Voice-Leading Schema Recognition using Rhythm and Pitch Features

ISMIR 2020, 12–15 October

Christoph Finkensiep, Ken Déguernel, Markus Neuwirth, and Martin Rohrmeier
christoph.finkensiep@epfl.ch

Overview

Go to the interactive version!

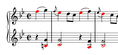
What is a Schema?

A voice-leading schema is a type of musical pattern. It consists of

- a fixed number of **voices**
- that go through several **stages**,

and it has a characteristic **interval structure**.

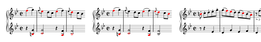
Occurrences of a schema in a piece of music are called **instances**. They consist of the same intervals, but the notes are **temporally displaced** and **ornamented** with other notes.



Finding Schemata

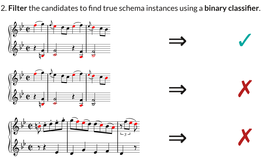
We can find schema instances in two steps:

1. **Enumerate** all note combinations (within a certain time window) with the correct interval structure using a skipgram-based algorithm. We call these note combinations **candidates**.



The candidates include
   - **true instances**,
   - **variants** of true instances with some wrong notes,
   - and **incidental** note combinations with the right interval structure.

2. **Filter** the candidates to find true schema instances using a **binary classifier**.



Classification of Schema Candidates

We classify schema candidates based on a set of **features**. These features are designed to measure the recognizability in three aspects:

- regularity
- complexity
- salience

The features are combined using **logistic regression**.

Regularity

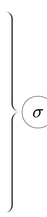
```
rreg
mreg
preg
```

Complexity

```
rdsums
pdsums
rdsumv
pdsumv
vdist
onsets
```

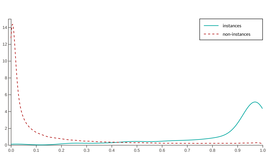
Salience

```
dur
mweight
profiledist
```



Classification Performance

The classifier is generally able to distinguish true instances and false candidates. Instances are mostly rated close to 1, while non-instances are mostly rated close to 0.



Distribution of the classification score for instances and non-instances.

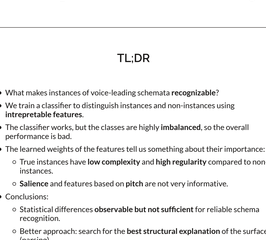
However, the two classes are **very imbalanced** (ca. 1:1000), with many *more* non-instances than instances. the actual recognition performance is not very good. Candidates with high ratings are more likely to be non-instances than instances, as we can see when we adjust the above plot.

Because of this strong imbalance, an extremely good classification performance is needed, which seems to be **impossible with heuristic features**.

Interpreting Feature Importance

We can interpret the **importance of features** for the classifier through the learned weights, scaled by the average feature value. We observe that:

- true instances are marked by **low complexity** and **high regularity** ,
- **salience** is *not* very important,
- features based on **pitch** are *not* important.

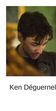


TL;DR

- What makes instances of voice-leading schemata **recognizable**?
- We train a classifier to distinguish instances and non-instances using **interpretable features**.
- The classifier works, but the classes are highly **imbalanced**, so the overall performance is bad.
- The learned weights of the features tell us something about their importance:
  - True instances have **low complexity** and **high regularity** compared to non-instances.
  - **Salience** and features based on **pitch** are not very informative.
- Conclusions:
  - Statistical differences **observable but not sufficient** for reliable schema recognition.
  - Better approach: search for the **best structural explanation** of the surface (parsing).

Authors and Acknowledgements

Christoph Finkensiep



Ken Déguernel

Markus Neuwirth

Martin Rohrmeier

from the Digital and Cognitive Musicology Lab

Corresponding author: Christoph Finkensiep

This research was kindly supported by Claude Latour and the Volkswagen Foundation.