

## A. Types of Chords in Tonal Music

A *chord* is a group of notes that form a cohesive harmonic unit to the listener when sounding simultaneously (Aldwell et al., 2011). We design our system to handle the following types of chords: triads, augmented 6th chords, suspended chords, and power chords.

### A.1 Triads

A *triad* is the prototypical instance of a chord. It is based on a root note, which forms the lowest note of a chord in standard position. A third and a fifth are then built on top of this root to create a three-note chord. Inverted triads also exist, where the third or fifth instead appears as the lowest note. The chord labels used in our system do not distinguish among inversions of the same chord. However, once the basic triad is determined by the system, finding its inversion can be done in a straightforward post-processing step, as a function of the bass note in the chord. The quality of the third and fifth intervals of a chord in standard position determines the mode of a triad. For our system, we consider three triad modes: *major* (maj), *minor* (min), and *diminished* (dim). A major triad consists of a major third interval (i.e. 4 half steps) between the root and third, as well as a perfect fifth (7 half steps) between the root and fifth. A minor triad has a minor third interval (3 half steps) between the root and third. Lastly, a diminished triad maintains the minor third between the root and third, but contains a diminished fifth (6 half steps) between the root and fifth. Figure 7 shows these three triad modes, together with three versions of a C major chord, one for each possible type of added note, as explained below.

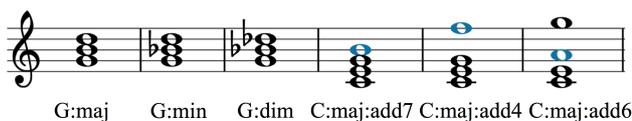


Figure 7: Triads in 3 modes and with 3 added notes.

A triad can contain an added note, or a fourth note. We include three possible added notes in our system: a fourth, a sixth, and a seventh. A fourth chord (add4) contains an interval of a perfect fourth (5 half steps) between the root and the added note for all modes. In contrast, the interval between the root and added note of a sixth chord (add6) of any mode is a major sixth (9 half steps). For seventh chords (add7), the added note interval varies. If the triad is major, the added note can form a major seventh (11 half steps) with the root, called a major seventh chord. It can also form a minor seventh (10 half steps) to create a dominant seventh chord. If the triad is minor, the added seventh can again either form an interval of a major seventh, creating a minor-major seventh chord, or a minor seventh, forming a minor seventh chord. Finally, di-

inished triads most frequently contain a diminished seventh interval (9 half steps), producing a fully diminished seventh chord, or a minor seventh interval, creating a half-diminished seventh chord.

### A.2 Augmented 6th Chords

An *augmented 6th chord* is a type of chromatic chord defined by an augmented sixth interval between the lowest and highest notes of the chord (Aldwell et al., 2011). The three most common types of augmented 6th chords are *Italian*, *German*, and *French* sixth chords, as shown in Figure 8 in the key of A minor. In a minor scale, Italian sixth chords can be seen as iv chords with a sharpened root, in the first inversion. Thus, they can be created by stacking the sixth, first, and sharpened fourth scale degrees. In minor, German sixth chords are iv<sup>7</sup> (i.e. minor seventh) chords with a sharpened root, in the first inversion. They are formed by combining the sixth, first, third, and sharpened fourth scale degrees. Lastly, French sixth chords are created by stacking the sixth, first, second, and sharpened fourth scale degrees. Thus, they are ii<sup>ø7</sup> (i.e. half-diminished seventh) chords with a sharpened third, in second inversion.



Figure 8: Common types of augmented 6th chords, shown for the A minor scale. The same notes would also be used for the A major scale.

### A.3 Suspended and Power Chords

Both *suspended* and *power chords* are similar to triads in that they contain a root and a perfect fifth. They differ, however, in their omission of the third. As shown in Figure 9, *suspended second chords* (sus2) use a second (2 half steps) with the root, while *suspended fourth chords* (sus4) employ a perfect fourth as replacement (Taylor, 1989). The suspended second and fourth often resolve to a more stable third. In addition to these two kinds of suspended chords, our system considers suspended fourth chords that contain an added minor seventh, forming a *dominant seventh suspended fourth chord* (7sus4).

