

Representing Antônio Carlos Jobim's Harmonic Progressions

Robert Willey¹, Giordano Cabral²

¹School of Music
University of Louisiana at Lafayette

²Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie

willey@louisiana.edu, Giordano.Cabral@lip6.fr

Abstract. *A large number of songs composed by Antônio Carlos Jobim were transcribed in preparation for statistical analysis, creating a data base representing chord progressions as well as general information regarding each song. We released this database for the scientific community with the aim of encouraging research projects on the subject. This paper describes this database in order to assist researchers interested in using it. We also present some applications we developed to illustrate its potential, and make a comparison with a similar database describing the songs by the Beatles.*

1. Motivation

Antônio Carlos Jobim (Willey 2005) is widely considered to be one of the most important songwriters of the 20th century. We are interested in objectively analyzing his work with the aim of making some interesting observations about Jobim's methods. The publishing of authoritative scores created an opportunity to develop a database of chord symbols representing the harmonic progressions in his music for subsequent statistical analysis. A previous study by (Johansson 1999) of the songs by the Beatles' (Fujita 1993) allowed comparison with mainstream popular music and served as a model for our study.

Apart from allowing many types of study related to Jobim's music, we believe this kind of material can be helpful to the academic community if used as a benchmark for MIR techniques (Downie et al 2005). We think benchmarks are crucial to accelerate scientific progress, especially in the computer music domain.

2. Chord symbols

The chord progressions for this study were obtained from the Cancioneiro Jobim collection (Jobim, 2001), approximately 250 songs in five volumes arranged chronologically of Jobim's complete recorded works, written for piano by Antônio Carlos Jobim, Paulo Jobim, and a group of arrangers under their supervision.

The Jobims' goal was to be as accurate as possible in the notation of melody, accompaniment, and counter melody in the piano scores. "The rule was to transcribe the arrangements and/or Jobim's piano recorded performance as exactly as possible. That means we did not write arrangements, but transcriptions...We were also given the choice of omitting from the chord symbol a high interval when it occurred in the

melody. If a melody note *E* is harmonized by a *G* dominant chord, we could write just *G9* or add the 13th too” (Mauro, 2006).

2.1. Handling Errors

The first phase of the data entry process was to compare the chord symbols with the notes of the piano part, and some mistakes were found, as would be expected in any publication of over 1100 pages. Any alterations to the chord symbols were noted in comments in the database.

2.2. Conversion

Other discrepancies between the chord symbols and the piano parts were believed to be the result of a desire by the arranger to simplify the notation. In these cases, since one of the intended applications of the database was to be the analysis of harmonic progressions, the chords when entered were converted to be more meaningful from a functional harmony perspective. For example, *Edim/B* leading to a *Bm7* in the scores was replaced with *A#dim/B* to reflect the function of the first chord as a leading tone diminished chord. In some of the introductions the chord symbols were missing, because there were a number of different introductions on the recordings (Cardim, 2006) or because a guitarist was likely to wait for the main body of the song to begin. In these cases chord symbols were created for the database. The convention in popular music scores is to indicate a chord symbol only when the harmony changes. However, at times one appears to have been omitted because there is no standard chord symbol to use, as when the piano part gradually dissolves into a single note or octave. These were indicated in the score with an “x”, and rests by the letter “r”.

Slash chords, that is with “/” separating a chord in the treble from a bass note in the bass, were entered as such in the database. It is up to the analysis software to decide how to interpret them. Depending on the goal of the analysis, a chord symbol such as *G/C* could be replaced by *G*, since it is just an inversion, and *F/G* converted to *Gsus7(9)*. Other combinations from the scores, such as *F#7/D* are more difficult to reduce. In places in the score where the music is more dissonant one senses the arranger’s dilemma in consolidating the piano part into a standard chord symbol, and when no alternative could be found making more sense from a theoretical perspective the chord symbol was accepted and entered into the database as written.

