (1 = C; 0 = others); Anacrusis (1 = starting with upbeat; 0 = starting

directly); Regularity (1 = regular phrases with same number of beats; 0 = irregular phrases); Final chord denotes Chord I of the tonality of the chorale (1 = Chord I; 0 = others).

CHOSEN FRAGMENTS. Only chorales that matched all five of the just mentioned criteria were chosen. In these, the three last phrases were selected—in previous experiments, it was observed that the first phrases of a chorale, which are used for establishing or grounding the tonal framework, are usually of a more predictable character than the following ones, and tend to not to produce as much useful data as subsequent and less stable phrases. Since time constraints are one of the keys for the sessions of this experiment containing three individual exercises, it was decided to select only the three last phrases of the pertinent chorales.

ECOLOGICAL VALIDITY. I decided to transcribe and slightly modify chorales by Bach, rather than to write new pieces, for two reasons: (1) to use materials that are similar to the ones used in previous experiments; (2) to address possible criticisms in terms of the ecological validity of the materials. It could be adduced that the differences in performance drawn out by the modified scores become a product of the musical idiom, or of the musical design and gestures. By using very well known and widely circulated musical materials (albeit not so much amongst percussionists as amongst other instrumentalists and singers) any concerns about a possible bias in the design of the music itself that would favour a more segmented or analytical approach to the reading task can be at least counterbalanced with the widespread use of the original sources. Furthermore, these chorales are not only well known pieces, but they also have a normative quality in the realms of harmonic writing and voice leading in European tonal music that should temper any suspicions on the pieces being selected from an obscure repertoire to fit the agenda of the hypothesis. With music composed specifically for the experiments these critiques would be more difficult to tackle, and the compositional choices could always be debatable.

B. STANDARDISING THE PIECES

Further changes were introduced in the scores in order to standardise the number of onsets and number of chromatic notes in *pairs of pieces*. After surveying the whole Richter collection, a list of 58 chorales was created, all of which met the conditions stated above, and had furthermore an identical number of 24 beats distributed in the last three phrases. The goal was to identify six pairs of scores that could be used for the three proposed Exercises (two pairs per Exercise, one for the pre-test, one for the test). To this end, the Pieces had to be ordered using the criteria that would reflect their difficulty most accurately: several orderings could be argued for, but it was decided to take the number of chromatic notes as first sorting value, and number of onsets in the Bass line as second (see document ‘Numbers of Onsets and Chromatic Notes in Selection of Chorales’, APPENDIX 7.H.). In a sight-reading situation, more than the total number of events to be tackled, the clarity of the grammar or structure over which the symbols are presenting their discourse (the key or tonal functions, in this particular case) is a more determining factor

of the difficulty for the performer to decode the score. The numbers of chromatic notes in these fragments are a good indicator of the level of tonal stability and predictability of the passages. As for the number of notes in the Bass line, they are a good indication of the density or imbrication of the voices, and it is for the purpose of creating standardised pairs of pieces a more useful value than the total number of onsets (in all voices), since it will make it easier to choose two chorales that have Bass lines with very similar numbers of events. This is advantageous since the Bass line is present in the materials for all four tiers: 1-voice, Bass solo; 2-voice, Bass + Sop.; 3-voice, Bass + Sop. + Alt.; 4-voice, Bass + Sop. + Alt. + Ten. (Although as mentioned, in the end only Tier 2 —two voices— was used for this particular set of participants.)

SELECTION OF 12 CHORALES. Taking into account that some of these chorales were incompatible with others since they had the same main melodic line in the Soprano (they are variations on a same tune), a total of 12 chorales could be selected from the list (see APPENDIX 7.H.).

SIX PAIRS OF VERY SIMILAR CHORALES. Thus six pairs of chorales could be formed: in each pair the number of chromatic notes (and therefore the difficulty for the performer in terms of ‘orientation’) are the same, and the number of notes in the Bass line very similar, thereby needing only minor tweaking in what is harmonically the most important voice. See full examples in APPENDICES 7.A.–7.F.

7.2.3. THE THREE EXERCISES AND THEIR INDEPENDENT VARIABLES

The design of the experiment with three separate exercises is aimed at including only one independent variable in each exercise to focus on (apart from order effects or other unwanted errors). With differences in performance expected to be minor, and a relatively limited amount of participants, the statistical models will not be able to support an extensive array of variables per exercise. In previous experiments the independent variable was ‘type of design’ (‘Conventional’ / ‘Modified’); here, I proposed that the exercises should look at a single component of the modified designs used in previous experiments:

1. Exercise 1: only the component determining the use of one phrase per line is applied
2. Exercise 2: only the component determining the use of gaps between articulation units is applied
3. Exercise 3: only the component determining the use of proportional notation is applied

It was decided to assign a fixed pair of chorales to each pre-test and test. Ideally, the Pieces should rotate through all the exercises, but the possible combinations would be disproportionate to the predicted number of participants, even in the most cautious of set-ups: the four exercises can be presented in 24 different orders; in each of these orders, if the Pieces (4 sets of chorales, assuming a fixed grouping of pre-test and test materials) are to be rotated, that means 24 combinations per order, and a total of 576

combined permutations of exercise order and Pieces used in each exercise. The Pieces have been thoroughly selected and polished to fit the experimental design, so it should be possible to factor them out. Thus, only the order in which the exercises are performed needs to be considered (the order can be affecting ceiling effects, since all exercises are relatively similar), and 24 participants would be an ideal (and feasible) number for the experiment.

A. EXERCISE 1: PHRASE SPACING

Fragments from chorales 260 and 359 (Pair 4) have been chosen for the pre-test, and from chorales 025 and 266 (Pair 3) for the test (see examples in APPENDICES 7.A. and 7.B.). All the shortlisted chorales have the same phrasing structure, so it was not relevant which Pair to choose: these are the pairs left after assigning fragments to all other exercises, for which there were relevant criteria to prioritise the use of certain chorales.

CONVENTIONAL RULES. In conventional score design, the bar is the indivisible unit used to organise the materials that fit into a line of musical text. Breaking a bar over two lines is therefore not an option, regardless of whether this is in concordance with the phrasing of the discourse or not. Another reason for the distribution of phrases and discourse subdivisions over lines in conventional music publishing is that it is considered unacceptable to leave a line incomplete. This leads to the tradition of distributing the material evenly so that it creates full lines from beginning to end of the Piece, irrespective of the actual musical subdivisions of the discourse.

Countless examples can be found in any edition of the collections of chorales by Bach of musical phrases being interrupted by a change of line. The majority of Bach's chorales are anacrusal (268 out of 389, 69%, in the Richter collection), and in these cases it is very frequent to not only divide a phrase between two lines but furthermore to end a line with the first note of a new phrase instead of carrying it over to the next line. In the less frequent cases where the phrasing is not anacrusal, the distribution of phrases tends to make the end of a phrase coincide with the end of a line, but this is by no means a consistent criterion, and there are numerous examples of the contrary. At any rate, anacrusal fragments were used in previous experiments, and in the proposed experiment presented here, I will use similar material, since the aim is to elucidate what factors were mostly influential in the modified designs presented in those experiments. The exact distribution of phrases chosen here is very frequent in the Richter edition: three regular phrases over two lines, with the second phrase spread between the two lines. All the chorales chosen for this experiment (021, 317, 140, 283, 025, 266, 260, 359, 008, 012, 286, 348, 079, 294, 033 & 052), feature this kind of distribution in their last three phrases in the Richter edition.

NOVEL RULES. In our Modified Version, it is the musical phrasing that determines what is included in each line. Therefore, in this exercise, keeping all other parameters alike to the Conventional Version, each line

in the modified Version corresponds with one phrase of the musical discourse. To this end, bars are presented incomplete where necessary, and no attempt is done at filling up a line completely from margin to margin. The width (horizontal space that they occupy) of a phrase in the novel Version is equal to the width of the last phrase in the Conventional one.

B. EXERCISE 2: SUB-PHRASAL SPACING

I use the term ‘sub-phrasal’ for the units of organization next below the musical phrase in rank. It is a purposefully generic term, simply indicating any possible sub-divisions of a musical phrase (see SECTION 2.6 on the complex and debatable relation between musical discourse and lexico-semantic divisions). Also, please note that ‘musical phrase’ would not correspond to what is defined by the term ‘phrase’ in the language domain (a small group of words standing together as a conceptual unit). In its most common musical usage, ‘phrase’ would perhaps be closer to what is defined by ‘sentence’ in linguistic studies (a longer set of words that is complete in itself).

Fragments from chorales 140 and 383 (Pair 2) have been chosen for the pre-test, and from chorales 317 and 021 (Pair 1) for the test (see examples in APPENDICES 7.C. and 7.D.). The criterion in this case was to find fragments that included at least two sub-divisions of phrases (not all do, since they are taken from the endings of the chorales, see below).

CONVENTIONAL RULES. In the case of the phrase, the articulation mark most commonly used for its delimitation is the *fermata*, whereas the sub-divisions of a phrase are either not marked at all (as is the case in all editions of the chorales known to this author, including the Richter–Breitkopf one) or, in certain idioms, signified with a slur. More importantly, and in a similar manner to what was described for the phrases, there is normally a prioritisation in music editing of the use of whole bars and of filling up the lines to the margin over trying to use spacing to separate or group components of the text, also at sub-phrasal level. This means that sub-phrasal units are completely ignored both when distributing the text over lines, and when spacing the text within a line.

One of the reasons for the Bach chorales to be considered exemplar or even normative for many future generations of composers up to our days is that Bach shows a complete command of the idiom that he is writing in: he can play with, transform, or adjust the materials according to his compositional choices, always within an astonishing command of the consistency and the cohesiveness of the results. To what extent this was due to Bach being a great recompiler of a series of traditions that culminate in him, or to his personal genius, or a combination of both, is not very important at this point, but what is relevant is that, in the chorales, there are clear compositional choices by Bach indicating intentional sub-divisions of the phrases, which can by no means be ascribed to randomness or carelessness.

Thus, and certainly in the less dense chorales, it is frequent to find that the first word/s of a phrase are set with a crotchet without any further embellishments (Bach could perfectly include passing notes, changes of position or harmonic adornments, but he chooses not to do so), and that the discourse flows in an uninterrupted chain of quavers after that (*e. g.*, the beginning of the fragment of chorale 286 chosen for our exercises). More in general, as a chorale advances, there is a tendency for the phrases to be more dense and flowing (*i. e.*, with an uninterrupted chain of quavers), whereas the first phrase/s is/are normally more sparse and constituted by shorter sub-phrasal units. For consistency purposes, all the fragments chosen for these exercises are from the final phrases of the original chorales (see SECTION 7.2.2.A., particularly CHOSEN FRAGMENTS), so in a few of them there is not much sub-division of phrases. Surveying the shortlisted pairs of chorales, the more apt for the measurement of the effect of sub-phrasal spacing on sight-reading seem pairs 1 (chorales 021 & 317) and 2 (chorales 140 & 383), in which all of the chorales include at least two places where Bach is clearly indicating with his setting that a sub-phrasal division is intended in the discourse.

NOVEL RULES. Sub-phrasal divisions will be signified, as in previous experiments, by the use of small gaps in the stave. These gaps separate two contiguous notes visually, and also add an extra amount of space between them. The distance added should be the object of research in future experiments to determine its optimal width in various circumstances (idiom, tempo, entrainment with other players, *etc.*). For the present experiment, in order not to add extra factors in the analytical models, and in order to reproduce the design choices featured in previous research, the distance will be of a semiquaver. In the music editing program used for the writing and printing out of the pieces (Sibelius 7), this is achieved by including a hidden semiquaver under the sub-phrasal gaps, and by using hidden time signatures that include the exact number of semiquavers resulting per line.

C. EXERCISE 3: NOTE SPACING

Fragments from chorales 079 and 294 (Pair 7) have been chosen for the pre-test, and from chorales 033 and 052 (Pair 8) for the test (see examples in APPENDICES 7.E. and 7.F.). Supposedly, proportional notation should be helpful for the apprehension of passages that could be rhythmically ambiguous (it could perhaps also enhance the identification of certain recurring melodic contours, but these chorales contain very few cases of recurring motifs), so the criterion was to look for pairs of chorales that would have a similar number of short notes and syncopations.

CONVENTIONAL RULES. The note spacing rule used in the Conventional Versions of the scores will try to imitate as closely as possible the appearance of the Richter edition. Therefore, the distance between figures with the same name (*e. g.*, between quavers) will not be constant nor proportional to the distances between figures of a different rank. Since in conventional editions the traditions of including only full bars in a line, and of justifying all lines both right and left (*i. e.*, filling them up from margin to margin) can

not be broken, this means that the distances between notes have to be adjusted in relation to the density of the material, and can not be kept consistent.

It is perhaps worth noting that the convention of having right hand justification for all musical systems is not to be found in the original edition of the collection of chorales by C. Ph. E. Bach for the publisher Birnstiel in Berlin (1765), where the last line of a piece is only filled as far as it is needed, leaving the rest of the stave unwritten. This was a by-product of the use of fixed types for the setting of the music, which allowed for less flexibility in the distances between notes or, in any case, did not promote a vision of the music setting that used the conceptual value of a figure of a certain rank further than its more or less fixed visual or physical value on the page. As a consequence, and with the exception of the bars where a fermata on a single chord is present (which are reduced to the minimum possible width), the bars in the Bach–Birnstiel edition are far more regular and of a stable width throughout a piece than those in modern editions (*e. g.*, the Richter edition, or the Schubert edition), where the need to fill up the last line leads to stretched or compressed bars according to the circumstances.

NOVEL RULES. In this exercise, a proportional note spacing rule will be used, leaving all other parameters unaltered. This will also mean (necessarily) that no right-hand justification will be possible, since both lines of text do not have the same amount of beats: in all our examples in this experiment, the first line is 13 beats/crotchets long, and the second line is 11 beats/crotchets long.

7.2.4. PROCEDURE

The experimental procedure was approved by the Faculty of Music’s Research Ethics Committee.

Figure 7.1. shows the schematic timeline for the experiment. The order of presentation of the exercises and, within them, the order of Pieces and Versions, were counterbalanced, although not fully, due to the number of participants available.

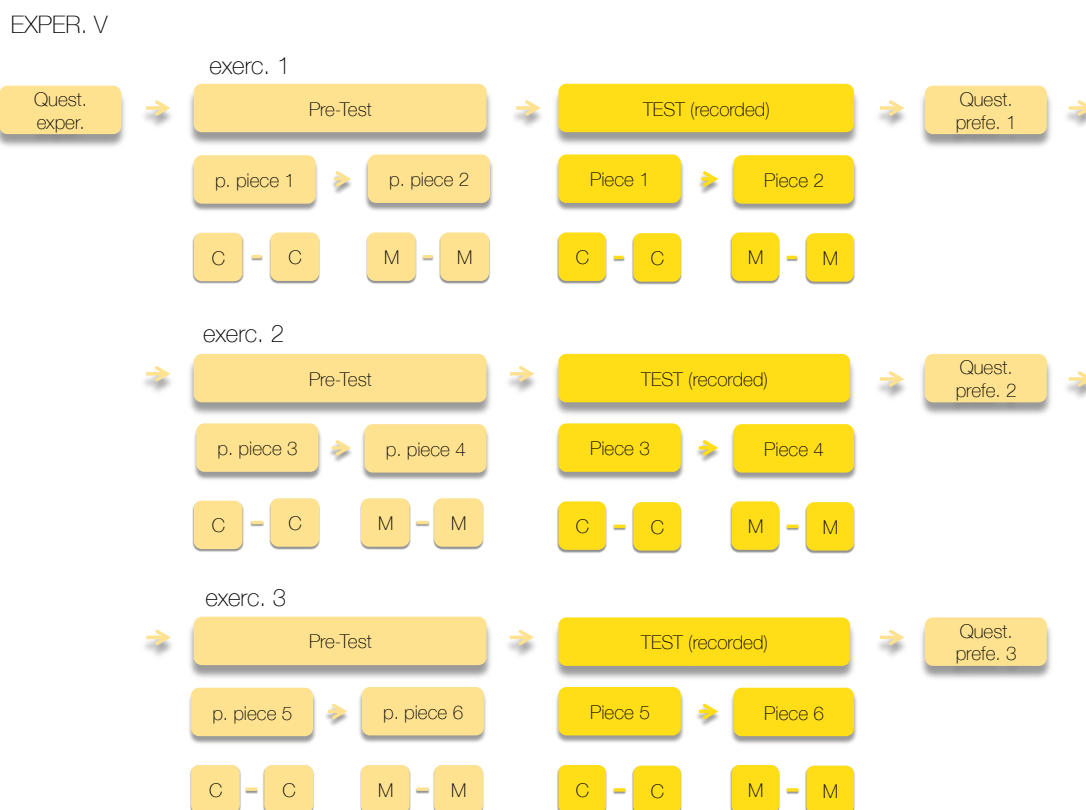


Figure 7.1.

Schematic timeline for EXPERIMENT V: the participant is questioned on musical experience and reading ability predictors, and then completes three reading exercises. Each exercise consists of a Pre-test and the recorded (and coded) Test; at the end of the respective exercise, the participant is questioned on preferences (Conventional or Modified Versions).

Approximate duration: 60 min.

7.2.5. FINAL CONSIDERATIONS ABOUT SPACING

It could be the case that, contrary to what has been hypothesised both in previous experiments and in this one, the main factor increasing the legibility of the novel Versions is not related to the substantiation of underlying grammatical rules or proportions —making more evident the arrangement of and relations between phrases, sub-phrase units, or notes— but more simply to the reduction in visual flanker confusions (a reduction of crowding effects) produced by the separation of the materials. Future research, once it is clearer which of the design components are more relevant, would have to address changes in spacing (at all pertinent levels) that would *not* be related to the clarification of structural units, but that would be added randomly, or even *against* the delimitation of these units. Including these kinds of spacings against the structure as a possible variable in the present experiment would bring more complications (logistics, statistical models, *etc.*; see ‘General design’, p. 5, and ‘Independent variables’, p.

12) than benefits. The design of an experiment including non-structural spacings could benefit from the design of the present experiment, and a series of short exercises could be adapted to its requirements, following the same plan presented here.

7.3. RESULTS: SIGNIFICANT EFFECT OF SEPARATION BUT ONLY AT SPECIFIC LEVELS

Based on comments by several participants (stating that certain keys were more difficult to play for them than others), the means of mistakes per Piece were explored, and notable differences were found (see Table 7.1). Apparently —although not in relation to any specific instrumental idiosyncrasies of mallet percussion (as was confirmed by the teachers)— this set of participants found the Piece in a-minor easier to perform than the Piece in C-Major (Exercise 1, see APPENDIX 7.A.), the Piece in e-minor easier than the Piece in d-minor (Exercise 2, see APPENDIX 7.B.), and the Piece in A-Major easier than the Piece in Eb-Major (Exercise 3, see APPENDIX 7.C.) (see Table 7.1). A recurring comment at the sessions was that ‘music with flats is more difficult’ but, as stated, this is not generalizable, nor can it be related to any limitation or specificity of mallet instruments or their repertoire.

Table 7.1.

Means of Mistakes: Variability per Piece

Exer.	Piece	Key	PITCH		RHYTHM		TOTALS	
			<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
1	1	a-minor	2.85 ± 0.63		1.35 ± 0.33		3.55 ± 0.77	
	2	C-Major	4.45 ± 0.76		2.65 ± 0.51		6.45 ± 1.10	
2	1	d-minor	4.38 ± 0.75		4.08 ± 0.67		7.50 ± 1.24	
	2	e-minor	4.43 ± 0.67		1.73 ± 0.45		5.20 ± 0.91	
3	1	A-Major	3.40 ± 0.65		1.88 ± 0.60		4.88 ± 1.13	
	2	Eb-Major	9.68 ± 1.30		4.48 ± 0.70		13.75 ± 1.68	

Exer. = Exercise

At any rate, this meant that Piece had to be integrated in the analysis as an Offset Variable (with its coefficient not estimated by the model). Since each case would have different levels of exposure to Piece difficulty, in each domain (Pitch, Rhythm, and Total Numbers) the Ln of the difficulty of each Piece (accounted in Means of mistakes) was used as to balance other effects (see also SECTION 6.3.1.). Reliable main effects and interaction effects could hereby be calculated, but for the pairwise or multiple comparisons, Estimated Marginal Means will thus be based on logarithmic values, and therefore rankings or outcomes rather than specific values are reported (as logarithmic values do not immediately help to visualize the differences).

7.3.1. EXERCISE 1: SEPARATION OF LARGE STRUCTURAL UNITS

A. NUMBERS OF PITCH MISTAKES

MAIN EFFECTS. A Generalized Linear Model assuming Poisson Probability Distribution and a Log Link Function, with Difficulty of Piece as Offset Variable, and using as categorical factors Version (Conventional or Modified), Reading (1st or 2nd Reading) and Order of Presentation (Conventional then Modified, or *vice versa*) showed a marginal significance for Version [$\chi^2(1, N = 21) = 3.46, p = .063$], high significance for the effect of Reading [$\chi^2(1, N = 21) = 43.91, p < .001$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = .25, p = .612$]. Contrary to all other experiments, the marginal significance for Version in this Exercise reflected slightly better performances in terms of pitch mistakes with the *Conventional* Version.

INTERACTION. The Model found that the effect of Version by Reading was significant [$\chi^2(1, N = 21) = 9.41, p = .002$], with the 2nd Reading showing much clearer differences between the Versions than the 1st Reading; it was in the 2nd Readings where the Conventional Versions elicited smaller number of mistakes than the Modified Versions (see *Figure. 7.2.a.*).

B. NUMBERS OF RHYTHM MISTAKES

MAIN EFFECTS. A similar Generalized Linear Model (see SECTION 7.3.1.A.) showed high significance for Version [$\chi^2(1, N = 21) = 13.26, p < .001$], a significant effect of Reading [$\chi^2(1, N = 21) = 3.93, p = .047$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = 1.08, p = .298$]. The significance for Version reflected better performances, in terms of rhythm mistakes, with the *Conventional* Version.

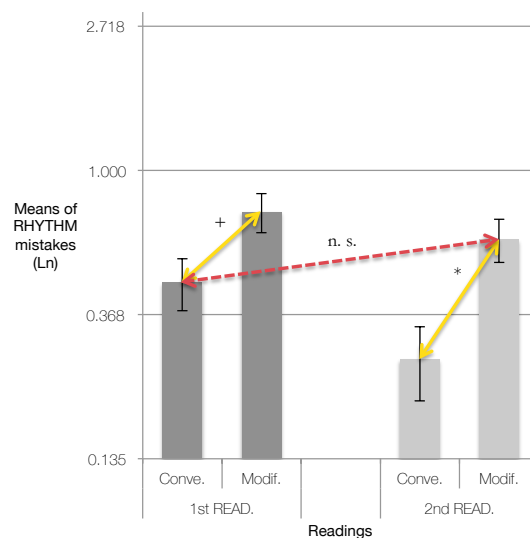
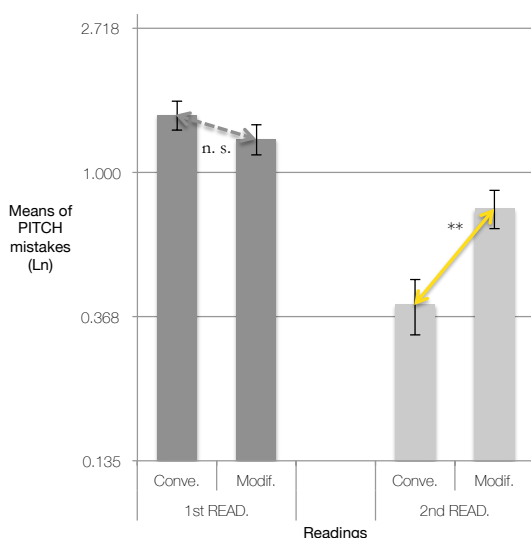
INTERACTION. The effect of Version by Reading was not significant [$\chi^2(1, N = 21) = .87, p = .350$], since both the 1st and 2nd Readings showed differences between the Versions, with the Conventional Versions eliciting in both cases smaller number of mistakes than the Modified Versions. Also, there were no significant differences between the *2nd Readings* of the Modified Versions and the *1st Readings* of the Conventional Versions (see *Figure. 7.2.b.*).

C. TOTAL NUMBERS OF MISTAKES

MAIN EFFECTS. A Generalized Linear Model (see SECTION 7.3.1.A.) showed high significance for Version [$\chi^2(1, N = 21) = 15.15, p < .001$], a highly significant effect of Reading [$\chi^2(1, n = 21) = 48.95, p < .001$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = .02, p = .865$]. The significance for Version reflected better performances, in terms of total numbers of mistakes, with the *Conventional* Version.

INTERACTIONS. The effect of Version by Reading was highly significant [$\chi^2(1, N = 21) = 12.54, p < .001$], with the 2nd Reading showing clear differences between the Versions whereas the 1st Readings showed

no significant difference; in the 2nd Readings the Conventional Versions elicited smaller number of mistakes than the Modified Versions, with the difference being highly significant (see *Figure. 7.2.c.*).



[a.]

[b.]