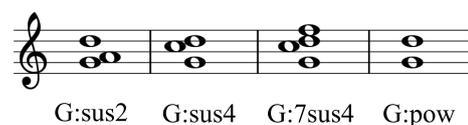


Figure 9: Suspended and power chords.

In contrast with suspended chords, power chords (pow) do not contain a replacement for the missing

third. They simply consist of a root and a perfect fifth. Though they are not formally considered to be chords in classical music, they are commonly referred to in both rock and pop music (Denyer, 1992).

#### A.4 Chord Ambiguity

Sometimes, the same set of notes can have multiple chord interpretations. For example, the German sixth chord shown in Figure 8 can also be interpreted as an F dominant seventh chord. Added notes can also lead to other types of ambiguity, for example {D, F, A, C} could be an F major sixth chord (i.e. F major with an added sixth) or a D minor seventh chord (i.e. D minor chord with an added minor seventh). Human annotators can determine the correct chord interpretation based on cues such as inversions and context. The semi-CRF model described in this paper captures inversions through the bass features (Appendix C.3), whereas context is taken into account through the chord bigram features (Appendix C.4). This could be further improved by adding other features, such as determining how notes in the current chord resolve to notes in the next chord.

### B. Figuration Heuristics

We designed a set of heuristics to determine whether a note  $n$  from a segment  $s$  is a figuration note with respect to a candidate chord label  $y$ . The heuristic rules shown below discover four types of figurations: passing and neighbor notes (Figure 10), and suspensions and anticipations (Figure 11).



Figure 10: Examples of figuration notes, in red.

**Passing:** There are two anchor notes  $n_1$  and  $n_2$  such that:  $n_1$ 's offset coincides with  $n$ 's onset;  $n_2$ 's onset coincides with  $n$ 's offset;  $n_1$  is one scale step below  $n$  and  $n_2$  is one step above  $n$ , or  $n_1$  is one step above  $n$  and  $n_2$  one step below;  $n$  is not longer than either  $n_1$  or  $n_2$ ; the accent value of  $n$  is strictly smaller than the accent value of  $n_1$ ; at least one of the two anchor notes belongs to segment  $s$ ;  $n$  is non-harmonic with respect to chord  $y$ , i.e.  $n$  is not equivalent to the root, third, fifth, or added note of  $y$ ; both  $n_1$  and  $n_2$  are harmonic with respect to the segments they belong to.

**Neighbor:** There are two anchor notes  $n_1$  and  $n_2$  such that:  $n_1$ 's offset coincides with  $n$ 's onset;  $n_2$ 's onset coincides with  $n$ 's offset;  $n_1$  and  $n_2$  are both either one step below or one step above  $n$ ;  $n$  is not longer than either  $n_1$  or  $n_2$ ; the accent value of  $n$  is strictly smaller than the accent value of  $n_1$ ; at least one of the two anchor notes belongs to segment  $s$ ;  $n$  is non-harmonic with respect to chord  $y$ ; both anchor notes

are harmonic with respect to the segments they belong to.



Figure 11: Examples of figuration notes, in red.

**Suspension:** Note  $n$  belongs to the first event of segment  $s$ . There is an anchor note  $m$  in the previous event (last event in the previous segment) such that:  $m$  and  $n$  have the same pitch;  $n$  is either tied with  $m$  (i.e. held over) or  $m$ 's offset coincides with  $n$ 's onset (i.e. restruck);  $n$  is not longer than  $m$ ;  $n$  is non-harmonic with respect to chord  $y$ , while  $m$  is harmonic with respect to the previous chord.

**Anticipation:** Note  $n$  belongs to the last event of segment  $s$ . There is an anchor note  $m$  in the next event (first event in the next segment) such that:  $n$  and  $m$  have the same pitch;  $m$  is either tied with  $n$  (i.e. held over) or  $n$ 's offset coincides with  $m$ 's onset (i.e. restruck);  $n$  is not longer than  $m$ ;  $n$  is non-harmonic with respect to chord  $y$ , while  $m$  is harmonic relative to all other notes in its event.