Over time Jobim began to use more polychords in his compositions, indicated by one chord symbol over another, separated by a horizontal line, as in the case of *Eb* over *C#m7*, for which *Eb_C#m7* was be entered in the database. Songs lacking chord symbols altogether were omitted from the database.

3. Data format

The database was initially written in five Microsoft Word files corresponding to the volumes of the Cancioneiro series, with songs represented in tables, one measure per cell, and one row per score staff line. These files were subsequently converted into simple text files, to facilitate the writing and reading software routines. These text files follow the format described in next section. Examples from the data are available online, along with other materials from the study (<http://willshare.com/jobim>). The data from the five volumes can be obtained from the authors for research purposes.

3.1 BNF Description

The figure below describes the database file format in BNF notation (Knuth 1964). Each file encloses a number of songs, each one having a header and a body. The header contains general information such as song title, key, time signature, volume, composer, etc. The body contains its chord progression. The chord progression in the body presents chords and key changes. The chord type is particularly flexible, allowing different syntaxes. The user can edit this particular field, in order to specify precise spellings.

```

<File>::=
  <Song>*

<Song>::=
  <Header>
  <Body>

<Header>::=
  <Header_Field>*
  "Page"

<Header_Field>::=
  <Title_Field> |
  <Tempo_Field> |
  <Comment_Field> |
  <Key_Field> |
  <Time_Field> |
  <Composer_Field> |
  <Arranged_Field> |
  <Volume_Field> |
  <Page_Field> |
  <NumBars_Field>

<Title_Field>::=
  "title"
  <Name>

<Tempo_Field>::=
  "tempo"
  <Name>

<Comment_Field>::=
  "comment"
  <Name>

<Key>::=
  "key"
  <Root>
  <Mode>

<Time_Field>::=
  "time"
  <Number>/"<Number>

<Composer_Field>::=
  "composer"
  <Name>

<Arranged_Field>::=
  "arranged"
  <Name>

<Volume_Field>::=
  "volume"
  <Number>

<Page_Field>::=
  "page"
  <Number>

<NumBars_Field>::=
  "bars"
  <Number>

<Root>::=
  <Note>

<Mode>::=
  "major" |
  "minor" |

<Note>::=
  <BaseNote> |

<BaseNote>::=
  "A" |
  "B" |
  "C" |
  "D" |
  "E" |
  "F" |
  "G"

<Accident>::=
  "#" |
  "b"

<event>::=
  <Chord> |
  <Key> |
  <Special>

<Chord>::=
  <Note><ChordType><Inversion>

<ChordType>
  "" |
  "m" |
  "m7" |
  etc...

<Inversion>::=
  "" |
  "/"<Note>

<Special>::=
  "/" |
  "r" |
  "." |
  "x"

```

Figure 1 - BNF Notation of the file format.

Our database explicitly includes repeats, unlike the work of Johansson with the Beatles. In his study, chords were grouped into blocks representing song form sections, such as introductions, verses, and choruses. Since these were not present in Jobim's music, all measures and repeats were expanded and entered in the database. While a field was created in the datafile to indicate the form, it was usually left blank, as Jobim often intentionally avoided standard forms like *AABA*.

4. Experiments

We performed some experiments which illustrate the potential of use of this database, not only to the analysis of music but also to the creation of interactive musical systems, and composition tools.

The first and main interest in creating the database was to be able to analyze Jobim's harmony. The tool developed to make this analysis (HarmIn) is described in (Cabral and Willey 2007). It converted all chords to one of six possible classes. The classification procedure is a way of interpreting the database. It is necessary for any analysis study, but the definition of how the classification will take place is left to the user. The complete results will appear in a future publication, but some of them are anticipated in this paper.

Figures 2a and 2b show graphs made of the frequency chords appear in songs in major keys made by Johansson of Beatles songs, followed by those of Jobim. Following Johansson’s idea, all the songs were transposed to *C* major or *C* minor. The predominance of the *ii* chord in Jobim compared with the *IV* chord in the Beatles comes from a greater use of the standard jazz progression *ii V I*. Also apparent is Jobim’s richer harmonic palette, seen in the greater variety of frequently used chords.

