CHALLENGES FOR EYE-TRACKING IN NATURALISTIC MUSIC-READING SETUPS: A PILOT STUDY

Charlotte SCHNEIDER (n11912429@students.meduniwien.ac.at) (0009-0001-3864-8852) [1]

[1] *Faculty of Music*, **University of Cambridge**, 11 West Rd, Cambridge, CB3 9DP UK

ABSTRACT

Music-reading research rarely employs naturalistic, corpus-based stimuli that are standard in language-reading research. To better understand how musicians interact with scores, a corpus-derived stimulus database for sight-singing experiments was developed, designed to balance experimental control with ecological validity by providing structured, idiom-specific materials that better reflect the musical knowledge trained performers bring to score reading. Preliminary results from an exploratory pilot eye-tracking study suggest that singers tended to remain visually close to the current score position rather than reading ahead, in contrast to common findings from language-reading research. Correlation analyses revealed no consistent associations between gaze measures and choir experience or professional singing experience. The only significant association was a negative correlation between formal singing training and number of fixations, although this effect appeared to be influenced by two individual participants and should therefore be interpreted cautiously. A further non-significant trend suggested that greater formal singing training might be associated with longer fixation durations. Overall, the findings indicate that established measures such as fixation duration, fixation count, and eye-voice span may not transfer straightforwardly to musical settings. The study thus provides both a methodological resource and an exploratory foundation for future research.

1. INTRODUCTION

Corpus-based databases are standard in reading research because they allow controlled stimulus selection based on variables such as frequency, length, and syllable count (e.g., BNC, ANC, COHA, Lexique). The present database applies the same principle to music by drawing on stylistically consistent material and including both syllable counts and note counts. This is important due to frequency strongly affecting processing efficiency in reading and word recognition [1,2], accounting for substantial variance in lexical tasks [3–6]. Focusing on one type of musical source therefore allows greater control over the stimuli. In addition, the use of real musical texts improves ecological validity, since musical syntax engages auditory, motor, and visual networks in everyday music cognition [7, 8], processes that may be less active with isolated or artificially constructed stimuli [9]. Rather than introducing new eye-tracking metrics, the present study evaluates whether established measures of gaze behaviour remain informative when applied to ecologically valid music-performance settings. In particular, we examine fixation duration, fixation count, and eye–voice span (EVS), three widely used measures in reading and music research, in a naturalistic sight-singing paradigm. Unlike many previous music-reading studies, the present paradigm deliberately required participants to actively sight-sing rather than merely inspect or evaluate musical notation. Sight-singing fundamentally changes the cognitive demands of music reading because visual decoding must be continuously transformed into vocal production while simultaneously maintaining auditory monitoring and temporal synchronisation with the accompanying ensemble. Consequently, eye movements reflect not only visual processing but the continuous coordination of perception and performance. Investigating gaze behaviour under these ecologically valid conditions is therefore essential for understanding real-world music reading and cannot be directly inferred from paradigms involving passive score inspection alone. The present pilot study therefore asks whether established eye-tracking measures remain informative in ecologically valid sight-singing and whether corpus-derived stimuli provide a suitable basis for such investigations.

2. GENERATING A CORPUS-BASED DATABASE

2.1 Fayrfax as a candidate for a stylistically coherent corpus

Fayrfax is a particularly well-suited source for generating our stimuli because his works show strong idiomatic and stylistic consistency, closely aligned with early Tudor conventions. As an academic and court musician, deemed a "court favourite" [10], his compositions were most likely shaped by regulated compositional circumstances and stable performance resources, including the musicians available at the leading Abbey of England where he served as organist [10]. On this basis, we treat his pieces as broadly comparable. Rather than sampling across Fayrfax's entire oeuvre, we deliberately restricted the corpus to masses, excluding other genres that differ in compositional function

Current affiliation: Imaging based Functional Brain Diagnostics and Therapy, Laboratory for Experimental Brain Stimulation, Department of neurology, Medical University of Vienna, Währinger Gürtel 18-20, Vienna, 1090 Austria

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and musical style.

