Exploring early vocal music and its lute arrangements

Using F-TEMPO as a musicological tool

Tim Crawford*
David Lewis*
t.crawford@gold.ac.uk
d.lewis@gold.ac.uk
Goldsmiths University of London
London, UK

Alastair Porter*
alastair.porter@upf.edu
Universitat Pompeu Fabra
Barcelona, Spain

ABSTRACT

In its earliest state, F-TEMPO (Full-Text searching of Early Music Prints Online) enabled searching in the musical content of about 30,000 page-images of early printed music from the British Library's Early Music Online collection (GB-Lbl). The images were processed using the Optical Music Recognition (OMR) program, Aruspix, whose output is saved in the MEI (Music Encoding Initiative) format. To enable fast searches of the MEI, we adopted an indexing strategy that is both scalable and substantially robust to the inevitable errors in the process. In this paper we show how searches using these indexes may be used as a first step in two useful musicological tasks without exhaustively processing the full encodings.

The F-TEMPO resource has subsequently been augmented to about 500,000 images including a large number from the Bavarian State Library in Munich (D-Mbs), and other libraries (D-Bsb, PL-Wn and F-Pn). Most recently, a new and more robust system architecture is in development, together with a new interface conforming better to modern web standards.

The simple, yet robust, indexing method we use can be applied to scores encoded in any format from which strings of pitches each corresponding to a voice or instrument in the score can be derived. In addition to page-images, in its current form F-TEMPO now includes a collection of over 10,000 scores encoded in MusicXML, largely of early music, from the online Choral Public Domain Library (CPDL).

To show the potential for F-TEMPO as a tool for musicologists to explore the full-text content of the collections, we look at two simple tasks: (a) finding pages which contain similar music to a given query page; and (b), given a query representing an approximation to the highest-sounding voice from a lute arrangement of a popular vocal item from the 16th century, finding a likely vocal model within the F-TEMPO index.

CCS CONCEPTS

• **Information systems** → Information systems applications; *Digital libraries and archives*;

*All authors contributed equally to this research.

DLfM 2023, November 10, 2023, Milan, Italy

ACM ISBN 979-8-4007-0833-6/23/11.
https://doi.org/10.1145/3625135.3625142

KEYWORDS

full-text searching, early music, F-TEMPO, indexing, encoded scores, lute tablature

ACM Reference Format:
Tim Crawford, David Lewis, and Alastair Porter. 2023. Exploring early vocal music and its lute arrangements: Using F-TEMPO as a musicological tool. In *10th International Conference on Digital Libraries for Musicology (DLfM 2023), November 10, 2023, Milan, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3625135.3625142

1 BACKGROUND AND MOTIVATION

Large-scale Music Information Retrieval (MIR) [6] is most often restricted to document-level retrieval (finding the item in a collection most likely to contain musical content relevant to a given query) rather than passage-level retrieval (finding all locations within all items in the collection matching the query). It is obvious that passage-level retrieval is the more arduous process, both in terms of time (the whole contents of the collection needs to be searched exhaustively), and of space (the entire contents of the collection must be available in machine-readable form and instantly accessed at search time).

For many purposes, metadata provides a surrogate for document-level retrieval, and faceted metadata searches may satisfy many of the needs of musicologists and their students, or professional or amateur performers, who can, for example, then make use of retrieved scores (usually in the form of PDF paged image files which they can print out or display on a computer screen or hand-held tablet) as they wish.[1] Any subsequent analysis of the scores will be done by eye, or by ear as the music is played. Such metadata searches will, in fact, be carried out using indexes at some level, usually invisible to the user and buried within a database system on a remote server, greatly reducing the amount of computational effort needed to perform the search.

Within the discipline of historical musicology, there is a long-standing tradition of using indexes based on musical incipits.[2] Incipits can be encoded using text-based codes such as Plaine And Easie,[2] which was developed specifically for this purpose; the musical items within many thousands of musical sources have been indexed in this manner over the past 50 years in RISM.[3] The RISM database contains over 2,000,000 encoded incipits, of which a large proportion are from manuscripts of polyphonic music from the 15th-17th centuries. These are discoverable through the RISM's

[1] https://imslp.org
[2] https://www.iaml.info/plaine-easie-code
[3] https://rism.info

OPAC, which has recently been enhanced and provided with a new user interface.[4]

However, incipits in themselves offer very limited possibilities for the large-scale analysis of music analogous to the 'distant reading' paradigm familiar from work in the Digital Humanities.[5] For this, we need access to full-text encodings of musical scores. Because of the great amount of expert effort involved, the concomitant cost, and the likelihood of errors due to fatigue, a 'manual' encoding solution is simply not scalable to large collections of scores.