Furthermore, because the weak semi-CRF features shown in Equation ?? do not have access to the candidate label  $y_{k-1}$  of the previous segment  $s_{k-1}$ , we need a heuristic to determine whether an anchor note is harmonic whenever the anchor note belongs to the previous segment. The heuristic simply looks at the other notes in the event containing the anchor note: if the event contains 2 or more other notes, at least 2 of them need to be consonant with the anchor, i.e. intervals of octaves, fifths, thirds, and their inversions; if the event contains just one note other than the anchor note, it has to be consonant with the anchor.

We emphasize that the rules mentioned above for detecting figuration notes are only approximations. We recognize that correctly identifying figuration notes can also depend on subtler stylistic and contextual cues, thus allowing for exceptions to each of these rules.

### C. Chord Recognition Features

Given a segment  $s$  and chord  $y$ , we will use the following notation:

- $s.Notes$ ,  $s.N$  = the set of notes in the segment  $s$ .
- $s.Events$ ,  $s.E$  = the sequence of events in  $s$ .
- $e.len$ ,  $n.len$  = the length (i.e. duration) of event  $e$  or note  $n$ , in quarters.
- $e.acc$ ,  $n.acc$  = the *accent value* of event  $e$  or note  $n$ , as computed by the `beatStrength()` function in Music21<sup>1</sup>.

<sup>1</sup>Link to Music21: <http://web.mit.edu/music21>

- $y.root$ ,  $y.third$ , and  $y.fifth$  = the triad tones of the chord  $y$ .
- $y.added$  = the added note of chord  $y$ , if  $y$  is an added tone chord.
- $s.Fig(y)$  = the set of notes in  $s$  that are *figuration* with respect to chord  $y$ .
- $s.NonFig(y) = s.Notes - s.Fig(y)$  = the set of notes in  $s$  that are not *figuration* with respect to  $y$ .

Note that a note may contain multiple events, as such the note length  $n.len$  can be seen as the sum of the length of all events that span the duration of that note. For example, the first G3 in the bass of Figure 2 has a length of a quarter – it corresponds to the G3 in measure 2 of Figure 1 and is shown as a tied note to simplify the description. Therefore its  $n.len = 1$ . Each of the two events that span its duration have a length of an eighth, hence  $e_1.len = e_2.len = 0.5$ .

The *accent value* is determined based on the metrical position of a note or event, e.g. in a song written in a 4/4 time signature, the first beat position would have a value of 1.0, the third beat 0.5, and the second and fourth beats 0.25. Any other eighth note position within a beat would have a value of 0.125, any sixteenth note position strictly within the beat would have a value of 0.0625, and so on. To determine whether a note  $n$  from a segment  $s$  is a *figuration* note with respect to a candidate chord label  $y$ , we use a set of heuristics, as detailed in Appendix B.

The duration and accent-weighted segment-level features introduced in this section have real values. Given a real-valued feature  $f(s, y)$  that takes values in  $[0, 1]$ , we discretize it into  $K+2$  Boolean features by partitioning the  $[0, 1]$  interval into a set of  $K$  subinterval bins  $\mathcal{B} = \{(b_{k-1}, b_k] | 1 \leq k \leq K\}$ . For each bin, the corresponding Boolean feature determines whether  $f(s, y) \in (b_{k-1}, b_k]$ . Additionally, two Boolean features are defined for the boundary cases  $f(s, y) = 0$  and  $f(s, y) = 1$ . For each real-valued feature, unless specified otherwise, we use the bin set  $\mathcal{B} = [0, 0.1, \dots, 0.9, 1.0]$ .