2.2 Compositional considerations

As a first step, a database of all Fayrfax vocal works was compiled, yielding 1,013 excerpts. For each segment, defined by changes in time or key, the database recorded the title of the piece, instrumentation, voices, vocal range, composition type, number of voices, subdivisions, time signature, and number of bars. A review of this initial database suggested that composition type (e.g., mass or part-song) influences stylistic choices, meaning that similar musical material may be perceived differently across genres. This was especially relevant because the participants, choir singers from the University of Cambridge, were familiar with the repertoire. A second, refined database was therefore restricted to Fayrfax's masses. Two works, Sponsus amat sponsam and Missa O quam glorifica, were excluded because they differ substantially from the remaining masses. Missa O quam glorifica, composed in 1504 for Fayrfax's doctorate at the University of Cambridge rather than for liturgical use, is unusually elaborate, relies much more extensively on five-voice writing than the remaining masses, uses relatively little imitation, and is rhythmically distinctive [10]. Sponsus amat sponsam, possibly written for the marriage of Henry VIII and Catherine of Aragon, also appears atypical: only fragments survive, it is written for four rather than the predominantly five voices of the remaining masses, lacks a motto, and is less imitative [11]. The remaining masses form a more coherent corpus that reflects shared stylistic expectations in early sixteenth-century England. As Warren notes, Fayrfax's harmonic practice across these masses is relatively stable, with 52% major triads, 47% minor triads, and only 0.67% diminished triads, excluding the two atypical works [11]. The final corpus therefore consisted of Missa Albanus, Missa O bone Jesu, Missa Regali ex progenie, and Missa Tecum principium.

2.3 Extracting Excerpts

From the selected masses, excerpts from the Gloria movements were extracted. Changes in time and key signatures were used to define subparts within the Gloria movement of each mass. To evaluate and compare candidate excerpts, a second excerpt-level database was built. Because the Tobii Pro Fusion system can only display two lines of score at once, the number of notes visible per two lines for each notation design was calculated. For every voice in every excerpt, notes and syllables were counted as well. This procedure yielded 106 potential excerpts. Given that Fayrfax's vocal ranges are relatively contained and large intervals are uncommon, vocal range was not treated as a selection factor. Instead, excerpts were ranked first by number of notes, which primarily determined visual complexity, and secondly by number of syllables to ensure comparable linguistic demands across excerpts. The aim was to select materials that were equally demanding. Difficulty was set high by maximising note density, retaining the original keys, and adjusting tempo to challenge participants.

2.4 Overlapping sensory-cognitive networks and distinct neural substrates in language and music

Language and music share important parallels that affect reading. Lesion studies and research on specific impairments, particularly aphasia and dyslexia, provide useful insights into these relationships. Some forms of aphasia affect singing as well as speech: expressive aphasia, typically linked to damage in the left inferior frontal gyrus, disrupts propositional speech and volitional prosody, while automatic or familiar singing may remain intact [13]. This suggests that producing new songs recruits neural resources beyond those used in singing from memory, including right-hemisphere homologues of Broca's area, the right insula, subcortical networks [12], and brainstem vocal-initiation circuits such as the PAG [13, 14]. In dyslexia, overlap between music and language processing becomes particularly visible. Developmental dyslexia has been associated with atypical eye movements and longer fixation times in music reading, mirroring difficulties in written language processing [15]. Sight-singing is especially demanding because it requires the integration of visual notation decoding with auditory-motor and somatosensory processes. This may help explain why music reading produces eye-tracking patterns that differ from language reading. Unlike written language, which is organised mainly from left to right, musical notation must be processed both horizontally, for rhythm and temporal progression, and vertically, for pitch relations.

2.5 Historical paralells in music and language reading

Music and language share historical parallels in the development of reading. Saenger describes how the shift from scriptura continua to spaced words and paragraphs transformed reading from an oral, performative practice into the faster, silent, and individualised activity that characterises literacy today [16]. Earlier texts in Greece and Rome contained few visual cues and were typically read aloud by professional readers [16–19]. Music evolved similarly. Early chant notation provided limited visual guidance, requiring performers to rely on memory and oral tradition. Later developments such as clearer note shapes, spacing conventions, and bar lines made written music easier to process and improved sight reading [20]. As in language, these conventions reduced the need for preparation [21, 22]. Research has shown that spacing lowers fixation counts, shortens processing time, and improves reading accuracy across different writing systems [18, 23–28].