In the domain of audio recordings, a vast quantity of digital music data is, in principle, available for this kind of 'distant listening', limited mainly by issues to do with rights ownership. But the state-of-the-art of audio to score transcription has not yet reached the point where any but the least complex score can be recreated from the clearest of audio signals.[1] For this reason, audio transcription is still not an option for building large corpora of encoded scores.

The remaining option for building large corpora of encoded music is to use Optical Music Recognition (OMR)[7], running it as an unsupervised batch process over any collections of suitable images to be indexed. However, the current state-of-the-art in OMR is by no means perfect, and recognition errors are frequent; for this reason we have devised an indexing method for F-TEMPO which is reasonably robust to the most serious of these errors.

In this paper, we investigate the use of F-TEMPO as a tool for musicologists, allowing retrieval from full-text sources of early modern (sixteenth-century) music, whether OMR or manually encoded, and whether vocal music (already explicitly in voices) or lute music (with voicing being implicit at best). We test the performance for the same or similar music in a vocal corpus, and then for searches based on lute tablature queries.

2 F-TEMPO

2.1 OMR and Indexing

The OMR system we use in our specialised early-music case, Aruspix,[5] was developed for recognising early printed typeset music. Aruspix saves its recognised output in various formats, including a version of the MEI mensural module.[6]

Like all OMR systems, Aruspix produces errors. These may be divided into two classes: (i) those that affect just a single musical object in the score (errors in a note's diatonic pitch or an accidental, for example); or (ii) those which may affect all subsequent notes (a durational error in a single note-value will have the effect of displacing all subsequent notes in time; an incorrect clef at the beginning of a staff alters the pitches of all following notes).

Despite these errors (typical of those encountered in all OMR systems) Aruspix can often achieve near-perfect recognition of the main features of music in high-quality scans taken from clear original prints in good condition. Our indexing method, designed for retrieval rather than for for detailed musical analysis, must produce useable results in the presence of both kinds of error.

The relative prevalence of type ii errors in note-values makes the capture of musical rhythm highly unreliable, so we do not attempt this at all in the F-TEMPO indexing method. On the other hand, type ii errors introduced by incorrect clef-recognition can be sidestepped by using note-names, i.e. diatonic, rather than chromatic pitch, and encoding diatonic intervals between notes instead of absolute values. Although the use of intervals introduces its own problem (a note-name error on a single note is likely to affect two intervals, one before and one after the note) it has the advantage that this allows the matching of music printed at different pitches in different editions, which does occur from time to time.

As our corpus consists of monophonic lines, transcribed from partbooks each intended for a single singer, we can adopt string-matching methods designed originally for bioinformatics, running them over our diatonic interval strings encoded using alphabetic characters. For this, we use a simple code originally devised by RISM for rapid searching of musical incipits. Upper-case letters (A-Z) represent ascending intervals and lower-case (a-z) descending ones; a hyphen ('-') is used for the case where a note is repeated at the same pitch. Rests are ignored altogether.

From the MEI output of Aruspix we encode all the intervals on each page as described above. Using the method described in [4], we derive a set of Minimal Absent Words (MAWs) from the resulting 'codestrings' and save these together with an ID value representing the page-image from which the codestring was derived. The search index is stored in Solr[7].

2.2 Queries and Searches

Queries to the system can be carried out using the F-TEMPO website or via an HTTP API, either by submitting a query codestring (usually itself derived automatically) or by entering the ID corresponding to a page in the corpus, in which case the system uses the corresponding codestring for that page as a query. Digital images can also be uploaded, in which case F-TEMPO performs the OMR recognition and indexing process in a few seconds and carries out the search automatically.

