Judah Daniels

Inferring Harmony from Free Polyphony

Computer Science Tripos – Part II

Clare College

July, 2023

Declaration of originality

I, Judah Daniels of Clare College, being a candidate for Part II of the Computer Science Tripos, hereby declare that this dissertation and the work described in it are my own work, unaided except as may be specified below, and that the dissertation does not contain material that has already been used to any substantial extent for a comparable purpose. I am content for my dissertation to be made available to the students and staff of the University.

Signed Judah Daniels

Date May 5, 2023

Proforma

Candidate Number:	**2200D**
Project Title:	**Inferring Harmony from Free Polyphony**
Examination:	**Computer Science Tripos – Part II, July, 2023**
Word Count:	**8900**[1]
Code Line Count:	**2672**[2]
Project Originator:	**Christoph Finkensiep**
Supervisor:	**Dr Peter Harrison**

Original Aims of the Project

Work Completed

Special Difficulties

There were no special difficulties encountered in this project

[1]This word count was computed by `texcount -v1 -sum -sub=chapter main.tex`
[2]This code line count was computed by using `cloc`

Contents

Acknowledgements

I'd like to thank Christoph Finkensiep and Peter Harrison for being awesome.

Chapter 1

Introduction

The problem of inferring the harmonic structure of a piece of music, represented by a sequence of chords, is a fundamental task in the analysis and understanding of Western tonal music. Free polyphony refers to one of the most general forms of western music, wherein multiple independent melodic lines (or voices) are combined without adhering to strict rules or constraints. Symbolic approaches to Automatic Chord Estimation (ACE) usually solve such problems by taking a sequence of notes as the input, and generating a sequence of chord labels as the output. In this project, the input, called the *surface*, is a sequence of notes with precise descriptions of when each note begins (onset) and ends (offset). The output is a sequence of chord labels, each describing a *chord segment*, a group of notes that sound simultaneously or in close succession.

This project proposes a novel approach to ACE that integrates the proto-voice model [15], a recent model of note-level structure, with a chord segment reduction process. This contributes an **interpretable** framework for inferring sequences of chord labels by providing an explicit explanation of how those labels relate to the surface notes. Furthermore, I design and implement a **novel fitness function** and **efficient heuristic search algorithms** to improve on the computational complexity of the parsing algorithm provided in the proto-voice paper [15] from **exponential to linear time** with respect to the length of the piece.

1.1 Previous Work

Automatic chord estimation systems have ranged from handcrafted grammar/rule-based approaches [34] [60], to the development of optimisation algorithms [41]. In more recent years, deep learning methods have risen in popularity, exploiting large datasets and improved compute power. Examples of deep architectures used include recurrent neural networks (RNNs) [1] and long short-term memory (LSTM) networks [5]. Some systems also make use of audio rather than symbols as an input, utilising convolutional neural networks (CNNs), and most recently convolutional recurrent neural networks (CRNNs) [61] [9]. These architectures have found success due to their ability to capture temporal dependencies in the music.

A broad limitation of these existing ACE approaches is that they do not specify the precise

relationship between the surface and the inferred chord labels. Deep-learning approaches typically involve a black box inference of the chord labels based on the raw input. For example, a recent chord classification model described by McLeod [35] consists of a neural network that takes a sequence of notes represented as one-hot feature vectors and outputs a distribution over a set of 1540 chord labels. Earlier probabilistic optimisation approaches consider the surface as being generated by a noisy process. Pardo and Birmingham's influential algorithm compares the notes in each segment to a set of templates describing common chords [41]. There is no explicit relationship described between the notes in each segment and the nearest template.

The recent paper *Modeling and Inferring Proto-voice Structure in Free Polyphony* [15] presents the proto-voice model, a generative model that represents a musical piece as a result of recursively applying primitive operations on notes. The combination of these operations forms a hierarchical structure which encodes explicit relations between all the notes in a piece. The paper presents a *chart parsing* algorithm which can parse a piece of music according to the grammar outlined by the model, returning a list of possible *derivations*, that is, all the ways the piece of music could have been created through the recursive application of these simple rules.

In this project, I build upon the proto-voice model by creating a new *proto-voice harmony parser* to enable the generation of a *reduction* of the musical surface that can be used to directly infer chord labels. In music analysis, a reduction refers to the process of simplifying a complex musical surface to reveal its underlying harmonic structure. The new parser is used to reduce the piece, with the goal of finding a reduction that preserves the harmony of the piece, while eliminating extraneous notes. This reduction facilitates the inference of descriptive chord labels and provides an *explicit explanation* through the corresponding proto-voice derivation, which has not been attempted before.

While the chart parsing algorithm provided [12] could in theory be used to generate harmonic annotations, the naive exhaustive parse strategy would be prohibitively time-consuming in practice for all but the shortest musical extracts; one half measure can have over 100,000 valid derivations [13]. Moreover, not all derivations are created equally. Each one corresponds to a specific interpretation of the note-level structure within a given piece of music. These reductions align with intuitive or theoretical musical interpretations to differing extents, with most being implausible. Determining the relative plausibility of derivations remains an *open question*.

I provide answers to these questions by proposing and implementing a **novel fitness function** for proto-voice derivations and address the **exponential search space** using heuristics and pruning techniques to significantly reduce the time and space complexity from $\boldsymbol{\mathcal{O}(2^{n \cdot m})}$ to $\boldsymbol{\mathcal{O}(n \cdot m)}$, where n is the length of in piece and m is the number of notes in each segment.

1.2 Contributions of my Project

The implementation of this project achieved all of the original aims. In doing so, the key contributions include:

1. Developing a novel approach to Automatic Chord Estimation that integrates the proto-voice model with a chord segment reduction process (Section 3). This not only provides

an interpretable and explainable framework for Automatic Chord Estimation but also has the potential to contribute to a more comprehensive understanding of the underlying structure of Western tonal music.

2. Proposing and implementing an informed heuristic-based fitness function to evaluate proto-voice derivations, which decomposes into a estimated score for each reduction step (Section 3.2.1). This enables the identification of musically meaningful interpretations of the piece while discarding implausible derivations.

3. Designing and developing the proto-voice harmony parser which can generate reductions of the musical surface to facilitate chord label inference (Section 3.3). This has the benefit over previous approaches of providing an explicit explanation through the proto-voice derivation, with potential to be incorporated in more complex systems to improve their interpretability.

4. Introducing a novel heuristic search strategy to efficiently explore the large space of possible derivations, using a beam search and pruning techniques to significantly reduce the computational complexity (Section 3.6).

5. Demonstrating the effectiveness of this approach on a diverse set of musical pieces in terms of accuracy, performance and interpretability (Section 4).

Chapter 2

Preparation

This Chapter provides an overview of the essential concepts, models and representations that are necessary to understand the implementation of this project. First, the representation used for pitches is described, along with a representation of chord labels commonly used in ACE (Section 2.2). Next, the proto-voice model is introduced, explaining the structures and operations used to manipulate the musical surface (Section 2.3). Subsequently, an outline of the existing parser is presented (Section 2.4), highlighting the limitations and innovations required. Finally, a discussion and analysis of project requirements is provided (Section 2.5) followed by a description of the software engineering techniques used throughout the project (Section 2.6).

2.1 Starting Point

This project builds on the codebase of Finkensiep, the `protovoices-haskell` repository [12], which contains an implementation of the proto-voice model, types and functions for representing and working with protovoice derivations, and an implementation of a chart parser and greedy parser. Everything else in the codebase is written by me, including the new harmony parser, a harmony module, the fitness function and heuristic search algorithms.

2.1.1 Relevant courses and experience

IB Data Science and *Artificial Intelligence* provided some of the Machine Learning background required and *Formal Models of Language* introduced some of the ideas and terminology used in the proto-voice model.

I had only a few months' experience with Haskell before starting this project, being introduced to the language during an internship in the summer of 2022. The existing `protovoices-haskell` repository is a large and complex codebase, and the parser provided is non-trivial. Completing this project involved learning many features of Haskell, including gaining familiarity with Monad Transformers, the subtleties of lazy execution and its rich type system.

I had ample experience coding in Python from personal projects as well as the 1A *Scientific Computing Practical Course*.

2.2 Inferring Harmony

The approach taken to infer harmony is to take a *multi-set* of pitches as an input, describing the counts of the pitches present in a chord segment, and predict the chord label that best describes that chord segment. In this project, each label is inferred independently for each chord segment, ignoring contextual information.

Definition 2.2.1 (Multi-set)**.** A *multi-set* is a set that allows multiple instances for each of its elements, formally defined as an ordered pair (A, m) where A is the *underlying set* of the multiset, and $m : A \to \mathbb{Z}^+$ gives the *multiplicity*, such that the number of occurrences of a in (A, m) is given by $m(a)$.

Chord Labels

Each chord label comprises a root note and a chord-type (e.g. major, minor), and are thus described as a pair $\mathcal{L} = (\mathcal{P}, \mathcal{C})$ where $\mathcal{P}$ is the set of pitches and $\mathcal{C}$ is the set of chord-types. The set of chord-types used in this project corresponds to the Digital Cognitive Musicology Lab (DCML) annotation standard, and consists of 14 unique types. As a result, there are $14 \times 29 = 406$ unique chord labels considered. I provide an implementation of chord labels using these chord-types as part of the Harmony model.

Pitches

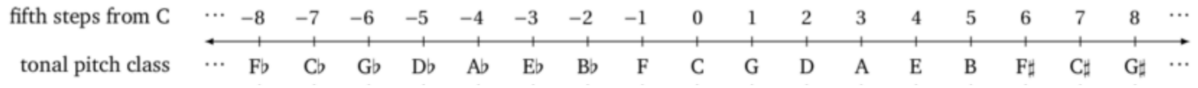


Figure 2.1: Tonal pitch class representation. The reference pitch class is C.

The representation used for pitches is the *tonal pitch class*, of which an implementation is provided in the `haskell-musicology` library [11]. The central object of this representation is the *interval*, the 'distance' between pitches, which is described by an integer representing the distance between pitches along the *line of fifths* (See Figure 2.1). Pitches are then derived from intervals by interpreting them with respect to a reference pitch. This is similar to the relation between vectors (intervals) and points (pitches) [11]. Although the line of fifths is theoretically infinite, in this project the reference pitch is set to C, and the intervals can range from -14 to 14, resulting in 29 unique tonal pitch classes (octaves are ignored).

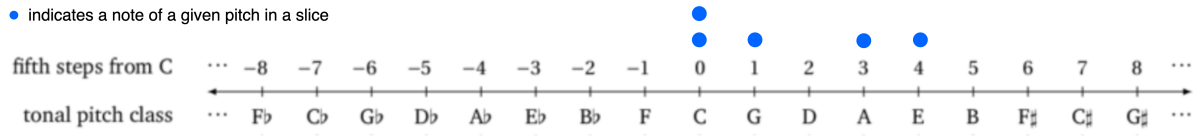


Figure 2.2: A representation of a slice showing the counts of each pitch.

Slices

Slices are multi-sets of pitches, used describe sets of pitches that sound simultaneously or in close succession, i.e. a set of pitches that sound within a 'slice' of time, or single time-frame.

Definition 2.2.2 (Slice)**.** A *slice* $s \in \mathcal{S}$ is defined as a multi-set of pitches $(\mathcal{P}, m)$.

$$s = \left\{ p_1^{m(p_1)}, \ldots, p_n^{m(p_n)} \right\} \tag{2.1}$$

Chord Profiles

Chord profiles, also known as chord templates, are often used in Automatic Chord Estimation (ACE) systems to find the chord label which best matches a given slice.

Chord profiles describe the *intervals* with respect to the *root note* of the chord that are typically present within a segment described by the given chord type. Figure 2.3 gives example chord profiles for major, minor and minor 7 chord-types, and shows how the profiles describe which notes are present for a given chord, by interpreting the intervals of the chord profile with respect to the root note. Intuitively, for a given chord-type, shifting the corresponding profile along the axis shifts the root-note of the chord. In Figure 2.3, the notes in the slice match the chord profile for an *A minor 7* chord exactly, and almost matches the profile for a *C Major* chord.

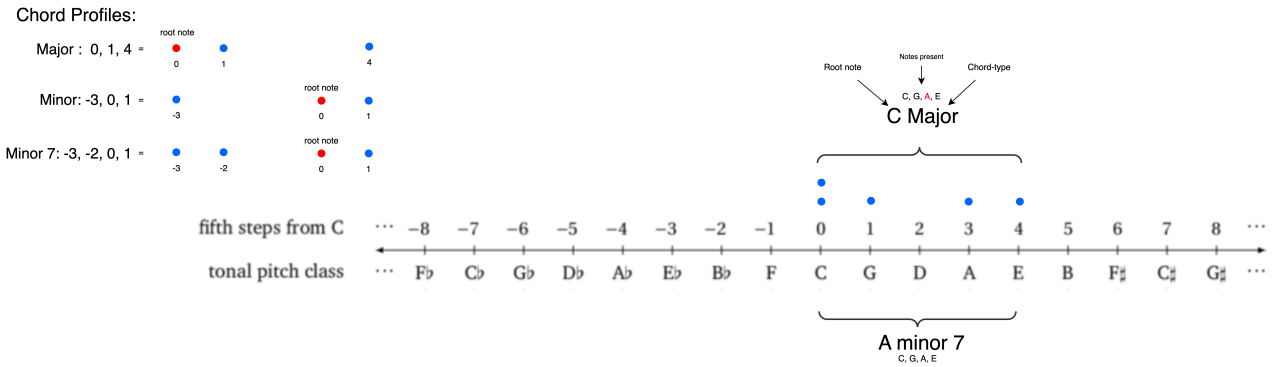


Figure 2.3: Application of chord profiles. The dots represent the presence of a pitch.

Chord labeling is inherently ambiguous; different chord labels can consist of similar (or even identical) groups of notes, and not all notes directly relate to harmony. There are cases where the notes in a given segment match a chord exactly, but an analyst would decide that one of those notes is in fact an ornament and the remaining notes exemplify a different chord. For example, the notes A, C, E and G match the profile for an *Am7* chord exactly, but there are many cases where the correct interpretation would decide that the A is an ornament, and the chord-tones C, E and G exemplify a C major chord.

Probabilistic Chord Profiles

Probabilistic chord profiles aim to address this limitation of standard chord profiles by assigning each note a probability value indicating the likelihood of its presence in a particular chord. In the paper *A Bayesian Model of Extended Chord Profiles* [14], Finkensiep extends this idea by inferring two different distributions for each chord type. Notes in a given segment are categorised into chord-tones or ornaments, where the chord-tones are the notes that directly relate to the chord. The paper presents a *chord-tone distribution* and an *ornament distribution* for each chord-type, inferred from a labelled dataset. In Figure 2.4 the most common intervals are 0, 1 and 4, which corresponds exactly to the major chord profile shown in Figure 2.3.