3. FIRST PILOT

A pilot study was conducted to assess the feasibility of the naturalistic eye-tracking setup for choir singers. Questionnaires measured choir singing experience, formal singing training, and professional singing experience. The study was approved by the Ethics Committee of Arts and Humanities at the University of Cambridge and was part of a larger project funded by grant G11383 and approved by the Research Operations Office (X5,47435). Participants

were informed about the procedure and data handling, familiarised with the protocol in a pretest, and gave written informed consent prior to participation.

3.1 Participants

Participants were recruited via the newsletter of the Centre for Music Performance, the University of Cambridge German Society, the University of Cambridge Austrian Society, and the choir of Darwin College. Interested singers registered online and were subsequently screened by e-mail. To minimise familiarity effects, individuals who had previously performed any of Robert Fayrfax's masses were excluded from participation. Individuals with psychiatric disorders or neurological impairments were also excluded. The final sample consisted of 14 participants (5 male, 9 female; mean age = 20.4 ± 2.1 years). Choir experience ranged from 0 to 15 years (mean = 8.3 ± 4.5); four participants reported professional singing experience and two formal singing training.

3.2 Setup

Two microphones were used to improve audio quality, and a separate camera recorded the session for use in other studies. Playback was presented through two Logitech X-140 loudspeakers and generated in Sibelius (version 24.6.1.4481), with organ sounds replacing the other vocal parts. Stimuli were shown on a Dell G15-5530 laptop with a 15.6-inch screen (34.54 × 19.43 cm; 16:9). Participants sat 60–70 cm from the screen, in line with Tobii recommendations. All sessions were conducted in the studio at 11 West Road, Cambridge, under consistent lighting conditions. The Tobii Pro Fusion eye tracker sampled at 250 Hz and was recalibrated after each trial.



Figure 1. Eye-tracking setup at the recording studio

3.3 Methods and procedure

Before the session, participants gave written consent for the recording of their singing, eye movements, and video data. Each session began with eye-tracker calibration and a sound check, followed by a pretest to familiarise participants with the setup. During this pretest, playback tempo was adjusted individually so that sight-reading remained challenging while still ecologically valid. As part of a larger study on notation design and music reading, participants viewed three stimuli, including one in conventional notation, presented once each in counterbalanced order. All stimuli were accompanied by MIDI playback simulating a choir setting, and participants received two beats before playback onset to time their entry. Between trials, a fixation screen with a black X on a grey background was displayed; the next stimulus appeared after a mouse click, allowing for questions or recalibration if needed. Sessions lasted about one hour in total and were scheduled between 10:00 and 17:00.

use	**model**
Eye-tracker	Tobii Pro Fusion 250
Eye-tracking Laptop	Dell G15-5530
Computer to start playback	MacBook Air 13" using Sibelius version 24.6.1.4481
Camera	Panasonic HC-X2000
Speakers	Logitech X-140
Interface	Focusrite Scarlett 8i6 (3rd gen)
Microphones	AKG 190 microphone (connected to computer for internal Tobii tracking) to enhance audio and an AKG C1000 microphone for recordings on the HD camera.

Table 1. Technical equipment used for the pilot study

3.3.1 Formatting and Presentation of Stimuli

To format the excerpts, the design of the commonly used Fairfax editions by Humphrey Thompson, a specialist on Tudor music, was replicated. His edition is freely available at the Choral Public Domain Library (CPDL). This webpage not only provides information on the pieces and composers but also shares XML files of musical text. These files have formed the foundation for transforming the musical text for our stimuli. The selected XML document was imported into Sibelius. The excerpts were then used to fill a template with set distances. The stimuli were converted to a horizontal layout (as opposed to the vertical version in Thompson's edition) and converted to a PDF to fit the screen of the Lenovo laptop used in the Tobii Pro Fusion eye-tracking system. The original keys used by Fayrfax were reapplied so as to increase difficulty for the participants.