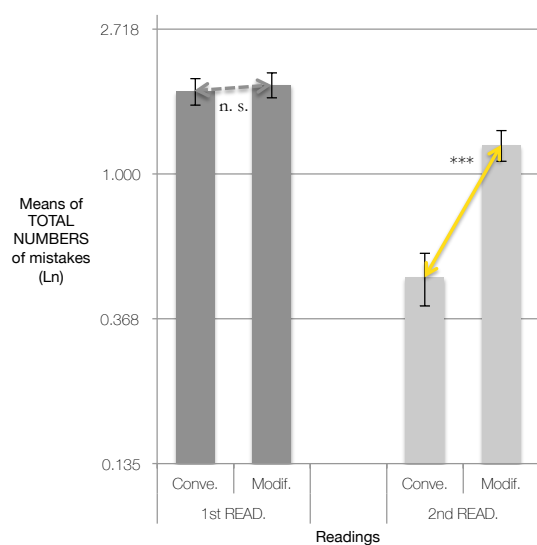


Figure 7.2.

EXERCISE 1: Pairwise Comparisons of Estimated Marginal Means of [a] PITCH, [b] RHYTHM, and [c] TOTAL NUMBERS of mistakes, for performances using Conventional Versions (Conve.) or Modified Versions (Modif.) of the scores. In each panel, from left to right, comparisons of mistakes made in the 1st Readings and in the 2nd Readings. The vertical axes represent logarithmically transformed values (Ln), having used the Ln of the difficulty of each Piece as an offset variable

Error bars represent Standard Error.

Significance using the Sidak adjustment for multiple comparisons:

- n. s. non-significant
- + $p < .100$ (marginal)
- * $p < .050$
- ** $p < .010$
- *** $p < .001$

[c.]

ORDER EFFECT. For the Total Numbers of Mistakes, the interaction between Version, Reading, and Order of Presentation was also analysed, and found to be generally significant [$\chi^2(3, N = 21) = 9.82, p = .020$]; the comparisons of estimated marginal means showed a trend towards the Versions performing better when presented later in the test (both the Conventional and the Modified Versions eliciting fewer mistakes when the other Version had been read first, and this trend being more marked in the 2nd Reading); although no clear significance was attained in specific pairwise comparisons [all $ps > .050$], marginal

significance was found when comparing 2nd Readings of the Modified Versions [$p = .086$] (see Figure 7.3.).

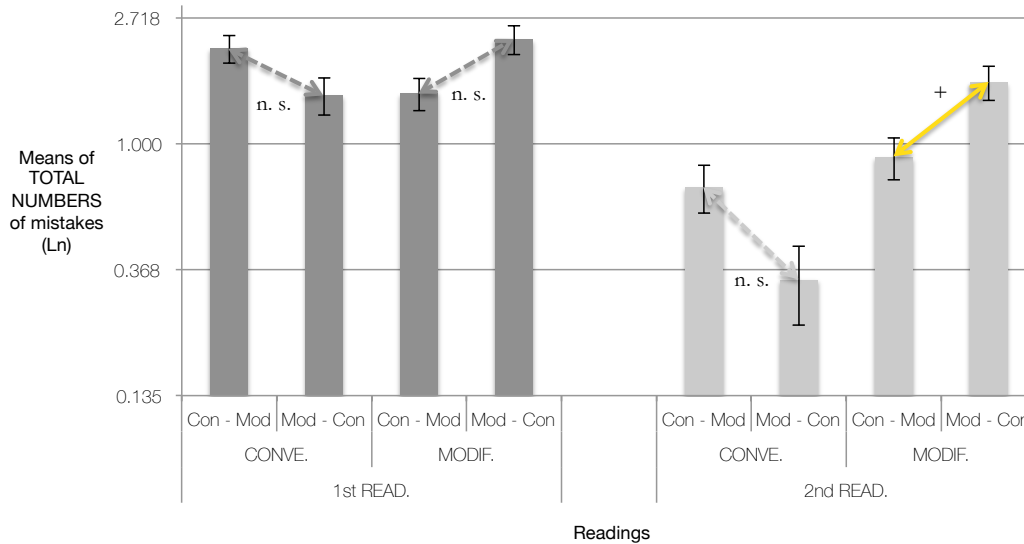


Figure 7.3.

EXERCISE 1: Pairwise comparisons of Total Numbers of mistakes for Readings using Conventional (Conve.) or Modified (Modi.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then Conventional). The vertical axis represents logarithmically transformed values (Ln).

Significance using the Sidak adjustment for multiple comparisons:
 n. s. non-significant
 + $p < .100$

7.3.2. EXERCISE 2: SEPARATION OF SUB-PHRASAL DISCOURSE UNITS

Generalized Linear Models with the same setups (Dependent Variables, Probability Distributions, Link Functions, Offset Variables, and Factors) as for Exercise 1 (see SECTION 7.3.1.) were implemented.

A. NUMBERS OF PITCH MISTAKES

MAIN EFFECTS. The model showed high significance for Version [$\chi^2(1, N = 21) = 16.27, p < .001$], a highly significant effect of Reading [$\chi^2(1, N = 21) = 45.33, p < .001$], and significance for the Order of Presentation [$\chi^2(1, N = 21) = 10.58, p = .001$]. As hypothesized, the significance for Version reflected better performances with the Modified Versions.

INTERACTION. The Model found that the effect of Version by Reading was significant [$\chi^2(1, N = 21) = 5.55, p = .018$], with the 2nd Reading showing much clearer differences between the Versions than the 1st Reading; it was in the 2nd Readings where the reduction in number of pitch mistakes with the Modified Versions became highly significant (see Figure 7.4.a.).

B. NUMBERS OF RHYTHM MISTAKES

MAIN EFFECTS. The Model showed high significance for Version [$\chi^2(1, N = 21) = 14.26, p < .001$], a highly significant effect of Reading [$\chi^2(1, N = 21) = 35.09, p < .001$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = 2.40, p = .121$]. As hypothesized, the significance for Version reflected better performances with the Modified Versions.

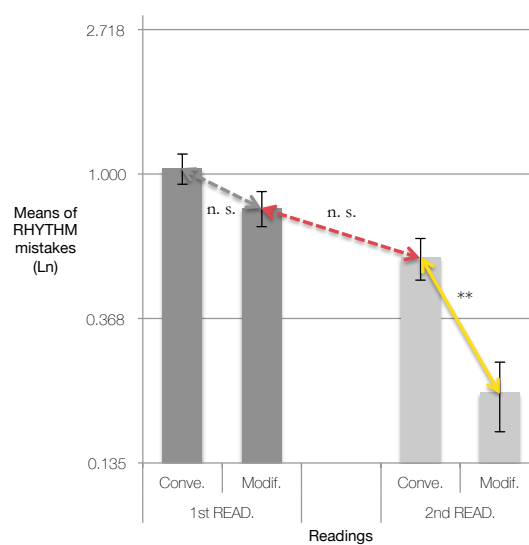
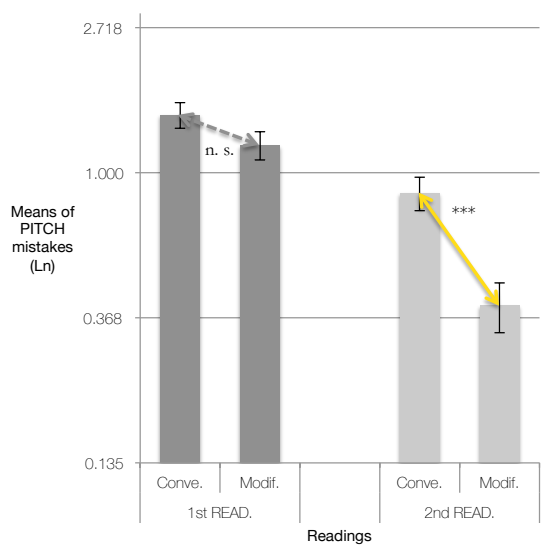
INTERACTION. The effect of Version by Reading was significant [$\chi^2(1, N = 21) = .435, p = .037$], with the 2nd Readings showing much clearer differences between the Versions than the 1st Readings; it was in the 2nd Readings where the reduction in number of rhythm mistakes with the Modified Versions became significant (see *Figure 7.4.b.*). Also, there were no significant differences between the 2nd Readings of the Conventional Versions and the 1st Readings of the Modified Versions.

C. TOTAL NUMBERS OF MISTAKES

MAIN EFFECTS. The Model showed high significance for Version [$\chi^2(1, N = 21) = 37.33, p < .001$], a highly significant effect of Reading [$\chi^2(1, N = 21) = 89.24, p < .001$], and significance for the Order of Presentation [$\chi^2(1, N = 21) = 8.97, p = .003$]. As hypothesized, the significance for Version reflected better performances with the Modified Versions.

INTERACTIONS. The effect of Version by Reading was highly significant [$\chi^2(1, N = 21) = 13.65, p < .001$], with the 2nd Reading showing highly significant differences between the Versions whereas the 1st Readings showed significant differences, but not to the same high degree (see *Figure 7.4.c.*).

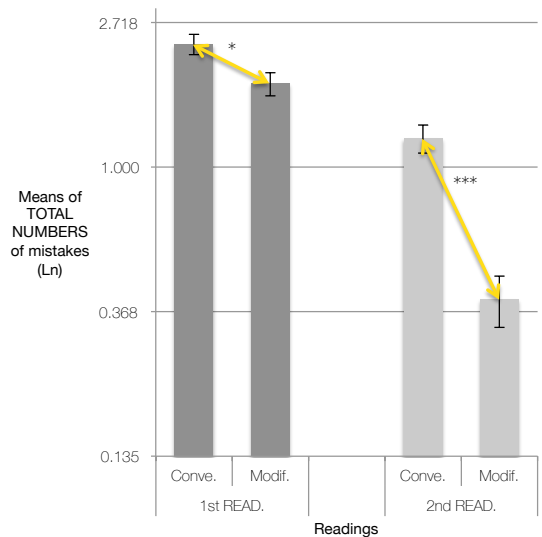
ORDER EFFECT. For the Total Numbers of Mistakes, the interaction between Version, Reading, and Order of Presentation was also analysed, and found to be generally significant [$\chi^2(3, N = 21) = 8.60, p = .035$]; the comparisons of estimated marginal means showed no particular trend, however, with differences being insignificant [all p s $> .100$], except for the 2nd Readings of the Modified Versions, showing slightly better results when the test had started with a Conventional Version [$p = .065$] (see *Figure 7.5.*).



[a.]

[b.]

Figure 7.4.



[c.]

EXERCISE 2: Pairwise Comparisons of Estimated Marginal Means of [a] PITCH, [b] RHYTHM, and [c] TOTAL NUMBERS of mistakes, for performances using Conventional Versions (Conve.) or Modified Versions (Modif.) of the scores. In each panel, from left to right, comparisons of mistakes made in the 1st Readings and in the 2nd Readings. The vertical axes represent logarithmically transformed values (Ln), having used the Ln of the difficulty of each Piece as an offset variable.

Error bars represent Standard Error.

Significance using the Sidak adjustment for multiple comparisons:

n. s. non-significant

* $p < .050$

** $p < .010$

*** $p < .001$

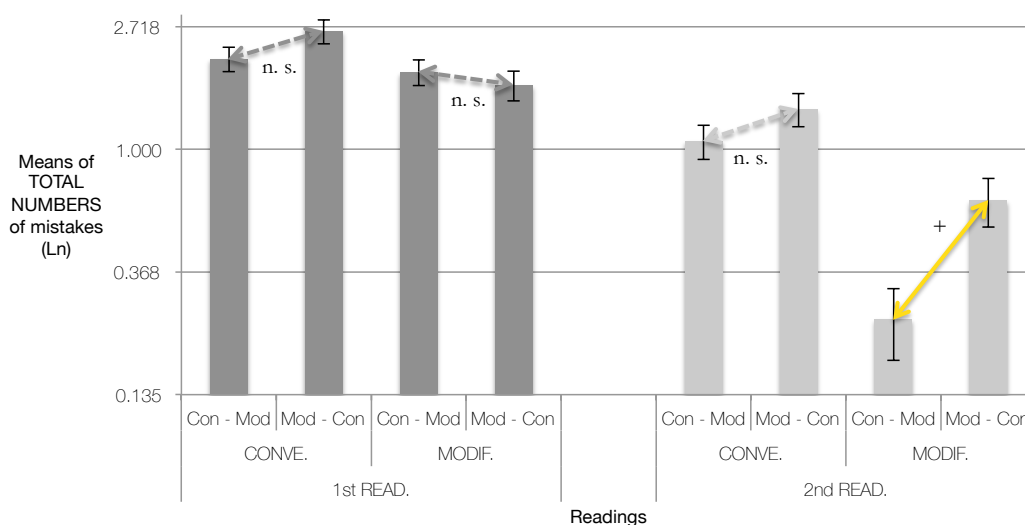


Figure 7.5.

EXERCISE 2: Pairwise comparisons of Total Numbers of mistakes for Readings using Conventional (Conve.) or Modified (Modif.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then Conventional). The vertical axis represents logarithmically transformed values (Ln).

Significance using the Sidak adjustment for multiple comparisons:

n. s. non-significant

+ $p < .100$

7.3.3. EXERCISE 3: SEPARATION OF GRAPHEMES

Generalized Linear Models with the same setups as for Exercise 1 (see details in SECTION 7.3.1.) were implemented.

A. NUMBERS OF PITCH MISTAKES

MAIN EFFECTS. The model showed *no* significance for Version [$\chi^2(1, N = 21) = .33, p = .856$], a highly significant effect of Reading [$\chi^2(1, N = 21) = 48.74, p < .001$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = 1.44, p = .230$].

INTERACTION. The Model found that the effect of Version by Reading was non significant [$\chi^2(1, N = 21) = .34, p = .560$], with first and 2nd Readings showing equally insignificant differences (see Figure 7.6.a.).

B. NUMBERS OF RHYTHM MISTAKES

MAIN EFFECTS. The model showed *no* significance for Version [$\chi^2(1, N = 21) = .05, p = .812$], a significant effect of Reading [$\chi^2(1, N = 21) = 5.51, p = .019$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = 1.44, p = .229$].

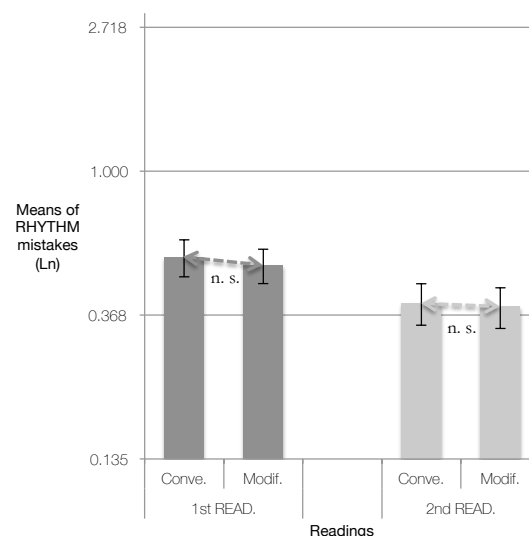
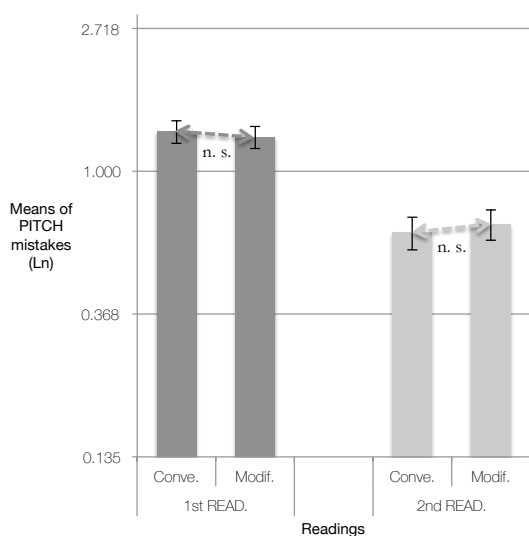
INTERACTION. The effect of Version by Reading was non significant [$\chi^2(1, N = 21) = .01, p = .917$], with first and 2nd Readings showing equally insignificant differences (see *Figure 7.6.b.*).

C. TOTAL NUMBERS OF MISTAKES

MAIN EFFECTS. The model showed *no* significance for Version [$\chi^2(1, N = 21) = .01, p = .930$], a significant effect of Reading [$\chi^2(1, N = 21) = 54.58, p < .001$], and no significance for the Order of Presentation [$\chi^2(1, N = 21) = .82, p = .364$].

INTERACTION. The effect of Version by Reading was non significant [$\chi^2(1, N = 21) = .54, p = .459$], with first and 2nd Readings showing equally insignificant differences (see *Figure 7.6.c.*).

ORDER EFFECT. The interaction between Version, Reading, and Order of Presentation was also analysed, and found to be generally *highly* significant [$\chi^2(3, N = 21) = 31.95, p < .001$], but with the variability coming only [with all other *ps* > .100] from the 2nd Readings of the Conventional Version where the effect of starting the test with a Conventional or a Modified Version was highly significant, with test starting with the Conventional Version yielding clearly better results, and no other trends being apparent (see *Figure 7.7.*).



[a.]

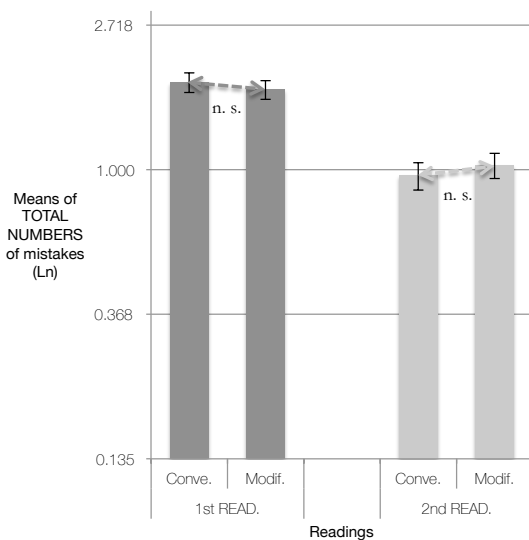
[b.]

Figure 7.6.

EXERCISE 3: Pairwise Comparisons of Estimated Marginal Means of [a] PITCH, [b] RHYTHM, and [c] TOTAL NUMBERS of mistakes, for performances using Conventional Versions (Conve.) or Modified Versions (Modif.) of the scores. In each panel, from left to right, comparisons of mistakes made in the 1st Readings and in the 2nd Readings. The vertical axes represent logarithmically transformed values (Ln), having used the Ln of the difficulty of each Piece as an offset variable.

Error bars represent Standard Error.

Significance using the Sidak adjustment for multiple comparisons:
n. s. non significant



[c.]

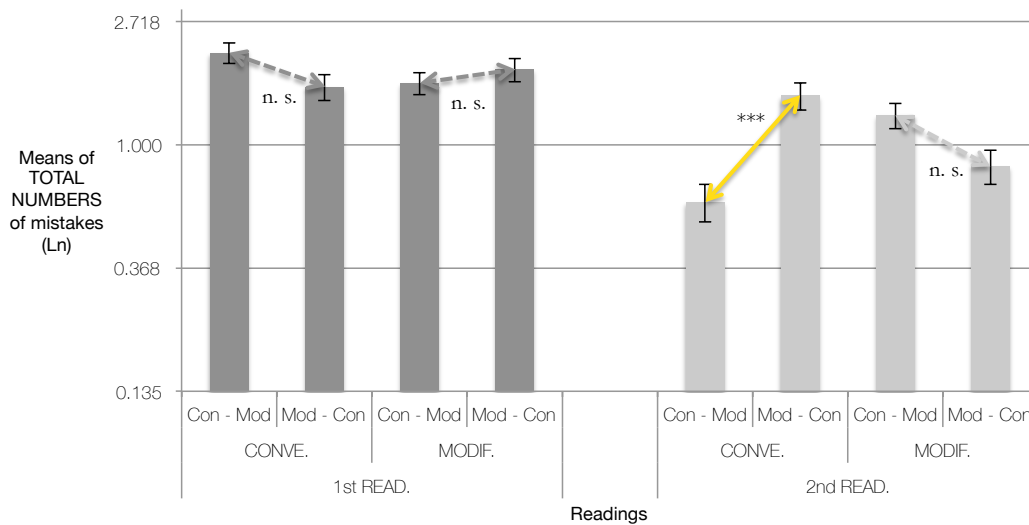


Figure 7.7.

EXERCISE 3: Pairwise comparisons of Total Numbers of mistakes for Readings using Conventional (Conve.) or Modified (Modi.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then Conventional). The vertical axis represents logarithmically transformed values (Ln).

Significance using the Sidak adjustment for multiple comparisons:

n. s. non-significant

*** $p < .001$

7.3.4. PREFERENCES BY PARTICIPANTS

Participants had been asked to state their preferred Version in a five-point Likert scale (5 = Clearly the Modified; 4 = Maybe the Modified; 3 = Equal; 2 = Maybe the Conventional; 1 = Clearly the Conventional). The probability distribution of the preferences as dependent variable for analysis was thus set as multinomial: responses could fit into five categories, without possible quantification of the differences between them, beyond an ordinal arrangement (it is not possible to establish if the difference between, *e. g.*, 'Equal' and 'Maybe the Conventional' is the same as, or smaller, or bigger than between 'Maybe the Conventional' and 'Clearly the Conventional'). The Link function used was a cumulative logit, as the outcomes had nonetheless an ordinal relation between them. All responses were furthermore divided by five (as an analysis weight value related to the variance of the response).

Since it had been detected that certain Pieces had unexpectedly posed more difficulties than others for this specific population (see Table 7.1.), these differences had to be integrated into the analysis of the preferences stated by the participants. Which Version was using the easier Piece was included as a factor. A Generalized Linear Model was run, with preference as dependent variable, multinomial probability distribution assumption, a cumulative logit link function, and a weighted scale of response. Factors were Exercise (1, 2 or 3) and 'Version Facilitated by Easier Piece' (Conventional or Modified).

MAIN EFFECTS. The model showed a main effect of Exercise [$\chi^2(2, N = 21) = 7.70, p = .021$], but no main effect of ‘Facilitation by Easier Piece’ [$\chi^2(1, N = 21) = 2.40, p = .121$].

INTERACTIONS. The interaction of ‘Facilitation by Easier Piece’ and Exercise, however, was highly significant [$\chi^2(2, N = 21) = 35.26, p < .001$], with the preferences not being affected in Exercise 2, slightly affected in the Exercise 1, and highly affected in the Exercise 3. Due to the multinomial assumption (*i. e.*, the differences between responses not being strictly scalable beyond order) it is not pertinent to quantify the significance of the differences in means, but the dissimilarities in evaluation can be visualized in *Figure 7.8*.

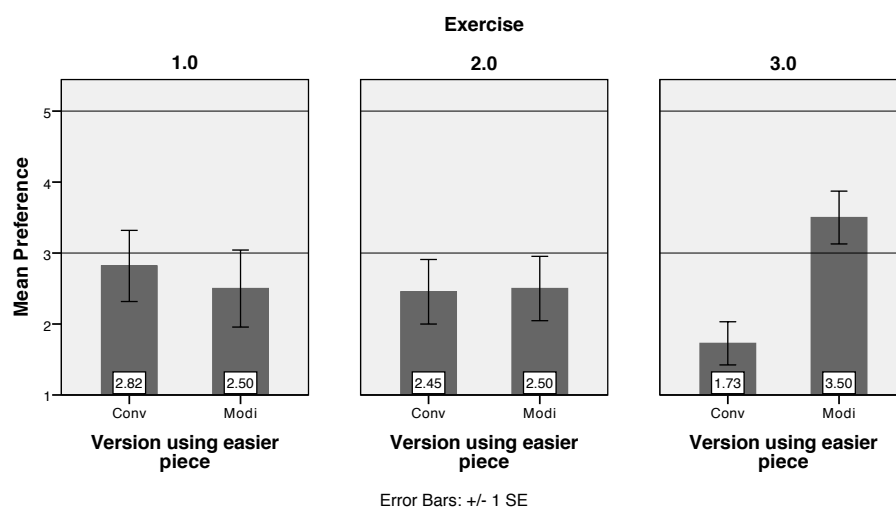


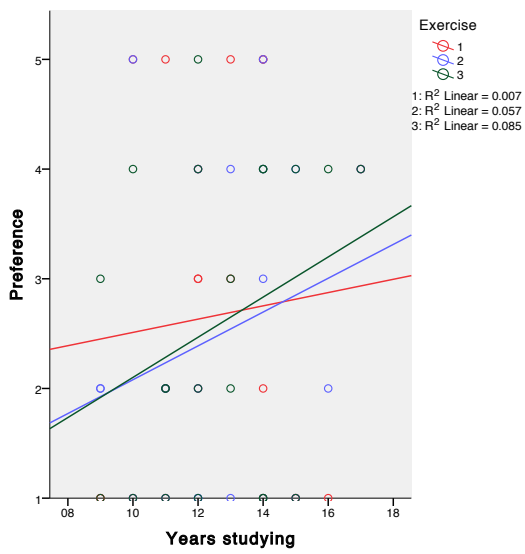
Figure 7.8.

STATED PREFERENCES as a factor of the Exercise and the Version that coincided with the easier Piece. The scale in the Y axis signifies: 5 = Clearly the Modified; 4 = Maybe the Modified; 3 = Equal; 2 = Maybe the Conventional; 1 = Clearly the Conventional. Multinomial Probability Distribution is assumed for these values —there are therefore no quantifiable comparisons of marginal means. The Graph shows nonetheless a different perception for materials in Exercise 2 compared to Exercise 1 and particularly to Exercise 3.