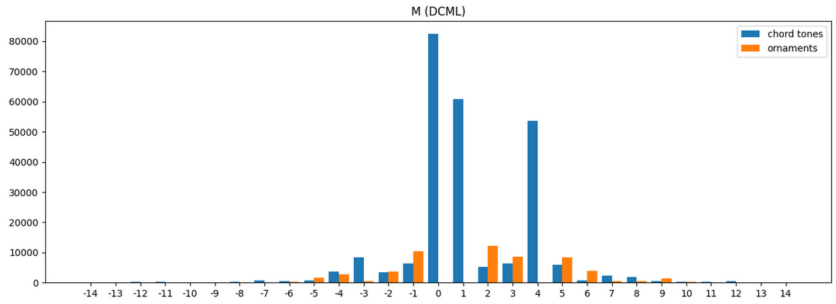


Figure 2.4: Relative counts of intervals relative to the root note in a major chord segment, distinguished by chord-tones and ornaments. Graph taken from the `chord-types and ornamentation` repository [14]

2.3 The Protovoice Model

The novel approach taken by this project is to integrate the protovoice model with a chord segment reduction process. The protovoice model is used to generate a sequence of *slices* for each chord segment to be labelled, wherein each slice clearly denotes the underlying harmony, and extraneous notes are *explained away*. By performing this reduction, the chord label inference is guided towards the most plausible interpretation for each segment.

The protovoice model is a generative model for tonal music based on a set of primitive operations that are applied recursively to create a hierarchical structure representing a musical surface. These operations are derived from a set of intuitive and theoretically motivated transformations, and was proposed by Finkensiep in the paper *Modeling and Inferring Protovoice Structure in Free Polyphony* [15]. The model is primarily concerned with the analysis of Western Classical music, although its expressiveness and generality means it could be applied to different musical styles including jazz or some popular western music [13].

In this project, I build on the existing open-source parser for the model [12] to perform the segment reduction process, allowing it to search for reductions that preserve and emphasise the harmony of the surface, rather than finding complete derivations.

The next two sections provide a description of the data structures used to represent the musical surface in the protovoice model, followed by a description of the generative protovoice operations.

2.3.1 Outer Structure: Slices and Transitions

The notation most commonly used for Western tonal music is called a *score*, a symbolic abstraction of a piece of music based on a *2-dimensional axis*. The marks on the score represent notes, with the *pitch* of the note represented by its position on the vertical axis, and the notes' placement in time represented by its position on the horizontal axis.

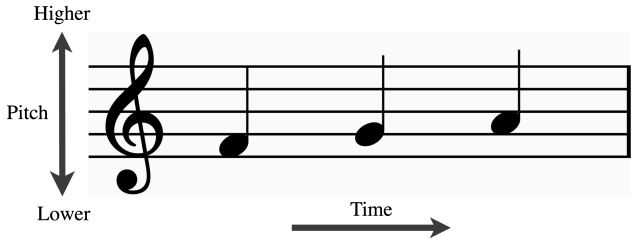


Figure 2.5: An example of music notation showing an ascending stepwise sequence.

Slices

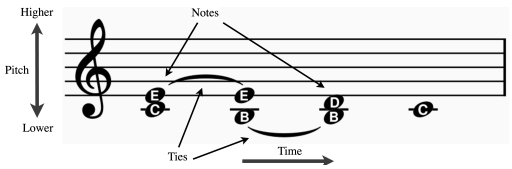
Groups of notes are stored as *slices*, initially representing the maximal durations in which a single group of pitches sound, shown in Figure 2.6b. Figure 2.6a shows a score representing a short musical phrase. Notes have an onset describing when they begin, and a duration describing how long they sound for. Ties are used to indicate that the same note continues sounding, rather than ending and starting again subsequently. As notes have different durations, notes that are simultaneous with several non-simultaneous notes are split among the corresponding slices, such as the note E in Figure 2.6a. In this case *edges* connect each of these notes, which ensures that a single surface note is generated through a single generation process [15], as shown in Figure 2.6c.

Transitions

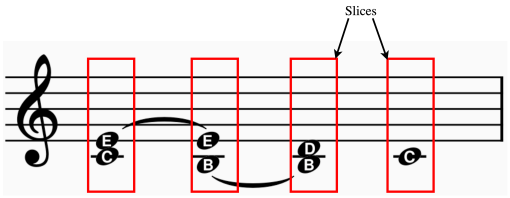
Transitions are introduced to relate adjacent slices with a configuration of edges, connecting notes in the adjacent slices. The unreduced surface contains *regular* edges, in particular, *repetitions* which connect single surface notes that span multiple slices, such as the edges shown in Figure 2.6c. Other types of edges exist only during the generation (or reduction) of a piece, and are described in the Section 2.3.2. An unreduced surface (e.g. Figure 2.6d) contains only *frozen* transitions. Frozen transition are *terminal* (analogous to CFGs), meaning no more generative operations can be applied. When all transitions are frozen, this means the generative process has completed, which is the case with the original surface.

Definition 2.3.1 (Transition)**.** A *transition* $t \in \mathcal{T}$ relates two adjacent slices, s_l and s_r, with a configuration of edges e.

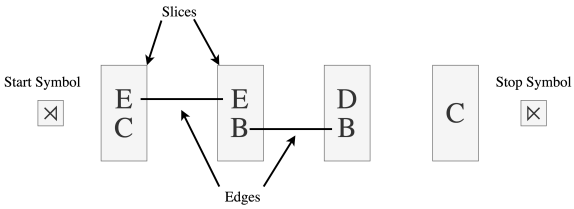
$$t = (s_l, e, s_r) \tag{2.2}$$



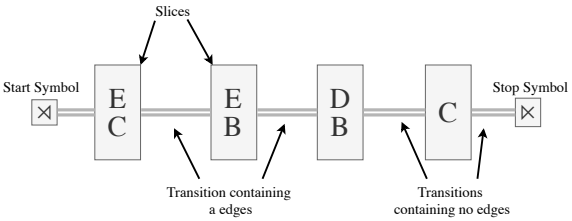
(a) The original score.



(b) The score divided into slices.



(c) Slices and edges.



(d) Slices and transitions.

The slices and transitions form a graph given slices as nodes and transitions as edges. However, as transitions only relate sequentially adjacent slices, the outer structure is in fact a *path graph* and can thus be represented as a alternating list of slices and transitions.

Definition 2.3.2 (Path)**.** A `path` is the data structure used to represent an alternating sequence of transitions and slices, defined inductively as:

$$P = t \mid t\ s\ P \qquad t \in \mathcal{T},\ s \in \mathcal{S} \tag{2.3}$$

Paths are implemented in a similar fashion to linked lists, where the first transition, the *head* of the path, is accessible in constant time.

As slices and transitions contain notes and edges respectively, slices and transitions are called *outer structure*, and the notes and edges contained therein are called *inner structure*.

2.3.2 Inner Structure: Notes and Edges

The following is a restatement of the core of the proto-voice model, as described in the original paper by Finkensiep [15].

Internally, proto-voices are represented as a directed graph with one vertex for each note contained in a slice, a vertex each for the beginning ($\rtimes$) and end ($\ltimes$) of the piece, and edges that indicate step-wise connections between notes, which are contained within transitions. A *proto-voice* is a path within this graph. The protovoice model is characterised by stepwise generative operations on notes. *Regular edges* indicate a sequential connection between two notes, which may be *elaborated* by introducing a repetition or a neighbour of either parent note, or both if the parents have the same pitch. The interval along a regular edge is always within a range of a step. *Passing edges* indicate connections between two notes with an interval larger than a step, introducing a new subordinate proto-voice. These passing edges must be filled with stepwise passing notes from either end [15].

The generation of a piece begins with the empty piece ($\rtimes \to \ltimes$) and is defined by the recursive application of elaboration rules.

These elaboration rules relate new child notes to one or two existing *parent* notes.

Single-sided rules attach a new repetition or neighbour note with an edge connected to a single parent note:

$$\begin{aligned} x &\Longrightarrow n \to x && \texttt{left-neighbor} \\ x &\Longrightarrow x \to n && \texttt{right-neighbor} \\ x &\Longrightarrow x' \to x && \texttt{repeat-before} \\ x &\Longrightarrow x \to x' && \texttt{repeat-after} \end{aligned} \tag{2.4}$$

Double-sided rules are represented by *edge replacement*.

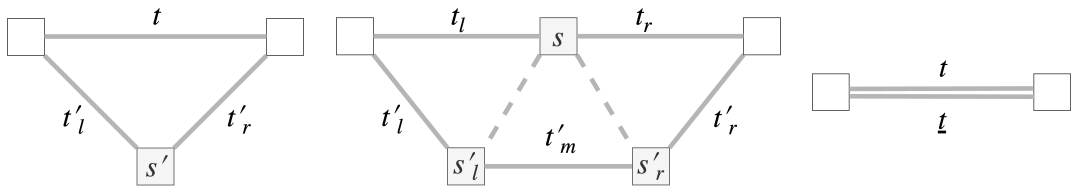
$$\begin{aligned} \rtimes \to \ltimes &\Longrightarrow \rtimes \to x \to \ltimes && \texttt{root-note} \\ x_1 \to x_1 &\Longrightarrow x_1 \to x' \to x_2 && \texttt{full-repeat} \\ x \to y &\Longrightarrow x \to y' \to y && \texttt{repeat-before'} \\ x \to y &\Longrightarrow x \to x' \to y && \texttt{repeat-after'} \\ x_1 \to x_1 &\Longrightarrow x_1 \to n \to x_2 && \texttt{full-neighbor} \end{aligned} \tag{2.5}$$

Passing rules, finally, fill passing edges with notes from the left or right until the progression is fully stepwise.

$$\begin{aligned} x \dashrightarrow y &\Longrightarrow x \to p \dashrightarrow y && \texttt{passing-left} \\ x \dashrightarrow y &\Longrightarrow x \dashrightarrow p \to y && \texttt{passing-right} \\ x \dashrightarrow y &\Longrightarrow x \dashrightarrow p \to y && \texttt{passing-final} \end{aligned} \tag{2.6}$$

The inner structure provided by protovoices captures the sequential and functional organisation of notes, but does not capture when notes are simultaneous. To model simultaneity, notes and edges are integrated into the outer structure of slices and transitions as described in Section 2.3.1.

2.3.3 Generative Proto-voice Operations



(a) `split` (b) `spread` (c) `freeze`

Figure 2.7: The three operations on outer structure. The original slices and transitions are shown at the top, while the generated structure is shown underneath. Figure reproduced from original paper [13]

The outer structure is generated through the recursive application of three production rules:

- A **split** replaces a parent transition t by inserting a new slice s' and two surrounding transitions t_l and t_r. One or more inner operations can be applied to each of the edges in t, and the resulting edges can be discarded or kept to form the new edges of t_l and t_r.

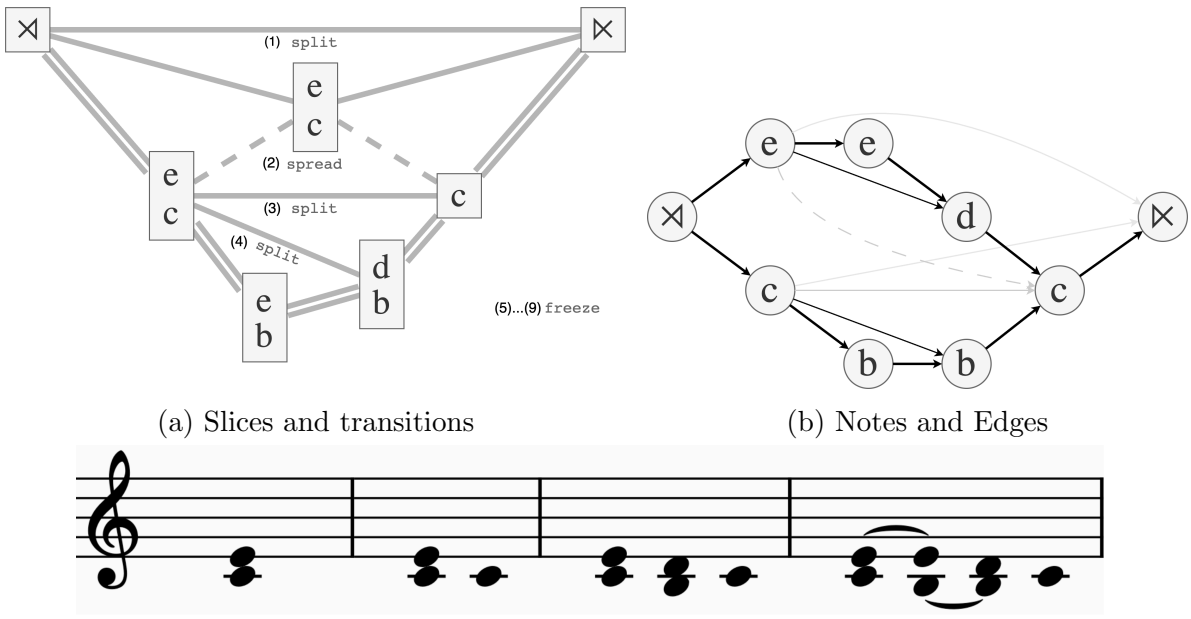
$$t \rightarrow t'_l \; s' \; t'_r \tag{2.7}$$

- A **spread** replaces a parent slice s by distributing its notes to two child slices s'_l and s'_r. This allows a vertical configuration of notes (i.e a slice) to become sequential, thereby generating implied harmonies such as those produced by broken chords, chords played without all the notes beginning simultaneously. These latent harmonies can also exist within a single melodic line, thus a single melody can be generated from an implied harmony. During a spread, passing edges can be introduced between any two of the child slices.

$$t_l \; s \; t_r \rightarrow t'_l \; s'_l \; t'_m \; s'_r \; t'_r \tag{2.8}$$

- A **freeze** marks a transition as terminal, preventing any further application of operations to its edges.

$$t \rightarrow \underline{t} \tag{2.9}$$



(a) Slices and transitions

(b) Notes and Edges

(c) Abstract notation

Figure 2.8: An example derivation of a short phrase.

