FOUR DIFFICULT LESSONS ON AUTOMATIC CHORD ESTIMATION

First author
Affiliation1
`author1@ismir.net`

Second author
Retain these fake authors in
submission to preserve the formatting

Third author
Affiliation3
`author3@ismir.net`

ABSTRACT

Automatic chord estimation (ACE) is now a hallmark research topic in content-based music informatics, but like many other tasks, system performance appears to be converging to yet another glass ceiling. Recently, two different large-vocabulary ACE systems were developed in the hopes that complex, data-driven models might significantly advance the state of the art. While arguably achieving some of the highest results to date, both approaches plateau at the same level, well short of having solved the problem. Therefore, this work explores the behavior of these two systems as a means of understanding obstacles and limitations in chord estimation, arriving at four difficult lessons: one, music recordings that invalidate tacit assumptions about harmony and tonality result in erroneous and even misleading performance; two, standard lexicons and comparison methods struggle to reflect the natural relationships between chords; three, conventional approaches conflate the competing goals of recognition and transcription to some undefined degree; and four, the perception of chords in real music can be highly subjective, making the very notion of "ground truth" annotations tenuous. Synthesizing these observations, this paper offers possible remedies going forward, and concludes with some perspectives on the future of ACE research.

1. INTRODUCTION

This template includes all the information about formatting manuscripts for the ISMIR 2015 Conference. Please follow these guidelines to give the final proceedings a uniform look. Most of the required formatting is achieved automatically by using the supplied style file (LaTeX) or template (Word). If you have any questions, please contact the Program Committee (`ismir2015-papers@ismir.net`). This template can be downloaded from the ISMIR 2015 web site (`http://ismir2015.ismir.net`).

2. METHODOLOGY

The regulations for the maximal paper length have changed. Instead of the strict limit of six pages (as used for ISMIR 2014), we adopt a "(6+1)-page policy" for ISMIR 2015. This means, the paper may have a maximum of 6 pages for technical content including figures and possible references with one additional optional 7th page containing only references. Note that this is a strict requirement. The seventh page (if used at all) must not contain any other material except for references.

2.1 Automatic Systems

3. PAGE SIZE

The proceedings will be printed on portrait A4-size paper (21.0cm x 29.7cm). All material on each page should fit within a rectangle of 17.2cm x 25.2cm, centered on the page, beginning 2.0cm from the top of the page and ending with 2.5cm from the bottom. The left and right margins should be 1.9cm. The text should be in two 8.2cm columns with a 0.8cm gutter. All text must be in a two-column format. Text must be fully justified.

4. TYPESET TEXT

4.1 Normal or Body Text

Please use a 10pt (point) Times font. Sans-serif or non-proportional fonts can be used only for special purposes, such as distinguishing source code text.

The first paragraph in each section should not be indented, but all other paragraphs should be.

4.2 Title and Authors

The title is 14pt Times, bold, caps, upper case, centered. Authors' names are omitted when submitting for double-blind reviewing. The following is for making a camera-ready version. Authors' names are centered. The lead author's name is to be listed first (left-most), and the co-authors' names after. If the addresses for all authors are the same, include the address only once, centered. If the authors have different addresses, put the addresses, evenly spaced, under each authors' name.

4.3 First Page Copyright Notice

Please include the copyright notice exactly as it appears here in the lower left-hand corner of the page. It is set in 8pt Times.

© First author, Second author, Third author. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** First author, Second author, Third author. "Four Difficult Lessons on Automatic Chord Estimation", 16th International Society for Music Information Retrieval Conference, 2015.

String value	Numeric value
Hello ISMIR	2015

Table 1. Table captions should be placed below the table.

4.4 Page Numbering, Headers and Footers

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

5. FIRST LEVEL HEADINGS

First level headings are in Times 10pt bold, centered with 1 line of space above the section head, and 1/2 space below it. For a section header immediately followed by a subsection header, the space should be merged.

5.1 Second Level Headings

Second level headings are in Times 10pt bold, flush left, with 1 line of space above the section head, and 1/2 space below it. The first letter of each significant word is capitalized.