### C.1 Segment Purity

The segment purity feature  $f_1(s, y)$  computes the fraction of the notes in segment  $s$  that are harmonic, i.e. belong to chord  $y$ :

$$f_1(s, y) = \frac{\sum_{n \in s.Notes} \mathbf{1}[n \in y]}{|s.Notes|}$$

The duration-weighted version  $f_2(s, y)$  of the purity feature weighs each note  $n$  by its length  $n.len$ :

$$f_2(s, y) = \frac{\sum_{n \in s.Notes} \mathbf{1}[n \in y] * n.len}{\sum_{n \in s.Notes} n.len}$$

The accent-weighted version  $f_3(s, y)$  of the purity feature weighs each note  $n$  by its accent weight  $n.acc$ :

$$f_3(s, y) = \frac{\sum_{n \in s.Notes} \mathbf{1}[n \in y] * n.acc}{\sum_{n \in s.Notes} n.acc}$$

The 3 real-valued features are discretized using the default bin set  $\mathcal{B}$ .

#### C.1.1 Figuration-Controlled Segment Purity

For each segment purity feature, we create a figuration-controlled version that ignores notes that were heuristically detected as *figuration*, i.e. replace  $s.Notes$  with  $s.NonFig(y)$  in each feature formula.

### C.2 Chord Coverage

The chord coverage features determine which of the chord notes belong to the segment. In this section, each of the coverage features are non-zero only for major, minor, and diminished triads and their added note counterparts. This is implemented by first defining an indicator function  $y.Triad$  that is 1 only for triads and chords with added notes, and then multiplying it into all the triad features from this section.

$$y.Triad = \mathbf{1}[y.mode \in \{\text{maj}, \text{min}, \text{dim}\}]$$

Furthermore, we compress notation by showing the mode predicates as attributes of the label, e.g.  $y.maj$  is a predicate equivalent with testing whether  $y.mode = \text{maj}$ . Thus, an equivalent formulation of  $y.Triad$  is as follows:

$$y.Triad = \mathbf{1}[y.maj \vee y.min \vee y.dim]$$

To avoid clutter, we do not show  $y.Triad$  in any of the features below, although it is assumed to be multiplied into all of them. The first 3 coverage features refer to the triad notes:

$$f_4(s, y) = \mathbf{1}[y.root \in s.Notes]$$

$$f_5(s, y) = \mathbf{1}[y.third \in s.Notes]$$

$$f_6(s, y) = \mathbf{1}[y.fifth \in s.Notes]$$

A separate feature determines if the segment contains all the notes in the chord:

$$f_7(s, y) = \prod_{n \in y} \mathbf{1}[n \in s.Notes]$$

A chord may have an added tone  $y.added$ , such as a 4th, a 6th, or a 7th. If a chord has an added tone, we define two features that determine whether the segment contains the added note:

$$f_8(s, y) = \mathbf{1}[\exists y.added \wedge y.added \in s.Notes]$$

$$f_9(s, y) = \mathbf{1}[\exists y.added \wedge y.added \notin s.Notes]$$

Through the first feature, the system can learn to prefer the added tone version of the chord when the segment contains it, while the second feature enables the

system to learn to prefer the triad-only version if no added tone is in the segment. To prevent the system from recognizing added chords too liberally, we add a feature that is triggered whenever the total length of the added notes in the segment is greater than the total length of the root:

$$\begin{aligned} alen(s, y) &= \sum_{n \in s.Notes} \mathbf{1}[n = y.added] * n.len \\ rlen(s, y) &= \sum_{n \in s.Notes} \mathbf{1}[n = y.root] * n.len \\ f_{10}(s, y) &= \mathbf{1}[\exists y.added] * \mathbf{1}[alen(s, y) > rlen(s, y)] \end{aligned}$$

The duration-weighted versions of the chord coverage features weigh each chord tone by its total duration in the segment. For the root, the feature would be computed as shown below:

$$f_{11}(s, y) = \frac{\sum_{n \in s.Notes} \mathbf{1}[n = y.root] * n.len}{\sum_{n \in s.Notes} n.len}$$

Similar features  $f_{12}$  and  $f_{13}$  are computed for the third and the fifth. The corresponding accent-weighted features  $f_{14}$ ,  $f_{15}$ , and  $f_{16}$  are computed in a similar way, by replacing the note duration  $n.len$  in the duration-weighted formulas with the note accent value  $n.acc$ .