Figure 2.8 gives an example derivation of a short phrase. In Figure 2.8a, nine outer operations have been applied to generate the surface. The generation of the piece begins with an initial split operation (1) which comprises two elaborations of the initial regular edge, both using the `root-note` rule, introducing child notes E and C. Next, a spread operation (2) distributes the notes in the parent slice to the two child slices, with the first slice inheriting both notes, and the second child slice inheriting just the C. This spread introduces a passing edge between E and C, shown as the dashed line in Figure 2.8b. The next `split` (3) elaborates the passing edge between E and C with the `passing-final` rule, introduce a child passing note, D. The

regular edge between the two C's is elaborated using the `full-neighbour` rule, generating the B. The final `split` (4) introduces a *repetition*, E, and a *neighbour*, B, both through double-sided regular operations. Finally, the five surface transitions are marked as terminal (5-9), indicating the surface has been fully generated. Note that these freeze operations could've been applied in many different permutations, with the only constraint being that no more operations are applied after a transition is frozen.

Note that the vertically aligned notes in Figure 2.8b correspond to the notes in the surface slices in Figure 2.8a, and the final surface shown on the right of Figure 2.8c.

2.4 Overview of Project Aims

This project aims to expand on the proto-voice model by implementing a new harmony parser to perform a chord segment reduction process, dealing with the **exponential state-space**.

There are two input dimensions in this problem. First is the number of notes in each slice, referred to as the size of the slice, m. The worst case is dominated by the number of possible `unsplit` reductions at a single step, which corresponds to to the number of configurations of inner operations involved. The number of possible inner operations is dominated in the worst case by the number of double-sided operations, $\mathcal{O}(|parents|^2 \cdot |children|) = \mathcal{O}(m^3)$. As a `split` operation can consist of any configuration of these inner operations, the number of `split` operations is $\mathcal{O}(2^{m^3}) = \mathcal{O}(2^m)$ where m is the size of the slices involved and c is a constant. The second input is the number of slices in the original surface, referred to as the length of the surface, n. In the worst case, the total number of complete reductions is given by $\mathcal{O}(2^{m^n}) = \mathcal{O}(2^{m \cdot n})$. Thus the number of possible reductions is **exponential** with respect to **two inputs**, the size of each slice m and the length of the surface n.

I propose and implement a novel stochastic beam search algorithm to handle these dimensions of complexity in Section 3.6. In order to achieve this, a fitness function is developed that provides an approximate score for each partial reduction. This fitness function can be decomposed into each reduction step, allowing it to be used to score each reduction step, giving more musically meaningful reductions a greater fitness score. Finally, a new harmony parser is developed that adds constraints to ensure chord segments are preserved throughout the reduction.

2.5 Requirements Analysis

The Success Criteria are given in the Project Proposal. During the preparation phase, the Success Criteria were refined in light of increased clarity from reading the literature related to the project. Concretely, the final Success Criterion given in the Proposal describing the development of heuristic search algorithms has been divided into two criteria, developing the heuristic, and developing the search algorithms.

This project will be deemed a success given it achieves:

- A harmony module that can use a reduced surface to infer chord labels, and quantitatively evaluate its accuracy against the ground truth annotations.

- A heuristic fitness function that can be used to judge the relative plausibility of reductions.
- An implementation of a new parser for the protovoice model which finds possible reductions of a musical surface.
- Extension: One or more search algorithms that make use of the developed heuristic to inform the search, dealing with the two dimensions of exponential complexity.

Table 2.1: Project Deliverables

ID	Deliverable	Priority	Risk
`core1`	Harmony Model	High	Low
`core2`	End to End Pipeline	High	Medium
`core3`	Proto-voice Harmony Parser	High	High
`base1`	Templating Algorithm	High	Low
`algo1`	Random Parse	High	Low
`ext1`	Heuristic Design	Medium	High
`ext2`	Heuristic Search	Low	Very High

Risk Analysis

This project has a high general risk factor as it involves a **novel** approach to inferring harmony. Table 2.1 shows a list of project deliverables with associated priorities and risk, denoted qualitatively. The task with the greatest risk attached is designing and implementing the protovoice harmony parser, as this requires understanding and building on a complex existing codebase, and my proposed adaptation of the proto-voice parser has not been implemented before. Designing a bottom-up heuristic has a high risk factor as this is also a novel task, requiring creativity, research and iterative development. The baseline inference deliverable implements a standard chord templating method used commonly for ACE [41], posing minimal risk. Finally, the design and optimisation of efficient search algorithms is also a substantial task, which runs the risk of sinking a practically infinite amount of time.

In order to mitigate the risks posed by this project, the heuristic design and heuristic search tasks were set as extensions, so that only one high risk task was in the core part of this project.

2.6 Software Engineering Techniques

2.6.1 Development model

Based on the risk analysis (Table 2.1), a plan was created describing which modules to implement in which order, with a list of milestones on a 2 week basis. Notion was used to maintain a list of core tasks and corresponding subtasks with associated priorities, facilitating the selection of the next tasks to work on. The development strategy chosen drew from the Agile methodology, involving two-week long sprints with regular re-evaluations of the plan informed

by experimental data and testing. GitHub's continuous integration features were used to run a test suite on the repository after every commit.

2.6.2 Languages, libraries and tools

Table 2.2 shows a justified list of the key languages, libraries and tools used in the project. The licensing agreements for all the tools used in the project were determined and analysed. For the most part, these are all permissive licenses, guaranteeing freedom to use, modify and redistribute as well as permitting proprietary derivative works.

2.6.3 Hardware, version control and backup

The code was developed using Vim for Haskell and Visual Studio Code for Python notebook development, on my personal laptop (16' MacBook Pro 2022, M1 Max, 32GB). Algorithms were first run on my laptop, then later run on a server provided by the EPFL Digital Cognitive Musicology Lab (Dell PowerEdge R740XD Server, 2x Xeon Gold 625R, 768GB), using Jupyter notebooks to conduct the evaluation. I used GitHub for all my notes, development and dissertation writing. Finally, this dissertation was written in Vim with VimTeX.

Table 2.2: Languages, libraries and tools

Tool	Purpose	Justification	License
Languages			
Haskell	Main language used for the core, baseline and extension implementations	Protovoice model implementation is in Haskell. Functional and amenable to parser development.	GHCL
Python	Secondary language for experiments and analysis	Powerful library ecosystem for running experiments and creating plots	PSFL
Libraries			
Musicology	Haskell Library with data-types for pitches	Contains a robust implementation of spelled pitch classes, which would be tedious to reimplement.	BSD-3.0
Timeit	Lighweight wrapper to show the used CPU time of a monadic computation	This is used to measure the runtime of the algorithms as part of analysis	BSD-3.0
Dimcat	Python library: DIgital Musicology Corpus Analysis Toolkit	This library was written to work with the datasets used in this project	GPL-3.0
Numpy	Python library used for preprocessing and analysis	Powerful standard library that is used in conjunction with Seaborn to run analysis and visualise data	BSD-3.0
Pandas	Python library for preprocessing and analysis	This is a standard library for data manipulation and processing	BSD-3.0
Seaborn	Python data visualisation library used for analysis	Creates high quality graphs and charts	BSD-3.0
Tools			
Docker	Containerised software service used to run repeatable experiments	Protects code from breaking changes and allows code to be executed on different devices without manually installing dependencies	Free/Paid
Git	Version Control, Continuous Integration	Provides natural backups and allows for reverts to previous commits if necessary	GPL-3.0
GitHub	Hosting source code	Free, reliable hosting	GPL-3.0
GHC	Compiling and profiling.	This is the standard Haskell compiler.	BSD-3.0
Stack	Haskell building and testing	Creates reliable builds, and includes a powerful testing framework.	BSD-3.0
Undotree	Vim Plugin: stores all past actions as a tree	Solves the problem of linear undo history being lost. Protects code between commits.	BSD-3.0
MuseScore	Music notation software	The raw inputs are in the MusicXML format, which is used by MuseScore 3	GPL-3.0
PAT	Protovoice Annotation Tool, Used to view protovoice derivations on a web browser	The protovoice derivations are huge and very complex, so it's vital to have a viewing tool for use in analysis and iterative development	GPL-3.0

Chapter 3

Implementation

This chapter provides a high-level overview of the project structure (Section 3.1), followed by a detailed description of the key components of the implementation. First, the design and implementation of the proto-voice harmony parser is discussed (Section 3.3), followed by an exposition of the harmony module (Section 3.4) used to infer chord labels from reduced surfaces extracted by the proto-voice harmony parser. The implementation of the core algorithms is then discussed (Section 3.5), including an implementation of the template matching algorithm [41] as a baseline. Subsequently, the heuristic design and implementation is presented (Section 3.2.1), followed by a description of the heuristic search algorithms developed (Section 3.6). Finally, the test strategies used evaluate the effectiveness of the proposed algorithms are outlined (Section 3.7).

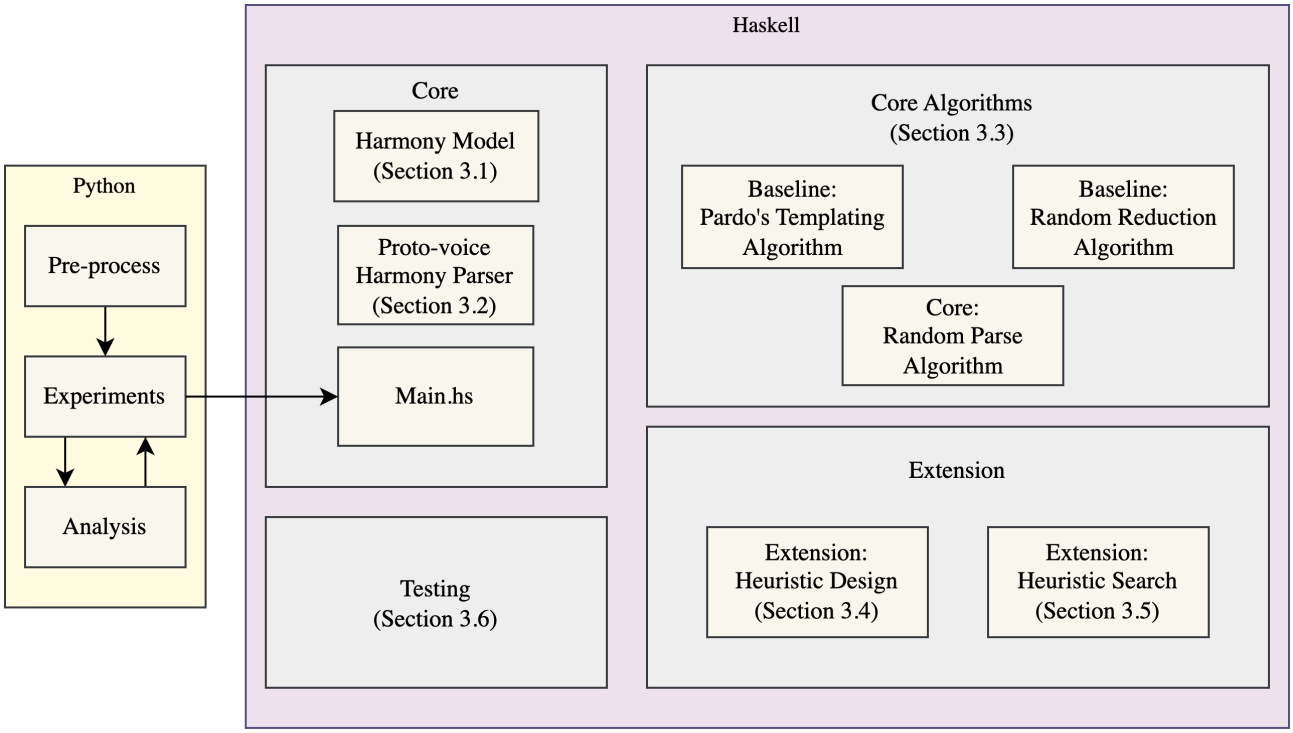


Figure 3.1: Diagram of project components

3.1 Repository Overview:

Table 3.1: Repository Overview

File/Folder	Description	LOC
protovoices-haskell/	Root directory	2272
├─src/		
│ ├─HarmonyParser.hs	Harmony Parser (Section 3.3)	470
│ ├─Harmony.hs	Harmony Model (Section 3.4)	121
│ ├─Harmony/		
│ │ ├─ChordLabel		
│ │ └─Profiles		
│ ├─Algorithm.hs	Core Algorithms (Section 3.5)	383
│ ├─Algorithm/		
│ │ ├─ TemplateMatching.hs, RandomSample.hs, InformedReduction.hs		188
│ │ ├─RandomParse.hs	Core Algorithm (Section 3.5)	
│ │ ├─HeuristicSearch/	Extension Algorithms (Section 3.6)	
│ │ │ └─BestFirst.hs, Beam.hs, StochasticBeam.hs		
│ ├─Heuristics.hs	Heuristic Implementation (Section 3.2)	431
│ ├─FileHandling.hs	Utilities	115
│ ├─Probability.hs		
│ ├─Common.hs	Existing code	611
│ ├─GreedyParser.hs		
│ ├─PVGrammar.hs, PVGrammar/		
│ └─Display.hs		
├─experiments/	Experiment Pipeline (Section 3.7)	
│ ├─preprocess.ipynb		
│ ├─experiments.ipynb		
│ ├─analysis.ipynb		
│ ├─inputs/		
│ └─outputs/		
├─app/		
│ └─MainFullParse.hs		
└─test/	Unit Tests (Section 3.7)	

Repository Justification

The repository has broadly been split into five main modules, as illustrated in Figure 3.1. Everything shown has been written during this project except for the files shown in blue.

- Firstly, the `HarmonyParser.hs` contains the first core contribution, an adaptation of the greedy parser (shown in blue) which integrates the protovoice model with the chord segment reduction process.
- Next, the `Harmony` module contains functions and data-types for dealing with chord labels, and profiles, as well as performing probabilistic inference.
- The `Algorithm` module provides a generic framework to run chord segment reduction algorithms. This module contains implementations of all the core algorithms, including Pardo and Birmingham's template maching algorithm [41]. Within this, the `HeuristicSearch/..` algorithms are those that make use of the heuristic fitness function.
- `Heuristics.hs` contains the novel fitness function developed.
- The `experiments/` folder contains all the python code that is used for this project. Experiments were conducted by running the Haskell executable parameterised by the input piece and algorithm to use, and results are stored using `Json`. The experiments consist of three stages, as described by the three main files: `preprocess.py`, `experiments.py` and `analysis.py`. Splitting these stages up prevents wasteful computation, as all the pre-processing can be done just once, while experiments are run on the processed data iteratively alongside algorithm development.
- Finally, the `test/` folder contains unit tests for all of the Haskell modules, using the `Spec` testing framework which is used in Continuous Integration.