5.1.1 Third and Further Level Headings

Third level headings are in Times 10pt italic, flush left, with 1/2 line of space above the section head, and 1/2 space below it. The first letter of each significant word is capitalized.

Using more than three levels of headings is highly discouraged.

6. FOOTNOTES AND FIGURES

6.1 Footnotes

Indicate footnotes with a number in the text. [1] Use 8pt type for footnotes. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a 0.5pt horizontal rule.

6.2 Figures, Tables and Captions

All artwork must be centered, neat, clean, and legible. All lines should be very dark for purposes of reproduction and art work should not be hand-drawn. The proceedings are not in color, and therefore all figures must make sense in black-and-white form. Figure and table numbers and captions always appear below the figure. Leave 1 line space between the figure or table and the caption. Each figure or table is numbered consecutively. Captions should be Times 10pt. Place tables/figures in text as close to the reference as possible. References to tables and figures should be capitalized, for example: see Figure 1 and Table 1. Figures and tables may extend across both columns to a maximum width of 17.2cm.

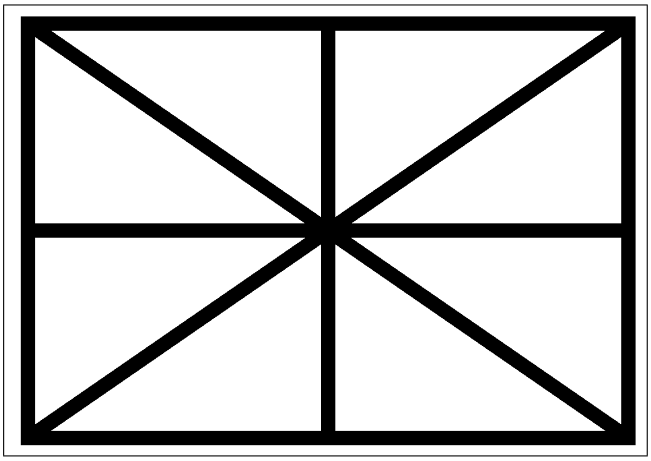

Figure 1. Figure captions should be placed below the figure.

7. SUMMARY

In this work, the application of deep learning to large-vocabulary ACE is thoroughly explored, advancing the state of the art using standard evaluation methods. Arguably of more importance, both the behavior of the resulting systems and the data used for development are explored in rigorous detail. Our results show that the state of the art may have truly hit a glass ceiling, due to the conventional assumption that “ground truth” data can be obtained for what is, at times, an unavoidably subjective task. This challenge is further compounded by approaches to prediction and evaluation, which attempt to perform flat classification of a hierarchically structured chord taxonomy. Thus, while there certainly remains room for improvement, error analysis indicates that the vast majority of error in modern chord recognition systems is a result of invalid assumptions baked into the very question being asked.

Notably, four issues with current chord estimation methodology have been identified in this work. One, it seems necessary that computational models, and especially those that estimate a large number of chord types, embrace structured outputs; one-of-K class encoding schemes introuduce unnecessary complexity between what are naturally hierarchical relationships. Two, there is value in distinguish between the two tasks at hand, being chord recognition —I am playing this *exact* chord shape on guitar— and chord transcription —finding the best chord label to describe this harmonically homogenous region of music— and how this intent is conveyed to the authors of reference annotations. Three, as championed by [**?**], chord transcription would certainly seem to benefit from explicit segmentation, rather than letting such boundaries between regions of harmonic stability result implicitly from post-filtering algorithms, i.e. Viterbi. Lastly, the all-too-often subjective nature of chord labeling needs to be acknowledged in the process of curating reference data, and the human labeling task should average or combine multiple perspectives rather than attempt to yield canonical “expert” references.

[1] This is a footnote.

8. REFERENCES

[1] E. Author. The title of the conference paper. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 000–111, 2000.

[2] A. Someone, B. Someone, and C. Someone. The title of the journal paper. *Journal of New Music Research*, A(B):111–222, 2010.

[3] X. Someone and Y. Someone. *The Title of the Book*. Editorial Acme, Porto, 2012.