A search using MAWs is performed by simply counting the absent words in common between the query and each document in the corpus in turn. This has the great benefit that, owing to a theoretically-established property of minimal absent words, the order of characters in the strings from which they are derived is preserved.[3] Final results are ranked by decreasing similarity before being presented to the user; in early testing we found that the best results are obtained using the widely-used Jaccard distance measure,[8] which favours matches with codestrings of similar length to those of the query.

3 THE F-TEMPO CORPUS

3.1 Early Printed Music

The Early Music Online collection (EMO) originated in a JISC Rapid Digitisation project in 2011; about 300 books of typeset music printed before 1600 from the British Library were catalogued and their page-images digitised from high-quality archival microfilms and made freely available online.[9] The books in EMO are all 'anthologies' containing mostly vocal music by more than a single

[4]https://rism.online
[5]https://www.aruspix.net
[6]https://music-encoding.org/guidelines/dev/modules/MEI.mensural.html
[7]https://solr.apache.org
[8]https://en.wikipedia.org/wiki/Jaccard_index
[9]https://www.royalholloway.ac.uk/research-and-teaching/departments-and-schools/music/research/research-projects-and-centres/early-music-online/

composer, largely comprising part-books for single singers, plus a few scores and choir-books, which present all the voice-parts on a single double-page opening. F-TEMPO's index includes around 30,000 separate page-images from the total of approximately 10,000 works in the collection.

In 2019 we were granted access to a much larger collection of high-quality colour digital images of printed music from a similar period held by the Bavarian State Library in Munich, Germany. Although this also included books of music theory, or texts otherwise containing music examples, and a large quantity of hymn-books, all of which are outside the immediate scope of our project, we have incorporated all of this material (just over 7,000 books) into our indexes, adding about 400,000 pages altogether. At the same time, we were able to incorporate digital images for just under 500 similar books from the Polish National Library in Warsaw. Although many of these are incomplete (often single surviving part-books from collections of vocal or instrumental music), this collection includes a significant number of unique items. We have also downloaded smaller selections of similar material from the *Bibliothèque nationale* in Paris and the Berlin State Library.

3.2 Encoded Scores of 16c vocal music

During the TROMPA project (2017-20)[10] we downloaded about 11,000 encoded four-voice scores of vocal music from the Choral Public Domain Library (CPDL)[11] in MusicXML format. Each vocal line from each score is indexed in the same way as the rest of the collection. The total number of pages in the F-TEMPO index currently stands at 482,486, and we intend that this will increase significantly in the near future.

4 RETRIEVAL TESTS

4.1 Page-based musical similarity matching

A straightforward test of the F-TEMPO system is to assess its ability, given a page-image's ID as query, to return pages containing music 'similar' to that in the query page. In this test, the general level of error in query-encoding is likely to be similar to that in F-TEMPO's internal indexes, subject to the OMR program used in each case to convert from an image into music.

We used anonymised user-generated data from early testers of F-TEMPO on the British Library's (GB-Lbl) EMO collection. The online GUI offers users an option to provide simple relevance judgements from a menu for the images matched to a query: 'duplicate' images are different photographs of the same page from the same book; 'same music' images are from different original editions or issues of the music, whose layout (on printed systems or pages) may often be very different from that in the query, or those with different texts (including *contrafacta* or translations); and the more subjective category of 'relevant' music often arises from a different voice part of the same polyphonic work in imitative style, or from a derivative work.

Query IDs from 50 'same music' and 50 'relevant music' judgements were randomly selected from the log, and re-run as queries over the current entire F-TEMPO database. Since the now greatly enlarged collection often contains matching pages from libraries other than GB-Lbl, we do not expect meaningful precision estimates to emerge from the results, and recall is impossible to estimate without detailed page-level metadata. For this reason, we simply visually checked each search, counting the number of 'same music' and 'relevant music' items in the top ten rankings, in each case omitting the first, 'identity' match of a query page with itself. We also ran 50 queries randomly selected from the larger database (Table 1).

Queries	Same		Relevant	
	Total	Avg	Total	Avg
50_EMO_same	170	3.4	28	0.6
50_EMO_relv	105	2.1	99	2.0
50_Random	52	1.0	5	0.1

Table 1: Number of matches in top 10 for 'same' and 'relevant' music, and averages for each of 50 queries.