3.2 Proto-voice Fitness Function

To address the **exponential time** complexity of the naive parser and enable the efficient identification of musically meaningful proto-voice derivations, I propose a **novel fitness function** that evaluates a plausibility score for each derivation. The fitness function decomposes into a score for each reduction step and is used to guide the efficient heuristic search algorithm designed in Section 4.3, allowing **excellent solutions** to be found in **linear time**.

3.2.1 Overview of Approach

The proto-voice fitness function is motivated by defining a probability distribution for proto-voice derivations, $P(\vec{d})$, factored by each derivation step. Given a derivation $\vec{d} = d_0 \dots d_N$ define:

$$P(\vec{d}) = \frac{1}{Z} \prod_i \phi(d_i) \tag{3.1}$$

Where Z is the normalisation constant, and ϕ is a heuristic function that defines a relative probability for each reduction step, given by:

$$\phi(d_i) = \phi_{\texttt{parents}}(d_i) \cdot \phi_{\texttt{children}}(d_i)^{\alpha} \tag{3.2}$$

The heuristic function consists of two factors computed using probabilistic chord profiles, corresponding to the generative proto-voice operations. The `parents` factor evaluates the relative plausibility of the parent notes chosen to be elaborated in the operation, and the `children` factor evaluates the relative plausibility of the child notes introduced from those parents.

Conditioning this distribution on the input sequence S, it follows from Bayes' Rule:

$$\begin{gathered} P(\vec{d}|S) = \frac{P(S|\vec{d}) \cdot P(\vec{d})}{P(S)} = P(S|\vec{d}) \frac{1}{P(S)} \frac{1}{Z} \prod_i \phi(d_i) \\ \text{where } P(S|\vec{d}) = \begin{cases} 1 & \text{if } \vec{d} \text{ produces S} \\ 0 & \text{otherwise} \end{cases} \end{gathered} \tag{3.3}$$

It is guaranteed (see Section 3.7) that any derivation $\vec{d}$ found by the proto-voice harmony parser produces S, so $P(S|\vec{d})$ is always 1. Computing the normalisation constant, $\frac{1}{P(S) \cdot Z}$, is intractable, but for the purpose of maximisation it can be ignored. It therefore suffices to compute $P(\vec{d}) = \frac{1}{Z} \prod_i \phi(d_i)$.

Justification

It is informative to consider how a true generative probability distribution for proto-voice derivations would be structured. The probability of a derivation $\vec{d}$ in the generative direction

is naturally factored into each derivation step as follows:

$$p(\vec{d}) = \prod_{i}^{N} p(d_i | d_0, \ldots, d_{i-1}) \tag{3.4}$$

Where d_0 is the first generation step. A generative model would factorise this expression into conditionals that correspond to the generation steps, thus providing a guess of the distribution $p(\vec{d})$. Instead, the fitness function approximates this overall distribution based on local plausibility properties, corresponding to ϕ. This approximation allows the fitness function to remain computationally tractable while capturing essential aspects of the true generative distribution.

3.2.2 Scoring Split Operations

Split operations elaborate parent edges by introducing new child notes that are within an interval of a *step* from one or two parent notes. The plausibility scores for both parent and child notes are computed by assessing how well they match the appropriate chord-tone or ornament profiles of the slices they reside in.

Recall the split rule :

$$t \rightarrow t'_l \; s' \; t'_r$$

The split operation takes the non-terminal $t = (s_l, e, s_r)$ and substitutes it with $t'_l \; s' \; t'_r$ where $t_l = (s_l, e^l, s')$ and $t_r = (s', e^r, s_r)$. Each $e = (e_{reg}, e_{pass})$ contains a set of *regular* edges e_{reg} and a multi-set of *passing* edges, e_{pass}.

The generative process of an single-sided elaboration contained in a `split` operation is modelled as follows. The parent slice has a latent chord label $l \in \mathcal{L}$ which is not known. The parent notes $\vec{p}$ are drawn based on this chord label, then the child notes $\vec{n}$ are drawn based on the parent notes $\vec{p}$ as well as the latent chord label l. The joint probability distribution $P(\vec{p}, \vec{n})$ is factored as:

$$\begin{aligned} P(\vec{p}, \vec{n}) &= \sum_{L} P(\vec{n} | \vec{p}, L) \cdot P(\vec{p} | L) \cdot P(L) \\ &\approx P(\vec{n} | \vec{p}, \hat{L}) \cdot P(\vec{p} | \hat{L}) \end{aligned} \tag{3.5}$$

Where the distribution is approximated by replacing the marginalisation sum with the maximum-likelihood estimate of L, given by $\hat{L}$.

The term $P(\vec{p}|\hat{L})$ expresses the likelihood of the parent notes exemplifying chord-tones of the guessed chord label $\hat{L}$. This is computed by sampling the multinomial distribution at the value given by the parent slice vector, parametrised by the chord-tone profile vector of the guessed chord label.

Secondly, the term $P(\vec{n}|\vec{p}, \hat{L})$ expresses the plausibility of the child notes being generated from the parent notes $\vec{p}$. In this case, an approximation is made for each child note, where instead of evaluating the proability of the note being generated from the specific parent, the distinction is made whether child is an ornament or a chord-tone. The term $P(\vec{n}|\vec{p}, \hat{L})$ is computed by

sampling the multinomial distribution at the value given by the note slice vector, parametrised by the chord-tone or ornament profile vector of the guessed chord label.

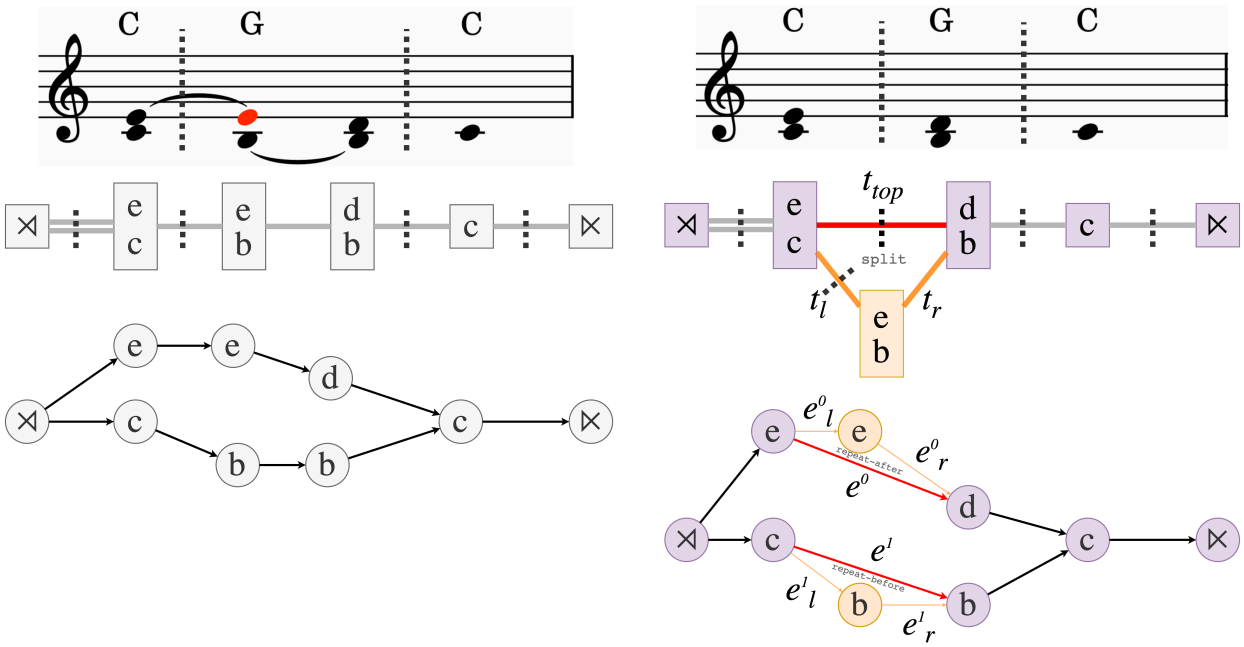
The plausibility of a split operation is given by evaluating each of the inner operations individually and taking geometric mean of the individual plausibilities. This way, the plausibility is independent of the number of inner operations.

Assumptions

The assumptions made in order to calculate ϕ are as follows. First, the operation plausibilities are conditioned on the assumption that the transition $t = (s_l, e, s_r)$ relates two slices that exemplify their guessed chord labels, $\hat{L}_{s_l}$ and $\hat{L}_{s_r}$. Second, the parent note is assumed to be a chord-tone of the parent slice. Third, when a child note is a repetition of the parent, the child note is assumed to be a chord-tone. Finally, when a child note n is a *step* away from the parent note p, such as a neighbour or passing note, it is assumed to be an *ornament* with respect to the parent chord label.

3.2.3 Scoring Spread Operations

Recall that the proto-voice model comprises an inner and outer structure. The outer structure provides an abstraction of the inner structure, allowing the parsing algorithm to be designed using operations applied to slices and transitions. In order to evaluate the plausibility of operations, the inner configuration of notes and edges are considered.



(a) Before (b) After

Figure 3.2: Single unsplit reduction

The core of the proto-voice model is a set of generative operations on notes. Each operation is first considered separately.

Scoring Split Operations

The last three terms are *independent* of the parent and child notes, so their product is stored within the slice data structure to avoid redundant computation.

For single-sided operations, such as `left-neighbor`: $p \implies n \to p$, calculating $P(n, p|\hat{L}_s)$, requires and calculating the likelihood of n being generated as an *ornament*(or chord tone in the case of a repetition), using the categorical probability density function parameterised by the per note ornament probabilities of the estimated chord of $\hat{L}_s$, evaluated at index corresponding to n. For double-sided operations, the plausibility is calculated using a mixture model based the ornament or chord tone probability vectors for each parents, with equal weighting on either side.

In order to avoid a bias towards single sided operations, the plausibility is calculated using a geometric mean of each of the involved notes' plausibilities, $P(n, p_l, p_r|\hat{L}_{s_l}, \hat{L}_{s_r}) = (\prod_{x \in \{n, p_l, p_r\}})^{\frac{1}{3}}$

Scoring Spread Operations

Consider the Spread rule :

$$t_l s_r \to t'_l s_l t'_m s_r t'_r$$

The spread operation has the music theoretical function of *prolonging* the parent slice, that is, the child slice is just a repetition of the parent slice, with the same notes associated chord label, but with some notes dropped, and some ornamented.

The plausibility of a spread operation is conditioned on both parents consisting of only chord-tones, and the constraint that they have the same chord label. The inner operations of a spread are all double-sided, so the same approach is taken as for the double-sided case of the split operation.

Scoring Freeze Operations

The freeze operation $t \to t'$ marks a transition as *terminal*. Within the context of the parse, the unfreeze shifts the context to the left. No judgement is made about freeze operations, so they are assigned a plausibility of 1, and it is left to the search algorithm to handle this case. During the parse, it is not permitted to unfreeze a transition across a segment boundary, so assigning a plausibility of 1 will still result in a full parse, even using a greedy algorithm.

Combining Plausibilities

The plausibilities of each of the operations cannot be compared across different operation types as they are of different distributions, so the decision between which operation to do use needs to be considered. In practice, the distributions lead to plausible results as is, but further work would build on this, establishing a prior over the different generative operations.

3.3 The Proto-Voice Harmony Parser

This section first describes the existing `GreedyParser` before explaining how it has been adapted to create the `HarmonyParser`, finding a chord segment reduction rather than a full derivation.

3.3.1 The Greedy Parser

[TODO: make this very succinct + clearer.]

The Greedy Parser is a bottom-up parser for the protovoice model that enumerates all the valid proto-voice derivations that produce a given score $S \to [D]$. Recall that within the context of the proto-voice model, the original score is called the *surface*, and the surface is *reduced* by parsing, removing non-chord-tones. A derivation is described as a sequence of rule applications in *left-most derivation order*, applied to a fully reduced surface, $\mathcal{R}(S)$.

Formally, a derivation D is defined as a pair (top, ops), where the $surface$, S, is derived by starting with the fully reduced top and applying each operation in ops in order.

It is informative to consider the *generation order* and *parse order* of a proto-voice derivation. In generation order, we begin with the reduced surface top, consisting of only chord-tones, with a pointer at the left-most transition and apply each `split` or `spread` operation to the two left-most non-terminal transitions in the path graph, generating new slices and transitions. The `freeze` operation marks a transition as terminal thus shifts the pointer to the right. Operations are applied until the pointer is right-most, and all transitions are terminal, resulting in the fully elaborated $surface$ and a derivation $D = (top, ops)$, where ops is the sequence of operations applied, $ops = [d_N \dots d_0]$, resulting in the $surface = d_0 \circ \dots \circ d_N(top)$.

Parsing is the inverse of generation: the parse begins with the elaborated surface S consisting of only frozen transitions with a pointer at the right-most frozen transition. During the parse, we can either `unfreeze` the transition at the pointer, shifting the pointer to the left, or apply a `unsplit` or `unspread` reduction to the two transitions to the right of the pointer. Once the pointer is left-most, all transitions are unfrozen and there is a *single reduced slice* for each chord segment at the top of the derivation, the resulting derivation is $D = (top, ops)$, where the reduced surface is $\mathcal{R}(S) = top = (d_0^{-1} \circ \dots \circ d_N^{-1})(surface)$.

This is similar to a *shift-reduce* parser, making one pass across the surface, right to left. The state of the parse is represented by `parseState`$= (top, ops)$ and implements `expand: parseState → [parseState]` such that we can use search algorithms to conduct the parse.

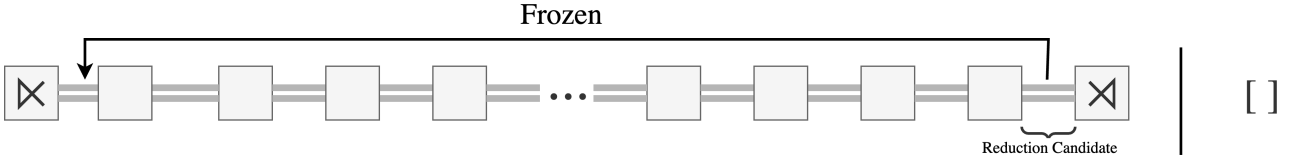


Figure 3.3: `pSFrozen`

In the initial parse state, `pSFrozen`, the surface is represented as a `path` from the end to the beginning of the piece, with the right-most transition at its head. The start($\rtimes$) and stop($\ltimes$) symbols are not explicitly stored in the `path`, but are implied.