The duration-weighted feature for the added tone is computed similarly:

$$f_{17}(s, y) = \frac{\mathbf{1}[\exists y.added] * \sum_{n \in s.Notes} \mathbf{1}[n = y.added] * n.len}{\sum_{n \in s.Notes} n.len}$$

Furthermore, by replacing  $n.len$  with  $n.acc$ , we also obtain the accent-weighted version  $f_{18}$ .

An alternative definition of duration-weighted features is based on the proportion of the segment time that is covered by a particular chord note. The corresponding duration-weighted feature for the chord root is shown below:

$$f_{19}(s, y) = \frac{\sum_{e \in s.Events} \mathbf{1}[y.root \in e] * e.len}{\sum_{e \in s.Events} e.len}$$

Similar duration-weighted features normalized by the segment length are defined for thirds, fifths, and added notes.

All duration-weighted and accent-weighted features are discretized using the default bin set  $\mathcal{B}$ .

### C.2.1 Chord Coverage for Augmented 6th Chords

We label each note appearing in an augmented 6th chord as follows:

- $y.bass$  = the lowest note.
- $y.3rd$  = the note that is a third above  $y.bass$ .
- $y.6th$  = the note that is an augmented sixth above  $y.bass$ .

- $y.5th$  = the note that forms a perfect fifth (for German 6th chords) or a diminished fifth (for French) above  $y.bass$ .

The features defined in this section are non-zero only for augmented 6th chord labels. Similar to Appendix C.2, we define an indicator function  $y.AS$  that is 1 only for augmented 6th chords and implicitly multiply this into each of the features from this section.

$$\begin{aligned} y.AS &= \mathbf{1}[y.mode \in \{it6, fr6, ger6\}] \\ y.AS &= \mathbf{1}[y.it6 \vee y.fr6 \vee y.ger6] \end{aligned}$$

We define an additional indicator function  $y.FG$  that is 1 only for French and German 6th chords.

$$y.FG = \mathbf{1}[y.fr6 \vee y.ger6]$$

The coverage features for augmented 6th chords are overall analogous to the ones for triad chords.

$$\begin{aligned} as_1(s, y) &= \mathbf{1}[y.bass \in s.Notes] \\ as_2(s, y) &= \mathbf{1}[y.3rd \in s.Notes] \\ as_3(s, y) &= \mathbf{1}[y.6th \in s.Notes] \\ as_4(s, y) &= \mathbf{1}[y.FG \wedge y.5th \in s.Notes] \end{aligned}$$

The duration-weighted versions are as follows:

$$as_5(s, y) = \frac{\sum_{n \in s.Notes} \mathbf{1}[n = y.bass] * n.len}{\sum_{n \in s.Notes} n.len}$$

$$as_6(s, y) = \mathbf{1}[y.FG] * \frac{\sum_{n \in s.Notes} \mathbf{1}[n = y.5th] * n.len}{\sum_{n \in s.Notes} n.len}$$

As before, we replace  $n.len$  with  $n.acc$  to obtain the accent-weighted versions of  $as_5$  and  $as_6$ . We also define segment-based duration-weighted features:

$$as_7(s, y) = \frac{\sum_{e \in s.Events} \mathbf{1}[y.bass \in e] * e.len}{\sum_{e \in s.Events} e.len}$$

### C.2.2 Chord Coverage for Suspended and Power Chords

As before, we define the features in this section to be non-zero only for suspended or power chord labels by implicitly multiplying them with an indicator function  $y.SP$ .

$$y.SP = \mathbf{1}[y.sus2 \vee y.sus4 \vee y.7sus4 \vee y.pow]$$

The coverage features for suspended and power chords are also similar to the ones defined for triad chords.