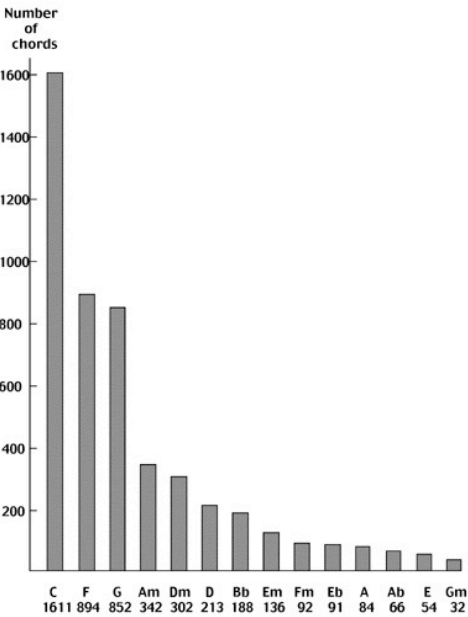


Figure 2a. Johansson’s chord distribution data for Beatles’ songs in major

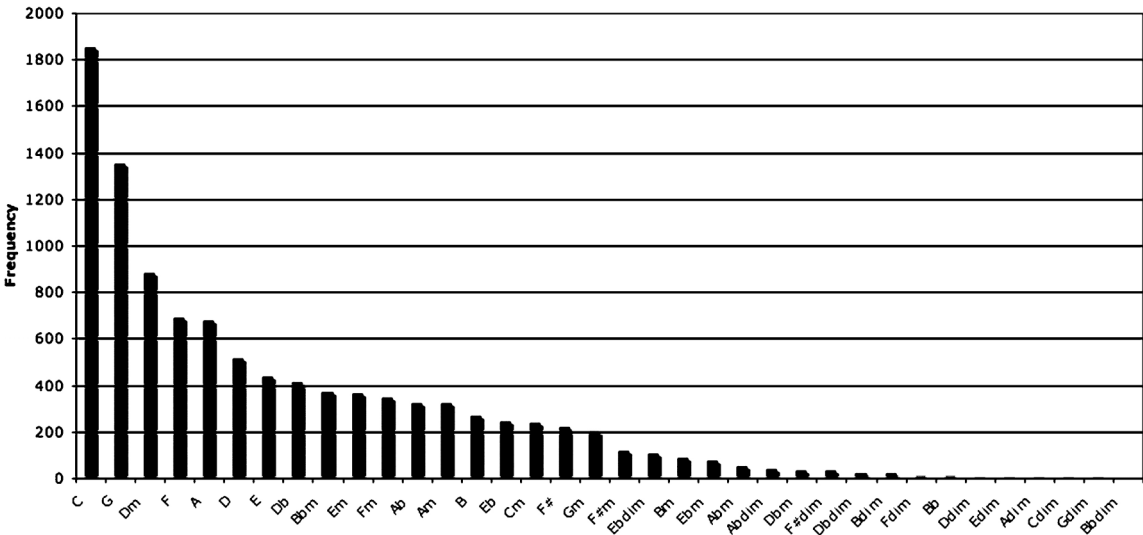
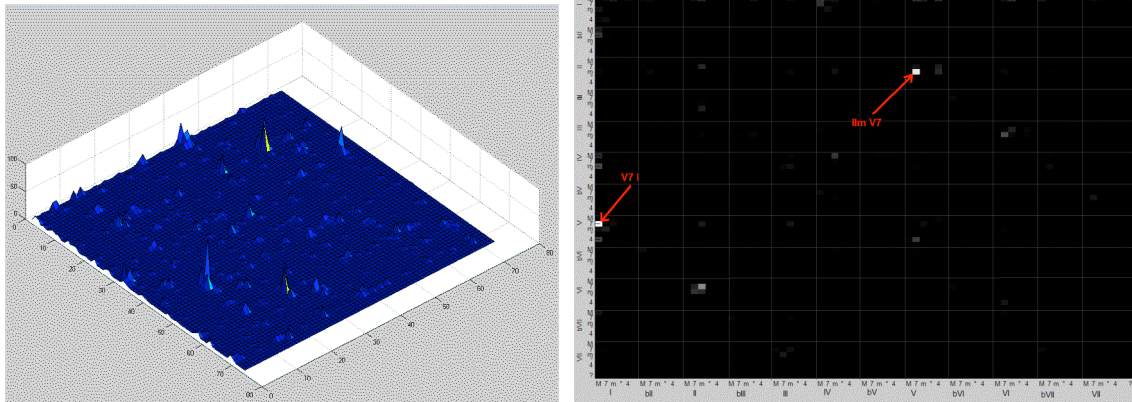