All transitions are initialised as frozen, indicated by the double lines in Figure 3.3. The `path` begins on the right. In this initial state the only option is to unfreeze the rightmost transition, moving to the `pSSemiOpen` state.

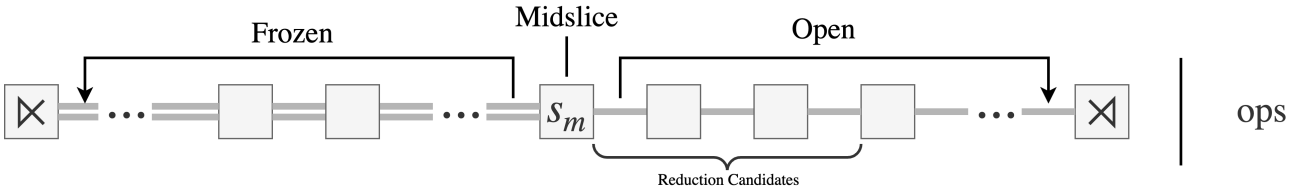


Figure 3.4: `pSSemiOpen`

The `pSSemiOpen` represents the majority of the parse. The pointer is represented by `midSlice`, and points to the rightmost frozen transition. The `open` path contains all unfrozen transitions (denoted by a single line) from the right of the pointer to the end of the piece, and the `frozen` path contains all frozen transitions from the pointer to the start of the piece.

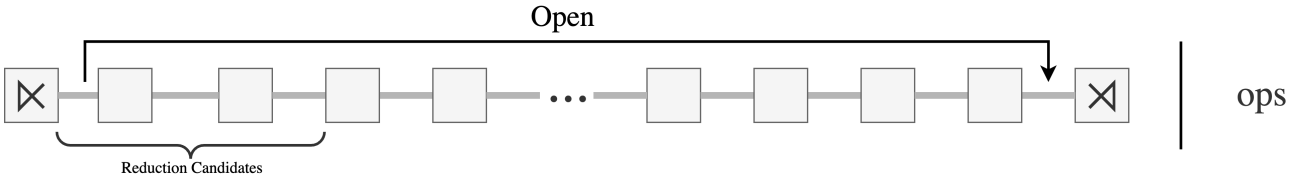


Figure 3.5: `pSOpen`

In the final state, `pSOpen`, the pointer is at the beginning of the piece, and comprises a single path `open` containing only unfrozen transitions from the beginning to the end of the piece.

Each state also stores a list of operations, *ops* that when applied to the current reduction, results in the original surface. This is empty at the beginning of the parse; for each reduction applied, the corresponding generative operation *op* is consed to the list: $ops' := op : ops$.

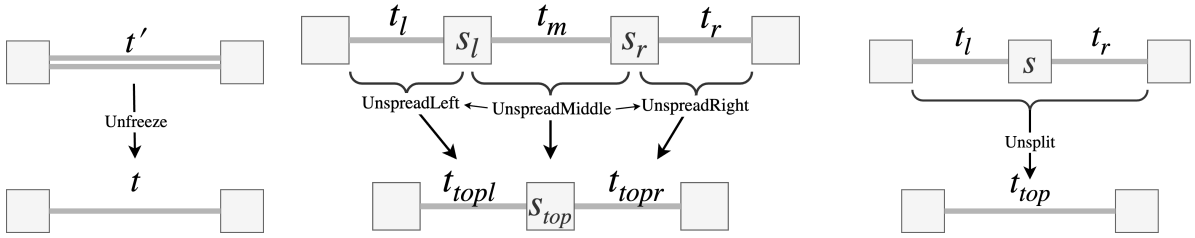
This is not the only way to parse according to the proto-voice model; the proto-voice reduction rules can be applied to any pair of open transitions. Allowing reductions at any point along the surface could allow for different parse strategies to be employed that may have advantages. I choose not to pursue this due to the associated combinatorial explosion, and the benefits of being able to represent a proto-voice derivation as a sequence of left-most reductions.

Parsing Operations

The `protoVoiceEvaluator` [15] provides methods that are used to enumerate the possible operations for each reduction type [1]. It also includes an implementation of a greedy parser, from which ideas were adapted and expanded upon to create the proto-voice parser. This allows the parser to consider only the outer structure of slices and transitions while parsing without exposing the internal notes and edges.

[1]Note that the notation used here for functions combines type definitions with variable names. For example, the function `evalUnsplit` has type `evalUnsplit` :: $(transition, slice, transition) \rightarrow [(transition, operation)]$, but type names (e.g. *transition*) have been substituted by variable names (e.g. t_l, t_r, t_{top}) for expository purposes.

$$
\begin{aligned}
&\texttt{evalUnfreeze}: && t \to [(t', op)] \\
&\texttt{evalUnsplit}: && (t_l, s_l, t_m) \to [(t_{top}, op)] \\
&\texttt{unSpreadLeft}: && (s_l, t_l) \to s_{top} \to [t_{topl}] \\
&\texttt{unSpreadRight}: && (s_r, t_r) \to s_{top} \to [t_{topr}] \\
&\texttt{unSpreadMiddle}: && (s_l, t_{top}, s_r) \to \texttt{Maybe}\ (s_{top}, op)
\end{aligned}
\tag{3.6}
$$



(a) `Unfreeze` (b) `Unspread` (c) `Unsplit`

Figure 3.6: Reduction operations.

The `evalUnfreeze` and `evalUnsplit` functions allow the associated generative operation, `freeze` and `split` to be determined trivially, but the possible `spread` operations need to derived using `unSpreadLeft`, `unSpreadRight`, and `unSpreadMiddle`. This is achieved using the *list monad*, wherein multiple branches of sequential computation appear to be executed simultaneously, returning the results of all possible computation branches in a list.

Algorithm 1 Enumerate unspread reductions

```
1: function CollectUnspreads(s_l t_l s_m t_r s_r)
2:     using List Monad do
3:     (s_top, op) ← maybeToList $ evalUnspreadMiddle (s_l, t_m, s_r)
4:     t_topl ←evalUnspreadLeft (s_l, t_l) s_top
5:     t_topr ← evalUnspreadRight (s_r, t_r) s_top
6:     return (l_top, s_top, r_top)
7: end function
```

3.3.2 Conserving Segment Boundaries

Recall that the goal of the parse is to reduce the original surface such that there is one slice per segment, containing only chord-tones. In order to achieve this, constraints are imposed on the reduction operations dependent on adjacent segment boundaries.

Each transition has a boolean *boundary* value which indicates if the transition is across a segment boundary, denoted by the dashed vertical line in Figure 3.8. Let $B_t : t \to \{\textbf{True}, \textbf{False}\}$, such that B_f, B_l, B_m, and B_r denote the boundary values of t_f, t_l, t_m and t_r respectively, as shown in Figure 3.9. As an `unspread` operation ($t_l\ s_l\ t_m\ s_r\ t_r \to t_{topl} s_{top} t_{topr}$) merges two slices s_l and s_r, removing t_m, the constraint $\neg B_m$ is imposed to prevent two segments from merging. Similarly, an `unsplit` operation (eg. $t_l\ s_l\ t_m \to t_{top}$) combines two transitions, thus we impose the constraint $\neg(B_l \wedge B_m)$. Finally, the `unfreeze` operation shifts the context to the left. If a boundary transition is unfrozen, no more reduction operations can be applied to its right, hence

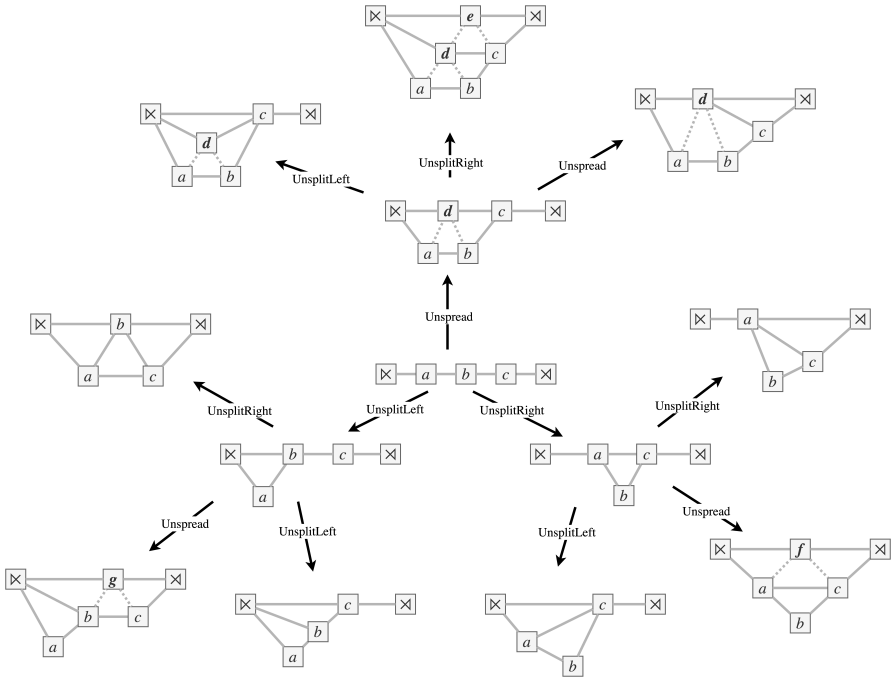


Figure 3.7: Parse state transitions for an open segment with three slices.

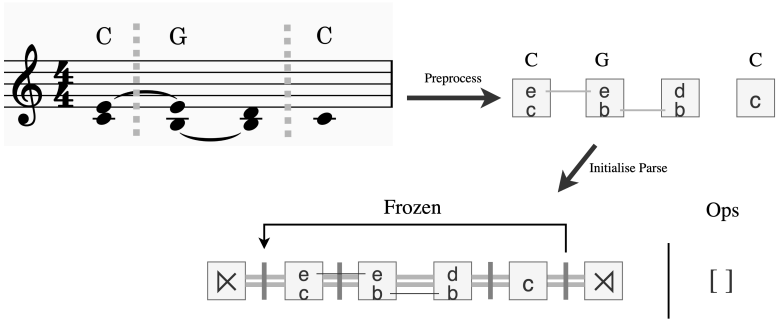


Figure 3.8: Parse Initialisation

it a necessary condition that the segment has been fully reduced. Thus the constraint imposed on an `unfreeze` operation is $\neg B_f \vee (B_l \wedge B_m \wedge B_r)$. Note that there are some configurations that are *unreachable*, such as $B_f \wedge (B_l \wedge B_m \wedge \neg B_r)$, thus the `unfreeze` constraint could be simplified to $\neg B_f \vee (B_l \wedge B_m)$, but I have chosen not to as it would obfuscate the semantics, and the cost of an additional logic check is negligible.

Figure 3.9 provides an illustration of these constraints. The boundary must also be propagated to parent transitions following a reduction, defined by the union of each child transitions' boundary values for a given reduction.

Figure 3.7 shows the 9 possible outer configurations of slices after two reduction steps, shown for a single segment reduction consisting of three slices and four transitions. One of these configurations, `UnsplitLeft` ∘ `UnsplitRight`, is redundant as it results in the same configuration as `UnsplitLeft` ∘ `UnsplitLeft`, so it is prohibited.

The parse states define a **directed acyclic graph** where the parseStates are nodes and the

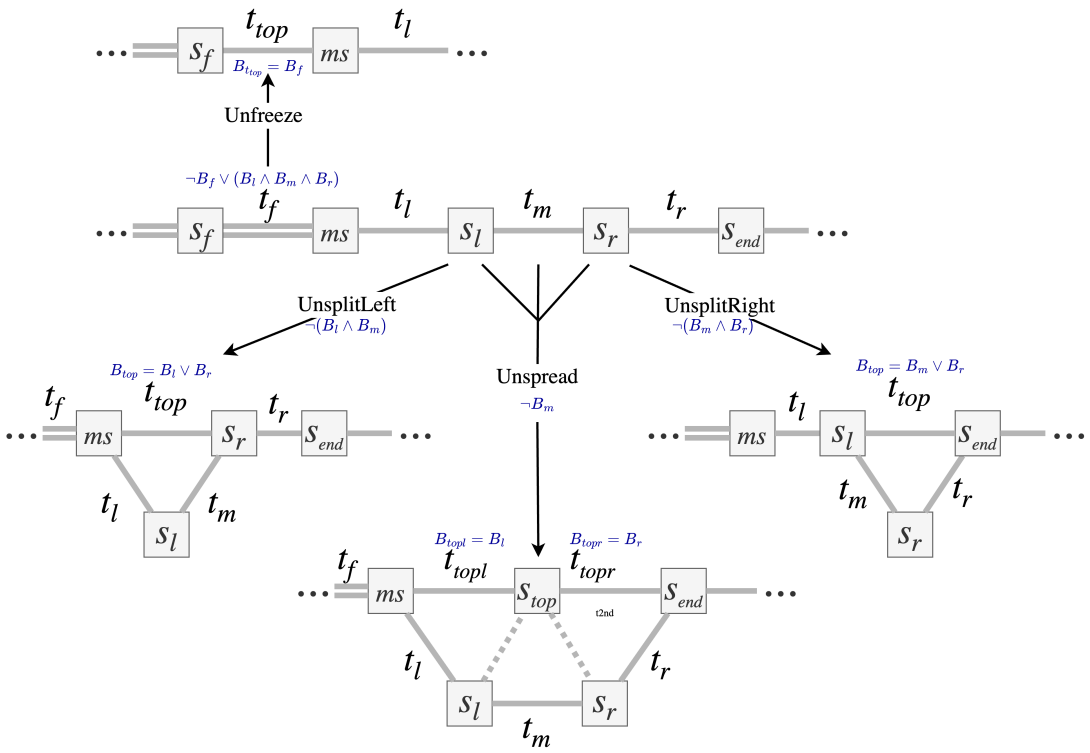


Figure 3.9: Boundary reduction conditions and propagation

parse operations are edges. Traversing this graph results in either reach a valid derivation or dead state.

3.4 Harmony Module

[Short and sweet]

3.4.1 Chord Labels

[TODO: write]

3.4.2 Chord Profiles

[TODO: write]

Chord-tone Profiles

[TODO: rewrite]

An alternative approach would be to marginalise over all chords find the ”expected” chordness.

Ornamentation Profiles

[TODO: rewrite]

The original model $P(\mathcal{R}(S), L)$ is extended to account for both chord-tones and non-chord-tones, $P(S, L)$, by introducing a *mixture model*.

Notes that are not chord-tones are referred to as *ornaments* as music theory states that they often have a *function* with respect to an adjacent chord-tone, ornamenting that chord-tone. This phenomenon is captured in the proto-voice model through *neighbour* notes. Here, a neighbour note and ornamentation are considered synonymous.