$$\begin{aligned} sp_1(s, y) &= \mathbf{1}[y.root \in s.Notes] \\ sp_2(s, y) &= \mathbf{1}[y.sus2 \wedge y.2nd \in s.Notes \vee \\ &\quad (y.sus4 \vee y.7sus4) \wedge y.4th \in s.Notes] \\ sp_3(s, y) &= \mathbf{1}[y.5th \in s.Notes] \\ sp_4(s, y) &= \mathbf{1}[y.7sus4 \wedge y.7th \in s.Notes] \\ sp_5(s, y) &= \mathbf{1}[y.7sus4 \wedge y.7th \notin s.Notes] \end{aligned}$$

$$\begin{aligned}
alen(s, y) &= \sum_{n \in s.Notes} \mathbf{1}[n = y.7th] * n.len \\
rlen(s, y) &= \sum_{n \in s.Notes} \mathbf{1}[n = y.root] * n.len \\
sp_6(s, y) &= \mathbf{1}[y.7sus4 \wedge alen(s, y) > rlen(s, y)]
\end{aligned}$$

The duration-weighted versions are as follows:

$$\begin{aligned}
sp_7(s, y) &= \frac{rlen(s, y)}{\sum_{n \in s.Notes} n.len} \\
sp_8(s, y) &= \mathbf{1}[y.7sus4] * \frac{alen(s, y)}{\sum_{n \in s.Notes} n.len}
\end{aligned}$$

We also define accent-weighted versions of  $sp_7$  and  $sp_8$ , as well as segment-based duration-weighted features:

$$sp_9(s, y) = \frac{\sum_{e \in s.Events} \mathbf{1}[y.root \in e] * e.len}{\sum_{e \in s.Events} e.len}$$

### C.2.3 Figuration-Controlled Chord Coverage

For each chord coverage feature, we create a figuration-controlled version that ignores notes that were heuristically detected as figuration, i.e. replace  $s.Notes$  with  $s.NonFig(y)$  in each feature formula.

### C.3 Bass

The bass note provides the foundation for the harmony of a musical segment. For a correct segment, its bass note often matches the root of its chord label. If the bass note instead matches the chord's third or fifth, or is an added dissonance, this may indicate that the chord is inverted. Thus, comparing the bass note with the chord tones can provide useful features for determining whether a segment is compatible with a chord label. As in Appendix C.2, we implicitly multiply each of these features with  $y.Triad$  so that they are non-zero only for triads and chords with added notes.

There are multiple ways to define the bass note of a segment  $s$ . One possible definition is the lowest note of the first event in the segment, i.e.  $s.e_1.bass$ . Comparing it with the root, third, fifth, and added tones of a chord results in the following features:

$$\begin{aligned}
f_{20}(s, y) &= \mathbf{1}[s.e_1.bass = y.root] \\
f_{21}(s, y) &= \mathbf{1}[s.e_1.bass = y.third] \\
f_{22}(s, y) &= \mathbf{1}[s.e_1.bass = y.fifth] \\
f_{23}(s, y) &= \mathbf{1}[\exists y.added \wedge s.e_1.bass = y.added]
\end{aligned}$$

An alternative definition of the bass note of a segment is the lowest note in the entire segment, i.e.

$\min_{e \in s.E} e.bass$ . The corresponding features will be:

$$\begin{aligned}
f_{24}(s, y) &= \mathbf{1}[y.root = \min_{e \in s.E} e.bass] \\
f_{25}(s, y) &= \mathbf{1}[y.third = \min_{e \in s.E} e.bass] \\
f_{26}(s, y) &= \mathbf{1}[y.fifth = \min_{e \in s.E} e.bass] \\
f_{27}(s, y) &= \mathbf{1}[\exists y.added \wedge y.added = \min_{e \in s.E} e.bass]
\end{aligned}$$

The duration-weighted version of the bass features weigh each chord tone by the time it is used as the lowest note in each segment event, normalized by the duration of the bass notes in all the events. For the root, the feature is computed as shown below:

$$f_{28}(s, y) = \frac{\sum_{e \in s.Events} \mathbf{1}[e.bass = y.root] * e.len}{\sum_{e \in s.Events} e.len}$$