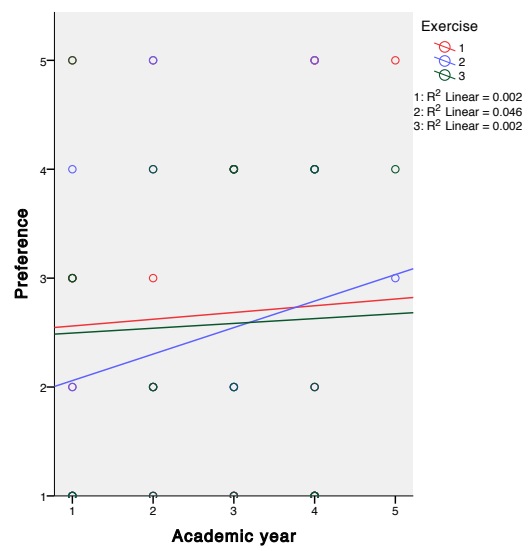
INTERACTIONS WITH COVARIATES. Seeing that there was a difference in assessments between the Exercises, the interaction of these responses with covariates possibly measuring expertise was further investigated. Covariates introduced were ‘Years Studying the Instrument’, ‘Year in the Academic Programme’ and ‘Age of Participant’. The interaction of Exercise with ‘Years Studying’ had a significant effect on Preference [$\chi^2(3, N = 21) = 9.21, p = .027$] (see *Figure 7.9.a.*), as had the interaction with ‘Academic Year’ [$\chi^2(3, N = 21) = 11.70, p = .008$] (see *Figure 7.9.b.*); the interaction with Age was highly significant [$\chi^2(3, N = 21) = 23.22, p < .001$] (see *Figure 7.9.c.*). As can be seen in the relevant Figures (7.9.a., b., c.; again, due to the multinomial assumption, significances not quantified), there is in all cases a positive correlation with expertise predictors for the preference of the materials in Exercise 2: participants that had been studying longer, that had been longer engaged with an academic degree, or were older in age all preferred the

novelties presented in Exercise 2 more than participants that had studied for fewer years, were fewer years into their degree, or were younger in age. This trend was shared with the preferences for novelties presented in Experiment 1 when the interaction with the number of years was accounted for. Notably, when the number of ‘Academic Years’ were accounted, the trend for Exercise 2 was different to the one shown for Exercises 1 and 3, with the preferences for these remaining neutral through the spectrum. And remarkably, the tendencies were inverted, when the students’ age was accounted for, with older (allegedly more experienced) students showing *less* preference for the Modified Versions, as presented in Exercises 1 and 3.

Finally, it should also be noted that contrary to what was implemented in the previous experiment (with violinists, see SECTION 6.3.6.), the Tempo of Performance was not analysed for interactions in this experiment (in spite of being the strongest predictor of preference for innovations in Experiment IV), since the Tempi used by the Spanish students were all, apart from rather slow, in a very narrow range of beats per minute (all within 40–45 bpm, with only one outlier at 60 bpm). When tested, the results were indeed flat through the range, and showed no significance.



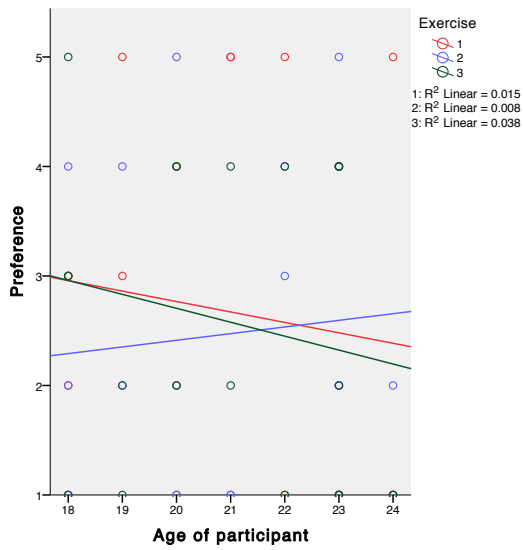
[a.]



[b.]

Figure 7.9.

Grouped scatterplots showing PREFERENCES STATED by participants in relation to [a] Number of years studying percussion, [b] Academic year (1–4 BMus; 5 MMus), [c] Age. The scale in the Y axis signifies: 5 = Clearly the Modified; 4 = Maybe the modified; 3 = Equal; 2 = Maybe the Conventional; 1 = Clearly the Conventional.



[c.]

7.4. DISCUSSION: ASSESSMENTS OF LEVELS OF SEPARATION

7.4.1. EXERCISE 1: SEPARATION ONLY AT PHRASE LEVEL

The separation of large structural units (phrases), including only one unit per line, elicited a reaction contrary to what has been observed in all other experiments: in this case, the performances with conventional versions produced fewer mistakes than the performances with modified versions. The differences were more marked, and attained statistical significance in the second readings, meaning that performances with the conventional scores also gained more from reassurance as the tests progressed.

If we imagine a similar situation with language scripts, the results seem understandable: on top of the effect of familiarity discussed above (see SECTION 6.4.) which, in equal conditions of legibility, will make readers attune faster to the recognisable patterns of the better known styles [see Sloboda *et al.*, 1998], there is no gain in presenting a text with no separation of basic units of information (words, or cells/motifs) and only implementing an unfamiliar separation of the major divisions of the discourse.

It might be the case, though, that with professional performers very accustomed to the idiomatic conventions of a certain style, the predictability of certain lines and their grammatical (harmonic) implications would mean that the encoding could encompass whole phrases. Future research could look at the relation between expertise and ideal length of integration/separation (see SECTION 2.4.3., on the relation of perceptual span to expertise).

7.4.2. EXERCISE 2: SEPARATION AT MANAGEABLE INFORMATIONAL LEVEL

Exercise 2, where the separation was introduced at cell/motif level, showed patterns of behaviour similar to what we have encountered in all previous experiments: significant differences in numbers of mistakes in favour of the modified scores, with the differences becoming highly significant in the second readings and, in the case of rhythm mistakes, no significant differences between a first performance with the modified score and a second performance with the conventional one.

7.4.3. EXERCISE 3: SEPARATION AT MINIMUM UNIT LEVEL

A. SEPARATION OF GRAPHEMES

As would have been the case in a language script, simply separating the constituent graphemes of the system did not yield an increase in legibility: analysis models showed no significance for the effect of Version.

However, and contrary to what was found in the order effects of the other two exercises (see Figures 7.2. and 7.3.), where there was no remarkable effect of the order in which the versions were presented on the numbers of mistakes, except a marginal significance for the effect of habituation on the modified scores towards the end of the session (predictably, perhaps, as the players would be more comfortable) — contrary to that, a highly significant difference appeared here in the second readings. Namely, in the case of the conventional scores, the numbers of mistakes were markedly higher in the second reading when the test had started with the modified scores, that is, in the 4th and last place in the sequence of readings (Modi. - Modi. - Conv. - CONV.). It could be hypothesized that this is signifying a decrease in concentration, as the conventional scores of these predictable pieces do not gather the focus and interest of the participant. The marginal effects on the last reading of the sequence (of four readings) for the modified scores found in the other two exercises was in the opposite direction: readers made *fewer* mistakes with the 2nd reading of the modified scores if this was the last reading in the sequence.

At any rate, and more importantly, it has been found in our experiments (see, *e. g.*, Experiment I —SECTION 3.3.4.; Experiment II —SECTION 4.3.3.; Experiment III —SECTION 5.3.3., and particularly Experiment IV —SECTION 6.3.3.) that the effect of the modified designs is either maintained or tending to slightly increase as the test and the numbers of readings progress.

The effect of order or more generally of tiredness on decreasing concentration with conventional scores —particularly in long sessions as implemented here (three exercises with four pre-test and four test readings each) should also be investigated further in future research.

B. TEMPORAL VALUE OF MUSICAL GRAPHEMES

On the other hand, the graphemes used in the music writing system are imbued with a precise timing value that has no clear parallel in language scripts. Many languages (*e. g.*, Latin) play with the length of accented and non-accented syllables in their prosody, but never in a strictly proportionate or quantified way as in music [see, however Glushko *et al.*, 2016, on shared mechanisms for music and language prosody; also Jentschke, 2016, for an overview of positions on the relationship between music and language].

The fact that in this exercise the clarification of the exact proportions between figures (*e. g.*, the distance, both visually and in timing, between two crotchets being exactly two times the distance between two quavers) had no effect on performance could be seen as a further confirmation of the [classically assumed, at least since the end of the XIXth cent.; see Cattell, 1886a, 1886b.; Pillsbury, 1897; Erdmann & Dodge, 1898] integrative or abstractive quality of visual encoding of text —rather than it being regulated by a continuous scanning process.

However, the pieces presented here used extremely simple rhythmic writing, and were not in the least challenging at this specific level for percussion students, as the proportions signified in the text were basic and overlearned. Future research should test whether visually proportional notation would have a clarifying effect and enhance legibility of rhythmically less predictable texts (*e. g.*, including triplets, syncopations, or additive rhythms).

7.4.4. CONCURRENCE OF PARTICIPANTS' PREFERENCES

The stated preferences by the participants concurred in showing a different pattern for Exercise 2 than for the other two exercises: the analysis model showed a main effect of exercise on the preferences, and an analysis of the interaction of preference with the coincidence with the pieces that were easier to perform in each pair demonstrated that preferences in Exercise 2 (where the design modifications introduced had had a detectably positive influence on legibility) were kept independent of this difference in intrinsic difficulty of the pieces —whereas in Exercise 1 and particularly in Exercise 3 the use of an easier piece altered the preferences.

Finally, the preference for the modifications introduced at a manageable informational level in Exercise 2 were positively correlated with the numbers of years studying, the level of academic progress, and the age of the participants (this last covariate showed in fact a negative correlation with the preference for the modifications introduced in the other two exercises)

In sum, the analysis of numbers of mistakes and of the preferences by the participants (especially when taking into account their expertise) confirmed that the level at which the separation of information is introduced will affect its facilitatory effect and its positive perception. The parallels of music reading with language reading were supported once more, in that the most decisive level of separation in musical scripts is equivalent to the highly influential separation of words —a separation that has been furthermore proposed to have had an enormous cultural, social and developmental impact on humans.

ACKNOWLEDGEMENTS

I am grateful to the participants, all of whom engaged selflessly and enthusiastically in the experiment. For this last study —the most logistically challenging— I am largely grateful to *Maestro* Cesar Peris at the *Escuela Superior de Música de Aragón*, in Saragossa, and *Maestro* Lorenzo Ferrandiz, at the *Escuela Superior de Música de Cataluña*, in Barcelona, for their invaluable help in the organisation of the tests and recording sessions.

*General Discussion
and Future Research
Directions*

CHAPTER 8

General Discussion and Future Research Directions

8.1. TRENDS THROUGH THE SERIES OF EXPERIMENTS

8.1.1. READING FLUENCY AND PERFORMANCE MODULATED BY DESIGN

Experiments were conducted with musicians reading scores using different designs, showing that the performance was affected—in a highly significant way, in most cases—by the implementation of legibility cues in the scores. Five experiments have been reported, in all of which the reading and re-reading of music was affected positively by the use of visual cues structuring the spacing and layout of the notation. Several trends could be observed throughout the experiments, when musicians used the novel scores: (a) readers made fewer mistakes in total, and also separately in the pitch and the rhythm domains; (b) the tempo was kept steadier and with fewer hesitations; (c) slight increases in dynamic range were observed; (d) the performances were evaluated more positively by experts; (e) in most cases, the evaluation of the scores was positive by the players as well, certainly after using them.

The fact that percussionists (Exp. I, II, III, & V) and violinists alike (Exp. IV) performed better with the novel scores is in accordance with the independence of mistakes from mode of execution that has been hypothesized in sight-reading literature. Music reading expertise can be independent of performance, with expert musicians being able to competently read music without an instrument [*e. g.*, Drai–Zerbib, Baccino, & Bigand, 2012; Drai–Zerbib & Baccino, 2013]. If music reading is facilitated by visual structuring, the results, as seen here, should be independent of instrumental idiosyncrasies.

8.1.2. MARKED INCREASES IN READING FLUENCY WITHIN THE RHYTHM DOMAIN

A. THE RHYTHM DOMAIN AS PREDICTOR OF ABILITIES

Further, both groups of instrumentalists showed more significant increases in reading fluency within the rhythm domain. In the (language) reading literature, temporal awareness, and more specifically rhythm perception and production are considered solid predictors of reading abilities. Conversely, learning disabilities affecting reading acquisition or performance (frequently independently of general intelligence levels) have been associated with poor rhythmic perception. The correlation has been amply studied in the last years, for instance in children with non-comorbid dyslexia [Flaugnacco *et al.*, 2014; Bonacina *et al.*, 2015; Cancer *et al.*, 2015; Flaugnacco *et al.*, 2015], children with benign epilepsy with centrotemporal spikes [Amaral *et al.*, 2015], children with specific language impairments [R. Cumming *et al.*, 2015], or patients with Parkinson's disease [Biswas *et al.*, 2016].

As a consequence, the recent neuroscientific literature includes numerous arguments supporting the use of musical training, with special emphasis on rhythmic perception and production, as a therapeutic tool for reading-related disorders or impairments [Flaugnacco *et al.*, 2014; Bonacina *et al.*, 2015; Cancer *et al.*, 2015; Flaugnacco *et al.*, 2015; Habib *et al.*, 2016; for a larger framework also including dysphasia and deafness see Schon & Tillmann, 2014]. The robustness of the use of rhythmic training as systematic therapeutic and instructional practice for dyslexic children has been shown by Habib *et al.* [2016], with their study revealing improvements in phonological awareness (importantly, at syllable integration level, not only at word recognition level), reading abilities, and the repetition of pseudo-words —with improvements persisting after a long untrained period (six weeks).

It has been hypothesized that rhythmic training improves reading skills by influencing phonological awareness [Schon & Tillmann, 2014] —although other authors have suggested that theories stressing the role of phonemic sensitivity in the form of word recognition for fluent reading need to be modified to include the role of *symbol* processings [Bowers, 1993, 1995; Young & Bowers, 1995; Young, Bowers, & MacKinnon, 1996]. Phonological awareness, however, has been shown to encompass not only high- or low-frequency words, but also, crucially, short and long pseudowords [Bonacina *et al.*, 2015; Habib *et al.*, 2016]. Since phonological awareness is not necessarily restricted to word recognition, and since it is influenced by rhythmic proficiency, any patterns loaded with plausible meaning within a given system could be more fluently processed as a result of increased rhythmic awareness.

It has also been hypothesized, more specifically, that there is a universal relationship between rhythmic *regularity detection* and reading skill. In experimental work with native speakers of several different native languages, Bekius, Cope, and Grube [2016] have found results supporting this relationship, which was furthermore robust even accounting for differences in fluid intelligence, musical expertise, and language-specific differences in speech rhythm.

Similarly, studies of skill acquisition in music performance have revealed that when learning to perform a novel piece, musicians show improvements in relative timing (temporal continuity, underlying beat, metrical structure) with skill and practice, which gradually compensate for the tendency by less proficient musicians to have a relatively high incidence of simultaneous pitch/time errors [Drake & Palmer, 2000]. With the modified scores, musicians not only made fewer rhythm mistakes, but also showed higher temporal stability and adherence to an underlying beat (as measured in the nPVI values). That is, they momentarily performed with the traits of enhanced expertise.

In more general terms, investigations into changes in cognitive capacities that occur as musicians acquire performance skills have shown that expertise (including music-reading) is positively correlated with time-related capacities, rather than directly with acuity or perceptive enhancements. These capacities manifest in quicker detection and correction of errors, more anticipatory and less perseveratory behaviour, and a larger range of planning [Palmer & Drake, 1997]. Participants in our experiments commented recurrently how the separation of sub-phrasal units in the modified scores acted as a firewall that afforded quicker correction of errors, eliminated perseveratory behaviour, and helped planning the performance of the next informational unit.

B. LIMITATIONS BY THE USE OF SIMPLE PROTOTYPICAL TEMPORAL STRUCTURES

Nonetheless, it could also be hypothesized that the salient reading facilitation in the rhythm domain was due to a much simpler fact: particularly for the percussionists, the rhythmic figures presented were very basic, and did not in any way represent a processing or dexterity challenge. Once the spacing and separation had facilitated the comprehension of the simple sequence of events and the mostly two-level recursive combination forming it (three maximum: from semiquavers to quavers to crotchets; but some pieces did not even include semiquavers), the processing of rhythm would be automatized and cognitive effort would be applied only to other parameters in the script. The rhythms used presented characteristics that have been shown in psychophysical research [Drake, 1993] to correspond with a simple prototypical temporal structure, functional already in childhood: binary subdivision; two rather than three different levels or durations, and intensity accents on important hierarchical positions. Drake [*id.*] showed that for adults and children, reproduction was easier for rhythms with the mentioned characteristics. The processing of prototypical structures could be different from the processing of more complex hierarchical structures, which would perhaps not be facilitated by the separation and spacing proposed in the novel design, and might even be hindered by it.

However, in the violin pieces the rhythms, even if following a binary subdivision, did include up to four levels of recursive integration, and used syncopations and suggestions of polyphony that undermined accentuation predictions. Still, the results in terms of rhythmic facilitation were very similar to what was found in the experiments with percussionists, with the reduced number of mistakes being more significant

than in the pitch domain. Nonetheless, as mentioned above (see SECTION 6.2.4.) the use of broader criteria for coding correct notes in the pitch domain could be a confounding factor in this issue. Within the time and logistic constraints of the experiment with violinists, the simplified coding was inevitable, since it had to be done by hand.

8.2. FUTURE RESEARCH: NEEDED IMPROVEMENTS IN CODING AND IN EXPANDING METHODS

8.2.1. PROBLEMATIC CODING OF MUSICAL PERFORMANCE MISTAKES

Future research would possibly benefit from standardising the criteria for mistake coding and also, wherever possible, use an automated (or at least algorithmic) approach. But the coding of mistakes in musical performance is complicated, in respect to language, with the imposition of additional task constraints beyond fluency, namely temporal constraints [Drake & Palmer, 2000] (pitch detection proceeds in a methodologically similar way in language and music studies). Overall speed of performance is less important than producing the correct events at the correct point in time: musical performance must reflect the event duration categories specified in a score (usually simple integer ratio relationships). Performance is considered deficient if it does not respect these duration categories. Yet, at the same time, these proportions are expected to be stretched to allow for idiomatic timing variations. Using statistical averages and ranges of contextualised inter-note-onsets for coding, this can be overcome. However, the more problematic issue is the automatized coding of these mistakes and deviations against the prescriptions in a musical score. If a participant hesitates or stops during testing, for instance, all notes played after the interruption will be coded as wrong, since they will no longer match the scrolling score, even if the duration categories are defined statistically (whereas in language the utterances would not usually rely on these timing constraints).

Attempts at automatized coding have attempted to use systems derived from speech error coding [like the ones proposed by Garrett, 1975; Dell, 1986]. Palmer [1992], and later Palmer and Van de Sande [1993, 1995] modified these systems for the musical domain, but had to exclude certain types of errors (hesitations and cumulative errors) because they could not be coded unambiguously. These exclusions might not be ecologically valid —certainly not phenomenologically: it is precisely the ability to play fluently without hesitations (allowing for small mistakes if necessary), and to retain control of the performance after a mistake, that is desirable and valued in expert readers, and that qualitatively distinguishes one (sight-reading) performance from another.

Up to very recently, the available commercial products for mistake coding still had the limitation of requiring the participant to play without stopping or hesitations, and manual marking of errors has been utilised in many sight-reading studies [Banton, 1995; Betts & Cassidy, 2000; Kostka, 2000; Mainz & Hambrick, 2010; Penttinen & Huovinen, 2011; Penttinen, Huovinen, & Ylitalo, 2015].

Zhukov [2014] has now developed a custom-made software, with the aid of a small team of consultants with expertise in data analysis and music, for the study of mistakes in sight-reading. The algorithm developed for the programme is able to detect if the playing is interrupted, or a beat missed or skipped. It can also detect extra notes played that are not in the score, and notes in the score missing from performance (all of which would normally have consequences on the coding of all subsequent notes) [see also Zhukov *et al.*, 2016].

8.2.2. EXTENDING PARTICIPANT SAMPLES AND USE OF DIGITAL INTERFACES

Even though the literature on sight-reading is now considerable, one of its limitations is the use of small samples (a meagre maximum of 21 participants in my experiments), and the related lack of replicated large scale studies, as Zhukov herself has lamented [Zhukov, 2006; 2014; Zhukov *et al.*, 2016]. An important factor in this limitation is (or has been) the coding of mistakes, which is time consuming and cumbersome. The development of customized software for the task is therefore very welcome in the field, and could help progressing the scope, refinement, and validity of findings. Nonetheless, one important limitation remains unsolved: for the analyses using the Zhukov application, MIDI data (a widely used standard for interconnecting electronic musical instruments and computers) has to be imported from digitally captured performances. Similarly, other sight-reading or performance assessments based on computer algorithms have largely used MIDI data (one of the few exceptions being the audio signal based Sonic Visualiser, which we used for the nPVI coding). This in itself poses another restriction: few instruments beyond the keyboard have reliable MIDI transfer systems, and consequently the vastly prevailing (and somehow limiting) use of pianists as subjects for these studies can not be transcended yet.

Specifically, in the percussion ambit, MIDI instruments have been available for some time, either by attachment of pick-up pads to analogue instruments or, more reliably, by the use of keypads with simulated sounds, but these remain experimental models, and therefore difficult to find and expensive to use. The *Conservatorium van Amsterdam*, where we implemented three of the experiments, has one of the best equipped percussion studios in the world, but did not own a MIDI mallet instrument, and the conditions they offered for the rental of a MIDI keypad (which had to include transport, insurance, and instruction) were just not affordable within the framework of a PhD project.

Although the numerical simulation of impacted bars (the basis of a mallet instrument) has been already attained in successful models [*e. g.*, Doutaut, Matignon, & Chaigne, 1998; Henrique & Antunes, 2003/Part I] and translated into synthesized impact bars with manipulable perceptual dimensions [Henrique & Antunes, 2003; McAdams, Chaigne, & Roussarie, 2004] the construction of a novel affordable and durable percussion instrument with integrated electronic connectivity (via MIDI or other standards) has not yet been completed. However, a rather promising (and very ingenious) development has been recently

presented by Tavares *et al.* [2012] for the real-time music transcription of performance on mallet instruments (specifically a vibraphone). The authors have shown how the added information from a video camera can be used to impose constraints on the algorithmic figuration search (*e. g.*, deciding between a quaver or a crotchet) based on the gestures of the performer. This multi-modal approach has led to significant reduction of false positives (the programme therefore being able to distinguish hesitations or false starts, for instance) and generally increased accuracy of automatized performance transcription.

8.2.3. LIMITATIONS IN THE CHOICES OF REPERTOIRE, PARTICIPANTS AND USAGES

All experiments shared certain limitations, most of them the product of logistic and methodologic considerations. More specifically, all of them used: (a) music students as participants, when legibility cues have different facilitatory effects depending on expertise; (b) short pieces of music; (c) music with a steady beat, and stylistically homogeneous; (d) performances in solo settings, without peer interaction.

For instance, due to logistic and methodological reasons, short pieces of music were used in all of the experiments. In using visual separation between information units it was hypothesized that one of the benefits would be the detection or facilitated processing of recurrent patterns. The literature on segmental pattern discovery in music, however, uses algorithms (based on suffix trees) over large corpora of musical pieces. One such large collection of relatively homogeneous pieces is the Bach chorales. It is possible that the separation of recurring patterns would be incorporated into a performer's idiomatic lexicon (for instance, when performing a chorale) only after exposure to numerous instances of similar pieces, and that the fluency facilitation might be modulated by lexicon acquisition. Levin and Kaplan [1970], for instance, have shown that young inexpert readers (in the language domain) are not affected by the deletion of interword spaces, and the patterns of their ocular movements remain unchanged in the absence of cues the use of which they have not yet mastered.

Again, if certain fundamental logistic limitations (cumbersome coding, unstandardised audio signals, fatigue of participants) can be overridden, it would be desirable to expand the research to include longer sessions and ideally long-term effects of the proposed innovations and investigate the effects on different levels of expertise, levels of tempo flexibility or the effect on inter-subject entrainment and idiomatic cohesiveness.

Similarly, population samples would also benefit from expansion, as the experiments so far have focused on participants (Conservatoire and University students) that might be more likely than other groups to be open to challenge, and to respond to novelty with increased levels of concentration and interest. Results from self-assessment questionnaires seem to preclude a placebo effect for this particular population, as an *a priori* evaluation of the modified scores (by participants not using them for performance) was rather

negative (see SECTION 5.3.6., specially *Figure 5.11.*). The increases in fluency and precision when performing with the modified scores occurred thus *in spite of* the initially negative perception of the design novelties. However, it could be the case that other (perhaps more conservative, or busier) populations will not engage similarly with the novelties, and whether equivalent results will be found beyond Conservatoire and University students needs to be investigated.

With access to bigger datasets, the questions of what *kinds* of mistakes are being made when reading a given piece, and of *where* in the score they are made, would also be relevant issues for further investigation. It would be of particular interest to analyse how the mistakes relate to structure or form (*e. g.*, would there be fewer mistakes in presentational or developmental sections?), and specially to the underpinnings of musical grammar (*e. g.*, in tonal music, to harmonic priming).