It is assumed that each note in a chord l is generated as either a chord-tone or ornament according to a Bernoulli distribution $t \sim \text{Bernoulli}(\theta_l)$. Then each note is generated from a categorical distribution $n \sim \text{Categorical}(n, \mathbf{\Phi}_{\boldsymbol{l}}^{(\boldsymbol{p})})$ where Φ is either for the chord tone distribution or ornament distribution depending on t [2]

Mixture Profiles

[TODO: motivate and explain implmentation] [for double-sided rules: mixture of experts. disjunction vs conjunction.]

3.4.3 Chord Label Inference

[TODO: rewrite]

Recall that input is a sequence of slices, groups of notes, that are only chord-tones. Each note has an associated pitch $p \in \mathcal{P}$. A chord label is defined as a tuple of *root note* pitch and chord-quality $\mathcal{L} = (\mathcal{P}, \mathcal{C})$, where the chord qualities are taken from the DCML chord label annotation standard as used in the datasets: $\mathcal{C} = \{\text{M, m, Mm7, om, o7, mm7, \%7, MM7, +, Ger, It, Fr, mM7, +7}\}$.

Given a slice $s = \left\{p_1^{m(p_1)}, \ldots, p_n^{m(p_n)}\right\}$ where $p_i \in \mathcal{P}$ and $n = |\mathcal{P}|$, the goal is to find the most plausible chord label:

$$\hat{l} = \arg\max_l P(l|s) \tag{3.7}$$

We choose to find $P(l|x)$ rather than $P(s|l)$ as it provides a *confidence* in the prediction, given by $P(\hat{l}|s)$. The generation of a slice s of size n from a label l is modelled as follows:

$$\begin{aligned} s \sim &\text{Multinomial}(n, |\mathcal{P}|, \mathbf{\Phi}_l) \\ \text{where } \boldsymbol{\phi}_l^{(p)} &= P(\text{Pitch} = p | \text{Label} = l) \end{aligned} \tag{3.8}$$

[2] todo formalise

The *slice vector* is defined $\boldsymbol{v}(s) = [\ m(p_1)\ m(p_2)\ \ldots\ m(p_k)\ ]$ where $k = |\mathcal{P}|$. Then $P(l|s)$ is given by: [3]

$$\begin{aligned} P(l|s) &= P(s|l) \cdot P(l) \\ &= f(\boldsymbol{v}(s), \boldsymbol{\phi}_l) \cdot P(l) \\ &\text{where } f \text{ is the multinomial probability density function} \\ f(x_1, \ldots, x_k; p_1, \ldots, p_k) &= \frac{\Gamma\left(\sum_i x_i + 1\right)}{\prod_i \Gamma\left(x_i + 1\right)} \prod_{i=1}^{k} p_i^{x_i} \end{aligned} \tag{3.9}$$

The log probability $\log P(l|s)$ is given by

$$\begin{aligned} \log P(l|s) &= \log P(s|l) + \log P(l) \\ &= \cdots \\ &= \log\left(\Gamma\left(\sum_i x_i + 1\right)\right) + \sum \left(x_i \log p_i^{x_i} - \log\Gamma(x_i + 1)\right) + \log P(l) \end{aligned} \tag{3.10}$$

So to find the most likely label l given slice s take:

$$\hat{l} = \arg\max_l \left(\log\left(\Gamma\left(\sum_i x_i + 1\right)\right) + \sum \left(x_i \log p_i - \log\Gamma(x_i + 1)\right) + \log P(l)\right) \tag{3.11}$$

The full derivation has been omitted for the sake of brevity. Given that $|\mathcal{L}| = |\mathcal{P}| \times |\mathcal{C}| = 29 \times 12 = 348$, the equation can be solved directly.

3.4.4 Optimisations

[TODO:]

SliceWrapped data-type storing the MAP estimate for each slice within the datatype. Parametrised SliceWrapper, either storing MAP estimate or mariginalising over the top chord types.

3.5 Core Algorithms

The goal of proto-voice parser is to find a derivation $D = (top, ops)$ which results in a top sequence of slices *top* consisting of only chord-tones.

The `ParseAlgo` type-class provides an interface for running these algorithms. The result is wrapped in a `Maybe` as it is possible for the search to get stuck, and is in the `IO` monad to allow non-determinism.

[3] Struggling with naming here. Using P for pitches conflicts the P for probability.

```
class (Show algo) => ParseAlgo algo where
  runParse :: algo -> AlgoInput -> IO (Maybe AlgoResult)
```

Listing 3.1: Algorithm type-class

3.5.1 Baseline: Template Matching

[TODO: explain implementation with pseudocode]

3.5.2 Baseline Reductions

[TODO: short motivation]

Informed Reduction

[TODO: explain: using the actual ground truth to do the chord segment reduction. This gives a theoretical bound on the expected performance of the chord segment reduction process.]

Random Reduction

As a second baseline, consider the simplest form of reduction, removing notes by randomly guessing if they are chord tones or non-chord-tones. This is the **RandomReduction** algorithm. It is equivalent to combine all the slices in a given segment and take a random sample without replacement of n notes from the combined slice. The number of notes is sampled from a Poisson distribution, $n \sim \text{Poisson}(\lambda)$, where λ is learned from data.

3.5.3 Core: Random Parse

Next, the core algorithm is the **RandomWalk** algorithm using the protovoice parser. This algorithm takes a random walk through the parse states graph. The proto-voice model does not make judgements based on harmony or note function explicitly, but by comparing this algorithm to the other baselines, it can be determined whether the constraints enforced by using a proto-voice derivation has any effect on the quality of the reduction.

3.6 Extension: Heuristic Search

[TODO: entire section, justify and explain implemntation, use fitness funciton but without]

3.6.1 Best-first Search

Given the heuristic $\mathcal{H}_0$, the first option is to use a greedy search. We keep unfreeze operations as an option.

We find that the number of operations causes a combinatorial blowup, so the search doesn't end.

Problem of very large slices.
Segment by segment heuristic parse - avoids the problem, but is slightly hacky. Can we incorprate our knowledge regarding the relative proportion of chord tones and ornaments. Should we allow duplicates of notes in slices? Perhaps we should favour spreads more.

Algorithm 2 Greedy Search

initialise: $state \leftarrow initialState$
while *state* is not a goal state **do**
  $state \leftarrow \underset{s \in \textsc{Expand}(state)}{\arg\min} \mathrm{H}(s)$
  $freezes, spreads, splits \leftarrow$ Split $nextStates$ by *operation* type
  $open \leftarrow$ Take best β best states from $freezes \cup spreads \cup splits$
end while
return Best state in open

Always consider a certain number of slices and spreads.

Complexity Analysis

3.6.2 Beam Search

Step 2: Relax the heuristic search in order to reduce runtime/ lower complexity.

In the case that there are 85,000,000 options, perhaps we should sample the options rather than evaluating all of them.

This version of heuristic search should be able to parse full pieces (hopefully), so can be used to compare with the baselines on an entire corpus.

Beam of size n, with 1 for a freeze, k for spread, n-k-1

Dual Beam Search

It isn't clear how to determine how we should balance the costs of unspread and unsplit op

3.6.3 Stochastic Dual Beam Search

[TODO: Justify and correct pseudocode/ cut the bullshit. Genentic algorithm]

Algorithm 3 MultiBeam Search

hyper-parameters: $\beta \leftarrow BeamWidth, \gamma \leftarrow ReservoirSize$
initialise: $open \leftarrow (initialState, 0)$
while *open* does not contain any goal states **do**
$nextStates \leftarrow \bigcup_{(s,c) \in open} \{(s', c' + \text{H}(s')) : (s', c') \in \text{EXPAND}((s, c))\}$
$freezes, spreads, splits \leftarrow$ Split $nextStates$ by *operation* type
$open \leftarrow$ Take best β best states from $freezes \cup spreads \cup splits$
end while
return Best state in open

We propose sampling from the options, increasing the proportion that we ignore dependent on the number of options.

Algorithm 4 Reservoir MultiBeam Search

hyper-parameters: $\beta \leftarrow BeamWidth, \gamma \leftarrow ReservoirSize$
initialise: $open \leftarrow (initialState, 0)$
while *open* does not contain any goal states **do**
$nextStates \leftarrow \bigcup_{(s,c) \in open} \{(s', c' + \text{H}(s')) : (s', c') \in \text{EXPAND}((s, c))\}\}$
$freezes, spreads, splits \leftarrow$ Split $nextStates$ by *operation* type
$unFreezes \leftarrow \bigcup_{s \in freezes} (\text{RESERVOIRSAMPLE}(\gamma, s))$
$unSpreads \leftarrow \bigcup_{s \in spreads} (\text{RESERVOIRSAMPLE}(\gamma, s))$
$unSplits \leftarrow \bigcup_{s \in splits} (\text{RESERVOIRSAMPLE}(\gamma, s))$
$open \leftarrow$ Take best β best states from $unFreezes \cup unSpreads \cup unSplits$
end while
return Best state in open

Here H assigns a cost for moving from state s to s'. This is defined as the negated log likelihood in $\mathcal{H}$.

Complexity Analysis

[Show how complexity has been reduced wrt. the two input dimensions, compared to best-first]

3.7 Testing

[TODO:]

3.7.1 End to End Pipeline

The first step was to implement a full end-to-end pipeline from piece to chord label predictions. This was achieved by writing `preprocess.py`, `runExperiment.py` and `analysis.ipynb`. The

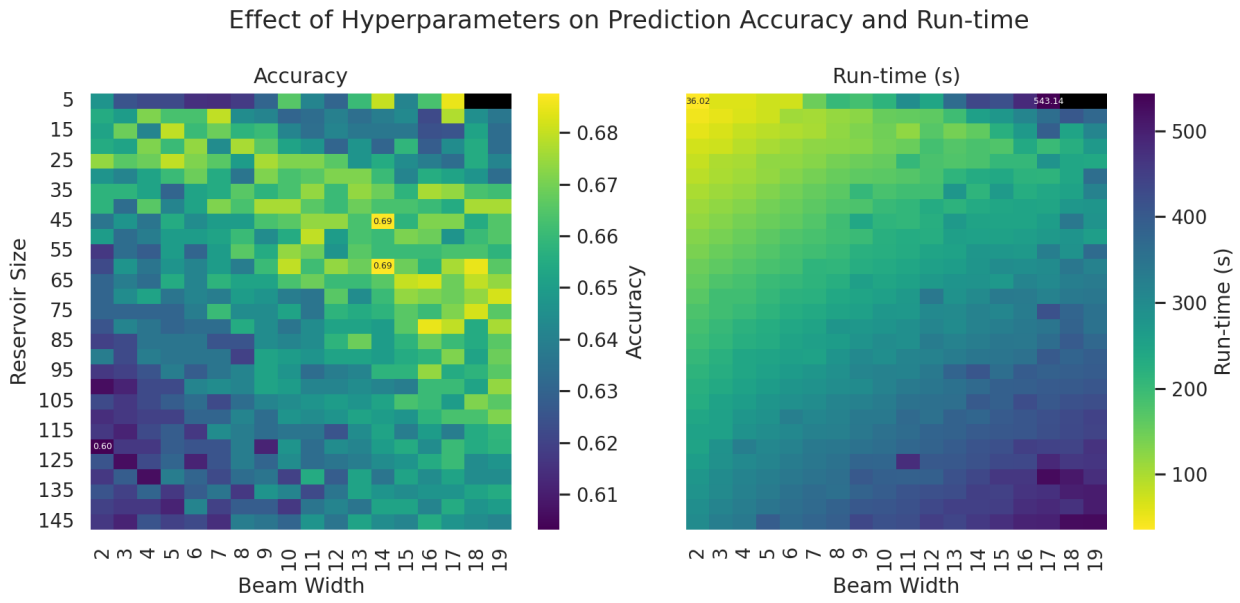


Figure 3.10

repository was then containerised using *Docker*, so that experiments could be run and analysed using a remote server with automated dependency resolution. Experiments were executed on a server, initialised via *ssh*, and *tmux* was used to maintain the server environment across connections, with automated scripts to pull the latest changes from GitHub. A *jupyter-lab* environment was set up on the server, and scripts written to fetch the results of the latest experiments and produce plots and tables, allowing analysis to be conducted through a browser.

Dataset selection

The datasets used contain bodies of work by a single composer as shown below, these were selected to exhibit a range of different styles. Each dataset contains an entire set work by a composer consisting of between $15 - 50$ annotated scores, each with varying lengths. Each dataset was split into training, validation, and test sets, using a 60:20:20 split with *stratified sampling* in order to maintain a balanced representation of the different composers. In order to choose hyperparameters, the training and validation sets from each of the five datasets were combined into a single training pool, for use in cross-validation. For the final evaluation, the performance of all developed algorithms will be evaluated on the hold-out test set (20%). This is to provide an unbiased estimate of each algorithm's performance on **unseen** data.

3.7.2 Unit Tests

3.7.3 Qualitative Tests

Throughout the design of the heuristic H_0 and the implementation of different search algorithms, a few segments we used as reccuring test examples - interpretability - how well the output lines up with a theoretical analysis.

Chapter 4

Evaluation

This chapter justifies all claims made in the introduction of this project by describing quantitative and qualitative evaluations of the work, demonstrating that the core and extension Success Criteria were met. This includes both the evaluation of the music theoretical assumptions behind the fitness function and harmony module, as well as the performance, scalability and interpretability of the developed algorithms.

Table 4.1: Evaluation Overview

Section	Deliverable	Axes of Evaluation
Section 4.1	Harmony Module	Accuracy
Section 4.2	Random Parse, Harmony Parser	Accuracy, Scalability
Section 4.3	Heuristic Search, Fitness function	Interpretability, Accuracy, Scalability

In order to show that the Success Criteria have been met, this chapter asks and answers the following questions:

- Does the Harmony Module effectively infer chord labels from multi-sets of corresponding chord-tones? (Section 4.1)
- Can the proto-voice model be used to perform a chord segment reduction? (Section 4.2)
- Does the fitness function developed provide a meaningful proxy for the true accuracy of a proto-voice reduction? (Section 4.3)
- Does the developed heuristic search algorithm achieve an improvement in accuracy compared to the random parse? (Section 4.3)
- Does the developed heuristic search algorithm achieve linear time complexity with respect to the input dimensions? (Section 4.3)

4.1 Harmony Module

Question Does the Harmony Module effectively infer chord labels from multi-sets of corresponding chord-tones?