Similar features  $f_{29}$  and  $f_{30}$  are computed for the third and the fifth. The duration-weighted feature for the added tone is computed as follows:

$$f_{31}(s, y) = \frac{\mathbf{1}[\exists y.added] * \sum_{e \in s.E} \mathbf{1}[e.bass = y.added] * e.len}{\sum_{e \in s.E} e.len}$$

The corresponding accent-weighted features  $f_{32}$ ,  $f_{33}$ ,  $f_{34}$ , and  $f_{35}$  are computed in a similar way, by replacing the bass duration  $e.bass.len$  in the duration-weighted formulas with the note accent value  $e.bass.acc$ .

All duration-weighted and accent-weighted features are discretized using the default bin set  $\mathcal{B}$ .

#### C.3.1 Bass Features for Augmented 6th Chords

Similar to the chord coverage features in Appendix C.2.1, we assume that the indicator  $y.AS$  is multiplied into all features in this section, which means they are non-zero only for augmented 6th chords.

$$\begin{aligned}
as_8(s, y) &= \mathbf{1}[s.e_1.bass = y.bass] \\
as_9(s, y) &= \mathbf{1}[s.e_1.bass = y.3rd] \\
as_{10}(s, y) &= \mathbf{1}[s.e_1.bass = y.6th] \\
as_{11}(s, y) &= \mathbf{1}[(y.fr6 \vee y.ger6) \wedge s.e_1.bass = y.5th] \\
as_{12}(s, y) &= \mathbf{1}[y.bass = \min_{e \in s.E} e.bass] \\
as_{13}(s, y) &= \mathbf{1}[y.3rd = \min_{e \in s.E} e.bass] \\
as_{14}(s, y) &= \mathbf{1}[y.6th = \min_{e \in s.E} e.bass] \\
as_{15}(s, y) &= \mathbf{1}[y.FG \wedge y.5th = \min_{e \in s.E} e.bass]
\end{aligned}$$

We define the following duration-weighted version for the augmented sixth bass and fifth.

$$\begin{aligned}
as_{16}(s, y) &= \frac{\sum_{e \in s.E} \mathbf{1}[e.bass = y.bass] * e.len}{\sum_{e \in s.E} e.len} \\
as_{17}(s, y) &= \mathbf{1}[y.FG] * \frac{\sum_{e \in s.E} \mathbf{1}[e.bass = y.5th] * e.len}{\sum_{e \in s.E} e.len}
\end{aligned}$$

### C.3.2 Bass Features for Suspended and Power Chords

The indicator  $y.SP$  is multiplied into all features in this section like in Appendix C.2.2, meaning they are non-zero only for suspended and power chords.

$$\begin{aligned}
sp_{10}(s, y) &= \mathbf{1}[s.e_1.bass = y.root] \\
sp_{11}(s, y) &= \mathbf{1}[y.sus2 \wedge s.e_1.bass = y.2nd \vee \\
&\quad (y.sus4 \vee y.7sus4) \wedge s.e_1.bass = y.4th] \\
sp_{12}(s, y) &= \mathbf{1}[s.e_1.bass = y.5th] \\
sp_{13}(s, y) &= \mathbf{1}[y.7sus4 \wedge s.e_1.bass = y.7th] \\
\\
sp_{14}(s, y) &= \mathbf{1}[y.root = \min_{e \in s.E} e.bass] \\
sp_{15}(s, y) &= \mathbf{1}[y.sus2 \wedge y.2nd = \min_{e \in s.E} e.bass \vee \\
&\quad (y.sus4 \vee y.7sus4) \wedge y.4th = \min_{e \in s.E} e.bass] \\
sp_{16}(s, y) &= \mathbf{1}[y.5th = \min_{e \in s.E} e.bass] \\
sp_{17}(s, y) &= \mathbf{1}[y.7sus4 \wedge y.7th = \min_{e \in s.E} e.bass]
\end{aligned}$$