Finally, a vast area of research could also open up if we were to assess the impact of score design not only on legibility and short-term transcoding and performance, but also on the most common of usages of scores in modern musical practices, that is, when they function as basis for a non-read performance (either overtly by heart, or, perhaps more commonly, using the score only as a guidance tool). In future attempts to create notation and score designs that would enhance the legibility and direct performance of a musical text, it should therefore be relevant to factor in the requirements for its subsequent memorisation and interpretation.

In any case, the results reported here could serve to open a debate on the conventions of music publishing as they stand, and are hopefully well placed to open new lines of research in legibility and design of music and language scripts.

ACKNOWLEDGEMENTS

ACKNOWLEDGMENTS

I am indebted to my supervisor, Prof. Ian Cross, with whom it has been an enormous privilege to work over the past years. I am equally grateful to my co-supervisor, Prof. Sarah Hawkins, for her invaluable guidance and support during these years.

I gratefully acknowledge the Cambridge Home and EU Scholarship Scheme, for generously funding my research. Gonville & Caius College through their Board of Tutors generously offered additional funding whenever it was needed. I am particularly grateful to my Tutors, Dr Melissa Calaresu and Dr Deborah Bowman for their help, support and personal advice. I am also sincerely grateful to the Board of the Faculty of Music for their financial support for travel and experimental setups.

The Faculty of Music and its Centre for Music and Science provided me with a friendly and inspiring environment to complete this dissertation, and I have benefitted greatly from exchanges with several colleagues. I would like to especially thank Dr Gabriela Pavarini for her friendship and complicity.

In the past years I was fortunate to supervise a good number of bright students at the Faculty of Music who helped me stay alert and focused: thank you all for your hard work and commitment.

My thanks are also due to all the people involved in organizing and participating in the experiments, without whom this research project would have been incomparably more burdensome.

I owe heartfelt thanks to several friends who kept me moving forward: Angel Moguel who helped me tremendously with the re-location to the UK; Garnell Fitz-Henley who has regularly and patiently listened to my rants; Carlos Sotomayor who has questioned things and reminded me of the value of humour and scepticism.

Finally, I wish to thank my parents, Aud and Antonio. Through your love and support I have survived the challenges of the past years. I owe you more than I have the capacity to express.

REFERENCES

REFERENCES

A

- ACKERMANN, B. J., KENNY, D. T., O'BRIEN, I., & DRISCOLL, T. R. [2014]. Sound practice: Improving occupational health and safety for professional orchestral musicians in Australia. *Frontiers in Psychology*, 5.
- ADORNO, T. W. [1997]. *Alban Berg: Master of the smallest link* (J. Brand, Trans.). Cambridge: University Press. (Original work published 1968)
- ADORNO, T. W. [2002]. *Aesthetic theory* (R. Hullot-Kentor, Trans.). London: Continuum. (Original work published 1969)
- ALDHELM, S. [2009]. *The prose works*. M. Lapidge, & M. W. Herren, Eds. Woodbridge: D. S. Brewer. (Original work published ca 640–709)
- ALLEN, R. J., BADDELEY, A. D., & HITCH, G. J. [2006]. Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology-General*, 135(2), 298–313.
- ALONSO, I., SAMMLER, D., VALABREGUE, R., DINKELACKER, V., DUPONT, S., & SAMSON, S. [2014]. Hippocampal sclerosis affects fMR-adaptation of lyric and melodies in songs. *Frontiers in Human Neuroscience*, 8.
- AMARAL, M. I. R., CASALI, R. L., BOSCARIOL, M., LUNARDI, L. L., GUERREIRO, M. M., & COLELLA-SANTOS, M. F. [2015]. Temporal auditory processing and phonological awareness in children with benign epilepsy with centrotemporal spikes. *Biomed Research International*.
- AMLUND, J. T., KARDASH, C. A. M., & KULHAVY, R. W. [1986]. Repetitive reading and recall of expository text. *Reading Research Quarterly*, 21(1), 49–58.
- ANDERSON, R. D., PARSONS, P. J., & NISBET, R. G. [1979]. Elegiacs by Gallus from Qasr Ibrim. *Journal of Roman Studies*, 69, 125–55.
- ANNETT, J. [1995]. Motor imagery: Perception or action. *Neuropsychologia*, 33(11), 1395–417.
- ANNETT, J. [1996]. On knowing how to do things: A theory of motor imagery. *Cognitive Brain Research*, 3(2), 65–9.
- ANONYMOUS. [1995]. *Musica enchiridis et Scholica enchiridis*. C. V. Palisca, Ed., Music Theory Translation Series. New Haven, USA: Yale University Press. (Original work published IXth cent.)
- ANONYMOUS. [2014]. *Rhetorica ad Herennium*. H. Caplan, Ed., Loeb Classical Library Series. Cambridge, USA: Harvard University Press. (Original work published ca 80 BC)
- ANONYMOUS. [before 692]. *Book of Mulling*. Illuminated manuscripts. [Ink on parchment]. Insular Irish Manuscripts, (IE TCD MS 60). Trinity College Dublin.
- ANONYMOUS. [Vth cent.]. *Vergilius Romanus*. Illustrated manuscript. [Vellum parchment]. Retrieved from http://digi.vatlib.it/view/MSS_Vat.lat.3867/0154 database (3867, 3v, 9r, 14r). Biblioteca Apostolica Romana.
- ANONYMOUS. [XVth cent.]. *Codex Faenza*. [containing keyboard music along with additional vocal pieces]. (I-FZc 117). Biblioteca Comunale, Faenza.
- ASHCRAFT, M. H., & CHRIST, K. S. [1995]. The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1–6. *Journal for Research in Mathematics Education*, 26(5), 396–421.
- ASHCRAFT, M. H., & GUILLAUME, M. M. [2009]. Mathematical cognition and the problem size effect. *Psychology of Learning and Motivation: Advances in Research and Theory*, 51, 121–51.

- AVANCINI, C., SOLTESZ, F., & SZUCS, D. [2015]. Separating stages of arithmetic verification: An ERP study with a novel paradigm. *Neuropsychologia*, 75, 322–9.
- B
- BACH, J. S. [1990]. *371 vierstimmige Choräle*. (K. Schubert Ed.). Wiesbaden: Edition Breitkopf. (Original work published 1784–87)
- BADDELEY, A. D., ALLEN, R. J., & HITCH, G. J. [2010]. Investigating the episodic buffer. *Psychologica Belgica*, 50(3-4), 223-43.
- BADDELEY, A. D., HITCH, G. J., & ALLEN, R. J. [2009]. Working memory and binding in sentence recall. *Journal of Memory and Language*, 61(3), 438-56.
- BAHLMANN, J., SCHUBOTZ, R. I., MUELLER, J. L., KOESTER, D., & FRIEDERICI, A. D. [2009]. Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research*, 1298, 161–70.
- BAI, X., YAN, G., LIVERSEDGE, S. P., ZANG, C., & RAYNER, K. [2008]. Reading spaced and unspaced chinese text: Evidence from eye movements. *Journal of Experimental Psychology-Human Perception and Performance*, 34(5), 1277–87.
- BALL, F., WOOD, C., & SMITH, E. E. [1975]. When are semantic targets detected faster than visual or acoustic ones. *Perception & Psychophysics*, 17(1), 1–8.
- BANNIARD, M. [1989]. *Genèse culturelle de l'Europe: Ve - VIIIe siècle*. Paris: Editions du Seuil.
- BANTON, L. J. [1995]. The role of visual and auditory feedback during the sight-reading of music. *Psychology of Music*, 23, 3–16.
- BARON, J., & THURSTON, I. [1973]. Analysis of word-superiority effect. *Cognitive Psychology*, 4(2), 207–28.
- BARRETT, S. [1997]. Music and writing: On the compilation of 'Paris Bibliothèque Nationale Lat. 1554' (Medieval music manuscripts). *Early Music History*, 16, 55–96.
- BASSETTI, B. [2009]. Effects of adding interword spacing on Chinese reading: A comparison of Chinese native readers and English readers of Chinese as a second language. *Applied Psycholinguistics*, 30(4), 757–75.
- BASSETTI, B., & LU, M. H. [2016]. Effects of interword spacing on native English readers of Chinese as a second language. *International Review of Applied Linguistics in Language Teaching*, 54(1), 1–22.
- BASSETTI, B., & MASTERSON, J. [2012]. Effects of removing morphemic information and adding interword spacing on reading in Chinese experienced and inexperienced readers. *Reading and Writing*, 25(9), 2291–314.
- BATTIER, M. [2015]. Describe, transcribe, notate: Prospects and problems facing electroacoustic music. *Organised Sound*, 20(1), 60–7.
- BECK, R. J., CESARIO, T. C., YOUSEFI, A., & ENAMOTO, H. [2000]. Choral singing, performance perception, and immune system changes in salivary immunoglobulin A and cortisol. *Music Perception*, 18(1), 87–106.
- BECKERS, G. J. L., BOLHUIS, J. J., OKANOYA, K., & BERWICK, R. C. [2012]. Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport*, 23(3), 139–45.
- BEGGS, W. D. A., & HOWARTH, P. N. [1985]. Inner speech as a learned skill. *Journal of Experimental Child Psychology*, 39(2), 396–411.
- BEKIUS, A., COPE, T. E., & GRUBE, M. [2016]. The beat to read: A cross-lingual link between rhythmic regularity perception and reading skill. *Frontiers in Human Neuroscience*, 10.
- BELLINI, P., FIORAVANTI, F., & NESI, P. [1999]. Managing music in orchestras. *Computer*, 32(9), 26–+.
- BENAVOT, A., ANTONINIS, M., BARRY, M., BELLA, N., BLANCHY, N. K., DELPRATO, M., . . . ZUBAIRI, A. (2006). *Global Education Monitoring Report*. (GMR06 - Target 4.6). Paris: UNESCO.
- BENT, I. D., HUGHES, D. W., PROVINE, R. C., RASTALL, R., KILMER, A., HILEY, D., . . . CHEW, G. [2001]. Notation. In *The New Grove Dictionary of Music and Musicians*. (Vol. 17). Oxford: Oxford University Press.
- BERG, D. H. [2008]. Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*, 99(4), 288–308.
- BERTIN, J. [1967]. *Sémiologie graphique: Les diagrammes, les réseaux, les cartes*. Paris: Gauthier-Villars.

- BERWICK, R. C. [1997]. Syntax Facit Saltum: Computation and the genotype and phenotype of language. *Journal of Neurolinguistics*, 10(2-3), 231–49.
- BERWICK, R. C., FRIEDERICI, A. D., CHOMSKY, N., & BOLHUIS, J. J. [2013]. Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, 17(2), 89–98.
- BERWICK, R. C., OKANOYA, K., BECKERS, G. J. L., & BOLHUIS, J. J. [2011a]. Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, 15(3), 113–21.
- BERWICK, R. C., PIETROSKI, P., YANKAMA, B., & CHOMSKY, N. [2011b]. Poverty of the Stimulus revisited. *Cognitive Science*, 35(7), 1207–42.
- BETTS, S. L., & CASSIDY, J. W. [2000]. Development of harmonisation and sight-reading skills among university class piano students. *Journal of Research in Music Education*, 48(2), 151.
- BEVAN, A., ROBINSON, G., BUTTERWORTH, B., & CIPOLOTTI, L. [2003]. To play 'B' but not to say 'B': Selective loss of letter names. *Neurocase*, 9(2), 118–28.
- BEVER, T. G., JANDREAU, S. M., BURWELL, R., KAPLAN, R., & ZAENEN, A. [1991]. Spacing printed text to isolate major phrases improves readability. *Visible Language*, 25(1), 74–87.
- BISWAS, A., HEGDE, S., JHUNJHUNWALA, K., & PAL, P. K. [2016]. Two sides of the same coin: Impairment in perception of temporal components of rhythm and cognitive functions in Parkinson's disease. *Basal Ganglia*, 6(1), 63–70.
- BLACKWELL, A. F. [2001]. Introduction. In A. F. Blackwell (Ed.), *Thinking with diagrams* (pp. 1–3). Dordrecht: Springer Netherlands.
- BLACKWELL, A. F. [2002]. Psychological perspectives on diagrams and their users. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning*. London: Springer.
- BLACKWELL, A. F. [2008, Sep]. *Cognitive dimensions of notations: Understanding the ergonomics of diagram use*. Poceedigs of the Vth International Conference on Diagrammatic Representation and Inference Location, Herrsching, Germany.
- BLACKWELL, A. F. [2013]. Visual representation. In M. Soegaard, & R. F. Dam (Eds.), *The Encyclopedia of Human-Computer Interaction* (2nd ed.). Aarhus: The Interaction Design Foundation.
- BLACKWELL, A. F., & ENGELHARDT, Y. [2002]. A meta-taxonomy for diagram research. In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic representation and reasoning*. London: Springer.
- BLUM, T., ABBOTT, C., AUSTIN, L., BATTIER, M., BEAUCHAMP, J., DASHOW, J., . . . STRAWN, J. [1983]. Report on the 1982 international computer music conference. *Computer Music Journal*, 7(2), 8–35.
- BOLGER, D. J., PERFETTI, C. A., & SCHNEIDER, W. [2005]. Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*, 25(1), 92–104.
- BOLHUIS, J. J., TATTERSALL, I., CHOMSKY, N., & BERWICK, R. C. [2014]. How could language have evolved? *Plos Biology*, 12(8).
- BONACINA, S., CANCER, A., LANZI, P. L., LORUSSO, M. L., & ANTONIETTI, A. [2015]. Improving reading skills in students with dyslexia: The efficacy of a sublexical training with rhythmic background. *Frontiers in Psychology*, 6.
- BOORMAN, S. [2001]. Petrucci, Ottaviano. In *The New Grove Dictionary of Music and Musicians*. (Vol. 19). Oxford: University Press.
- BORGES, J. L. [1989]. *Obra poética*. S. L. d. Carril, Ed. Barcelona Ediciones Destino. (Original work published 1925-1985)
- BOWERS, P. G. [1993]. Text reading and rereading : Determinants of fluency beyond word recognition. *Journal of Reading Behavior*, 25(2), 133–53.
- BOWERS, P. G. [1995]. Tracing symbol naming speeds unique contributions to reading disabilities over time. *Reading and Writing*, 7(2), 189–216.
- BRINGHURST, R. [1992]. *The elements of typographic style*. Vancouver, CAN: Hartley & Marks.
- BROCHARD, R., DUFOUR, A., & DESPRES, O. [2004]. Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, 54(2), 103–9.
- BROTONS, M. [1994]. Effects of performing conditions on music performance anxiety and performance quality. *Journal of Music Therapy*, 31(1), 63–81.

- BUCHNER, E. F. [1909]. Review of 'The Psychology and Pedagogy of Reading' by E. B. Huey. *Psychological Bulletin*, 6(147-150).
- BURNHILL, P., HARTLEY, J., FRASE, S., & YOUNG, M. [1975]. The typography of college prospectuses: A critique and case history. In L. Evans, & J. Leedham (Eds.), *Aspects of educational technology* (Vol. IX, pp. 233–45). London: Kogan Page.
- C
- CANCER, A., BONACINA, S., LORUSSO, M. L., LANZI, P. L., & ANTONIETTI, A. [2015, Sep]. *Rhythmic Reading Training (RRT): A computer-assisted intervention program for dyslexia*. Proceedings of the Vth International Symposium on Pervasive Computing Paradigms for Mental Health, Milan, Italy.
- CAPPELLETTI, M., WALEY-COHEN, H., BUTTERWORTH, B., & KOPELMAN, M. [2000]. A selective loss of the ability to read and write music. *Neurocase*, 6, 321–32.
- CARRUTHERS, M. J. [1993]. The poet as master-builder: Composition and locational memory in the middle-ages. *New Literary History*, 24(4), 881-904.
- CARRUTHERS, M. J. [2002]. The art of memory and the art of page layout in the Middle Ages. *Diogenes*(196), 20-30.
- CARRUTHERS, M. J. [2008]. *The book of memory: A study of memory in medieval culture* (2nd ed.). Cambridge: University Press.
- CARRUTHERS, M. J. [2009]. The Cherub Figure and the Arts of Memory. *Gesta-International Center of Medieval Art*, 48(2), 99-117.
- CARRUTHERS, M. J. [2014]. The sociable text of the "troilus frontispiece": A different mode of textuality. *Elb*, 81(2), 423-41.
- CARRUTHERS, M. J., & ZIOLOWSKI, J. M. (Eds.) [2004]. *The medieval craft of memory: An anthology of texts and pictures*. Material Texts Series. Philadelphia, USA: University Press.
- CATTELL, J. M. [1886a]. The time it takes to see and name an object. *Mind*, 11, 63–5.
- CATTELL, J. M. [1886b]. The time taken up by cerebral operations. *Mind*, 11, 220–42.
- CHAN, C., & ACKERMANN, B. [2014]. Evidence-informed physical therapy management of performance-related musculoskeletal disorders in musicians. *Frontiers in Psychology*, 5.
- CHOMSKY, N. [1966]. *Cartesian linguistics: A chapter in the history of rationalist thought*. New York, USA: Harper & Row.
- CHOMSKY, N. [2005]. Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22.
- CHOMSKY, N. [2006]. *Language and mind* (3rd ed.). Cambridge: University Press.
- CHOMSKY, N. [2007]. Biolinguistic explorations: Design, development, evolution. *International Journal of Philosophical Studies*, 15(1), 1–21.
- CHOMSKY, N. [2009]. *Cartesian linguistics: A chapter in the history of rationalist thought* (3rd ed.). Cambridge: University Press.
- CHOMSKY, N. [2012]. *The science of language: Interviews with J. McGilbray*. Cambridge: University Press.
- CHOMSKY, N. [2017]. The language capacity: Architecture and evolution. *Psychonomic Bulletin & Review*, 24(1), 200–3.
- CLARKE, E. F. [2006]. Making and hearing meaning in performance. *Nordisk Estetisk Tidsskrift*, 33–34, 24–47.
- COHEN, L., DEHAENE, S., NACCACHE, L., LEHERICY, S., DEHAENE-LAMBERTZ, G., HENAFF, M. A., & MICHEL, F. [2000]. The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123, 291–307.
- COLTHEART, M. [1980]. Iconic memory and visible persistence. *Perception & Psychophysics*, 27(3), 183–228.
- COLTHEART, M., LEA, C. D., & THOMPSON, K. [1974]. Defence of iconic memory. *Quarterly Journal of Experimental Psychology*, 26(NOV), 633–41.
- COULL, J. T. [2004]. fMRI studies of temporal attention: Allocating attention within, or towards, time. *Cognitive Brain Research*, 21(2), 216–26.
- CROSS, I. [2005]. Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical communication* (pp. 27–43). Oxford: University Press.
- CROSS, I. [2009]. The nature of music and its evolution. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (1st ed., pp. 3–13). Oxford: University Press.

- CROSS, I., & TOLBERT, E. [2009]. Music and meaning. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 24–34). Oxford: Oxford University Press.
- CROSS, I., & WOODRUFF, G. E. [2009]. Music as a communicative medium. In R. Botha, & C. Knight (Eds.), *The prehistory of language* (Vol. 1, pp. 113–44). Oxford: Oxford University Press.
- CUMMING, N. [1991]. Analogy in Leonard B. Meyer's theory of musical meaning. In J. Kassler (Ed.), *Metaphor: A musical dimension*. Sydney, AUS: Currency Press.
- CUMMING, R., WILSON, A., LEONG, V., COLLING, L. J., & GOSWAMI, U. [2015]. Awareness of rhythm patterns in speech and music in children with specific language impairments. *Frontiers in Human Neuroscience*, 9.
- D
- D'AUSILIO, A. [2007]. The role of the mirror system in mapping complex sounds into actions. *Journal of Neuroscience*, 27(22), 5847–8.
- D'AUSILIO, A. [2009]. Mirror-Like mechanisms and music. *The scientific world journal*, 9, 1415–22.
- D'AUSILIO, A., ALTENMUELLER, E., BELARDINELLI, M. O., & LOTZE, M. [2006]. Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, 24(3), 955–8.
- DAHL, P. [2013, Apr]. *Music reading as creative practice in music performance*. Proceedings of the 11th Performance Studies Network International Conference, Cambridge, UK.
- DAHL, P. R., & SAMUELS, S. J. [1977]. Teaching children to read using hypothesis-test strategies. *Reading Teacher*, 30(6), 603–6.
- DEHAENE, S., & COHEN, L. [2011]. The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–62.
- DEHAENE, S., COHEN, L., SIGMAN, M., & VINCKIER, F. [2005]. The neural code for written words: A proposal. *Trends in Cognitive Sciences*, 9(7), 335–41.
- DEHAENE, S., LE CLECH, G., POLINE, J. B., LE BIHAN, D., & COHEN, L. [2002]. The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3), 321–5.
- DEHAENE, S., NACCACHE, L., COHEN, L., LE BIHAN, D., MANGIN, J. F., POLINE, J. B., & RIVIERE, D. [2001]. Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4(7), 752–8.
- DEHAENE, S., NAKAMURA, K., JOBERT, A., KUROKI, C., OGAWA, S., & COHEN, L. [2010]. Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area. *NeuroImage*, 49(2), 1837–48.
- DEJERINE, J. J. [1892]. Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *Mémoires de la Société de Biologie*, 4, 61–90.
- DELL, G. S. [1986]. A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- DOUTAUT, V., MATIGNON, D., & CHAIGNE, A. [1998]. Numerical simulations of xylophones: Time-domain modeling of the resonator and of the radiated sound pressure. *Journal of the Acoustical Society of America*, 104(3), 1633–47.
- DOWNING, J. A., & LEONG, C. K. [1982]. *Psychology of reading*. London: Macmillan.
- DRABKIN, W. [2001]. Retrograde. In *The New Grove Dictionary of Music and Musicians*. (Vol. 21). Oxford: University Press.
- DRAI-ZERBIB, V., & BACCINO, T. [2013]. The effect of expertise in music reading: cross-modal competence. *Journal of Eye Movement Research*, 6(5), 10.
- DRAI-ZERBIB, V., BACCINO, T., & BIGAND, E. [2012]. Sight-Reading expertise: Cross modality integration investigated using eye-tracking. *Psychology of Music*, 40(2), 216–35.
- DRAKE, C. [1993]. Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, 53(1), 25–33.
- DRAKE, C., & PALMER, C. [2000]. Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, 74, 1–32.

- DROST, U. C., RIEGER, M., BRASS, M., GUNTER, T. C., & PRINZ, W. [2005a]. Action-effect coupling in pianists. *Psychological Research-Psychologische Forschung*, 69(4), 233–41.
- DROST, U. C., RIEGER, M., BRASS, M., GUNTER, T. C., & PRINZ, W. [2005b]. When hearing turns into playing: Movement induction by auditory stimuli in pianists. *Quarterly Journal of Experimental Psychology*, 58(8), 1376–89.
- DROST, U. C., RIEGER, M., & PRINZ, W. [2007]. Instrument specificity in experienced musicians. *Quarterly Journal of Experimental Psychology*, 60(4), 527–33.
- DRURY, J. E., BAUM, S. R., VALERIOLE, H., & STEINHAEUER, K. [2016]. Punctuation and implicit prosody in silent reading: An ERP study investigating English garden-path sentences. *Frontiers in Psychology*, 7.
- E
- EGERMANN, H., FERNANDO, N., CHUEN, L., & MCADAMS, S. [2015]. Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese Pygmies. *Frontiers of Psychology*, 5, 1–9.
- ENRICI, I., ADENZATO, M., CAPPÀ, S., BARA, B. G., & TETTAMANTI, M. [2011]. Intention processing in communication: A common brain network for language and gestures. *Journal of Cognitive Neuroscience*, 23(9), 2415–31.
- ERDMANN, B., & DODGE, R. [1898]. *Psychologische untersuchungen über das lesen auf experimentellen grundlage*. Halle: Niemeyer Publishers.
- EVANS, G. R. [1975a]. Duc-oculum: Aids to understanding in some medieval treatises on abacus. *Centaurus*, 19(4), 252–63.
- EVANS, G. R. [1975b]. The influence of quadrivium studies in the 11th-century and 12th-century schools. *Journal of Medieval History*, 1(2), 151–64.
- EVANS, P., & SCHUBERT, E. [2008]. Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12(1), 75–99.
- EVERAERT, M. B. H., HUYBREGTS, M. A. C., CHOMSKY, N., BERWICK, R. C., & BOLHUIS, J. J. [2015]. Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences*, 19(12), 729–43.
- F
- FADIGA, L., CRAIGHERO, L., & D'AUSILIO, A. [2008, Jun]. *Broca's Area in language, action, and music*. Proceedings of the III^d Conference on the Neurosciences and Music, Montreal, Canada.
- FEDORENKO, E., MCDERMOTT, J. H., NORMAN-HAIGNERE, S., & KANWISHER, N. [2012]. Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, 108(12), 3289–300.
- FIEBACH, C. J., & SCHUBOTZ, R. I. [2006]. Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca's area and ventral premotor cortex across domains? *Cortex*, 42(4), 499–502.
- FINE, P., BERRY, A., & ROSNER, B. [2006]. The effect of pattern recognition and tonal predictability on sight-singing ability. *Psychology of Music*, 34(4), 431–47.
- FISHER, D. F. [1976]. Spatial factors in reading and search. In R. A. Monty, & J. W. Senders (Eds.), *Eye movement and psychological processes*. Hillsdale, USA: Lawrence Erlbaum Associates.
- FISK, J. (Ed.) [1997]. *Composers on music*. Boston, USA: Northeastern University Press.
- FITCH, W. T., HAUSER, M. D., & CHOMSKY, N. [2005]. The evolution of the language faculty: Clarifications and implications. *Cognition*, 97(2), 179–210.
- FLAUGNACCO, E., LOPEZ, L., TERRIBILI, C., MONTICO, M., ZOIA, S., & SCHON, D. [2015]. Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *Plos One*, 10(9).
- FLAUGNACCO, E., LOPEZ, L., TERRIBILI, C., ZOIA, S., BUDA, S., TILLI, S., . . . SCHON, D. [2014]. Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, 8.
- FRIEDERICI, A. D. [2004]. The neural basis of syntactic processes. *Cognitive Neurosciences Iii, Third Edition*, 789–801.