Experiment setup

In order to test how well the harmony module can be used predict the ground truth labels, an *informed* reduction is performed, removing all non-chord-tones according to the chord-tone profiles of the ground truth labels. This is achieved by filtering all notes in each segment that are not in the chord tone profile of corresponding to the ground truth label. This method provides an upper bound on the expected performance of all subsequent algorithms for each piece.

The two metrics used are *accuracy* and *likelihood*. Accuracy is the proportion of correct label predictions, $\frac{TP}{TP+FP}$. The likelihood is the log probability of the reduced slice given the ground truth labels, based a multinomial sample from the corres is the intermediate value optimised by the harmony model when predicting the chord labels $\frac{\log P(L|\mathcal{R}(S))}{|L|}$, where the prediction is given by $\arg\max_L P(L|\mathcal{R}(S))$. The log-likelihood metric is divided by the number of segments to get the average number log-likelihood for each segment in a piece.

Discussion of Results

Table 4.2: Informed Reductions

Corpus	Accuracy	Log Likelihood
ABC	0.68 ± 0.11	-32.9 ± 11.4
Chopin Mazurkas	0.71 ± 0.16	-33.2 ± 15.9
Grieg Lyric Pieces	0.59 ± 0.15	-35.7 ± 14.4
Schumann kinderszenen	0.77 ± 0.13	-24.0 ± 6.62

Table 4.2: This table shows the results of the *Informed Reduction* algorithm described above. The accuracies are high, but reach a limit of around 0.7, even with the chord-tones known. This is likely due to the ambiguity of each segment, and that predictions are made locally, ignoring context. Furthermore, Gotham The variation between different corpuses indicate the different levels of ambiguity in each corpus, as well as other factors such as differing harmonic conceptions and note choices between composers. These will correspond to the DCML standard to differing extents.

The Pearson correlation coefficient was calculated between accuracy and likelihood, resulting in a coefficient of 0.823. This relationship is shown clearly in Figure 4.1, and provides evidence that the probabilistic chord-tone profiles are useful predictors of the chord labels when given the set of chord-tones that exemplify that label. This result is also relevant for the fitness function, as it relies on these probabilistic chord profiles to make estimations.

Need to sort out the axes here. Each row has been normalised, so the colour represents the proportion of each true label that has been predicted a specific label. E.G. Here we see a

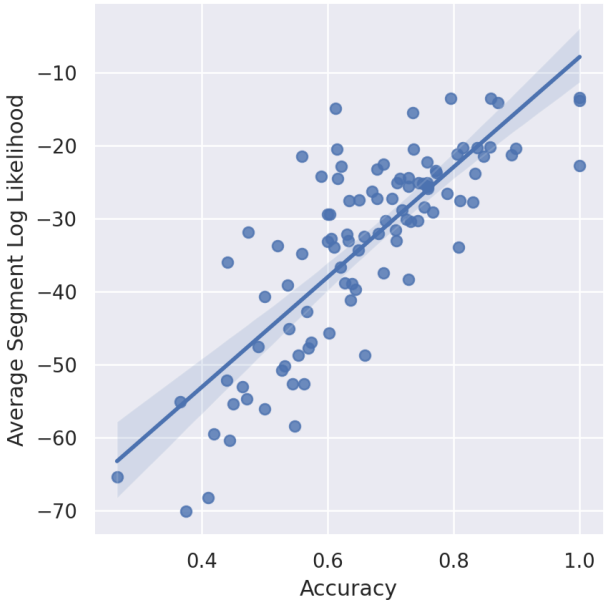


Figure 4.1: Accuracy against log likelihood

systematic problem with the prediction of mm7 chord as a major chord. Figure 4.2 shows the confusion matrix of the different predicted labels against the true labels the value plotted is the proportion of the true class data points. As expected, the classes with the greatest recall are those that appear the most in the datasets.

Figure **??**: IR represents incorrect root. NC represents no chord. Actually a problem with root note prediction. We are using a bag of spelled pitches, ignoring the octave information.

Limitations

Flat vs hierarchical classication.

Annotator subjectivity. DCML annotation standard aims to deal with this. [21]

4.2 Baseline Algorithms

Questions

Can the proto-voice model be used to perform a chord segment reduction? (Section 4.2)

Experiment Setup

It is possible for searches to get stuck. I ran the **RandomWalk** algorithm on the combined dataset consisting of 174 pieces, allowing 3 reruns per parse with a 300s timeout for each parse,

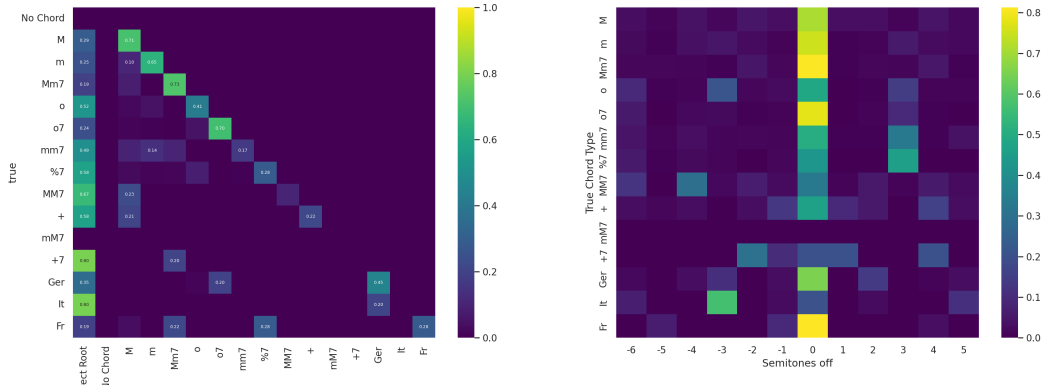


Figure 4.2: Chord type confusion matrices. Chord type prediction is shown on the left, and root note prediction is shown on the right.

and found that 118 pieces failed to be parsed, 57 pieces were able to be parsed successfully and 5 pieces had some failures and some successes.

Discuss why some pieces cannot be parsed. Two reasons. 1. Combinatorial explosion with the number of possible reductions, causing a timeout and memory overload. 2. Actually haven't no possilbe reductions available. Perhaps it would be good to show an example here.

As a result, tests were run on a stratified sample of the datasets consisting of only pieces that could be consistently parsed by the RandomWalk algorithm using the proto-voice Parser.

Discussion of results

It is possible for searches to get stuck. I ran the **RandomWalk** algorithm on the combined dataset consisting

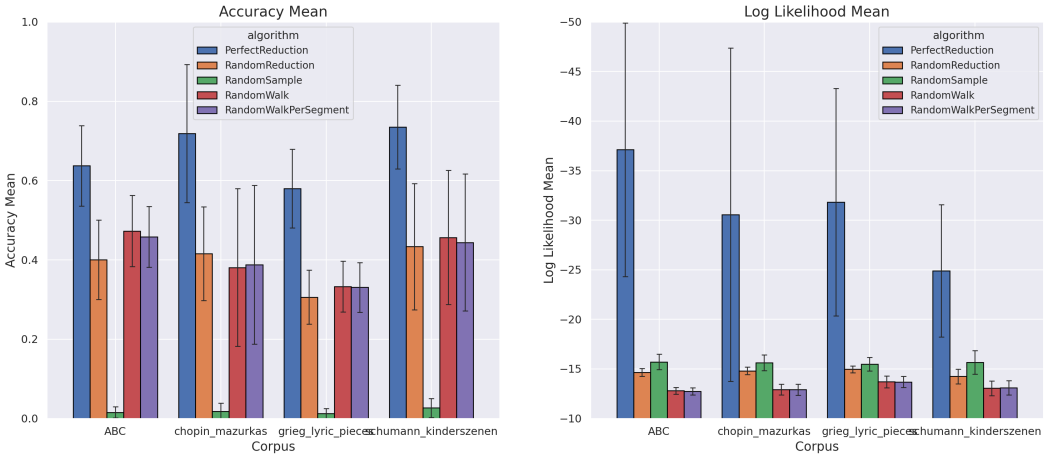


Figure 4.3

4.3 Extension: Heuristic Search

4.3.1 Interpretability

Question: Does the fitness function developed provide a meaningful proxy for the true accuracy of a proto-voice reduction? Show the derivation of the same segment, one from the random parse, one from the heuristic search
Heuristic search vs Random Search Quantititaively

Experiment Setup

Discussion of Results

Question: Does the developed heuristic search algorithm achieve an improvement in accuracy compared to the random parse?

Experiment Setup

Discussion of Results

4.3.2 Scalability

Question: Does the developed heuristic search algorithm achieve linear time complexity with respect to the input dimensions?

Experiment Setup

Discussion of Results

Experiment Setup

4.4 Success Criteria

4.5 Overview of Limitations

Chapter 5

Conclusions

This project aimed to [] and it has been demonstrated that this was achieved with great success. This chapter provides a reflection on the achievements of the project, the lessons learned, and possible avenues of future work.

5.1 Achievements

- Reiterate how each of the questions asked on the outset have been answered.

5.2 Lessons learned

- Use of visualisations for development very complex algorithm and output.
- Logging at different verbosity levels proved very helpful.

5.3 Future Work

- Different Domains
- Further heuristic algorithms/ genetic algorithms
- Context
- Non left-most parse
-

Bibliography

[1] Chen, T.-P., and Su, L. Functional Harmony Recognition of Symbolic Music Data with Multi-task Recurrent Neural Networks, Sept. 2018.

[2] Chen, T.-P., and Su, L. Harmony Transformer: Incorporating Chord Segmentation into Harmony Recognition. In *International Society for Music Information Retrieval Conference* (2019).

[3] Cohen, D. E. "The Imperfect Seeks Its Perfection": Harmonic Progression, Directed Motion, and Aristotelian Physics. *Music Theory Spectrum 23*, 2 (Oct. 2001), 139–169.

[4] Davidson-Pilon, C. Bayesian Methods for Hackers, Mar. 2023.

[5] Deng, J., and Kwok, Y.-K. Large vocabulary automatic chord estimation using bidirectional long short-term memory recurrent neural network with even chance training. *Journal of New Music Research 47*, 1 (Jan. 2018), 53–67.

[6] Dong, L., and Lapata, M. Language to Logical Form with Neural Attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Berlin, Germany, 2016), Association for Computational Linguistics, pp. 33–43.

[7] Ebcioğlu, K. An Expert System for Harmonizing Four-Part Chorales. *Computer Music Journal 12*, 3 (1988), 43–51.

[8] Fan, X., Liu, S., and Henderson, T. C. Explainable AI for Classification using Probabilistic Logic Inference, May 2020.

[9] Fard, H. H. Automatic Chord Recognition with Fully Convolutional Neural Networks.

[10] Feisthauer, L., Bigo, L., Giraud, M., and Levé, F. Estimating keys and modulations in musical pieces. In *Sound and Music Computing Conference (SMC 2020)* (June 2020).

[11] Finkensiep, C. Haskell-musicology: Scientific music types in Haskell. DCMLab, 2019.

[12] Finkensiep, C. Protovoices - A Model of Tonal Structure. Digital and Cognitive Musicology Lab, Nov. 2021.

[13] Finkensiep, C. *The Structure of Free Polyphony.* PhD thesis, EPFL, Lausanne, 2023.

[14] Finkensiep, C., Ericson, P., Klassmann, S., and Rohrmeier, M. Chord Types and Ornamentation - A Bayesian Model of Extended Chord Profiles. *Open Research Europe* (Apr. 2023).

[15] Finkensiep, C., and Rohrmeier, M. Modeling and Inferring Proto-voice Structure in Free Polyphony. In *Proceedings of the 22nd ISMIR Conference* (Online, Nov. 2021).

[16] Gjerdingen, R. Cognitive Foundations of Musical Pitch Carol L. Krumhansl. *Music Perception: An Interdisciplinary Journal 9* (July 1992), 476–492.

[17] Goodman, J. Semiring Parsing. *Computational Linguistics 25*, 4 (1999), 573–606.

[18] Gotham, M., Kleinertz, R., Weiss, C., Müller, M., and Klauk, S. What if the 'When' Implies the 'What'?: Human harmonic analysis datasets clarify the relative role of the separate steps in automatic tonal analysis. In *Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR 2021, Online, November 7-12, 2021* (2021), J. H. Lee, A. Lerch, Z. Duan, J. Nam, P. Rao, P. van Kranenburg, and A. Srinivasamurthy, Eds., pp. 229–236.

[19] Granroth-Wilding, M. Harmonic Analysis of Music Using Combinatory Categorial Grammar. *The University Of Edinburgh* (2013).

[20] Harasim, D. Harmonic Syntax in Time: Rhythm Improves Grammatical Models of Harmony. In *Proceedings of the 20th ISMIR Conference* (2019), T. J. O'Donnell and M. A. Rohrmeier, Eds., ISMIR.

[21] Humphrey, E. J., and Bello, J. P. Four Timely Insights on Automatic Chord Estimation.

[22] Islam, R., Andreev, A. V., Shusharina, N. N., and Hramov, A. E. Explainable Machine Learning Methods for Classification of Brain States during Visual Perception. *Mathematics 10*, 15 (Jan. 2022), 2819.

[23] Johannes, H. Ms3 - Parsing MuseScore 3. https://github.com/johentsch/ms3, 2021.

[24] Kidney, D. O., and Wu, N. Algebras for weighted search. *Proceedings of the ACM on Programming Languages 5*, ICFP (Aug. 2021), 1–30.

[25] Koops, H. V., de Haas, W. B., Bransen, J., and Volk, A. Automatic chord label personalization through deep learning of shared harmonic interval profiles. *Neural Computing and Applications 32*, 4 (Feb. 2020), 929–939.

[26] Koops, H. V., de Haas, W. B., Burgoyne, J. A., Bransen, J., Kent-Muller, A., and Volk, A. Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research 48*, 3 (May 2019), 232–252.

[27] Korzeniowski, F., and Widmer, G. Improved Chord Recognition by Combining Duration and Harmonic Language Models, Aug. 2018.

[28] Krumhansl, C. L., and Kessler, E. J. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review 89* (1982), 334–368.

[29] Lerdahl, F., and Jackendoff, R. *A Generative Theory of Tonal Music*, repr. ed. MIT Press, Cambridge, Mass., 2010.

[30] Lu, S., Mao, J., Tenenbaum, J. B., and Wu, J. Neurally-Guided Structure Inference, Aug. 2019.

[31] Marsden, A. Schenkerian Analysis by Computer: A Proof of Concept. *Journal of New Music Research 39*, 3 (Sept. 2010), 269–289.