3.3.2 Eye-tracking metrics

To characterise gaze behaviour, three eye-tracking measures were considered: eye–voice span (EVS), fixation duration, and fixation count. EVS refers to the distance the eyes remain ahead of the spoken or sung output. Originally quantified by Buswell in child readers [29], it was later extended to adult reading by Tinker [30], with Levin et al. showing that it varies across sentence structures [31]. EVS can therefore serve as an indicator of oculomotor–speech coupling and cognitive load. Fixations stabilise visual input on the fovea and permit detailed inspection. Their duration, typically between 150 and 600 ms, and their overall count are standard measures in reading research, reflect-

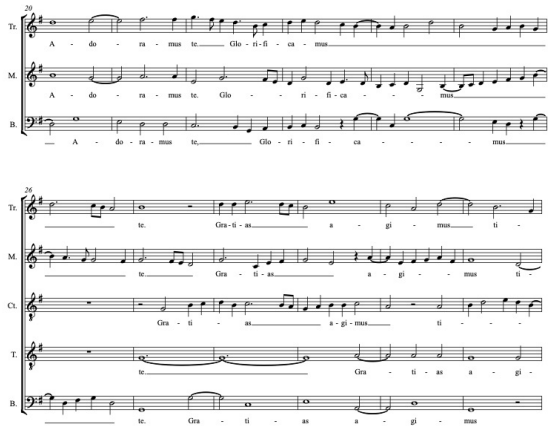


Figure 2. Example of a stimulus from *Missa Regali ex progenie*.

ing attention allocation and processing demands [32]. Saccades, that is, eye movements between fixations, are also commonly analysed, and their length can vary with visual and processing demands. However, because the larger experiment involved different notation layouts, including altered note spacing, saccade measures were likely to be distorted and were therefore not analysed here. The present evaluation thus focused on EVS, fixation duration, and fixation count in more and less experienced singers.

3.3.3 Processing of Eye-tracking data

To calculate EVS, we first examined transitions from the first to the second line of each musical stimulus, as these moments were expected to reveal how far participants read ahead. A time of interest (TOI) was defined in Tobii using the tick-end function rather than automatic first-hit detection, since first hits often reflected reorientation rather than stable reading. The onset of the TOI was the first fixation on a note in the second line before the gaze remained there, and the end was the corresponding voice onset. In practice, however, this approach did not yield usable EVS values, because participants made the most errors at these line transitions, often reorienting visually or reducing their singing to a mumble. We therefore adopted a second approach and focused on pauses during singing. For these events, recorded fixations were inspected frame by frame (every 4 ms). The onset of the interval was defined as the first fixation on a note after the pause, and the end as the re-onset of the voice, determined manually from the audio waveform. This procedure allowed EVS to be calculated for all but one participant, who produced no clearly identifiable pauses. Because participants sang at different tempi in order to ensure active sight-reading, EVS, fixation duration, and fixation count were corrected for tempo. Tempi ranged from 50 to 66 bpm (mean = 53.92 $\pm$ 4.75). To obtain tempo-normalised, dimensionless measures, each metric was divided by the duration of one beat. For EVS, this was calculated as follows:

$$\mathrm{EVS}_{\mathrm{beats}} = \frac{\text{EVS (in ms)}}{\text{Duration of one beat (in ms)}}. \qquad (1)$$

As shown in Equation (1), EVS in milliseconds was divided by the duration of one beat, also in milliseconds, to yield $\mathrm{EVS}_{\mathrm{beats}}$.

4. RESULTS

Pearson correlation analyses were conducted to examine whether fixation duration, number of fixations, and EVS were associated with years of choir experience, formal singing training, and professional singing experience. Overall, these exploratory analyses revealed little evidence for robust associations between the experience-related variables and the music-reading measures. Neither years of choir experience nor professional singing experience was significantly related to number of fixations, fixation duration, or EVS ($p > .05$). Likewise, formal singing training was not significantly associated with EVS. The only statistically significant association observed was a negative correlation between formal singing training and average number of fixations, $r = -.709$, $p = .004$, 95% CI $[-0.901, -0.287]$. This indicates that participants with more formal singing training tended to make fewer fixations during music reading. Such a pattern may suggest a less fragmented or more efficient visual scanning strategy in more highly trained singers. However, this finding should be interpreted cautiously, as inspection of the data revealed that the effect was driven by two individual participants (Table 2, Figure 3). In addition to this significant result, one non-significant trend was observed for the association between formal singing training and fixation duration. This relationship was positive and of moderate magnitude, r=.491, p=.074, 95% CI $[-0.053, 0.811]$, suggesting that greater training may be associated with longer fixation durations.