Figure 2b. Chord distribution data for Jobim’s songs in major

Figures 3 and 4 show the frequency of the chord transitions from the first to the second in a pair of chords. The graph on the left represents the first chord as the row and the second chord as the column. Higher points (in Figure 3) and brighter pixels (in Figure

4) indicate a higher incidence of that transition. For example, the pixels representing the transitions from *ii* to *V7* and from *V7* to *I* are clearly brighter, as annotated in Figure 4.



Figures 3 and 4. Distribution of chord transitions.

Figure 5 also shows the frequency of chord transitions, but in a more legible way as one can also follow the sequence of transitions. The little circles represent the chords, and the lines connecting them represent the transitions.

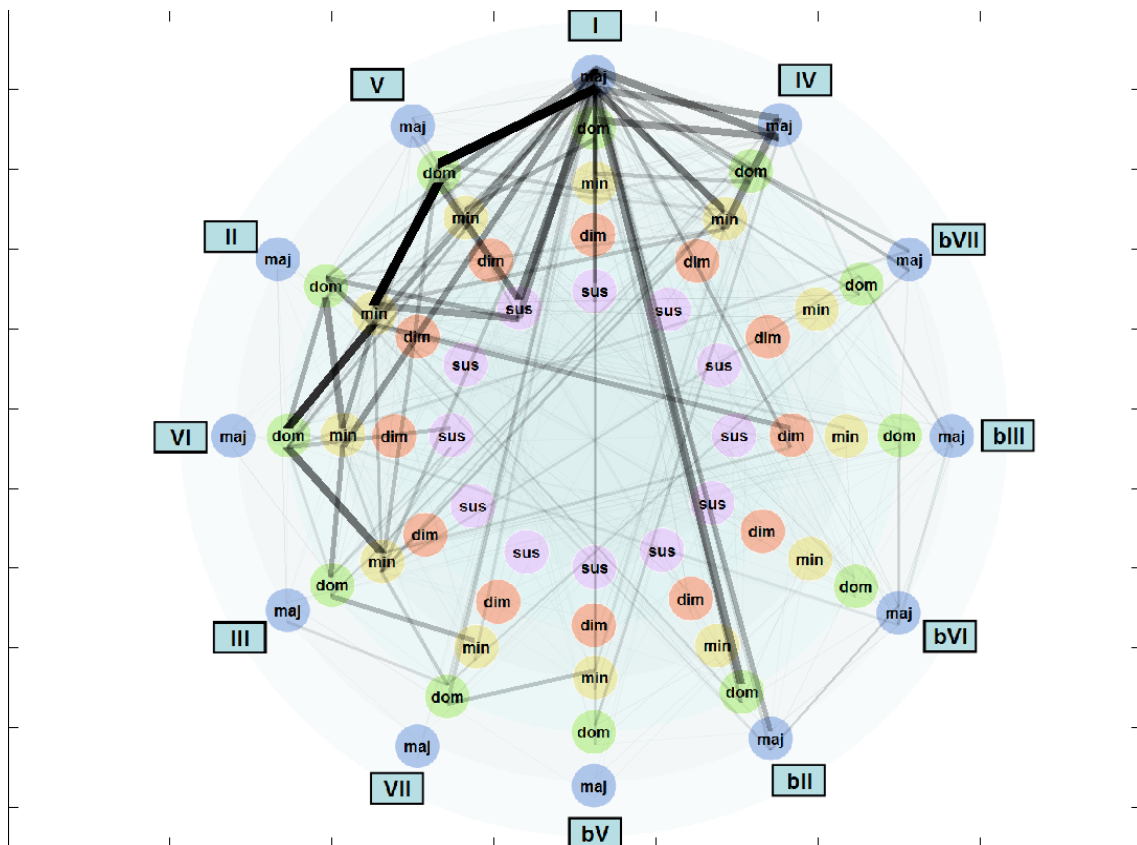


Figure 5. Harmin chord transition graph.

The chord roots are displayed following the circle of fifths, so that the final graph generally becomes simpler. The quality of the chord varies according to its distance from the center, and has a different color (in the case illustrated in Figure 6, major=blue, dominant=green, minor=yellow, diminished=salmon, suspended=purple). The thickness

and transparency of the connecting lines indicate the frequency of the chord transitions. Progressions that occurred less than 1% of the time were omitted to make the graph more readable. The direction of the transition is indicated by the position of a line's vertices inside the chord type circles. Each line starts in the upper part of one circle and ends in the lower part of another. For example, in Figure 5, the frequent transition from *V7* to *I* is indicated by a thick line starting above the label (*V*)*dom* and ending below the label (*I*)*maj*.