[32] Masada, K., and Bunescu, R. Chord Recognition in Symbolic Music: A Segmental CRF Model, Segment-Level Features, and Comparative Evaluations on Classical and Popular Music, Oct. 2018.

[33] Mauch, M., Mullensiefen, D., Dixon, S., and Wiggins, G. Can Statistical Language Models be used for the Analysis of Harmonic Progressions?

[34] Maxwell, H. J. An expert system for harmonizing analysis of tonal music. In *Understanding Music with AI: Perspectives on Music Cognition*. MIT Press, Cambridge, MA, USA, Aug. 1992, pp. 334–353.

[35] McLeod, A., and Rohrmeier, M. A Modular System for the Harmonic Analysis of Musical Scores using a Large Vocabulary. In *Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR 2021, Online, November 7-12, 2021* (2021), J. H. Lee, A. Lerch, Z. Duan, J. Nam, P. Rao, P. van Kranenburg, and A. Srinivasamurthy, Eds., pp. 435–442.

[36] Mcleod, A., Suermondt, X., Rammos, Y., Herff, S., and Rohrmeier, M. A. Three Metrics for Musical Chord Label Evaluation. In *Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation* (Kolkata India, Dec. 2022), ACM, pp. 47–53.

[37] Mearns, L. The Computational Analysis of Harmony in Western Art Music.

[38] Neuwirth, M., Harasim, D., Moss, F. C., and Rohrmeier, M. The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets. *Frontiers in Digital Humanities 5* (July 2018), 16.

[39] Ni, Y., Mcvicar, M., Santos-Rodriguez, R., and De Bie, T. An end-to-end machine learning system for harmonic analysis of music, July 2011.

[40] Oudre, L., Fevotte, C., and Grenier, Y. Probabilistic Template-Based Chord Recognition. *IEEE Transactions on Audio, Speech, and Language Processing 19*, 8 (Nov. 2011), 2249–2259.

[41] Pardo, B., and Birmingham, W. P. Algorithms for Chordal Analysis. *Computer Music Journal 26*, 2 (June 2002), 27–49.

[42] Pearl, J. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. The Addison-Wesley Series in Artificial Intelligence. Addison-Wesley Pub. Co, Reading, Mass, 1984.

[43] Pickens, J. *Harmony Modeling for Polyphonic Music Retrieval*. PhD thesis, University of Massachusetts, 2004.

[44] Quinn, I. Are Pitch-Class Profiles Really "Key for Key"? *Zeitschrift der Gesellschaft für Musiktheorie [Journal of the German-Speaking Society of Music Theory] 7* (Jan. 2010).

[45] Radicioni, D. P., and Esposito, R. BREVE: An HMPerceptron-Based Chord Recognition System. In *Advances in Music Information Retrieval*, J. Kacprzyk, Z. W. Raś, and A. A. Wieczorkowska, Eds., vol. 274. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 143–164.

[46] Raffel, C., McFee, B., Humphrey, E. J., Salamon, J., Nieto, O., Liang, D., and Ellis, D. P. W. Mir_eval: A Transparent Implementation of Common MIR Metrics. In *Proceedings of the 15th International Conference on Music Information Retrieval* (2014).

[47] Raphael, C., and Stoddard, J. Functional Harmonic Analysis Using Probabilistic Models. *Computer Music Journal 28*, 3 (Sept. 2004), 45–52.

[48] Rocher, T., Robine, M., Hanna, P., and Strandh, R. Dynamic Chord Analysis for Symbolic Music.

[49] Sapp, C. Visual hierarchical key analysis. *Computers in Entertainment 3* (Oct. 2005), 1–19.

[50] Sapp, C. 6 Computational Chord-Root Identification in Symbolic Musical Data: Rationale, Methods, and Applications.

[51] Sarica, A., Quattrone, A., and Quattrone, A. Explainable machine learning with pairwise interactions for the classification of Parkinson's disease and SWEDD from clinical and imaging features. *Brain Imaging and Behavior 16*, 5 (Oct. 2022), 2188–2198.

[52] Sidorov, K., Jones, A., and Marshall, D. MUSIC ANALYSIS AS A SMALLEST GRAMMAR PROBLEM.

[53] Temperley, D. An Algorithm for Harmonic Analysis. *Music Perception 15*, 1 (Oct. 1997), 31–68.

[54] Temperley, D. A Bayesian Approach to Key-Finding. In *Music and Artificial Intelligence*, G. Goos, J. Hartmanis, J. van Leeuwen, C. Anagnostopoulou, M. Ferrand, and A. Smaill, Eds., vol. 2445. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 195–206.

[55] Temperley, D. A Bayesian Approach to Key-Finding. In *Music and Artificial Intelligence* (Berlin, Heidelberg, 2002), C. Anagnostopoulou, M. Ferrand, and A. Smaill, Eds., Lecture Notes in Computer Science, Springer, pp. 195–206.

[56] Temperley, D. A Unified Probabilistic Model for Polyphonic Music Analysis. *Journal of New Music Research 38*, 1 (Mar. 2009), 3–18.

[57] Vijayakumar, A. K., Cogswell, M., Selvaraju, R. R., Sun, Q., Lee, S., Crandall, D., and Batra, D. DIVERSE BEAM SEARCH: DECODING DIVERSE SOLUTIONS FROM NEURAL SEQUENCE MODELS.

[58] Vijayakumar, A. K., Cogswell, M., Selvaraju, R. R., Sun, Q., Lee, S., Crandall, D., and Batra, D. DIVERSE BEAM SEARCH: DECODING DIVERSE SOLUTIONS FROM NEURAL SEQUENCE MODELS.

[59] Volk, A. Improving Audio Chord Estimation by Alignment and Integration of Crowd-Sourced Symbolic Music. 141–155.

[60] Winograd, T. Linguistics and the Computer Analysis of Tonal Harmony. *Journal of Music Theory 12*, 1 (1968), 2–49.

[61] Wu, Y., Carsault, T., and Yoshii, K. Automatic Chord Estimation Based on a Frame-wise Convolutional Recurrent Neural Network with Non-Aligned Annotations. In *2019 27th European Signal Processing Conference (EUSIPCO)* (A Coruna, Spain, Sept. 2019), IEEE, pp. 1–5.

Chapter A

Additional Information

Chapter B

Project Proposal

Inferring Harmony from Free Polyphony

Judah Daniels

May 5, 2023

DOS: Prof. Larry Paulson

Crsid: jasd6
College: Clare College

B.1 Abstract

A piece of music can be described using a sequence of chords, representing a higher level harmonic structure of a piece. There is a small, finite set of chord types, but each chord can be realised on the musical surface in a practically infinite number of ways. Given a score, we wish to infer the underlying chord types.

The paper *Modeling and Inferring Proto-voice Structure in Free Polyphony* describes a generative model that encodes the recursive and hierarchical dependencies between notes, giving rise to a grammar-like hierarchical system [15]. This proto-voice model can be used to reduce a piece into a hierarchical structure which encodes an understanding of the tonal/harmonic relations of a piece.

Christoph Finkensiep suggests in his paper that the proto-voice model may be an effective way to infer higher level latent entities, such as harmonies or voice leading schemata. Thus in this project I will ask the question: is this parsing model an effective way to annotate harmonies? By 'effective' we are referring to two things:

- Accuracy: can the model successfully emulate how experts annotate harmonic progressions in musical passages?
- Practicality: can the model be used to do this within a reasonable time frame?

While the original model could in theory be used to generate harmonic annotations, its exhaustive search strategy would be prohibitively time-consuming in practice for any but the shortest musical extracts; one half measure can have over 100,000 valid derivations [15]. My approach will be to explore the use of heuristic search algorithms to solve this problem.

B.2 Substance and Structure

B.2.1 Core: Search

The core of this project is essentially a search problem characterised as follows:

- The state space S is the set of all possible partial reductions of a piece along with each reduction step that has been done so far.
- We have an initial state $s_o \in S$, which is the empty reduction, corresponding to the unreduced surface of the piece. The score is represented as a sequence of slices grouping notes that sound simultaneously. We are also given the segmentation of the original chord labels that we wish to retrieve.
- We have a set of actions, A modelled by a function $action : A \times S \to S$. These actions correspond to a single reduction step.
  - The reduction steps are the inverses of the operations defined by the generative proto-voice model.

- Finally we have a goal test, $goal : S \to \{true, false\}$ which is true iff the partial reduction s has exactly one slice per segment of the input.
  - This means the partial reduction s contains a sequence of slices which start and end positions corresponding to the segmentation of the piece.
- At the first stage, this will be implemented using a random graph search algorithm, picking each action randomly, according to precomputed distributions.

B.2.2 Core: Evaluation

The second core task is to create an evaluation module that iterates over the test dataset, and evaluates the partial reduction computed by the search algorithm above. This will be done by comparing the outputs to ground truth annotations from the Annotated Beethoven Corpus.

In order to do this I will make use of the statistical harmony model from Finkensiep's thesis, *The Structure of Free Polyphony* [13]. This model provides a way of mapping between the slices that the algorithm generates and the chords in the ground truth. This can be used to empirically measure how closely the slices match the expert annotations.

B.2.3 Extension

Once the base search implementation and evaluation module have been completed, the search problem will be tackled by heuristic search methods, with different heuristics to be trialled and evaluated against each other. The heuristics will make use of the chord profiles from Finkensiep's statistical harmony model discussed above. These profiles relate note choices to the underlying harmony. Hence the heuristics may include:

- How the chord types relate to the pitches used.
- How the chord types relate which notes are used as ornamentation, and the degree of ornamentation.
- Contextual information about neighboring slices

B.2.4 Overview

The main work packages are as follows:

Preliminary Reading – Familiarise myself with the proto-voice model, and read up on similar models and their implementations. Study heuristic search algorithms.

Dataset Preparation – Pre-process the Annotated Beethoven Corpus into a suitable representation for my algorithm.

Basic Search – Implement a basic random search algorithm that takes in surface and segmentations, and outputting the sequence of slices matching the segmentations.

Evaluation Module – Implement an evaluation module to evaluate the output from the search algorithm.

End-to-end pipeline – Implement a full pipeline from the data to the evaluation that can be used to compare different reductions.

Heuristic Design – Extension – Trial different heuristics and evaluate their performance against each other.

Dissertation – I intend to work on the dissertation throughout the duration of the project. I will then focus on completing and polishing the project upon completion.

B.3 Starting Point

The following describes existing code and languages that will be used for this project:

Haskell – I will be using Haskell for this project as it is used in the proto-voice implementation. It must be noted that my experience with Haskell is limited, as I was first introduced to it via an internship this summer (July to August 2022).

Python – Python will be used for data handling. I have experience coding in Python.

Prior Research - Over the summer I have been reading the literature on computational models of music, as well as various parsing algorithms such as semi-ring parsing [17], and the CYK algorithm, which is used in the implementation of the proto-voice model.

Protovoices-Haskell – The paper *Modeling and Inferring Proto-Voice Structure in Free Polyphony* [15] includes an implementation of the proto-voice model in Haskell. A fork of this repository will form the basis of my project. This repository includes as parsing module which will be used to perform the actions in the search space of partial reductions. There is module that can exhaustively enumerate reductions of a piece, but this is infeasible in practice due to the blowup of the derivation forest.

MS3 – This is a library for parsing MuseScore Files and manipulating labels [23], which I will use as part of the data processing pipeline.

ABC – The *Annotated Beethoven Corpus* [38] contains analyses of all Beethoven string quartets composed between 1800 and 1826), encoded in a human and machine readable format. This will be used as a dataset for this project.

B.4 Success Criteria

This project will be deemed a success if I complete the following tasks:

- Develop a baseline search algorithm that uses the proto-voice model to output a partial reduction of a piece of music up to the chord labels.
- Create an evaluation module that can take the output of the search algorithm and quantitatively evaluate its accuracy against the ground truth annotations by providing a score based on a statistical harmony model.
- Extension: Develop one or more search algorithms that use additional heuristics to inform the search, and compare the accuracy with the baseline algorithm.

B.5 Timetable

Time frame	Work	Evidence
Michaelmas (Oct 4 to Dec 2)		
Oct 14 to Oct 24	*Oct 14*: Final proposal deadline. Preparation work: familiarise myself with the dataset and the proto-voice model implementation. Work on manipulating reductions using the proto-voice parser provided by the paper.	None
Oct 24 to Nov 7	Dataset preparation and handling.	Plot useful metrics about the dataset using Haskell
Nov 7 to Nov 21	Random Search implementation	None
Nov 21 to Dec 5	Evaluation Module. Continue with search implementation.	Evaluate a manually created derivation and plot results
Vacation (Dec 3 to Jan 16)		
Dec 5 to Dec 11	Evaluate performance of random search. Begin to work on extensions	Plot results
Dec 10 to Dec 21	Trial different heuristics. Implement an end-to-end pipeline from input to evaluation.	None
Dec 21 to Dec 27	None	None
Dec 27 to Jan 10	Continue trialing and evaluating heuristics	*Fulfill success criterion: At least one heuristic technique gives better performance than random search.*
Lent (Jan 17 to Mar 17)		
Jan 4 to Jan 20	Buffer Period to help keep on track	None
Jan 20 to Feb 3	*Feb 3*: Progress Report Deadline. Write progress report and prepare presentation. Write draft *Evaluation* chapter	Progress Report (approx. 1 page)
Feb 3 to Feb 17	Prepare presentation.	*Feb 8 – 15*: Progress Report presentation
Feb 17 to Mar 3	*Feb 17*: How to write a Dissertation briefing. Write draft Introduction and Preparation chapters. Incorporate feedback on Evaluation chapter.	Send draft Introduction and Preparation chapter to supervisor
Mar 3 to Mar 17	Write draft Implementation chapters. Incorporate feedback on Introduction and Preparation chapters.	Send draft Implementation chapters to Supervisor
Vacation (Mar 18 to		

B.6 Resources

I plan to use my own laptop for development: MacBook Pro 16-inch, M1 Max, 32GB Ram, 1TB SSD, 24-core GPU.

All code will be stored on a GitHub repository, which will guarantee protection from data loss. I will easily be able to switch to using university provided computers upon hardware/software failure.

The project will be built upon work that has been done in the DCML (Digital cognitive musicology lab) based in EPFL. The files are in their Github repository, and I have been granted permission to access their in-house datasets of score annotations, as well as software packages which are used to handle the data.

B.7 Supervisor Information

Peter Harrison, head of Centre for Music and Science at Cambridge, has agreed to supervise me for this. We have agreed on a timetable for supervisions for this year. I am also working with Christoph Finkensiep, a PHD student at the DCML, and originator of the proto-voice model. Professor Larry Paulson has agreed to be the representative university teaching officer.