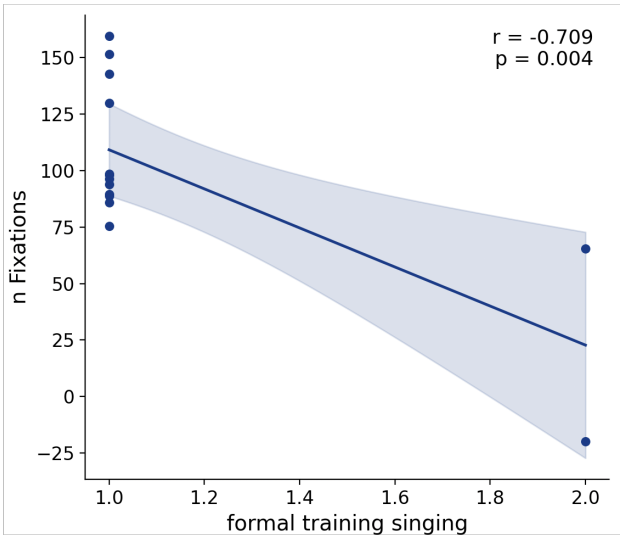


Figure 3. Scatterplot illustrating the correlation between average number of fixations and formal singing training

Although this effect did not reach conventional levels of significance, when considered together with the negative association between training and number of fixations, it may tentatively point toward a pattern in which more

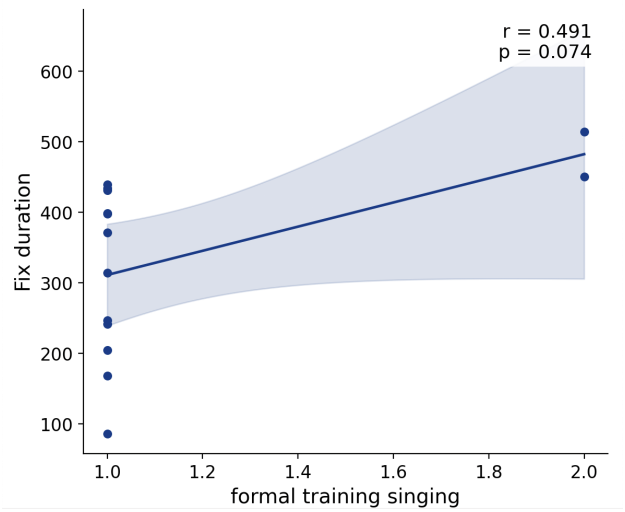


Figure 4. Scatterplot illustrating the non-significant correlation of fixation duration and formal training in singing

highly trained singers make fewer, but somewhat longer, fixations. Given the exploratory nature of the analysis and the limited sample size, this pattern should be regarded as preliminary. All remaining correlations were small to moderate in magnitude and non-significant. Number of fixations was unrelated to choir experience and professional singing experience, and fixation duration likewise showed no reliable association with either of these variables. EVS also did not correlate significantly with any of the experience-based measures. Taken together, these findings suggest that formal singing training, rather than broader performance experience, may be the experience variable most closely related to visual aspects of music reading in the present sample, although this conclusion remains tentative and warrants careful interpretation.