- FRIEDERICI, A. D., & BRAUER, J. [2009]. Syntactic complexity in the brain. In T. Givón, & M. Shibatani (Eds.), *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution* (pp. 491–505). Amsterdam: John Benjamins Publishing.
- FRIEND, R. [2011]. Bad education? *Strad*, 122(1452), 34–6.
- FURNEAUX, S., & LAND, M. F. [1999]. The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society of Biological Sciences*, 266(1436), 2435–40.
- G
- GABRIELSSON, A. [2009]. The relationship between musical structure and perceived expression. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 119–30). Oxford: University Press.
- GAILLARD, R., NACCACHE, L., PINEL, P., CLEMENCEAU, S., VOLLE, E., HASBOUN, D., . . . COHEN, L. [2006]. Direct intracranial fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron*, 50(2), 191–204.
- GARCÍA-CALVO, A. [1986]. *De la felicidad: Alabanza de lo bueno*. Zamora: Ediciones Lucina.
- GARCÍA-CALVO, A. [1993]. *Análisis de la sociedad del bienestar*. Zamora: Ediciones Lucina.
- GARCÍA-CALVO, A. [2002]. *Registro de recuerdos*. Zamora: Ediciones Lucina.
- GARCÍA-CALVO, A. [2013]. *Mentiras principales*. Zamora: Ediciones Lucina.
- GARRETT, M. F. [1975]. The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–77). New York, USA: Academic Press.
- GILREATH, C. T. [1993]. Graphic cueing of text, the typographic and diagraphic dimensions. *Visible Language*, 27(3), 336–61.
- GLEZER, L. S., EDEN, G., JIANG, X., LUETJE, M., NAPOLIELLO, E., KIMA, J., & RIESENHUBER, M. [2016]. Uncovering phonological and orthographic selectivity across the reading network using fMRI-RA. *NeuroImage*, 138, 248–56.
- GLEZER, L. S., JIANG, X., & RIESENHUBER, M. [2009]. Evidence for highly selective neuronal tuning to whole words in the "Visual Word Form Area". *Neuron*, 62(2), 199–204.
- GLEZER, L. S., KIM, J., RULE, J., JIANG, X., & RIESENHUBER, M. [2015]. Adding words to the brain's visual dictionary: Novel word learning selectively sharpens orthographic representations in the VWFA. *Journal of Neuroscience*, 35(12), 4965–72.
- GLORIE, F. (Ed.) [1968]. *Variae collectiones aenigmatum Merovingicae aetatis*. Corpus Christianorum Series. Turnhout: Brepols.
- GLUSHKO, A., STEINHAUER, K., DEPRIEST, J., & KOELSCH, S. [2016]. Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical expertise. *Plos One*, 11(5).
- GOODALE, M. A., & MILNER, A. D. [1992]. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–5.
- GOOLSBY, T. W. [1994a]. Eye-movement in music reading: Effects of reading-ability, notational complexity, and encounters. *Music Perception*, 12(1), 77–96.
- GOOLSBY, T. W. [1994b]. Profiles of processing: Eye movements during sight-reading. *Music Perception*, 12(1), 97–123.
- GOOLSBY, T. W. [1996]. Time use in instrumental rehearsals: A comparison of experienced, novice, and student teachers. *Journal of Research in Music Education*, 44(4), 286–303.
- GOOLSBY, T. W. [1999]. A comparison of expert and novice music teachers' preparing identical band compositions: An operational replication. *Journal of Research in Music Education*, 47(2), 174–87.
- GOUGH, P. B. [1972]. One second of reading. In J. F. Kavanagh, & I. G. Mattingly (Eds.), *Language by ear and eye*. Cambridge, USA: MIT Press.
- GOULD, E. [2011]. *Behid bars: The definitive guide to music notation*. London: Faber Music.
- GRABE, E., & LOW, E. L. [2000]. *Durational variability in speech and the rhythm class hypothesis*. Pceedings of the VIIth Conference on Laboratory Phonology, Nijmegen, The Netherlands.
- GREENFIELD, P. M. [1991]. Language, tools, and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14(4), 531–50.

- GREUSSAY, P., ARVEILLER, J., BATTIER, M., COLERE, C., DAMASSO, G., ENGLERT, G., & RONCIN, D. [1980]. Musical software: Descriptions and abstractions of sound generation and mixing. *Computer Music Journal*, 4(3), 40–7.
- GREY, T. [2001]. Hanslick, Eduard. In *The New Grove Dictionary of Music and Musicians*. (Vol. 10). Oxford: University Press.
- GRIFFITHS, P. [2001]. Messiaen, Olivier. In *The New Grove Dictionary of Music and Musicians*. (Vol. 16). Oxford: University Press.
- H
- HABIB, M., LARDY, C., DESILES, T., COMMEIRAS, C., CHOBERT, J., & BESSON, M. [2016]. Music and dyslexia: A new musical training method to improve reading and related disorders. *Frontiers in Psychology*, 7.
- HAENGGI, D., & PERFETTI, C. A. [1992]. Individual-differences in reprocessing of text. *Journal of Educational Psychology*, 84(2), 182–92.
- HALL, C., MOORE, J., ANNETT, J., & RODGERS, W. [1997]. Recalling demonstrated and guided movements using imaginary and verbal rehearsal strategies. *Research Quarterly for Exercise and Sport*, 68(2), 136–44.
- HANSLICK, E. [1905]. *The beautiful in music: A contribution to the revival of musical aesthetics* (G. Cohen, Trans.). London: Novello & Co., Ltd. (Original work published 1854)
- HARKINS, F. T., & VAN LIERE, F. (Eds.) [2012]. *Interpretation of scripture theory: A selection of works of Hugh, Andrew, Richard and Godfrey of St Victor*. Victorine Texts in Translation Series. Turnhout: Brepols.
- HARRIS, R. [1989]. How does writing restructure thought? *Language & Communication*, 9(2-3), 99–106.
- HART, C. E. [1971]. *The early charters of Essex*. Leicester: University Press.
- HARTLEY, J. [1978]. *Designing instructional text*. London: Kogan Page.
- HARTLEY, J. [1993]. Recalling structured text - does what goes in determine what comes out. *British Journal of Educational Technology*, 24(2), 84–91.
- HARTLEY, J. [1999]. Readers as text designers: Personalizing the layout of text. *Innovations in Education and Training International*, 36(4), 346–50.
- HARTLEY, J. [2004]. Designing instructional and informational text. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology*. Mahwah, USA: Lawrence Erlbaum.
- HARTLEY, J. [2014]. Current findings from research on structured abstracts: An update. *Journal of the Medical Library Association*, 102(3), 146–8.
- HARTLEY, J., & BETTS, L. [2007]. The effects of spacing and titles on judgement of the effectiveness of structured abstracts. *Journal of the American Society for Information Science and Technology*, 58(14), 2335–40.
- HARTLEY, J., & BURNHILL, P. [1976a]. Explorations in space: A critique of the typography of the British Psychological Society publications. *Bulletin of the British Psychological Society*, 29, 97–107.
- HARTLEY, J., & BURNHILL, P. [1976b]. *Textbook design: A practical guide*. Paris: UNESCO.
- HARTLEY, J., & BURNHILL, P. [1977]. Fifty guidelines for improving instructional text. *Programmed Learning & Educational Technology*, 14(1), 65–73.
- HARTLEY, J., BURNHILL, P., & FRASER, S. [1974]. Typographical problems of journal design. *Applied Ergonomics*, 5(1), 15–20.
- HAUSER, M. D., CHOMSKY, N., & FITCH, W. T. [2002]. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–79.
- HAUSER, M. D., & McDERMOTT, J. [2003]. The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, 6(7), 663–8.
- HAUSER, M. D., YANG, C., BERWICK, R. C., TATTERSALL, I., RYAN, M. J., WATUMULL, J., . . . LEWONTIN, R. C. [2014]. The mystery of language evolution. *Frontiers in Psychology*, 5.
- HAWKES, T. [1977]. *Structuralism and semiotics*. London: Methuen Publishers.

- HEBERT, S., BELAND, R., BECKETT, C., CUDDY, L. L., PERETZ, I., & WOLFORTH, J. [2008]. A case study of music and text dyslexia. *Music Perception*, 25(4), 369–81.
- HEBERT, S., & CUDDY, L. [2006]. Music-reading deficiencies and the brain. *Advances in Cognitive Psychology*, 2(2-3), 199–206.
- HENRIQUE, L. L., & ANTUNES, J. [2003]. Optimal design and physical modelling of mallet percussion instruments. *Acta Acustica United with Acustica*, 89(6), 948–63.
- HIROTANI, M., FRAZIER, L., & RAYNER, K. [2006]. Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54(3), 425–43.
- HOCH, L., POULIN-CHARRONNAT, B., & TILLMANN, B. [2011]. The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology*, 2.
- HOCHBERG, J. [1970]. Components of literacy: Speculations and exploratory research. In H. Levin, & J. Williams (Eds.), *Basic studies in reading*. New York, USA: Basic Books.
- HOGABOAM, T. W., & PERFETTI, C. A. [1978]. Reading skill and role of verbal experience in decoding. *Journal of Educational Psychology*, 70(5), 717–29.
- HOOPE, J. [1991]. *Peirce on signs: Writings on semiotics by Charles Sanders Peirce*. Chapel Hill, USA: University of North Carolina Press.
- HOWE, M. J. A. [1970]. Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology*, 61(3), 214–9.
- HOWE, M. J. A. [1972]. Repeated presentation and retention of meaningful information. *Psychological Reports*, 31, 840–2.
- HSU, S. H., & HUANG, K. C. [2000]. Interword spacing in Chinese text layout. *Perceptual and Motor Skills*, 91(2), 355–65.
- HUDSON, A. [1971]. A Lollard quaternion. *Review of English Studies*, 22, 442–9.
- HUEY, E. B. [1968]. *The psychology and pedagogy of reading: With a review of the history of reading and writing and of methods, texts, and hygiene in reading*. New York, USA: Macmillan. (Original work published 1908)
- I
- INHOFF, A. W., & WU, C. L. [2005]. Eye movements and the identification of spatially ambiguous words during Chinese sentence reading. *Memory & Cognition*, 33(8), 1345–56.
- IRRGANG, M., & EGERMANN, H. [2016]. From motion to emotion: Accelerometer data predict subjective experience of music. *PLoS One*, 11(7).
- J
- JACKSON, M. D. [1980]. Further evidence for a relationship between memory access and reading-ability. *Journal of Verbal Learning and Verbal Behavior*, 19(6), 683–94.
- JACKSON, M. D., & MCCLELLAND, J. L. [1979]. Processing determinants of reading speed. *Journal of Experimental Psychology-General*, 108(2), 151–81.
- JACOBSEN, O. I. [1941]. An analytical study of eye-movements in reading vocal and instrumental music. *Journal of Musicology*, 3, 1–22.
- JANCKE, L. [2012]. The relationship between music and language. *Frontiers in Psychology*, 3.
- JANDREAU, S. M., & BEVER, T. G. [1992]. Phrase-spaced formats improve comprehension in average readers. *Journal of Applied Psychology*, 77(2), 143–6.
- JANDREAU, S. M., MUNCER, S. J., & BEVER, T. G. [1986]. Improving the readability of text with automatic phrase-sensitive formatting. *British Journal of Educational Technology*, 17(2), 128–33.
- JARVELLA, R. J., & LUNDBERG, I. [1987]. Visual constraints in reading: Evidence from nonocular behavior. *Scandinavian Journal of Psychology*, 28(2), 93–103.
- JAVAL, L. E. [1878]. Essai sur la physiologie de la lecture. *Annales d'Oculistique*, 82, 243–53.

- JENTSCHKE, S. [2016]. The relationship between music and language. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd. ed., pp. 343–55). Oxford: Oxford University Press.
- JOBARD, G., CRIVELLO, F., & TZOURIO–MAZOYER, N. [2003]. Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage*, 20(2), 693–712.
- JUSLIN, P. N. [2009]. Emotional responses to music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 131–40). Oxford: University Press.
- K
- KALLINEN, K., & RAVAJA, N. [2006]. Emotion perceived and emotion felt: Same and different. *Musicae Scientiae*, 10(2), 191–213.
- KATZ, J., & PESETSKY, D. [2011]. *The identity thesis for language and music* (Unpublished research document). Institut Jean Nicod, Paris.
- KENDRICK, L. [1988]. *The game of love: Troubadour wordplay*. Berkeley, USA: University Press.
- KENNY, D. T., & ACKERMANN, B. [2015]. Performance-related musculoskeletal pain, depression and music performance anxiety in professional orchestral musicians: A population study. *Psychology of Music*, 43(1), 43–60.
- KENNY, D. T., DAVIS, P., & OATES, J. [2004]. Music performance anxiety and occupational stress amongst opera chorus artists and their relationship with state and trait anxiety and perfectionism. *Journal of Anxiety Disorders*, 18, 757–77.
- KEREN–HAPPUCH, E., CHEN, S. H. A., HO, M. H. R., & DESMOND, J. E. [2014]. A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Human Brain Mapping*, 35(2), 593–615.
- KINSLER, V., & CARPENTER, R. H. S. [1995]. Saccadic eye-movements while reading music. *Vision Research*, 35(10), 1447–58.
- KLEIN, G. A., & KLEIN, H. A. [1973]. Word identification as a function of contextual information. *American Journal of Psychology*, 86(2), 399–406.
- KOELSCH, S., FRITZ, T., SCHULZE, K., ALSOP, D., & SCHLAUG, G. [2005]. Adults and children processing music: An fMRI study. *NeuroImage*, 25(4), 1068–76.
- KOELSCH, S., KASPER, E., SAMMLER, D., SCHULZE, K., GUNTER, T., & FRIEDERICI, A. D. [2004]. Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–7.
- KOPIEZ, R., & LEE, J. I. [2006]. Towards a dynamic model of skills involved in sight-reading music. *Music Education Research*, 8(1), 97–120.
- KOPIEZ, R., & LEE, J. I. [2008]. Towards a general model of skills involved in sight-reading music. *Music Education Research*, 10(1), 41–62.
- KOPIEZ, R., WEIHS, C., LIGGES, U., & LEE, J. I. [2005]. Classification of high and low achievers in a music sight-reading task. *Psychology of Music*, 34(1), 5.
- KOSTKA, M. J. [2000]. The effects of error-detection practice on keyboard sight-reading achievement of undergraduate music majors. *Journal of Research in Music Education*, 48(2), 114–22.
- L
- LABBE, C., & GRANDJEAN, D. [2014]. Musical emotions predicted by feelings of entrainment. *Music Perception*, 32(2), 170–85.
- LABERGE, D., & SAMUELS, S. J. [1974]. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2), 293–323.
- LAND, M. F., & FURNEAUX, S. [1997]. The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 352(1358), 1231–9.
- LARSON, M. C., GUNNAR, M. R., & HERTSGAARD, L. [1991]. The effects of morning naps, car trips, and maternal separation on adrenocortical activity in human infants. *Child Development*, 62(2), 362–72.

- LAWRY, J. A., & LABERGE, D. [1981]. Letter and word code interactions elicited by normally displayed words. *Perception & Psychophysics*, 30(1), 71–82.
- LEARY, D. E. [1980]. 100 years of experimental-psychology: An American perspective. *Psychological Research-Psychologische Forschung*, 42(1-2), 175–89.
- LEBRUN, M. A., MOREAU, P., McNALLY–GAGNON, A., GOULET, G. M., & PERETZ, I. [2012]. Congenital amusia in childhood: A case study. *Cortex*, 48(6), 683–8.
- LEE, C. [1999]. Eye and head coordination in reading: roles of head movement and cognitive control. *Vision Research*, 39(22), 3761–8.
- LEE, J. I. [2003, Sep]. *The role of working memory and short-term memory in sight-reading*. Poceedigs of the Vth International Conference of the European Society for the Cognitive Sciences of Music, Hanover, Germany.
- LEHMANN, A. C., & ERICSSON, K. A. [1996]. Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, 15, 1–29.
- LEHMANN, A. C., & KOPIEZ, R. [2009]. Sight-Reading. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 334–51). Oxford: University Press.
- LEPSIEN, J., & NOBRE, A. C. [2006]. Cognitive control of attention in the human brain: Insights from orienting attention to mental representations. *Brain Research*, 1105, 20–31.
- LEVASSEUR, V. M., MACARUSO, P., PALUMBO, L. C., & SHANKWEILER, D. [2006]. Syntactically cued text facilitates oral reading fluency in developing readers. *Applied Psycholinguistics*, 27(3), 423–45.
- LEVASSEUR, V. M., MACARUSO, P., & SHANKWEILER, D. [2008]. Promoting gains in reading fluency: A comparison of three approaches. *Reading and Writing*, 21(3), 205–30.
- LEVIN, H., & ADDIS, A. B. [1979]. *The eye-voice span*. Cambridge, USA: MIT Press.
- LEVIN, H., & KAPLAN, E. L. [1970]. Grammatical structure and reading. In H. Levin, & J. Williams (Eds.), *Basic studies in reading*. New York, USA: Basic Books.
- LI, X. S., ZHAO, W. C., & POLLATSEK, A. [2012]. Dividing lines at the word boundary position helps reading in Chinese. *Psychonomic Bulletin & Review*, 19(5), 929–34.
- LIANG, K. Y., & ZEGER, S. L. [1986]. Longitudinal data-analysis using generalized linear-models. *Biometrika*, 73(1), 13–22.
- LIANG, K. Y., & ZEGER, S. L. [1995]. Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science*, 10(2), 158–73.
- LIANG, K. Y., ZEGER, S. L., & QAQISH, B. [1992]. Multivariate regression-analyses for categorical-data. *Journal of the Royal Statistical Society Series B-Methodological*, 54(1), 3–40.
- LITTON, M., FROST, A. A. M., & MOHR, P. B. [2003]. The prediction of musical performance anxiety. *Medical Problems of Performing Artists*, 18(3), 120–5.
- LIU, H. C., JIANG, Y., ZHANG, B., MA, L. F., HE, S., & WENG, X. C. [2013]. The orthographic sensitivity to written Chinese in the occipital-temporal cortex. *Experimental Brain Research*, 227(3), 387–96.
- LIVERSEDGE, S. P., DRIEGHE, D., LI, X., YAN, G. L., BAI, X. J., & HYONA, J. [2016]. Universality in eye movements and reading: A trilingual investigation. *Cognition*, 147, 1–20.
- LÖFSTEDT, B. (Ed.) [2003]. *Opera omnia: Virgilius Maro Grammaticus*. Bibliotheca scriptorum Graecorum et Romanorum Teubneriana Series. München: K. G. Sauer.
- LOW, L. E., GRABE, E., & NOLAN, F. [2000]. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43, 377–401.
- LU, Y., PARASKEVOPOULOS, E., HERHOLZ, S. C., KUCHENBUCH, A., & PANTEV, C. [2014]. Temporal processing of audiovisual stimuli is enhanced in musicians: Evidence from magnetoencephalography (MEG). *Plos One*, 9(3).

M

- MACEachREN, A. M. [1995]. *How maps work: Representation, visualization, and design*. London: Guilford Press.
- MADELL, J., & HEBERT, S. [2008]. Eye movements and music reading: Where do we look next? *Music Perception*, 26(2), 157–70.
- MAHALSKI, P. A. [1995]. What happens when students copy notes with different content and layout from an overhead screen - how often do they glance up and how accurate are they. *British Journal of Educational Technology*, 26(1), 5–15.
- MCADAMS, S., CHAIGNE, A., & ROUSSARIE, V. [2004]. The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America*, 115(3), 1306–20.
- MCCLELLAND, J. L., & OREGAN, J. K. [1981]. Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology-Human Perception and Performance*, 7(3), 634–44.
- MCCONKIE, G. W., & RAYNER, K. [1975]. Span of effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17(6), 578–86.
- MCCORMICK, C., & SAMUELS, S. J. [1979]. Word recognition by 2nd graders: Unit of perception and interrelationships among accuracy, latency, and comprehension. *Journal of Reading Behavior*, 11(2), 107–18.
- MCDERMOTT, J., & HAUSER, M. [2005]. The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, 23(1), 29–59.
- MCGOWAN, V. A., WHITE, S. J., JORDAN, T. R., & PATERSON, K. B. [2014]. Aging and the use of interword spaces during reading: Evidence from eye movements. *Psychonomic Bulletin & Review*, 21(3), 740–7.
- MEINZ, E. J., & HAMBRICK, D. Z. [2010]. Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science*, 21(7), 914–9.
- MERKLEY, R., WILKEY, E. D., & MATEJKO, A. A. [2016]. Exploring the origins and development of the visual number form area: A functionally specialized and domain-specific region for the processing of number symbols? *Journal of Neuroscience*, 36(17), 4659–61.
- MERRILL, J., SAMMLER, D., BANGERT, M., GOLDHAHN, D., LOHMANN, G., TURNER, R., & FRIEDERICI, A. D. [2012]. Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, 3, 13.
- MESSIAEN, O. [1949]. *Mode de valeurs et d'intensités*, in *Quatre Etudes de Rythme*. Paris: Editions Durand.
- MESSIAEN, O. [1953]. *Réveil des oiseaux*, pour piano solo et grande orchestre. Paris: Editions Durand. (Revised in 1988)
- MESSIAEN, O. [1956]. *Oiseaux exotiques*, pour piano solo et petit orchestre. Vienna: Universal Edition.
- MESSIAEN, O. [1956–58]. *Catalogue d'oiseaux*, pour piano solo. Paris: Editions Durand.
- MESSIAEN, O. [1960]. *Chronochromie*, pour grand orchestre. Paris: Editions Durand.
- METZGER, B. M. [1968]. When did scribes begin to use writing desks? In M. Metzger (Ed.), *Historical and literary studies*. Grand Rapids, USA: W. B. Eerdmans.
- MEYER, L. B. [1967]. *Music, the arts, and the ideas: Patterns and predictions in Twentieth-Century culture*. Chicago, USA: University Press.
- MEYER, L. B. [1973]. *Explaining music: Essays and explorations*. Berkeley, USA: University of California Press.
- MILNER, J. C., & REGNAULT, F. [1987]. *Dire le vers: Court traité à l'intention des acteurs et des amateurs d'alexandrins*. Paris: Editions du Seuil.
- MIYAGAWA, S., BERWICK, R. C., & OKANOYA, K. [2013]. The emergence of hierarchical structure in human language. *Frontiers in Psychology*, 4.
- MONGELLI, V., COHEN, L., & BARTOLOMEO, P. [2014]. *Music reading in the brain: How musical literacy shapes the ventral cortical visual system* (Master). Institut du Cerveau et de la Moelle Épinière, Paris.
- MULLER, M., BUDDÉ, A., & NETTER, P. [1992]. The influence of need for achievement and type and controllability of the stressor on responses of saliva cortisol. In C. Kirschbaum, G. F. Read, & D. H. Hellhammer (Eds.), *Assessment of hormones and drugs in saliva in biobehavioural research* (pp. 175–202). Seattle, USA: Hogrefe & Huber.

N

- NAKAMURA, K., DEHAENE, S., JOBERT, A., LE BIHAN, D., & KOUIDER, S. [2005]. Subliminal convergence of Kanji and Kana words: Further evidence for functional parcellation of the posterior temporal cortex in visual word perception. *Journal of Cognitive Neuroscience*, 17(6), 954–68.
- NATTIEZ, J. J. [1990]. *Music and discourse: Towards a semiology of music*. Princeton, USA: University Press.
- NIYOGI, P., & BERWICK, R. C. [1997]. Evolutionary consequences of language learning. *Linguistics and Philosophy*, 20(6), 697–719.
- NOBRE, A. C., COULL, J. T., MAQUET, P., FRITH, C. D., VANDENBERGHE, R., & MESULAM, M. M. [2004]. Orienting attention to locations in perceptual versus mental representations. *Journal of Cognitive Neuroscience*, 16(3), 363–73.

O

- OLIVER, M. (Ed.) [1999]. *Settling the score*. London: Faber & Faber Ltd.