These results are, predictably, highly dependent on the music that makes up the F-TEMPO collection; in general the proportion of musically-similar items cannot be predicted. However, it is likely that the number of 'same' matches will increase as the contents of more books from other libraries are included, due partly to the likely presence of multiple copies of works by the best-known composers, or of increasing numbers of books in different editions. This is also true, to a lesser extent, of 'relevant' matches, since in the predominant largely imitative style of 16th-century composition similar motifs are likely to be found in other voices of a work. But this is by no means universally the case, as the '50_Random' figures show: the number of 'relevant' matches is one tenth of the 'same' matches (though a far larger sample is needed to verify this).

4.2 Queries from Lute Tablatures

Many examples of arrangements in lute tablature ('intabulations') of well-known polyphonic vocal works from the 16th century exist, whether sacred (Latin motets or mass-extracts) or secular (Italian madrigals, French chansons or German *Lieder*). In many cases, the title or composer's name is either not given, incorrect or even garbled beyond readability. It would be desirable to be able to identify, from a collection such as F-TEMPO, possible vocal models for a given intabulation. Using crude pitch-based heuristics, it is possible to encode an approximation to the diatonic note-sequence of the highest voice in an intabulation (e.g. all notes above a certain threshold pitch at specified time-intervals) and use it to generate our queries. This is the basis of the experiment we report here.

As long as the tuning is known, the pitch for each of the notes in every chord can be extracted and used as a basis for indexing a collection of such music. But lute tablature - with almost no exceptions[12] - gives no information as to which voice any note belongs, leaving the expression of such voicing to the skill and taste of the player.

Although we hope that the robustness of our matching process to OMR errors can also serve to ensure a degree of success, there are two central problems with this simple process. The first is that the top note of a chord struck at a certain time may not in fact belong to the uppermost voice part in the model; besides the complication of

[10] https://trompamusic.eu
[11] https://www.cpdl.org/wiki/

[12] Sebastian Ochsenkun's *Tabulaturbuch auff die Lauten* (Heidelberg, 1558), carefully arranges the symbols of German lute tablature to show the individual voice parts in the model.

possible exchanges between the upper voices, a lower voice might in fact be the highest pitch sounding at that time (most obviously when the top voice has a rest). In each of these cases, every wrong note generates at least two errors in the pitch-interval sequence on which our method depends. So, as with OMR-derived queries, our codestrings are potentially highly subject to error, and we cannot expect perfect results.

The second problem, which is a general issue with arrangements of this period, is that when setting a vocal item for performance on the lute or a keyboard instrument, it was normal to add extra notes in the form of elaborations of the melody or other voice parts. This 'diminution' technique was an expected aspect of the virtuoso performer's skill, and only rarely (usually in the context of a didactic work) are 'plain', unadorned versions encountered. In the same way as described above, a single extra note in a melodic line introduces two errors in an interval representation; for this reason it is desirable to attempt to reduce the ornamented surface as near as possible to a 'vocal' melody which might correspond to the music printed in the vocal *cantus* or *superius*.

Since the encoded lute intabulations we work with are reliably accurate in terms of rhythm, we can calculate the time at which each chord is played. In order to mitigate the effects of both the uncertain voicing in lute-tablature derived queries, and the interference from ornamental notes, we extract our diatonic pitch-interval strings from the top note of successive chords which occur at a regular time-interval corresponding to one of the predominant note values in the repertory of vocal models. Here we use time-intervals equivalent to *fusa* (eighth note), *semiminima* (quarter) and *minima* (half) durations, and use the highest notes sounding at these times to extract three different codestrings for each piece as queries.

For the trials we report here, we used a freely-accessible online resource maintained by Sarge Gerbode containing around 20,000 tablature encodings from all periods, including several arrangements of modern popular music and basso continuo accompaniments.[13] For the most part, these are manually created from original sources, often dealing with the entire contents of a source (printed or manuscript); in this respect, this is an extremely rich resource of performing editions of music from the renaissance period which has been under-exploited for its musicological potential. However, some caution is needed since there is little by way of 'quality control' over the encodings, which can cause problems, though usually at the level of fine detail (where, for example, unreported editorial alterations are included, or repeats written out); these are not 'diplomatic' transcriptions from the source material.