Correlation	r	p	95% CI	z
training singing – n fixations	-0.709**	.004	[-0.901, -0.287]	-0.886
training singing – Fix duration	0.491	.074	[-0.053, 0.811]	0.538
training singing – EVS	-0.199	.514	[-0.676, 0.395]	-0.202
n fixations – choir years	-0.069	.823	[-0.597, 0.501]	-0.069
n fixations – prof. performance	-0.121	.693	[-0.630, 0.460]	-0.122
Fix duration – choir years	-0.210	.491	[-0.682, 0.385]	-0.213
Fix duration – prof. performance	0.296	.327	[-0.305, 0.728]	0.305
EVS – choir years	0.042	.896	[-0.545, 0.602]	0.042
EVS – prof. performance	-0.133	.681	[-0.657, 0.478]	-0.133

Table 2. Pearson's correlations between training singing, eye-tracking measures, EVS, and experience variables. Confidence intervals are 95% confidence intervals. Effect sizes are reported as Fisher's z.

5. DISCUSSION

Based on the literature, we expected greater musical experience and training to be associated with more efficient visual scanning during sight-singing. This expectation was only partially supported. Rather than years of choir or professional performance experience, formal singing training emerged as the only experience variable showing a relationship with gaze behaviour. Although this finding should be interpreted cautiously because of the small sample size and the influence of two participants, it may suggest that structured vocal training shapes visual strategies differently from accumulated performance experience. Future studies with larger samples will be required to determine whether this pattern is robust. Additionally, participants might have been focused on synchronising with the playback, as they would have in a choir performance. This could have hindered them from reading further ahead, greatly impacting EVS measures. This behaviour can be interpreted as readers being "stuck" on one part of the score. Not reading ahead has often been observed in less successful readers [29, 33–35]. In our case, however, this hovering could indicate something else, for example better synchronization.

Finally, and, for the sake of this pilot, importantly, music is inextricably understood in time. Language presents semantic units spatially on the page: readers can skip ahead, retrace, or view paragraphs holistically. For stylistically coherent musical excerpts, it might be worth reevaluating established parameters in reading research. Differences in language's "spatial semantics" and music's "temporal semiotics" have been well described [36,37] but have been neglected in experimental design. A sentence, once written, has a determinate meaning that can be parsed at leisure; by contrast, a musical phrase only achieves its contour and meaning in the very act of listening. As Peter Kivy notes, "Music's referential content is inextricable from its temporal flow" [38]. Our data could reflect this difference between music and language.

6. CONCLUSIONS

To draw definite conclusions on general ocular behaviour in naturalistic setups, a larger sample size is needed. However, even in this small setup, our results suggest that music reading, especially if accompanied, and language reading have to be treated differently. While music and language reading share neuronal resources as well as historical parallels, there remain key aspects that set them apart. There is, for instance, the "vertical pitch" component of music. Additionally, the time component of understanding and performing music, as outlined by Tchertov and Kivy, differs fundamentally from language [36, 38]. Another difference of paramount importance is the synchronous and joint element of music, especially choir music. During singing, performers have to continuously evaluate not only their own singing but also their synchronisation with other singers. This complicates direct comparisons with language reading or solo performances. At the same time, it offers an opportunity to understand how behaviour adapts in synchronisation tasks and how performers orient themselves in environments that require constant adjustment. These factors have been neglected in music reading research, although musical setups provide a unique opportunity to observe how humans integrate, understand,

and synchronise with a larger group. Current eye-tracking technology could support studies in ecologically valid setups that also involve group settings. Employing musical excerpts taken from real compositions, such as our excerpts, can further improve comparability and validity. However, even in the most ecologically valid setting, there remains a need to reevaluate the interpretation of usual eye-tracking metrics for music reading due to the unique cognitive demands of musical group settings. Overall, this work demonstrates the feasibility of corpus-derived naturalistic stimuli for music-reading eye-tracking research, provides a reusable stimulus database, and suggests that established eye-tracking metrics require careful reinterpretation in ecologically valid musical settings. Larger studies are now needed to validate these preliminary findings and to further refine the interpretation of gaze behaviour during naturalistic sight-singing.

Acknowledgments

I would like to acknowledge my supervisor Dr. Arild Stenberg, who hosted me at the Faculty of Music at the University of Cambridge. This project was made possible by a KWA-Stipend (Short-term grants abroad) awarded by the University of Vienna.