The duration-weighted version for the root and seventh are computed as follows:

$$\begin{aligned}
sp_{18}(s, y) &= \frac{\sum_{e \in s.E} \mathbf{1}[e.bass = y.root] * e.len}{\sum_{e \in s.E} e.len} \\
sp_{19}(s, y) &= \mathbf{1}[y.7sus4] * \frac{\sum_{e \in s.E} \mathbf{1}[e.bass = y.7th] * e.len}{\sum_{e \in s.E} e.len}
\end{aligned}$$

### C.3.3 Figuration-Controlled Bass

For each bass feature, we create a figuration-controlled version that ignores event bass notes that were heuristically detected as figuration, i.e. replace  $e \in s.Events$  with  $e \in s.Events \wedge e.bass \notin s.Fig(y)$  in each feature formula.

### C.4 Chord Bigrams

The arrangement of chords in chord progressions is an important component of *harmonic syntax* (Aldwell et al., 2011). A first-order semi-Markov CRF model can capture chord sequencing information only through the chord labels  $y$  and  $y'$  of the current and previous segment. To obtain features that generalize to unseen chord sequences, we follow Radicioni and Esposito (2010) and create chord bigram features using only the *mode*, the *added* note, and the interval in semitones between the roots of the two chords. We define the possible modes of a chord label as follows:

$$\begin{aligned}
\mathcal{M} &= \{\text{maj}, \text{min}, \text{dim}\} \\
&\cup \{\text{it6}, \text{fr6}, \text{ger6}\} \\
&\cup \{\text{sus2}, \text{sus4}, \text{7sus4}, \text{pow}\}
\end{aligned}$$

Other than the common major (maj), minor (min), and diminished (dim) modes, the following chord types have been included in  $\mathcal{M}$  as modes:

- Augmented 6th chords: Italian 6th (it6), French 6th (fr6), and German 6th (ger6).
- Suspended chords: suspended second (sus2), suspended fourth (sus4), dominant seventh suspended fourth (7sus4).
- Power (pow) chords.

Correspondingly, the chord bigrams can be generated using the feature template below:

$$\begin{aligned}
g_1(y, y') &= \mathbf{1}[(y.mode, y'.mode) \in \mathcal{M} \times \mathcal{M} \\
&\quad \wedge (y.added, y'.added) \in \{\emptyset, 4, 6, 7\} \times \{\emptyset, 4, 6, 7\} \\
&\quad \wedge |y.root - y'.root| \in \{0, 1, \dots, 11\}]
\end{aligned}$$

Note that  $y.root$  is replaced with  $y.bass$  for augmented 6th chords. Additionally,  $y.added$  is always none ( $\emptyset$ ) for augmented 6th, suspended, and power chords. Thus,  $g_1(y, y')$  is a feature template that can generate  $(3 \text{ triad modes} \times 4 \text{ added} + 3 \text{ aug6 modes} + 3 \text{ sus modes} + 1 \text{ pow mode})^2 \times 12 \text{ intervals} = 4,332$  distinct features. To reduce the number of features, we use only the  $(mode.added)-(mode.added)-interval$  combinations that appear in the manually annotated chord bigrams from the training data.

### C.5 Chord Changes and Metrical Accent

In general, repeating a chord creates very little accent, whereas changing a chord tends to attract an accent (Aldwell et al., 2011). Although conflict between meter and harmony is an important compositional resource, in general chord changes support the meter. Correspondingly, a new feature is defined as the accent value of the first event in a candidate segment:

$$f_{36}(s, y) = s.e_1.acc$$

## References

- Aldwell, E., Schachter, C., and Cadwallader, A. (2011). *Harmony and Voice Leading*. Schirmer, 4th edition.
- Denyer, R. (1992). *The Guitar Handbook: A Unique Source Book for the Guitar Player*. Knopf.
- Radicioni, D. P. and Esposito, R. (2010). BREVE: an HMPerptron-based chord recognition system. In *Advances in Music Information Retrieval*, pages 143–164. Springer Berlin Heidelberg.
- Taylor, E. (1989). *The AB Guide to Music Theory Part 1*. ABRSM.