P

- PADDISON, M. [2001]. Adorno, Theodor. In *The New Grove Dictionary of Music and Musicians*. (Vol. 1). Oxford: University Press.
- PALISCA, C. V. (Ed.) [1978]. *Hucbald, Guido and John on music: Three medieval treatises*. New Haven, USA: Yale University Press.
- PALMER, C. [1989]. Mapping musical thought to musical performance. *Journal of Experimental Psychology-Human Perception and Performance*, 15(2), 331–46.
- PALMER, C. [1992]. The role of interpretive preferences in music performance. In M. R. Jones, & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 249–62). Washington, USA: American Psychological Association.
- PALMER, C., & DRAKE, C. [1997]. Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale*, 51(4), 369–84.
- PALMER, C., & VAN DE SANDE, C. [1993]. Units of knowledge in music performance. *Journal of Experimental Psychology-Learning Memory and Cognition*, 19(2), 457–70.
- PALMER, C., & VAN DE SANDE, C. [1995]. Range of planning in music performance. *Journal of Experimental Psychology-Human Perception and Performance*, 21(5), 947–62.
- PARÁSSOGLOU, G. M. [1979]. Some thoughts on the postures of the ancient Greeks and Romans when writing on papyrus rolls. *Scrittura e Civiltà*, 3, 2–6.
- PARKES, M. B. [1993]. *Pause and effect: An introduction to the history of punctuation in the West*. Berkeley, USA: University Press.
- PATBERG, J. P., DEWITZ, P., & SAMUELS, S. J. [1981]. The effect of context on the size of the perceptual unit used in word recognition. *Journal of Reading Behavior*, 13(1), 33–48.
- PATEL, A. D. [1998]. Syntactic processing in language and music: Different cognitive operations, similar neural resources? *Music Perception*, 16(1), 27–42.
- PATEL, A. D. [2003]. Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674–81.
- PATEL, A. D., & DANIELE, J. R. [2003a]. An empirical comparison of rhythm in language and music. *Cognition*, 87(1), B35–B45.
- PATEL, A. D., & DANIELE, J. R. [2003b]. Stress-timed vs. syllable-timed music?: A comment on Huron and Ollen (2003). *Music Perception*, 21(2), 273–6.
- PATEL, A. D., GIBSON, E., RATNER, J., BESSON, M., & HOLCOMB, P. J. [1998]. Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717–33.
- PATEL, A. D., IVERSEN, J. R., WASSENAAR, M., & HAGOORT, P. [2008]. Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology*, 22(7-8), 776–89.
- PEARCE, M. T., & HALPERN, A. R. [2015]. Age-Related patterns in emotions evoked by music. *Psychology of Aesthetics Creativity and the Arts*, 9(3), 248–53.

- PEGADO, F., NAKAMURA, K., COHEN, L., & DEHAENE, S. [2011]. Breaking the symmetry: Mirror discrimination for single letters but not for pictures in the Visual Word Form Area. *NeuroImage*, 55(2), 742–9.
- PENTTINEN, M., & HUOVINEN, E. [2011]. The early development of sight-reading skills in adulthood: A study of eye movements. *Journal of Research in Music Education*, 59(2), 196–220.
- PENTTINEN, M., HUOVINEN, E., & YLITALO, A. K. [2015]. Reading ahead: Adult music students' eye movements in temporally controlled performances of a children's song. *International Journal of Music Education*, 33(1), 36–50.
- PERFETTI, C. A., GOLDMAN, S. R., & HOGABOAM, T. W. [1979]. Reading skill and the identification of words in discourse context. *Memory & Cognition*, 7(4), 273–82.
- PESETSKY, D. [2009]. Against taking linguistic diversity at "face value". *Behavioral and Brain Sciences*, 32(5), 464–+.
- PILLSBURY, W. B. [1897]. A study in apperception. *American Journal of Psychology*, 8, 315–93.
- PLATO. [2009]. *Phaedrus*, Oxford World's Classics Series. Oxford: University Press. (Original work published 370 BC approx.)
- Q
- QIAO, E., VINCKIER, F., SZWED, M., NACCACHE, L., VALABREGUE, R., DEHAENE, S., & COHEN, L. [2010]. Unconsciously deciphering handwriting: Subliminal invariance for handwritten words in the visual word form area. *NeuroImage*, 49(2), 1786–99.
- QUINTILLIANUS, M. F. [1970]. *Institutionis Oratoriae*. M. Winterbottom, Ed., Scriptorum Classicorum Bibliotheca Oxoniensis Series. Oxford: Clarendon Press. (Original work published ca 95 AD)
- R
- RASTALL, R. [2008]. *The notation of Western music: An introduction* (2nd rev. ed.). London: Travis & Emery.
- RAYNER, K., KAMBE, G., & DUFFY, S. A. [2000]. The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 53(4), 1061–80.
- RAYNER, K., & MCCONKIE, G. W. [1976]. What guides a readers eye-movements. *Vision Research*, 16(8), 829–37.
- RAYNER, K., & POLLATSEK, A. [1997]. Eye movements, the eye-hand span, and the perceptual span during sight-reading of music. *Current Directions in Psychological Science*, 6(2), 49–53.
- ROADS, C., BATTIER, M., BARLOW, C., BISCHOFF, J., BRUN, H., CHADABE, J., . . . VAGGIONE, H. [1986]. Symposium on computer music composition. *Computer Music Journal*, 10(1), 40–63.
- ROBERTS, C. H., & SKEAT, T. C. [1983]. *The birth of the codex*. London: Published for the British Academy by the Oxford University Press.
- ROSS, B. [2014]. *Music, language and syntactic integration* (Doctoral dissertation). University of Cambridge.
- ROUSSEAU, L., HEBERT, S., & CUDDY, L. L. [2001]. Impaired short temporal interval discrimination in a dyslexic adult. *Brain and Cognition*, 46(1-2), 249–54.
- S
- SAENGER, P. [1982]. Silent reading: Its impact on late Medieval script and society. *Viator*, 13, 366–14.
- SAENGER, P. [1989]. Books of hours and the reading habits of the later middle ages. In R. Chartier (Ed.), *The culture of print: Power and uses of print in early modern Europe* (pp. 143–73). Princeton, USA: University Press.
- SAENGER, P. [1990]. Coupure et séparation des mots sur le Continent au moyen age. In H.-J. Martin, & J. Venzin (Eds.), *Mise en page et mise en texte du livre manuscrit*. Paris: Editions du Cercle de la librairie-Promodis.
- SAENGER, P. [1996]. The impact of the early printed page on the history of reading. *Bulletin de Bibliophilie*, 2, 237–301.

- SAENGER, P. [1997]. *Spaces between words: The origins of silent reading*. Stanford, USA: Stanford University Press.
- SAINIO, M., HYONA, J., BINGUSHI, K., & BERTRAM, R. [2007]. The role of interword spacing in reading Japanese: An eye movement study. *Vision Research*, 47(20), 2575–84.
- SAMUELS, S. J., BEGY, G., & CHEN, C. C. [1976]. Comparison of word recognition speed and strategies of less skilled and more highly skilled readers. *Reading Research Quarterly*, 11(1), 72–86.
- SAMUELS, S. J., DAHL, P., & ARCHWAMETY, T. [1974]. Effect of hypothesis-test training on reading skill. *Journal of Educational Psychology*, 66(6), 835–44.
- SAMUELS, S. J., LABERGE, D., & BREMER, C. D. [1978]. Units of word recognition: Evidence for developmental-changes. *Journal of Verbal Learning and Verbal Behavior*, 17(6), 715–20.
- SAMUELS, S. J., MILLER, N. L., & EISENBERG, P. [1979]. Practice effects on the unit of word recognition. *Journal of Educational Psychology*, 71(4), 514–20.
- SCHON, D., & BESSON, M. [2002]. Processing pitch and duration in music reading: A RT-ERP study. *Neuropsychologia*, 40(7), 868–78.
- SCHON, D., & BESSON, M. [2005]. Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 17(4), 694–705.
- SCHON, D., & TILLMANN, B. [2014, May]. *Short- and long-term rhythmic interventions: Perspectives for language rehabilitation*. Proceedings of the Vth Conference on Neurosciences and Music, Dijon, France.
- SCHRIVER, K. A. [1997]. *Dynamics in document design: Creating texts for readers*. New York, USA: John Wiley & Sons.
- SCHUBERT, E. [2013]. Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology*, 4.
- SCHUBERT, K. (Ed.) [1990]. *J. S. Bach 371 vierstimmige Choräle*. Edition Breitkopf Series. Wiesbaden: Breitkopf & Härtel.
- SCHUBOTZ, R. I., FRIEDERICI, A. D., & VON CRAMON, D. Y. [2000]. Time perception and motor timing: A common cortical and subcortical basis revealed by fMRI. *NeuroImage*, 11(1), 1–12.
- SCHUBOTZ, R. I., & VON CRAMON, D. Y. [2001]. Interval and ordinal properties of sequences are associated with distinct premotor areas. *Cerebral Cortex*, 11(3), 210–22.
- SENDERS, J. W., FISHER, D. F., & MONTY, R. A. (Eds.) [1978]. *Eye movements and the higher psychological functions*. London: Distributed by Wiley.
- SEO, H., & LEE, C. [2002]. Head-free reading of horizontally and vertically arranged texts. *Vision Research*, 42(10), 1325–37.
- SERGENT, J. [1993]. Music, the brain and ravel. *Trends in Neurosciences*, 16(5), 168–72.
- SERGENT, J. [1994]. Brain-imaging studies of cognitive functions. *Trends in Neurosciences*, 17(6), 221–7.
- SERGENT, J., ZUCK, E., TERRIAH, S., & MACDONALD, B. [1992]. Distributed neural network underlying musical sight-reading and keyboard performance. *Science*, 257(5066), 106–9.
- SHAFFER, L. H. [1978]. Timing in motor programming of typing. *Quarterly Journal of Experimental Psychology*, 30(MAY), 333–45.
- SHAFFER, L. H. [1981]. Performances of chopin, bach, and bartok: Studies in motor programming. *Cognitive Psychology*, 13(3), 326–76.
- SHAFFER, L. H. [1982]. Rhythm and timing in skill. *Psychological Review*, 89(2), 109–22.
- SHAFFER, L. H. [1984a]. Motor programming in language production: A tutorial review. *Attention and Performance*, 10, 17–41.
- SHAFFER, L. H. [1984b]. Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 36(4), 577–95.
- SHEN, D. L., LIVERSEDGE, S. P., TIAN, J., ZANG, C. L., CUI, L., BAI, X. J., . . . RAYNER, K. [2012]. Eye movements of second language learners when reading spaced and unspaced Chinese text. *Journal of Experimental Psychology-Applied*, 18(2), 192–202.
- SHIEH, K. K., HSU, S. H., & LIU, Y. C. [2005]. Dynamic chinese text on a single-line display: Effects of presentation mode. *Perceptual and Motor Skills*, 100(3), 1021–35.

- SHUM, J., HERMES, D., FOSTER, B. L., DASTJERDI, M., RANGARAJAN, V., WINAWER, J., . . . PARVIZI, J. [2013a]. A brain area for the visual recognition of numerals. *Journal of Cognitive Neuroscience*, 262–3.
- SHUM, J., HERMES, D., FOSTER, B. L., DASTJERDI, M., RANGARAJAN, V., WINAWER, J., . . . PARVIZI, J. [2013b]. A brain area for visual numerals. *Journal of Neuroscience*, 33(16), 6709–15.
- SINGER, H. [1981]. Teaching the acquisition phase of reading development: An historical perspective. In O. J. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp. 10–28). Hillsdale, USA: Erlbaum.
- SIRAT, C. [1988]. The material conditions of the lateralization of the ductus. In D. d. Kerckhove, & C. J. Lumsden (Eds.), *The alphabet and the brain* (pp. 173–201). Berlin: Springer-Verlag.
- SLATTERY, T. J., & RAYNER, K. [2013]. Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention Perception & Psychophysics*, 75(6), 1275–92.
- SLATTERY, T. J., YATES, M., & ANGELE, B. [2016]. Interword and interletter spacing effects during reading revisited: Interactions with word and font characteristics. *Journal of Experimental Psychology-Applied*, 22(4), 406–22.
- SLEVC, L. R., ROSENBERG, J. C., & PATEL, A. D. [2009]. Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2), 374–81.
- SLOBODA, J. A. [1974a]. The eye-hand span: An approach to the study of sight reading. *Psychology of Music*, 2, 4–10.
- SLOBODA, J. A. [1974b]. Musician's eye - evidence for reorganization of visual memory. *Bulletin of the British Psychological Society*, 27(95), 181–.
- SLOBODA, J. A. [1976a]. Decision times for word and letter search: Wholistic word identification model examined. *Journal of Verbal Learning and Verbal Behavior*, 15(1), 93–101.
- SLOBODA, J. A. [1976b]. Effect of item position on likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology-Revue Canadienne De Psychologie*, 30(4), 228–37.
- SLOBODA, J. A. [1976c]. Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology*, 28(FEB), 1–16.
- SLOBODA, J. A. [1977]. Phrase units as determinants of visual processing in music reading. *British Journal of Psychology*, 68(FEB), 117–24.
- SLOBODA, J. A. [1978a]. Perception of contour in music reading. *Perception*, 7(3), 323–31.
- SLOBODA, J. A. [1978b]. The psychology of music reading. *Psychology of Music*, 6(2), 3–20.
- SLOBODA, J. A. [1978c]. Visual and phonological codes in spelling. *Bulletin of the British Psychological Society*, 31(FEB), 74–.
- SLOBODA, J. A. [1981]. The uses of space in music notation. *Visible Language*, 15(1), 86–110.
- SLOBODA, J. A. [1984]. Experimental studies of music-reading: A review. *Music Perception*, 2(2), 222–36.
- SLOBODA, J. A. [1985]. Expressive skill in 2 pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology-Revue Canadienne De Psychologie*, 39(2), 273–93.
- SLOBODA, J. A., CLARKE, E. F., PARNCUTT, R., & RAEKALLIO, M. [1998]. Determinants of finger choice in piano sight-reading. *Journal of Experimental Psychology-Human Perception and Performance*, 24(1), 185–203.
- SOKOLOV. [1972]. *Inner speech and thought* (G. T. Onischenko, & D. B. Lindsley, Trans.). New York, USA: Plenum Press.
- SOLOMON, J. A., & PELLI, D. G. [1994]. The visual filter mediating letter identification. *Nature*, 369(6479), 395–7.
- SPARSHOTT, F. E., & CUMMING, N. [2001]. Meyer, Leonard. In *The New Grove Dictionary of Music and Musicians*. (Vol. 16). Oxford: University Press.
- SPOEHR, K. T., & SMITH, E. E. [1973]. Role of syllables in perceptual processing. *Cognitive Psychology*, 5(1), 71–89.
- STERNBERG, S. [1970]. Memory scanning: Mental processes revealed by reaction time experiments. In J. S. Antrobus (Ed.), *Cognition and affect* Boston, USA: Little, Brown, & Co.
- SZUCS, D., & SOLTESZ, F. [2007]. Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia*, 45(14), 3190–202.

- SZUCS, D., & SOLTESZ, F. [2008]. The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: An ERP study. *Brain Research*, 1190, 143–58.
- SZUCS, D., & SOLTESZ, F. [2010]. Event-related brain potentials to violations of arithmetic syntax represented by place value structure. *Biological Psychology*, 84(2), 354–67.
- T
- TAIRA, M., MINE, S., GEORGOPOULOS, A. P., MURATA, A., & SAKATA, H. [1990]. Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Experimental Brain Research*, 83(1), 29–36.
- TAN, R. S. K., & LOW, E. L. [2014]. Rhythmic patterning in Malaysian and Singapore English. *Language and Speech*, 57(2), 196–214.
- TAVARES, T. F., ODOWICHUCK, G., ZEHTABI, S., TZANETAKIS, G., & IEEE. [2012, Sep]. *Audio-visual vibraphone transcription in real time*. Proceedings of the XIVth IEEE International Workshop on Multimedia Signal Processing Banff, Canada.
- TAYLOR, I., & TAYLOR, M. M. [1983]. *The psychology of reading*. London: Academic Press.
- TERRY, P., SAMUELS, S. J., & LABERGE, D. [1976]. Effects of letter degradation and letter spacing on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 15(5), 577–85.
- TETTAMANTI, M., ALKADHI, H., MORO, A., PERANI, D., KOLLIAS, S., & WENIGER, D. [2002a]. Neural correlates for the acquisition of natural language syntax. *NeuroImage*, 17(2), 700–9.
- TETTAMANTI, M., MORO, A., PERANI, D., CAPPÀ, S., & FAZIO, F. [2002b]. Specific brain responses in the acquisition of natural language grammatical rules. *Journal of Psychophysiology*, 16(4), 216–26.
- TETTAMANTI, M., ROTONDI, I., PERANI, D., SCOTTI, G., FAZIO, F., CAPPÀ, S. F., & MORO, A. [2009]. Syntax without language: Neurobiological evidence for cross-domain syntactic computations. *Cortex*, 45(7), 825–38.
- TETTAMANTI, M., & WENIGER, D. [2006]. Broca's area: A supramodal hierarchical processor? *Cortex*, 42(4), 491–4.
- THOMSON, S. H. [1969]. *Latin bookhands of the later Middle Ages, 1100-1500*. Cambridge: University Press.
- TILCH, K. [2011]. Neglected accompanist reply. *Strad*, 122(1455), 7–.
- TILLMANN, B. [2012]. Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science*, 4(4), 568–84.
- TILLMANN, B., KOELSEH, S., ESCOFFIER, N., BIGAND, E., LALITTE, P., FRIEDERICI, A. D., & VON CRAMON, D. Y. [2006]. Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, 31(4), 1771–82.
- TOOP, R. [2001]. Stockhausen, Karlheinz. In *The New Grove Dictionary of Music and Musicians*. (Vol. 22). Oxford: University Press.
- TREITLER, L. [1982]. The early history of music writing in the west. *Journal of the American Musicological Society*, 35(2), 237–79.
- TREITLER, L. [1989]. The beginnings of music-writing in the west: Historical and semiotic aspects. *Language & Communication*, 9(2-3), 193–211.
- TREITLER, L. [2007]. *With voice and pen: Coming to know medieval song and how it was made*. Oxford: University Press.
- TRUITT, F. E., CLIFTON, C., POLLASTEK, A., & RAYNER, K. [1997]. The perceptual span and the eye-hand span in sight-reading music. *Visual Cognition*, 4(2), 143–61.
- TULVING, E. [1966]. Subjective organisation and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behaviour*, 5(2), 193–7.
- TURNER, E. G. [1971]. *Greek manuscripts of the ancient world*. Princeton, USA: University Press.
- TURNER, E. G. [1977]. *The typology of the early codex*. Philadelphia, USA: University of Pennsylvania Press.
- TZENG, O. J. [1981]. Relevancy of experimental psychology to reading instruction. In O. J. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology*. Hillsdale, USA: Erlbaum.
- TZENG, O. J., & HUNG, D. L. [1979, Apr]. *Orthography and reading*. Proceedings of the Annual Meeting of the National Reading Conference, San Antonio, USA.

- TZENG, O. J., & HUNG, D. L. [1980]. Reading in a nonalphabetic writing system: Some experimental studies. In J. F. Kavanagh, & R. L. Venezky (Eds.), *Orthography, reading, and dyslexia* Baltimore, USA: University Park Press.
- TZENG, O. J., & HUNG, D. L. [1981]. Linguistic determinism: A written language perspective. In O. J. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp. 237–55). Hillsdale, USA: Erlbaum.
- TZENG, O. J., & SINGER, H. [1981]. *Perception of print: Reading research in experimental psychology*. Hillsdale, USA: Erlbaum.
- V
- VAN UYTFANGHE, M. [1976]. Le latin des hagiographies mérovingiens et la protohistoire du français: A quelle époque a-t-on cessé de parler Latin? *Romanica Gandesia*, 16, 57–75.
- VAN UYTFANGHE, M. [1983]. Histoire du latin, protohistoire des langues romanes, et histoire de la communication: À propos d'un recueil d'études. *Francia*, 11, 579–613.
- VENEZKY, R. I. [1977]. Research in reading processes: A historical perspective. *American Psychologist*, 34, 917–22.
- VERVAINIOTI, A., & ALEXOPOULOS, E. C. [2015]. Job-related stressors of classical instrumental musicians: A systematic qualitative review. *Medical Problems of Performing Artists*, 30(4), 197-+.
- VICTOR, G. J. [1980]. *Ars Rhetorica*. M. S. Celentano, & R. Giomini, Eds., Bibliotheca Scriptorum Graecorum et Romanorum Taubneriana Series. Leipzig: Teubner. (Original work published IVth cent.)
- VINCKIER, F., DEHAENE, S., JOBERT, A., DUBUS, J. P., SIGMAN, M., & COHEN, L. [2007]. Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron*, 55(1), 143–56.
- VOLANS, K. [2005]. *Piano Trio*. (A. Stenberg Ed.). Madrid: Ediciones DLH.
- W
- WANG, D. H., & LIU, H. S. [2014]. Functional Connectivity Architecture of the Human Brain: Not All the Same. *Neuroscientist*, 20(5), 432–8.
- WARRINER, A. B., & HUMPHREYS, K. R. [2008]. Learning to fail: Reoccurring tip-of-the-tongue states. *Quarterly Journal of Experimental Psychology*, 61(4), 535–42.
- WATERS, A. J., TOWNSEND, E., & UNDERWOOD, G. [1998]. Expertise in musical sight-reading: A study of pianists. *British Journal of Psychology*, 89(1), 123–49.
- WATERS, A. J., & UNDERWOOD, G. [1999]. Processing pitch and temporal structures in music reading: Independent or interactive processing mechanisms? *European Journal of Cognitive Psychology*, 11(4), 531–53.
- WATERS, A. J., UNDERWOOD, G., & FINDLAY, J. M. [1997]. Studying expertise in music reading: Use of a pattern matching paradigm. *Perception & Psychophysics*, 89(1), 123–49.
- WATKINS, J. G. [1942]. *Objective measurement of instrumental performance*. New York, USA: Columbia University Press.
- WATTS, L., & NISBET, J. [1974]. *Legibility in children's books: A review of research*. Windsor: NFER.
- WILLIAMS, J. D., RIPPON, G., STONE, B. M., & ANNETT, J. [1995]. Psychophysiological correlates of dynamic imagery. *British Journal of Psychology*, 86, 283–300.
- WINGO, E. O. [1972]. *Latin punctuation in the classical age*. The Hague: Mouton.
- WINSKEL, H., RADACH, R., & LUKSANEYANAWIN, S. [2009]. Eye movements when reading spaced and unspaced Thai and English: A comparison of Thai-English bilinguals and English monolinguals. *Journal of Memory and Language*, 61(3), 339–51.
- WOLF, T. [1976]. A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research*, 5(2), 143–71.
- WORTHY, M. D. [2003]. Rehearsal frame analysis of an expert wind conductor in high school vs college band rehearsals. *Bulletin of the Council for Research in Music Education*(156), 11–9.

WORTHY, M. D. [2006]. Observations of three expert wind conductors in college rehearsals. *Bulletin of the Council for Research in Music Education*(168), 51–61.

Y

YARBROUGH, C., & HENLEY, P. [1999]. The effect of observation focus on evaluations of choral rehearsal excerpts. *Journal of Research in Music Education*, 47(4), 308–18.

YOSHIE, M., KUDO, K., & OHTSUKI, T. [2008]. Effects of Psychological Stress on State Anxiety, Electromyographic Activity, and Arpeggio Performance in Pianists. *Medical Problems of Performing Artists*, 23(3), 120–32.

YOUNG, A. R., & BOWERS, P. G. [1995]. Individual difference and text difficulty determinants of reading fluency and expressiveness. *Journal of Experimental Child Psychology*, 60(3), 428–54.

YOUNG, A. R., BOWERS, P. G., & MACKINNON, G. E. [1996]. Effects of prosodic modeling and repeated reading on poor readers' fluency and comprehension. *Applied Psycholinguistics*, 17(1), 59–84.

YUMOTO, M., MATSUDA, M., ITOH, K., UNO, A., KARINO, S., SAITOH, O., . . . KAGA, K. [2005a]. Auditory imagery mismatch negativity elicited in musicians. *Neuroreport*, 16(11), 1175–8.

YUMOTO, M., UNO, A., ITOH, K., KARINO, S., SAITOH, O., KANEKO, Y., . . . KAGA, K. [2005b]. Audiovisual phonological mismatch produces early negativity in auditory cortex. *Neuroreport*, 16(8), 803–6.

Z

ZEGER, S. L., & LIANG, K. Y. [1986]. Longitudinal data-analysis for discrete and continuous outcomes. *Biometrics*, 42(1), 121–30.

ZEGER, S. L., & LIANG, K. Y. [1991]. Feedback models for discrete and continuous-time series. *Statistica Sinica*, 1(1), 51–64.

ZEGER, S. L., & LIANG, K. Y. [1992]. An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, 11(14-15), 1825–39.

ZEGER, S. L., LIANG, K. Y., & ALBERT, P. S. [1988]. Models for longitudinal data - a generalized estimating equation approach. *Biometrics*, 44(4), 1049–60.

ZHUKOV, K. [2006, Jul]. *Good sight-readers: Born or bred*. Proceedings of the VIIth Australasian Piano Pedagogy Conference, Adelaide, Australia.

ZHUKOV, K. [2014]. Evaluating new approaches to teaching of sight-reading skills to advanced pianists. *Music Education Research*, 16(1), 70–87.

ZHUKOV, K., VINEY, L., RIDDLE, G., TENISWOOD-HARVEY, A., & FUJIMURA, K. [2016]. Improving sight-reading skills in advanced pianists: A hybrid approach. *Psychology of Music*, 44(2), 155–67.

APPENDICES

APPENDICES

CHAPTER 3

APPENDIX 3.A.



Figure 3.A.1.

PIECE 1: J. S. Bach's Chorale, nr 075 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. Presented to participants in A3 paper size, in landscape orientation, with staff size of 13 mm.

(APPENDIX 3.A.)