7. REFERENCES

[1] K. A. Preston, "The speed of word perception and its relation to reading ability," *The Journal of General Psychology*, vol. 13, no. 1, pp. 199–203, 1935. [Online]. Available: https://doi.org/10.1080/00221309.1935.9917878

[2] S. Monsell, M. Doyle, and P. Haggard, "Effects of frequency on visual word recognition tasks: where are they?" *Journal of experimental psychology. General*, vol. 118, no. 1, p. 43—71, March 1989. [Online]. Available: https://doi.org/10.1037//0096-3445.118.1.43

[3] L. Ferrand, B. New, M. Brysbaert, E. Keuleers, P. Bonin, A. Méot, M. Augustinova, and C. Pallier, "The french lexicon project: Lexical decision data for 38,840 french words and 38,840 pseudowords," *"Behavior Research Methods"*, vol. 42, pp. 488–496, 2010.

[4] E. "Keuleers, K. Diependaele, and M. Brysbaert, "Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords," *Frontiers in Psychology*, vol. Volume 1 - 2010, 2010. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2010.00174

[5] E. Keuleers, P. Lacey, K. Rastle, and M. Brysbaert, "The british lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic english words," *Behavior Research Methods*, vol. 44, no. 1, pp. 287–304, mar 2012.

[6] M. J. Yap and D. A. Balota, "Visual word recognition of multisyllabic words," *Journal of Memory and Language*, vol. 60, no. 4, pp. 502–529, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0749596X09000254

[7] D. Schön and M. Besson, "Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study," *Journal of Cognitive Neuroscience*, vol. 17, no. 4, pp. 694–705, Apr 2005. [Online]. Available: https://doi.org/10.1162/0898929053467532

[8] ——, "Audiovisual interactions in music reading," *Annals of the New York Academy of Sciences*, vol. 999, pp. 193–198, nov 2003.

[9] D. Schön, J.-L. Anton, M. ROTH, and M. Besson, "An fmri study of music sight-reading," *Neuroreport*, vol. 13, pp. 2285–9, jan 2003.

[10] A. Hughes, "An introduction to fayrfax," *Musica Disciplina*, vol. 6, no. 1/3, pp. 83–104, 1952. [Online]. Available: http://www.jstor.org/stable/20531844

[11] E. B. Warren, "The masses of robert fayrfax," *Musica Disciplina*, vol. 12, pp. 145–176, 1958. [Online]. Available: http://www.jstor.org/stable/20531903

[12] A. Norton, L. Zipse, S. Marchina, and G. Schlaug, "Melodic intonation therapy: Shared insights on how it is done and why it might help," *Annals of the New York Academy of Sciences*, vol. 1169, pp. 431–6, 08 2009.

[13] K. Simonyan and B. Horwitz, "Laryngeal motor cortex and control of speech in humans," *The Neuroscientist*, vol. 17, no. 2, pp. 197–208, 2011, pMID: 21362688. [Online]. Available: https://doi.org/10.1177/1073858410386727

[14] S. P. Zhang, P. J. Davis, R. Bandler, and P. Carrive, "Brain stem integration of vocalization: role of the midbrain periaqueductal gray," *Journal of Neurophysiology*, vol. 72, no. 3, pp. 1337–1356, 1994, pMID: 7807216. [Online]. Available: https://doi.org/10.1152/jn.1994.72.3.1337

[15] S. Dehaene, G. Le Clec'H, J.-B. Poline, D. Le Bihan, and L. Cohen, "The visual word form area: a prelexical representation of visual words in the fusiform gyrus," *Neuroreport*, vol. 13, no. 3, p. 321—325, Mar 2002. [Online]. Available: https://doi.org/10.1097/00001756-200203040-00015

[16] P. Saenger, *Space Between Words: The Origins of Silent Reading*, ser. Figurae (Stanford, Calif.). Stanford University Press, 1997. [Online]. Available: https://books.google.at/books?id=w3vZaFoaa3EC

[17] C. Sirat, *The Material Conditions of the Lateralization of the Ductus*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1988, pp. 173–201. [Online]. Available: https://doi.org/10.1007/978-3-662-01093-8_11

[18] D. F. Fischer, *Spatial factors in reading and search*. Lawrence Erlbaum Associates, 1976, ch. Eye movement and psychological processes.