Since at the time of our tests we had no detailed (page/item-level) metadata for the F-TEMPO collection, for our first attempt we selected as queries to the system a small set of encoded intabulations of vocal items (identified from their titles within the Gerbode corpus) which we knew to have multiple corresponding page-images within F-TEMPO. For each of five well-known 4-part vocal works known to be represented in both the Gerbode intabulations and the F-TEMPO database, we collected all the lute versions we could identify by title, and tested the retrieval effectiveness in F-TEMPO at varying time-scales. The works tested in this way (a small fraction of the possible number) were: 'Ancor che col partire' (Cipriano da Rore); 'Il bianco e dolce cigno' (Jacques Arcadelt); 'Tant que vivray' (Claudin de Sermisy); 'Ung gay bergier' (Thomas Crecquillon); and 'Bonjour mon coeur' (Orlande de Lassus).

[13]https://www.lutemusic.org

Using our own topline-extraction method at the three time-intervals on each of the lute intabulations in turn, we counted the number of intabulations for which F-TEMPO correctly retrieved their respective vocal models within the top 10 results at any time interval (see Table 2). This would support the task of work identification. Where any relevant results were retrieved, there were usually many vocal versions returned.

Vocal item	Items in F-TEMPO	Lute versions	
		Queries	Successful
'Ancor che col partire'	15	3	0
'Il bianco e dolce cigno'	9	1	1
'Tant que vivray'	4	6	1
'Ung gay bergier'	5	4	1
'Bonjour mon coeur'	12	3	1

Table 2: Number of tablature queries that return their vocal model within the top ten F-TEMPO results.

Four out of the 15 lute intabulations could in fact be easily identified within the top ten F-TEMPO hits. This result is highly dependent on the time-interval at which notes are selected and the degree of ornamentation; an example is given in Figure 1, showing where we selected notes at time-steps 128 (quarter) and 256 (half). Also, the relationship between the tuning chosen for the instrument and the threshold below which notes are ignored (currently set at C4) is critical; in Figure 1, the notes coloured in red would be gathered using our defaults, and thus the 'top' voice has been contaminated with notes from a middle part. Some of the results in the table may seem disappointing, and this is probably due to these factors or other errors induced in our diatonic interval sequence extraction method. However, we have shown that by selecting the best results from multiple searches with different parameters, we can definitely identify the vocal models near the top of our rankings in all but one item; improving this will be the subject of future work.

Figure 1: Opening of 'Anchor che col partire' from F-TEMPO and transcription of opening of a typical lute intabulation (M. Newsidler, 1574). Notes are selected for indexing at time-steps 128 and 256 'ticks' as shown below the transcription.

ACKNOWLEDGMENTS

We are especially grateful to: The British Library; RISM Digital; Jürgen Diet and Bernhard Lutz at D-Mbs; Sonia Wronkowska at PL-Wn; Sarge Gerbode. The research was partially funded by the UK AHRC, The British Academy and JISC, EU Horizon 2020, and Goldsmiths, University of London.

REFERENCES

[1] Emmanouil Benetos, Simon Dixon, Zhiyao Duan, and Sebastian Ewert. 2019. Automatic Music Transcription: An Overview. *IEEE Signal Process. Mag.* 36, 1 (2019), 20–30.

[2] Barry S. Brook. 1973. A Tale of Thematic Catalogues. *Notes* 29, 3 (1973), 407–415.

[3] Supaporn Chairungsee and Maxime Crochemore. 2012. Using minimal absent words to build phylogeny. *Theor. Comput. Sci.* 450 (2012), 109–116.

[4] Tim Crawford, Golnaz Badkobeh, and David Lewis. 2018. Searching Page-Images of Early Music Scanned with OMR: A Scalable Solution Using Minimal Absent Words. In *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France, September 23-27, 2018*, Emilia Gómez, Xiao Hu, Eric Humphrey, and Emmanouil Benetos (Eds.). 233–239.

[5] Franco Moretti. 2000. Conjectures on world literature. (2000).

[6] Markus Schedl, Emilia Gómez, and Julián Urbano. 2014. Music Information Retrieval: Recent Developments and Applications. *Found. Trends Inf. Retr.* 8, 2-3 (2014), 127–261.

[7] Elona Shatri and György Fazekas. 2020. Optical Music Recognition: State of the Art and Major Challenges. In *International Conference on Technologies for Music Notation and Representation.*