The image displays a musical score for J. S. Bach's Chorale, nr 075, presented in a modified version. The score is organized into five systems, each consisting of a grand staff (treble and bass clefs). Vertical dashed lines are used to separate the major structural units. The music is in C major, 4/4 time, and consists of 16 measures. The first system contains measures 1-4, the second 5-8, the third 9-12, the fourth 13-15, and the fifth 16. The final measure (16) is a whole note chord.

Figure 3.A.2.

PIECE 1: J. S. Bach's Chorale, nr 075 in the Schubert collection; MODIFIED 1 Version, separating major structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 13 mm.

(APPENDIX 3.A.)

The image displays five systems of musical notation for a chorale. Each system consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. Vertical dashed lines are placed between measures to indicate phrasal boundaries. The notation includes various note values (quarter, eighth, and sixteenth notes), rests, and accidentals (sharps and naturals). The piece concludes with a repeat sign and a fermata over the final note in the fifth system.

Figure 3.A.3.

PIECE 1: J. S. Bach's Chorale, nr 075 in the Schubert collection; MODIFIED 2 Version, separating major and sub-phrasal units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 13 mm.

APPENDIX 3.B.

*Figure 3.B.1.*

PIECE 2: J. S. Bach's Chorale, nr 230 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. Presented to participants in A3 paper size, in landscape orientation, with staff size of 13 mm.

(APPENDIX 3.B.)

Figure 3.B.2.

PIECE 2: J. S. Bach's Chorale, nr 230 in the Schubert collection;
MODIFIED 1 Version, separating major structural units.
Presented to participants in A3 paper size, in portrait
orientation, with staff size of 13 mm.

(APPENDIX 3.B.)

The image displays five systems of musical notation for a chorale. Each system consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The music is written in G minor (one flat) and common time (C). Vertical dashed lines are placed between the staves in each system to delineate phrasal units. The first system shows a simple harmonic structure. The second system introduces a more complex rhythmic pattern in the bass line. The third system continues with similar harmonic and rhythmic elements. The fourth system shows a variation in the bass line's rhythm. The fifth system concludes the piece with a final cadence in both staves.

Figure 3.B.3.

PIECE 2: J. S. Bach's Chorale, nr 230 in the Schubert collection; MODIFIED 2 Version, separating major and sub-phrasal units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 13 mm.

APPENDIX 3.C.

*Figure 3.C.1.*

PIECE 3: J. S. Bach's Chorale, nr 336 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. Presented to participants in A3 paper size, in landscape orientation, with staff size of 13 mm.

(APPENDIX 3.C.)

The image displays five systems of musical notation for a chorale. Each system consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The key signature is G major (one sharp) and the time signature is common time (C). Vertical dashed lines are placed between the staves at the end of each measure, separating the five measures of the piece. The first measure begins with a common time signature 'C'. The notation includes various rhythmic values such as quarter notes, eighth notes, and sixteenth notes, along with rests and accidentals.

Figure 3.C.2.

PIECE 3: J. S. Bach's Chorale, nr 336 in the Schubert collection; MODIFIED 1 Version, separating major structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 13 mm.

(APPENDIX 3.C.)

The image displays five systems of musical notation for a chorale. Each system consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The key signature is one sharp (F#) and the time signature is common time (C). Vertical dashed lines are placed between measures to separate major and sub-phrasal units. The first system starts with a common time signature. The second system begins with a common time signature. The third system begins with a common time signature. The fourth system begins with a common time signature. The fifth system begins with a common time signature.

Figure 3.C.3.

PIECE 3: J. S. Bach's Chorale, nr 336 in the Schubert collection;
 MODIFIED 2 Version, separating major and sub-phrasal units.
 Presented to participants in A3 paper size, in portrait
 orientation, with staff size of 13 mm.

CHAPTER 4

APPENDIX 4.A.



Figure 4.A.1.

PIECE 1: J. S. Bach's Chorale, nr 031 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. Presented to participants in A3 paper size, in landscape orientation, with staff size of 12.8 mm.

(APPENDIX 4.A.)

The image displays five systems of musical notation for a chorale. Each system consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The key signature is G minor (two flats) and the time signature is common time (C). The notation includes various rhythmic values such as quarter, eighth, and sixteenth notes, as well as rests. Vertical dashed lines are placed between measures in each system to delineate structural units. The first system shows a simple harmonic structure. The second system introduces more complex rhythmic patterns in the bass line. The third system features a more active bass line with sixteenth notes. The fourth system continues with similar rhythmic complexity. The fifth system concludes with a final cadence in the bass line.

Figure 4.A.2.

PIECE 1: J. S. Bach's Chorale, nr 031 in the Schubert collection; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12.8 mm.

APPENDIX 4.B.

*Figure 4.B.1.*

PIECE 2: J. S. Bach's Chorale, nr 285 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. Presented to participants in A3 paper size, in landscape orientation, with staff size of 12.8 mm.

(APPENDIX 4.B.)

The image displays five systems of musical notation for a chorale in D major, 3/4 time. Each system consists of two staves: a treble clef staff and a bass clef staff. The music is divided into five measures by vertical dotted lines. The first measure is a whole note chord. The second measure contains a half note chord in the treble and a half note chord in the bass. The third measure contains a half note chord in the treble and a half note chord in the bass. The fourth measure contains a half note chord in the treble and a half note chord in the bass. The fifth measure contains a half note chord in the treble and a half note chord in the bass. The notation is a modified version of the original chorale, designed to separate major and minor structural units.

Figure 4.B.2.

PIECE 2: J. S. Bach's Chorale, nr 285 in the Schubert collection; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12.8 mm.

APPENDIX 4.C.

Re Mi O Mi Fa Mi Mi

✓ ✓ ✓ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✓

Bord Sib O

Re Mi Re Do

✓ ✓ ✗ ✗ ✓ ✗ ✗ ✓

✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗

Do ↓ O Sol#

Do# Re Re O

✓ ✓ ✗ ✗ ✓ ✗ ✗ ✓

✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓

✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗

Re

Adds a beat - Mi Do

✓ ✓ ✓ ✓ ✓ ✓ ✓ ✗ ✗ ✓

✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓

✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗

Mi Re Mi

Figure 4.C.

EXAMPLE OF MISTAKE CODING: first reading of Piece 2, Modified version, by participant 1.

CHAPTER 5

APPENDIX 5.A.

The image displays three systems of musical notation for J.S. Bach's Chorale nr 005. Each system consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The key signature is two flats (B-flat and E-flat), and the time signature is common time (C). The music is written in a conventional style with various dynamic markings and phrasing slurs.

System 1: Treble staff dynamics: *mp*, *p*, *mp*, *p mp*, *p mf*. Bass staff dynamics: *mp*, *p mp*, *p mp*, *p mf*.

System 2: Treble staff dynamics: *mp f*, *mf f*, *mf*, *f*, *mp*. Bass staff dynamics: *mp*, *mf*, *f*, *mp*.

System 3: Treble staff dynamics: *p*, *mp p*, *mf*, *mp mf*. Bass staff dynamics: *p*, *mp p*, *mf*, *mp mf*.

Figure 5.A.1.

PIECE 1 - TWO VOICES: J. S. Bach's Chorale, nr 005 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in *portrait* orientation, with staff size of 12.6 mm.

(APPENDIX 5.A.)

The image displays five systems of musical notation for a piano piece. Each system consists of a grand staff (treble and bass clefs) with dynamic markings below the notes. Vertical dashed lines separate the systems into major and minor structural units.

- System 1:** Treble clef starts with *mp*. Bass clef starts with *p*. Treble clef notes: G4, A4, B4, C5. Bass clef notes: G3, A3, B3, C4.
- System 2:** Treble clef notes: D5, E5, F5, G5. Bass clef notes: D4, E4, F4, G4.
- System 3:** Treble clef notes: A5, B5, C6, B5, A5. Bass clef notes: A3, B3, C4, B3, A3.
- System 4:** Treble clef notes: G5, F5, E5, D5. Bass clef notes: G3, F3, E3, D3.
- System 5:** Treble clef notes: C5, B4, A4, G4. Bass clef notes: C4, B3, A3, G3.

Detailed description of dynamics and structural units:

- System 1:** Treble clef starts with *mp*. Bass clef starts with *p*. Treble clef notes: G4, A4, B4, C5. Bass clef notes: G3, A3, B3, C4.
- System 2:** Treble clef notes: D5, E5, F5, G5. Bass clef notes: D4, E4, F4, G4. Dynamics: *p* (Treble), *p* (Bass).
- System 3:** Treble clef notes: A5, B5, C6, B5, A5. Bass clef notes: A3, B3, C4, B3, A3. Dynamics: *mp* (Treble), *p* (Bass).
- System 4:** Treble clef notes: G5, F5, E5, D5. Bass clef notes: G3, F3, E3, D3. Dynamics: *p* (Treble), *mp* (Bass).
- System 5:** Treble clef notes: C5, B4, A4, G4. Bass clef notes: C4, B3, A3, G3. Dynamics: *mp* (Treble), *mp* (Bass).

Figure 5.A.2.

PIECE 1 - TWO VOICES: J. S. Bach's Chorale, nr 005 in the Richter collection; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12.6 mm.

(APPENDIX 5.A.)

The image displays three systems of musical notation for a piano accompaniment of J.S. Bach's Chorale nr 005. Each system consists of a grand staff with a treble and bass clef. The music is in G minor (three flats) and common time (C). The first system contains five measures with dynamic markings *mp*, *p*, *mp*, *p*, and *mp* in the treble staff, and *p* and *mf* in the bass staff. The second system contains four measures with dynamic markings *mp*, *f*, *mf*, and *f* in the treble staff, and *mf*, *f*, and *mp* in the bass staff. The third system contains four measures with dynamic markings *p*, *mp*, *p*, *mf*, *mp*, and *mf* in the treble staff, and *mf*, *mp*, and *mf* in the bass staff. The notation includes various note values, rests, and phrasing slurs.

Figure 5.A.3.

PIECE 1 - THREE VOICES: J. S. Bach's Chorale, nr 005 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in *portrait* orientation, with staff size of 12.6 mm.

(APPENDIX 5.A.)

The musical score consists of five systems of music, each with a grand staff (treble and bass clefs). The key signature is B-flat major (two flats) and the time signature is common time (C). The dynamics are marked as follows:

- System 1: *mp*, *p*, *p*, *mp*, *p*, *mp*
- System 2: *mp*, *mp*, *p*
- System 3: *mf*, *mp*, *f*, *mf*, *f*
- System 4: *mf*, *f*, *mp*, *p*
- System 5: *p*, *mp*, *p*, *mf*, *mp*, *mf*

Vertical dashed lines separate the major and minor structural units.

Figure 5.A.4.

PIECE 1 - THREE VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12.6 mm.

APPENDIX 5.B.

The musical score consists of three systems, each with a treble and bass staff. The key signature is three sharps (F#, C#, G#) and the time signature is common time (C). Dynamics are indicated by letters in italics: *p* (piano), *mp* (mezzo-piano), *f* (forte), and *mf* (mezzo-forte). The first system has dynamics *p*, *mp*, *p*, *f*, *mf*, and *f*. The second system has dynamics *mf*, *f*, *mf*, *f*, and *mf*. The third system has dynamics *f*, *mp*, *p*, *f*, *mf*, *mp*, *p*, *mp*, and *p*. The score ends with a double bar line.

Figure 5.B.1.

PIECE 2 - TWO VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in *portrait* orientation, with staff size of 12.6 mm.

(APPENDIX 5.B.)

The figure displays five systems of musical notation for a two-voice chorale. Each system consists of a treble clef staff and a bass clef staff. The key signature is three sharps (F#, C#, G#) and the time signature is common time (C). Dynamic markings are placed below the notes in the treble staff. Vertical dashed lines indicate the boundaries of structural units. The dynamics and structural units across the systems are as follows:

- System 1:** Treble staff starts with *p*. The first unit (measures 1-2) is marked *p*. The second unit (measures 3-4) is marked *mp*. The third unit (measures 5-6) is marked *p*. The fourth unit (measures 7-8) is marked *f*. The fifth unit (measures 9-10) is marked *mf*.
- System 2:** Treble staff starts with *mf*. The first unit (measures 1-2) is marked *mf*. The second unit (measures 3-4) is marked *f*. The third unit (measures 5-6) is marked *mf*. The fourth unit (measures 7-8) is marked *mf*.
- System 3:** Treble staff starts with *mf*. The first unit (measures 1-2) is marked *mf*. The second unit (measures 3-4) is marked *f*. The third unit (measures 5-6) is marked *mf*. The fourth unit (measures 7-8) is marked *f*. The fifth unit (measures 9-10) is marked *mf*.
- System 4:** Treble staff starts with *mf*. The first unit (measures 1-2) is marked *mf*. The second unit (measures 3-4) is marked *f*. The third unit (measures 5-6) is marked *mp*. The fourth unit (measures 7-8) is marked *p*. The fifth unit (measures 9-10) is marked *f*. The sixth unit (measures 11-12) is marked *mf*.
- System 5:** Treble staff starts with *mf*. The first unit (measures 1-2) is marked *mf*. The second unit (measures 3-4) is marked *mp*. The third unit (measures 5-6) is marked *p*. The fourth unit (measures 7-8) is marked *mp*. The fifth unit (measures 9-10) is marked *p*.

Figure 5.B.2.

PIECE 2 - TWO VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12.6 mm.

(APPENDIX 5.B.)

The image displays three systems of musical notation for a piano accompaniment. Each system consists of a grand staff with a treble clef on the upper staff and a bass clef on the lower staff. The key signature is three sharps (F#, C#, G#) and the time signature is common time (C). The music is written in a conventional style with various dynamic markings and phrasing slurs.

System 1: Treble clef starts with a piano (*p*) dynamic. The first measure has a piano (*p*) dynamic. The second measure has a mezzo-piano (*mp*) dynamic. The third measure has a piano (*p*) dynamic. The fourth measure has a forte (*f*) dynamic. The fifth measure has a mezzo-forte (*mf*) dynamic. The sixth measure has a forte (*f*) dynamic. The bass clef starts with a piano (*p*) dynamic and continues with a mezzo-forte (*mf*) dynamic.

System 2: Treble clef starts with a mezzo-forte (*mf*) dynamic. The first measure has a mezzo-forte (*mf*) dynamic. The second measure has a forte (*f*) dynamic. The third measure has a mezzo-forte (*mf*) dynamic. The fourth measure has a forte (*f*) dynamic. The fifth measure has a mezzo-forte (*mf*) dynamic. The sixth measure has a forte (*f*) dynamic. The bass clef starts with a mezzo-forte (*mf*) dynamic and continues with a mezzo-forte (*mf*) dynamic.

System 3: Treble clef starts with a forte (*f*) dynamic. The first measure has a mezzo-piano (*mp*) dynamic. The second measure has a piano (*p*) dynamic. The third measure has a forte (*f*) dynamic. The fourth measure has a mezzo-forte (*mf*) dynamic. The fifth measure has a mezzo-piano (*mp*) dynamic. The sixth measure has a piano (*p*) dynamic. The seventh measure has a mezzo-piano (*mp*) dynamic. The eighth measure has a piano (*p*) dynamic. The bass clef starts with a mezzo-forte (*mf*) dynamic and continues with a mezzo-forte (*mf*) dynamic.

Figure 5.B.3.

PIECE 2 - THREE VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in *portrait* orientation, with staff size of 12.6 mm.

(APPENDIX 5.B.)

The musical score consists of five systems of music, each with a grand staff (treble and bass clefs). The key signature is three sharps (F#, C#, G#) and the time signature is common time (C). The dynamics are indicated by letters: *p* (piano), *mp* (mezzo-piano), *f* (forte), and *mf* (mezzo-forte). Vertical dashed lines separate major and minor structural units.

- System 1:** Treble clef starts with a piano (*p*) dynamic. The bass clef has a single note. The system ends with a mezzo-forte (*mf*) dynamic.
- System 2:** Treble clef starts with mezzo-forte (*mf*). The system ends with mezzo-forte (*mf*).
- System 3:** Treble clef starts with mezzo-forte (*mf*). The system ends with mezzo-forte (*mf*).
- System 4:** Treble clef starts with mezzo-forte (*mf*). The system ends with mezzo-forte (*mf*).
- System 5:** Treble clef starts with mezzo-forte (*mf*). The system ends with piano (*p*).

Figure 5.B.4.

PIECE 2 - THREE VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12.6 mm.

CHAPTER 6

APPENDIX 6.A.

The image shows a musical score for a Violin (Vln.) in G minor, 3/4 time. The score is divided into three staves, with measure numbers 5 and 8 indicated at the beginning of the second and third staves respectively. The first staff starts with a *p* dynamic marking. The second staff contains dynamic markings *mf*, *mp*, *mf*, and *mp*. The third staff contains dynamic markings *p*, *f*, *mf*, *f*, *p*, *mp*, *p*, *mf*, and *p*. The music features a mix of eighth and sixteenth notes, often beamed together, with some measures containing slurs and accents.

Figure 6.A.1.

PIECE 1: style composition imitating J. S. Bach's instrumental solo writing; CONVENTIONAL Version, using the default settings of a commercial notation software (Sibelius 7.1.3). Presented to participants in A3 paper size, in *portrait* orientation, with staff size of 12 mm.

(APPENDIX 6.A.)

The musical score consists of five staves of music in a 3/4 time signature. The key signature changes from G minor to C major in the second staff and back to G minor in the fifth staff. Vertical dashed lines separate the major and minor structural units. The dynamic markings are as follows:

- Staff 1: *p* (at the start), *mf* (at the end of the first unit), *mp* (at the end of the second unit).
- Staff 2: *mf* (at the start), *mf* (at the end of the first unit), *mp* (at the end of the second unit).
- Staff 3: *mp* (at the start), *mf* (at the end of the first unit), *mp* (at the end of the second unit), *mf* (at the end of the third unit), *mp* (at the end of the fourth unit).
- Staff 4: *f* (at the start), *mf* (at the end of the first unit), *mp* (at the end of the second unit), *p* (at the end of the third unit), *f* (at the end of the fourth unit), *mf* (at the end of the fifth unit).
- Staff 5: *f* (at the start), *p* (at the end of the first unit), *mp* (at the end of the second unit), *p* (at the end of the third unit), *mf* (at the end of the fourth unit), *p* (at the end of the fifth unit).

Figure 6.A.2.

PIECE 1: style composition imitating J. S. Bach's instrumental solo writing; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12 mm.

APPENDIX 6.B.

The image displays a musical score for a Violin (Vln.) in G major, 3/4 time. The score is presented in three staves, with dynamic markings indicated below the notes. The first staff begins with a *p* (piano) marking, followed by *mp* (mezzo-piano), *f* (forte), *mf* (mezzo-forte), *mp*, and *pp* (pianissimo). The second staff starts with *mp*, followed by *mf*, *p*, *pp*, *mp*, *p*, *mf*, *f*, and *mf*. The third staff begins with *ff* (fortissimo), followed by *f*, *mf*, *pp*, *p*, and *pp*. The score concludes with a double bar line.

Figure 6.B.1.

PIECE 2: style composition imitating J. S. Bach's instrumental solo writing; CONVENTIONAL Version, using the default settings of a commercial notation software (Sibelius 7.1.3.). Presented to participants in A3 paper size, in *portrait* orientation, with staff size of 12 mm.

(APPENDIX 6.B.)

The musical score consists of five staves of music in G major, C major, and G major. The score is divided into major and minor structural units. Dynamic markings include *p*, *mp*, *f*, *mf*, *pp*, and *ff*.

Staff 1: *p* *mp* *f* *mf* *mp*

Staff 2: *pp* *p* *pp*

Staff 3: *mf* *mp* *mf p* *pp* *mp p*

Staff 4: *mf f* *mf* *ff* *f* *mf*

Staff 5: *pp* *p* *pp*

Figure 6.B.2.

PIECE 2: style composition imitating J. S. Bach's instrumental solo writing; MODIFIED Version, separating major and minor structural units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 12 mm.

CHAPTER 7

APPENDIX 7.A.

The image displays a musical score for J.S. Bach's Chorale, nr 025, presented in a conventional layout. It consists of two systems of piano and bass staves. The first system contains three measures. The piano part (top staff) begins with a *pp* dynamic, followed by a *mp* dynamic, and ends with a *f* dynamic. The bass part (bottom staff) is marked with *pp*, *mp*, and *f* dynamics. The second system contains three measures. The piano part (top staff) is marked with *mp* and *pp* dynamics. The bass part (bottom staff) is marked with *mp* and *pp* dynamics. The score uses a common time signature (C) and features various musical notations including slurs, ties, and dynamic markings.

Figure 7.A.1.

EXERCISE 1 - PIECE 1: J. S. Bach's Chorale, nr 025 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

(APPENDIX 7.A.)

The musical score consists of three systems, each with a treble and bass staff. The first system begins with a piano (*pp*) dynamic in the bass staff and mezzo-piano (*mp*) in the treble staff. The second system starts with mezzo-piano (*mp*) in both staves and reaches a forte (*f*) dynamic in the treble staff. The third system begins with mezzo-piano (*mp*) in both staves and concludes with a pianissimo (*pp*) dynamic. The score is in common time (C) and features a variety of note values and rests, with phrasal units indicated by brackets and slurs.

Figure 7.A.2.

EXERCISE 1 - PIECE 1: J. S. Bach's Chorale, nr 025 in the Richter collection; MODIFIED Version, separating only *phrasal* units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

APPENDIX 7.B.

The image displays two systems of musical notation for a chorale by J.S. Bach. Each system consists of a grand staff with a treble clef on the upper staff and a bass clef on the lower staff. The time signature is common time (C). The first system spans three measures. The first measure is marked with a forte (*f*) dynamic. The second measure is marked with a mezzo-piano (*mp*) dynamic. The third measure is marked with a pianissimo (*pp*) dynamic, followed by a forte (*f*) dynamic in the final part of the measure. The second system also spans three measures. The first measure is marked with a mezzo-piano (*mp*) dynamic. The second measure is marked with a pianissimo (*pp*) dynamic. The third measure concludes with a final chord. The notation includes various note values, rests, and dynamic markings.

Figure 7.B.1.

EXERCISE 1 - PIECE 2: J. S. Bach's Chorale, nr 266 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

(APPENDIX 7.B.)

The image displays three systems of musical notation for a modified version of J.S. Bach's Chorale, nr 266. Each system consists of a grand staff (treble and bass clefs) with dynamic markings and phrasal units indicated by brackets and slurs.

- System 1:** Treble clef starts with a half note G4. Bass clef starts with a half note F3. Dynamics: *pp* (first measure), *f* (second measure). Phrasal units: first measure (bracketed), second measure (bracketed), and a slur over the final two measures.
- System 2:** Treble clef starts with a half note G4. Bass clef starts with a half note F3. Dynamics: *mp* (first measure), *pp* (second measure), *f* (third measure). Phrasal units: first measure (bracketed), second measure (bracketed), and a slur over the final two measures.
- System 3:** Treble clef starts with a half note G4. Bass clef starts with a half note F3. Dynamics: *mp* (first measure), *pp* (second measure). Phrasal units: first measure (bracketed), second measure (bracketed), and a slur over the final two measures.

Figure 7.B.2.

EXERCISE 1 - PIECE 2: J. S. Bach's Chorale, nr 266 in the Richter collection; MODIFIED Version, separating only *phrasal* units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

APPENDIX 7.C.

The image displays two systems of musical notation for a chorale by J.S. Bach. Each system consists of a treble and a bass staff. The first system begins with a treble staff marked *f* and a bass staff marked *pp*. The second system begins with a treble staff marked *mp* and a bass staff marked *pp*. The music is in C major, 4/4 time, and features a variety of note values and rests, with some notes tied across measures. The notation includes slurs, phrasing slurs, and dynamic markings.

Figure 7.C.1.

EXERCISE 2 - PIECE 1: J. S. Bach's Chorale, nr 021 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

(APPENDIX 7.C.)

The image displays two systems of musical notation for a piano piece. The first system consists of two measures. The first measure is marked with a forte dynamic (*f*), and the second measure is marked with a pianissimo dynamic (*pp*). The second system also consists of two measures. The first measure is marked with a mezzo-piano dynamic (*mp*), and the second measure is marked with a pianissimo dynamic (*pp*). The notation includes treble and bass staves with various note values, rests, and dynamic markings.

Figure 7.C.2.

EXERCISE 2 - PIECE 1: J. S. Bach's Chorale, nr 021 in the Richter collection; MODIFIED Version, separating only *sub-phrasal* units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

APPENDIX 7.D.

The image displays two systems of musical notation for a chorale. The first system consists of two staves (treble and bass clefs) with a common time signature (C). The treble staff begins with a dynamic marking of *f* (forte). The bass staff begins with a dynamic marking of *f*. The second system also consists of two staves. The treble staff begins with a dynamic marking of *mp* (mezzo-piano). The bass staff begins with a dynamic marking of *pp* (pianissimo). The music is written in a key signature of one sharp (F#) and a common time signature (C). The notation includes various note values, rests, and dynamic markings.

Figure 7.D.1.

EXERCISE 2 - PIECE 2: J. S. Bach's Chorale, nr 317 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

(APPENDIX 7.D.)