[19] J. Hochberg, *Basic studies in reading.* Basic Books, 1970, ch. Components of literacy: Speculations and exploratory research.

[20] A. Stenberg, "Legibility of musical scores and parallels with language reading," Ph.D. dissertation, University of Cambridge, 2017.

[21] M. Banniard, *Genèse culturelle de l'Europe: Ve-VIIIe siècle*, ser. Collection Points : série histoire. Editions du Seuil, 1989. [Online]. Available: https://books.google.at/books?id=r-IYAAAAYAAJ

[22] I. Taylor and M. Taylor, *The Psychology of Reading*. Academic Press, 1983. [Online]. Available: https://books.google.at/books?id=_og_qkv7TdUC

[23] J. Downing and C. Leong, *Psychology of Reading*. Macmillan, 1982. [Online]. Available: https://books.google.at/books?id=BL19AV457qYC

[24] J. W. Senders, D. F. Fisher, and R. A. Monty, Eds., *Eye Movements and the Higher Psychological Functions*, 1st ed. Routledge, 1978.

[25] J. Solomon and D. Pelli, "The visual filter mediating letter identification," *Nature*, vol. 369, pp. 395–397, 1994.

[26] R. J. Jarvella and I. Lundberg, "Visual constraints in reading: evidence from non-ocular behaviour," *Scandinavian Journal of Psychology*, vol. 28, no. 2, pp. 93–103, 1987. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9450.1987.tb00744.x

[27] R. Kliegl, E. Grabner, M. Rolfs, and R. Engbert, "Length, frequency, and predictability effects of words on eye movements in reading," *European Journal of Cognitive Psychology*, vol. 16, no. 1-2, pp. 262–284, 2004. [Online]. Available: https://doi.org/10.1080/09541440340000213

[28] X. Li, W. Zhao, and A. Pollatsek, "Dividing lines at the word boundary position helps reading in chinese," *Psychonomic Bulletin & Review*, vol. 19, pp. 929–934, 2012.

[29] G. T. Buswell, "The relationship between eye-perception and voice-response in reading," *Journal of Educational Psychology*, vol. 12, no. 4, pp. 217–227, 1921.

[30] M. A. Tinker, "Experimental study of reading," *Psychological Bulletin*, vol. 31, no. 2, pp. 98–110, 1934.

[31] H. Levin and E. L. Kaplan, "Eye-voice span (evs) within active and passive sentences," *Language and Speech*, vol. 11, no. 4, pp. 251–258, 1968, pMID: 5706565. [Online]. Available: https://doi.org/10.1177/002383096801100405

[32] D. Irwin, "Memory for position and identity across eye movements," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 18, pp. 307–317, 03 1992.

[33] J. M. Meixner, J. S. Nixon, and J. Laubrock, "The perceptual span is dynamically adjusted in response to foveal load by beginning readers," *Journal of Experimental Psychology: General*, vol. 151, no. 6, pp. 1219–1232, 2022.

[34] K. Rayner, M. H. Fischer, and A. Pollatsek, "Unspaced text interferes with both word identification and eye movement control," *Vision Research*, vol. 38, no. 8, pp. 1129–1144, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0042698997002745

[35] J. Sloboda, "The eye-hand span-an approach to the study of sight reading," *Psychology of Music*, vol. 2, no. 2, pp. 4–10, 1974. [Online]. Available: https://doi.org/10.1177/030573567422001

[36] L. Tchertov, "Spatial semiosis and time," *Sign Systems Studies*, vol. 33, no. 2, pp. 297–314, 2005.

[37] R. Jakobson, *Linguistics and Poetics*. Berlin, Boston: De Gruyter Mouton, 1981, pp. 18–51. [Online]. Available: https://doi.org/10.1515/9783110802122.18

[38] P. Kivy, *Sound Sentiment: An Essay on the Musical Emotions, Including the Complete Text of The Corded Shell*, ser. Arts and their philosophies. Temple University Press, 1989. [Online]. Available: https://books.google.at/books?id=3eBizKAPdmsC