The image displays a musical score for J.S. Bach's Chorale, nr 317, in a modified version. The score is presented in two systems, each with two staves (treble and bass clef). The first system shows a four-measure phrase. The first two measures are marked with a forte (*f*) dynamic, and the last two measures are marked with a mezzo-piano (*mp*) dynamic. The second system shows a four-measure phrase. The first two measures are marked with a pianissimo (*pp*) dynamic, and the last two measures are marked with a mezzo-piano (*mp*) dynamic. The score is written in G major (one sharp) and common time (C). The notes are: System 1, Measure 1: Treble (G4, A4, B4, C5), Bass (G3, A3, B3, C4); System 1, Measure 2: Treble (A4, B4, C5, B4), Bass (A3, B3, C4, B3); System 1, Measure 3: Treble (B4, C5, B4, A4), Bass (B3, C4, B3, A3); System 1, Measure 4: Treble (A4, B4, C5, B4), Bass (A3, B3, C4, B3); System 2, Measure 1: Treble (G4, A4, B4, C5), Bass (G3, A3, B3, C4); System 2, Measure 2: Treble (A4, B4, C5, B4), Bass (A3, B3, C4, B3); System 2, Measure 3: Treble (B4, C5, B4, A4), Bass (B3, C4, B3, A3); System 2, Measure 4: Treble (A4, B4, C5, B4), Bass (A3, B3, C4, B3). The score is written in a modified version, separating only sub-phrasal units.

Figure 7.D.2.

EXERCISE 2 - PIECE 2: J. S. Bach's Chorale, nr 317 in the Richter collection; MODIFIED Version, separating only *sub-phrasal* units. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

APPENDIX 7.E.

The image displays two systems of musical notation for a chorale by J.S. Bach. Each system consists of a grand staff with a treble clef on the upper staff and a bass clef on the lower staff. The key signature is three sharps (F#, C#, G#) and the time signature is common time (C). The first system begins with a dynamic marking of *mp* (mezzo-piano) and features a crescendo leading to a dynamic marking of *f* (forte). The second system begins with a dynamic marking of *pp* (pianissimo). The notation includes various note values, rests, and phrasing slurs across both staves of each system.

Figure 7.E.1.

EXERCISE 3 - PIECE 1: J. S. Bach's Chorale, nr 008 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

(APPENDIX 7.E.)

The image displays a musical score for J.S. Bach's Chorale, nr 008, in a modified version. The score is presented in two systems, each with a grand staff (treble and bass clefs). The first system consists of three measures. The first measure is marked *mp* (mezzo-piano) and the second and third measures are marked *f* (forte). The second system consists of three measures, all marked *pp* (pianissimo). The key signature is one sharp (F#) and the time signature is common time (C). The notation includes various note values, rests, and dynamic markings.

Figure 7.E.2.

EXERCISE 3 - PIECE 1: J. S. Bach's Chorale, nr 008 in the Richter collection; MODIFIED Version, separating only *symbols* ('proportional notation'). Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

APPENDIX 7.F.

The image displays two systems of musical notation for a chorale. Each system consists of a grand staff with a treble clef on the upper staff and a bass clef on the lower staff. The key signature is two flats (B-flat and E-flat), and the time signature is common time (C). The first system contains four measures. The first measure is marked *pp*, the second *mp*, the third *f*, and the fourth *pp*. Phrasing slurs are placed over the first two measures of the first system and the last two measures of the first system. The second system contains three measures, with a final double bar line at the end. Slurs are placed over the first two measures of the second system and the final measure of the second system.

Figure 7.F.1.

EXERCISE 3 - PIECE 2: J. S. Bach's Chorale, nr 012 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

(APPENDIX 7.F.)

The image displays a musical score for a chorale, presented in two systems. The score is written for piano and consists of two staves: a treble clef staff and a bass clef staff. The key signature is three flats (B-flat, E-flat, A-flat), and the time signature is common time (C). The first system contains four measures. The dynamic markings are *pp* (pianissimo) for the first measure, *mp* (mezzo-piano) for the second, *f* (forte) for the third, and *pp* for the fourth. The second system contains three measures. The notation includes various note values, rests, and phrasing slurs. The piece concludes with a double bar line.

Figure 7.F.2.

EXERCISE 3 - PIECE 2: J. S. Bach's Chorale, nr 012 in the Richter collection; MODIFIED Version, separating only *symbols* ('proportional notation'). Presented to participants in A3 paper size, in portrait orientation, with staff size of 14 mm.

APPENDIX 7.G.

Table 7.A.

List of Chorales in the Richter Edition

Chor.	TITLE	Repe.	Real	4 vo.	C	Anac.	Regu.	F = K	Chos.	24 b.
1	Ach bleib bei uns, Herr Jesu Christ	0	1	1	1	1	0	1	4	
2	Ach Gott, erhör mein Seufzen!	0	2	1	1	1	0	1	4	
3	Ach Gott und Herr	1	3	1	1	1	0	1	4	
4	Ach Gott und Herr	1	3	1	1	1	0	1	4	
5	Ach Gott, vom Himmel sieh' darein	1	4	1	1	1	1	0	4	
6	Ach Gott, vom Himmel sieh' darein	1	4	1	1	1	0	0	3	
7	Ach Gott, vom Himmel sieh' darein	1	4	1	1	1	0	0	3	
8	Ach Gott, wie manches Herzeleid	1	5	1	1	1	1	1	5*	1
9	Ach Gott, wie manches Herzeleid	1	5	1	0	0	1	1	3	
10	Ach, was soll ich Sünder machen	0	6	1	1	0	1	1	4	
11	Ach wie flüchtig, ach wie nichtig	0	7	1	1	0	1	1	4	
12	Allein Gott in der Höh sei Ehr'	1	8	1	1	1	1	1	5*	1
13	Allein Gott in der Höh sei Ehr'	1	8	1	1	1	1	1	5*	1
14	Allein Gott in der Höh sei Ehr'	1	8	0	1	1	1	1	4	
15	Allein zu dir, Herr Jesu Christ	1	9	1	1	1	0	1	4	
16	Allein zu dir, Herr Jesu Christ	1	9	1	1	1	0	1	4	
17	Alle Menschen müssen sterben	0	10	1	1	0	1	1	4	
18	Alle Menschen müssen sterben	0	11	1	1	0	0	1	3	
19	Alles is an Gottes Segen	0	12	1	1	0	1	1	4	
20	Als der gütige Gott	0	13	1	1	1	0	1	4	
21	Als Jesus Christus in der Nacht	0	14	1	1	1	1	1	5*	1
22	Als vierzig Tag' nach Ostern war'n	0	15	1	0	0	0	0	1	
23	An wasserflüssen Babylon	0	16	1	1	1	0	1	4	
24	Auf, auf, mein Herz, und du meiner ganzer	0	17	1	1	1	0	1	4	
25	Auf meinem lieben Gott	1	18	1	1	1	1	1	5*	1
26	Auf meinem lieben Gott	1	18	1	1	1	1	1	5*	1
27	Auf meinem lieben Gott	1	18	0	1	1	0	1	3	
28	Auf meinem lieben Gott	1	18	1	1	1	1	1	5*	1
29	Auf meinem lieben Gott	1	18	1	1	1	1	1	5*	1

First page of the list, including 389 chorales listed alphabetically by First Verse. There are many instances of repeated (slightly varied) chorales in the collection, so the real number of items is smaller (18 in this first page). The chorales were surveyed with the following criteria: 4-voices (1 = 4 voices; 0 = 5 or more voices); 'C' Time signature (1 = C; 0 = others); Anacrusis (1 = starting with upbeat; 0 = starting directly); Regularity (1 = regular phrases with same number of beats; 0 = irregular phrases); Final chord is Chord I in the Key (tonality) of the chorale (1 = Chord I; 0 = other chords). Only the Chorales that met all five criteria were chosen (*) for further analysis; in these, it was checked that the 24 last beats in the piece were regularly divided in three phrases.

APPENDIX 7.H.

Table 7.B.

Numbers of Onsets and Chromatic Notes in Selection of Chorales

Chorale	Sopr	Alto	Tenor	Bass	Total	Chro	Key	Mode	Incompatibilities	Sele	Pair
319	25	30	25	38	118	21	rem	Mi	316 317 318 320		
342	24	31	35	27	117	21	sim	Mi	348		
149	23	35	34	32	124	19	lam	Mi			
317	26	30	29	31	116	17	rem	Mi	316 318 319 320	1	1
21	22	31	29	30	112	17	rem	Mi		2	
316	25	26	30	29	110	17	rem	Mi	317 318 319 320		
318	25	28	26	26	105	16	rem	Mi	316 317 319 320		
216	31	34	31	32	128	14	dom	Mi			
61	24	34	28	29	115	14	SOLM	Ma			
320	26	30	30	28	114	14	dom	Mi	316 317 318 319		
388	27	30	30	37	124	13	lam	Mi	383 384 385 386		
140	23	39	35	33	130	13	solm	Mi		3	2
383	27	29	28	32	116	13	sim	Mi	384 385 386 388	4	
330	30	32	33	35	130	12	solm	Mi			
34	28	34	38	38	138	11	solm	Mi			
386	27	30	33	37	127	11	solm	Mi	383 384 385 388		
87	27	38	33	35	133	11	MIM	Ma	86 88 90		
384	27	30	31	34	122	11	lam	Mi	383 385 386 388		
43	28	28	35	32	123	11	lam	Mi			
385	29	30	36	36	131	10	lam	Mi	383 384 386 388		
146	24	33	33	33	123	10	LAM	Ma			
266	24	30	34	29	117	10	SOLM	Ma		5	3
25	25	30	32	29	116	10	lam	Mi	26 28 29	6	
260	26	28	33	36	123	9	rem	Mi		7	4
359	29	33	33	35	130	9	SOLM	Ma	358	8	
90	27	30	33	33	123	9	SOLM	Ma	86 87 88		
8	27	31	32	35	125	8	LAM	Ma		9	5
12	27	32	35	34	128	8	SOLM	Ma	13	10	
348	24	36	33	32	125	8	sim	Mi	342	11	6
26	22	26	28	32	108	8	solm	Mi	25 28 29	12	
286	23	31	30	31	115	8	MibM	Ma			
28	21	30	32	28	111	8	solm	Mi	25 26 28		
139	26	28	30	37	121	7	SOLM	Ma			
358	25	37	31	33	126	7	FAM	Ma	359		
80	24	32	35	33	124	7	SOLM	Ma			
29	25	27	29	32	113	7	fa#m	Mi	25 26 28		
294	24	31	33	30	118	7	LABM	Ma	289 290 291 292 293 295		

[...]

First page of the list of 58 pre-selected chorales.

Chorales are ordered by descending number of chromatic notes (**Chro.**) and then by number of onsets in the Bass line (**Bass**).

The mean number of Chromatic notes was 9.09; chorales with a number of chromatic notes more than two standard deviation units (4.67) above this mean (the three chorales at the beginning of the list) were considered unrepresentative. Taking into account incompatibilities (subsequent chorales based on the same melodic line; in grey), six pairs of chorales were selected, with both chorales in the pair having the same number of Chromatic notes and not too large differences (<2) in Bass voice onsets.

LIST OF FIGURES

LIST OF FIGURES

CHAPTER 1

- Figure 1.1.* The first datable word-separated manuscript: Dublin, Trinity College 60 (A. I. 15), the *Book of Mulling* (Anonymous, before 692). [11]
- Figure 1.2.* Folio 3 (recto) of the copy of the Vergilius Opera kept in the Biblioteca Apostolica Vaticana, Rome (Vat. Lat. 3867 - 3r). [13]

CHAPTER 2

- Figure 2.1.* First bars of the Second Movement of *Piano Trio* (2002, rev. 2005) by Kevin Volans. [41]
- Figure 2.2.* Three horizontal slices at different z coordinates (-12, -6, +6), showing various category-selective areas in the ventral visual cortex: [53]
- Figure 2.3.* Schematic model of the Dual Route Theory, both for language and music reading. [60]

CHAPTER 3

- Figure 3.1.* Schematic timeline for the PILOT EXPERIMENT [82]
- Figure 3.2.* Schematic timeline for EXPERIMENT I [89]
- Figure 3.3.* Mean Totals of errors in the sight-reading performances as a factor of the Version used: Conventional, Semi-modified, or Fully Modified. [92]

Figure 3.4. Boxplots of the distribution for each reading. [93]

Figure 3.5. Boxplots of datapoints for the factors affecting the number of mistakes. [95]

APPENDIX 3.A.

Figure 3.A.1. PIECE 1: J. S. Bach's Chorale, nr 075 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. [254]

Figure 3.A.2. Piece 1: J. S. Bach's Chorale, nr 075 in the Schubert collection; MODIFIED 1 Version, separating major structural units. [255]

Figure 3.A.3. PIECE 1: J. S. Bach's Chorale, nr 075 in the Schubert collection; MODIFIED 2 Version, separating major and sub-phrasal units. [256]

APPENDIX 3.B.

Figure 3.B.1. PIECE 2: J. S. Bach's Chorale, nr 230 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. [257]

Figure 3.B.2. PIECE 2: J. S. Bach's Chorale, nr 230 in the Schubert collection; MODIFIED 1 Version, separating major structural units. [258]

Figure 3.B.3. PIECE 2: J. S. Bach's Chorale, nr 230 in the Schubert collection; MODIFIED 2 Version, separating major and sub-phrasal units. [259]

APPENDIX 3.C.

Figure 3.C.1. PIECE 3: J. S. Bach's Chorale, nr 336 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. [260]

Figure 3.C.2. PIECE 3: J. S. Bach's Chorale, nr 336 in the Schubert collection; MODIFIED 1 Version, separating major structural units. [261]

Figure 3.C.3. PIECE 3: J. S. Bach's Chorale, nr 336 in the Schubert collection; MODIFIED 2 Version, separating major and sub-phrasal units. [262]

CHAPTER 4

Figure 4.1. The first system of Piece 1 in Conventional Version and its conversion to the Modified Version. [116]

Figure 4.2. Schematic timeline for EXPERIMENT II. [118]

Figure 4.3. Visualisation of the two first phrases and beginning of third phrase of a performance by participant 1. [120]

- Figure 4.4.* Pairwise Comparisons of Estimated Means of PITCH Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [126]
- Figure 4.5.* Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [128]
- Figure 4.6.* Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.), clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then conventional). [129]
- Figure 4.7.* Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [131]
- Figure 4.8.* Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.), clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then conventional). [132]

APPENDIX 4.A.

- Figure 4.A.1.* PIECE 1: J. S. Bach's Chorale, nr 031 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. [264]
- Figure 4.A.2.* PIECE 1: J. S. Bach's Chorale, nr 031 in the Schubert collection; MODIFIED Version, separating major and minor structural units. [265]

APPENDIX 4.B.

- Figure 4.B.1.* PIECE 2: J. S. Bach's Chorale, nr 285 in the Schubert collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1990 Edition. [266]
- Figure 4.B.2.* PIECE 2: J. S. Bach's Chorale, nr 285 in the Schubert collection; MODIFIED Version, separating major and minor structural units. [267]

APPENDIX 4.C.

- Figure 4.C.* Example of mistake coding: first reading of Piece 2, Modified version, by participant 1. [268]

CHAPTER 5

- Figure 5.1.* Schematic timeline for EXPERIMENT III. [150]
- Figure 5.2.* Pairwise Comparisons of Estimated Means of PITCH Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [153]

- Figure 5.3.* Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif). [155]
- Figure 5.4.* Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif). [156]
- Figure 5.5.* Histogram of the nPVI values, including a normal curve. [157]
- Figure 5.6.* Paired Samples Test of nPVI VALUES for performances using Conventional Versions (Conve.) or Modified Versions (Modif). [158]
- Figure 5.7.* Histogram of the Dynamic Range values, including a normal curve. [159]
- Figure 5.8.* Paired Samples Test of DYNAMIC RANGES for performances using Conventional Versions (Conve.) or Modified Versions (Modif). [160]
- Figure 5.9.* Overlay Scatterplot of MINIMUM AND MAXIMUM LOUDNESS for performances using Conventional Versions (Conv.) or Modified Versions (Modi.). [161]
- Figure 5.10.* Histograms of STATED PREFERENCES for scores using Conventional design (Con.) or Modified design (Mod.), measured in a five-point Likert scale (from ‘Clearly the Conventional one to ‘Clearly the Modified one’). [162]
- Figure 5.11.* Pairwise comparisons of Estimated Marginal Means of STATED PREFERENCES by participants for scores using Conventional design (Conv.) or Modified design (Modi.), depending on whether they only LOOKED at the scores or PLAYED them. [163]
- APPENDIX 5.A.
- Figure 5.A.1.* PIECE 1 - TWO VOICES: J. S. Bach’s Chorale, nr 005 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [270]
- Figure 5.A.2.* PIECE 1 - TWO VOICES: J. S. Bach’s Chorale, nr 005 in the Richter collection; MODIFIED Version, separating major and minor structural units. [271]
- Figure 5.A.3.* PIECE 1 - THREE VOICES: J. S. Bach’s Chorale, nr 005 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [272]
- Figure 5.A.4.* PIECE 1 - THREE VOICES: J. S. Bach’s Chorale, nr 033 in the Richter collection; MODIFIED Version, separating major and minor structural units. [273]
- APPENDIX 5.B.
- Figure 5.B.1.* PIECE 2 - TWO VOICES: J. S. Bach’s Chorale, nr 033 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [274]
- Figure 5.B.2.* PIECE 2 - TWO VOICES: J. S. Bach’s Chorale, nr 033 in the Richter collection; MODIFIED Version, separating major and minor structural units. [275]

Figure 5.B.3. PIECE 2 - THREE VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [276]

Figure 5.B.4. PIECE 2 - THREE VOICES: J. S. Bach's Chorale, nr 033 in the Richter collection; MODIFIED Version, separating major and minor structural units. [277]

CHAPTER 6

Figure 6.1. Schematic timeline for EXPERIMENT IV. [174]

Figure 6.2. Pairwise Comparisons of Estimated Means of PITCH Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [177]

Figure 6.3. Pairwise Comparisons of Estimated Means of RHYTHM Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [178]

Figure 6.4. Pairwise Comparisons of Estimated Means of TOTAL NUMBERS of Mistakes between performances using Conventional Versions (Conve.) or Modified Versions (Modif.). [180]

Figure 6.5. Pairwise comparisons of Total Numbers of mistakes for readings using Conventional (Conve.) or Modified (Modif.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then Conventional). [181]

Figure 6.6. Simple scatterplots showing the mapping of the participants' preferences (1= Conventional; 2 = Equal; 3 = Modified) to different covariates. [183]

APPENDIX 6.A.

Figure 6.A.1. PIECE 1: style composition imitating J. S. Bach's instrumental solo writing; CONVENTIONAL Version, using the default settings of a commercial notation software (Sibelius 7.1.3.). [278]

Figure 6.A.2. PIECE 1: style composition imitating J. S. Bach's instrumental solo writing; MODIFIED Version, separating major and minor structural units. [279]

APPENDIX 6.B.

Figure 6.B.1. PIECE 2: style composition imitating J. S. Bach's instrumental solo writing; CONVENTIONAL Version, using the default settings of a commercial notation software (Sibelius 7.1.3.). [280]

Figure 6.B.2. PIECE 2: style composition imitating J. S. Bach's instrumental solo writing; MODIFIED Version, separating major and minor structural units. [281]

CHAPTER 7

- Figure 7.1.* Schematic timeline for EXPERIMENT V. [203]
- Figure 7.2.* EXERCISE 1: Pairwise Comparisons of Estimated Marginal Means of [a] PITCH, [b] RHYTHM, and [c] TOTAL NUMBERS of mistakes, for performances using Conventional Versions (Conve.) or Modified Versions (Modif.) of the scores. [207]
- Figure 7.3.* EXERCISE 1: Pairwise comparisons of Total Numbers of mistakes for readings using Conventional (Conve.) or Modified (Modi.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then conventional). [208]
- Figure 7.4.* EXERCISE 2: Pairwise Comparisons of Estimated Marginal Means of [a] PITCH, [b] RHYTHM, and [c] TOTAL NUMBERS of mistakes, for performances using Conventional Versions (Conve.) or Modified Versions (Modif.) of the scores. [210]
- Figure 7.5.* EXERCISE 2: Pairwise comparisons of Total Numbers of mistakes for readings using Conventional (Conve.) or Modified (Modi.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then conventional). [211]
- Figure 7.6.* EXERCISE 3: Pairwise Comparisons of Estimated Marginal Means of [a] PITCH, [b] RHYTHM, and [c] TOTAL NUMBERS of mistakes, for performances using Conventional Versions (Conve.) or Modified Versions (Modif.) of the scores. [213]
- Figure 7.7.* EXERCISE 3: Pairwise comparisons of Total Numbers of mistakes for readings using Conventional (Conve.) or Modified (Modi.) Versions of the score, clustered by ORDER OF PRESENTATION (Conventional then Modified, or Modified then conventional). [214]
- Figure 7.8.* STATED PREFERENCES as a factor of the exercise and the version that coincided with the easier piece. [215]
- Figure 7.9.* Grouped scatterplots showing PREFERENCES STATED by participants in relation to [a] Number of years studying percussion, [b] Academic year (1–4 BMus; 5 MMus), [c] Age. [217]
- APPENDIX 7.A.
- Figure 7.A.1.* EXERCISE 1 - PIECE 1: J. S. Bach's Chorale, nr 025 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [282]
- Figure 7.A.2.* EXERCISE 1 - PIECE 1: J. S. Bach's Chorale, nr 025 in the Richter collection; MODIFIED Version, separating only *phrasal* units. [283]
- APPENDIX 7.B.
- Figure 7.B.1.* EXERCISE 1 - PIECE 2: J. S. Bach's Chorale, nr 266 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [284]

Figure 7.B.2. EXERCISE 1 - PIECE 2: J. S. Bach's Chorale, nr 266 in the Richter collection; MODIFIED Version, separating only *phrasal* units. [285]

APPENDIX 7.C.

Figure 7.C.1. EXERCISE 2 - PIECE 1: J. S. Bach's Chorale, nr 021 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [286]

Figure 7.C.2. EXERCISE 2 - PIECE 1: J. S. Bach's Chorale, nr 021 in the Richter collection; MODIFIED Version, separating only *sub-phrasal* units. [287]

APPENDIX 7.D.

Figure 7.D.1. EXERCISE 2 - PIECE 2: J. S. Bach's Chorale, nr 317 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [288]

Figure 7.D.2. EXERCISE 2 - PIECE 2: J. S. Bach's Chorale, nr 317 in the Richter collection; MODIFIED Version, separating only *sub-phrasal* units. [289]

APPENDIX 7.E.

Figure 7.E.1. EXERCISE 3 - PIECE 1: J. S. Bach's Chorale, nr 008 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [290]

Figure 7.E.2. EXERCISE 3 - PIECE 1: J. S. Bach's Chorale, nr 008 in the Richter collection; MODIFIED Version, separating only *symbols* ('proportional notation'). [291]

APPENDIX 7.F.

Figure 7.F.1. EXERCISE 3 - PIECE 2: J. S. Bach's Chorale, nr 012 in the Richter collection; CONVENTIONAL Version, using the layout and spacing of the Breitkopf 1976 Edition. [292]

Figure 7.F.2. EXERCISE 3 - PIECE 2: J. S. Bach's Chorale, nr 012 in the Richter collection; MODIFIED Version, separating only *symbols* ('proportional notation'). [293]

LIST OF TABLES

LIST OF TABLES

CHAPTER 3

- Table 3.1. *Data and Descriptive Statistics for the Students* [84]
- Table 3.2. *Sequences of Pieces and Versions Performed* [90]
- Table 3.3. *Tests of Normality for the Datapoints of the Numbers of Total Errors* [94]
- Table 3.4. *Descriptive Statistics for the First and Second Readings* [98]
- Table 3.5. *Tests of Normality for Original and Transformed Data of Mark Increases in the Second Reading* [99]
- Table 3.6. *Participants' Preferences and Best Performances* [103]

CHAPTER 4

- Table 4.1. *Sequences of Pieces and Versions Read by Participants in Each Group* [112]
- Table 4.2. *Data and Descriptive Statistics for the 15 Percussion Students* [113]
- Table 4.3. *Note Onset Detector Settings for Sonic Visualiser Programme* [121]
- Table 4.4. *List of Pairs of Recordings Sent to Referees for Judging* [122]
- Table 4.5. *Numbers of Pitch Mistakes: Shapiro-Wilk Tests of Normality* [124]
- Table 4.6. *nPVI Values: [a] 1st Tests of Normality; [b] 2nd Tests of Normality After log Transformation* [134]

- Table 4.7. *Assessment by Experts: Relations of Ranking Levels with Numbers of Mistakes* [135]
- Table 4.8. *Assessment by Experts: Comparison of Expert Markings and Coded Mistakes* [136]
- Table 4.9. *Evaluation by Participants: Preferred Versions and Comments* [137]
- Table 4.10. *Evaluation by Participants: Ordinal Ranking and Relation with Coded Mistakes* [138]
- Table 4.11. *Evaluation by Participants: Comparison of Markings* [138]

CHAPTER 5

- Table 5.1. *Questionnaire responses: Shapiro-Wilk Tests of Normality* [163]

CHAPTER 6

- Table 6.1. *Evaluation by Participants: Preferred Versions and Comments* [183]

CHAPTER 7

- Table 7.1. *Means of Mistakes: Variability per Piece* [205]
- Table 7.A. *List of Chorales in the Richter Edition* [294]
- Table 7.B. *Numbers of Onsets and Chromatic Notes in Selection of Chorales* [295]