We can gather non-visual information about the database as well. For example, the standard deviation of the chords, the most significant clusters and the most frequent subsequences. The latter can be seen in Figure 6. Note that sequences as *V7_I* and *ii_V7* occur more often than individual chords as *IV* and *iii*. Also, the chord *VI*, only appears after a long list of chord sequences.

I	1530	IIIm_VI7	151	IIm_V4	90	Im_I	57
IIm	932	bIII	146	II7_IIm	89	bVII_I	57
V7	867	IVm_I	144	IV_I	87	Vm_I	56
V7_I	483	V4_I	130	I_IIm	87	I_VIm_IIm	56
VI7	464	I_IV	121	VIm_IIm	84	VI4_VI7	56
IIm_V7	437	bII7_I	115	III7_VIm	83	II7_V7	54
IVm	418	V4_V7	114	VI7_II7	81	I_V4_I	54
IV	397	bII7	113	I_IV_I	80	V7_IIm	54
II7	393	VI7_IIm_V7	111	II7_V4	74	I_IVm	54
IIIm	388	IV_IVm	108	bII_I	72	I_bVII_I	53
VIm	334	IIIm_VI7_IIm	104	I_V7_I	70	bVIm	53
V4	334	I_VIm	103	bIII_IIm	70	IV_I_IV	52
VI7_IIm	259	VIm_II7	103	VIm_II7	68	I_IV_I_IV	52
IIm_V7_I	243	I_V7	101	V4_V7_I	66	V7_IIm_V7	50
III7	225	I7_IV	101	V7_IIm	64	IIm_V7_IIm	50
bVI	184	VI4	96	VI	63		
bII	168	III	96	I_VI7	59		
bVII7	162	I_V4	94	VI7_IIm_V7_I	58		

Figure 6. Most frequent subsequences.

The database was also integrated in the interactive system Ramin (Cabral 2005), which used this information to perform real time accompaniment, as well as to create new harmonies based on the learned transition data. Figure 7 shows one possible output.

Dm7(9) / G7(13) / Dm7(9) /	C#7(b13) / Bm7 / Bm / Bm7 /	Dm7(9) / / /
G7(13) / Dm7(9) / G7(13) /	E7(13) / Em7 / A7 A6(9) Dm7	G7/B / Gm7/Bb / Am7 / / /
Dm7(9) / G7(13) /	/	Am7 D7 Am7 D7 Am7 D7 Am7 D7
	/ / Gm7(9) / C7 / Gm7 /	
C7M(9) / G(add9) / G/A Em7	A7/E / Dm7(9) / / /	Dm7 G7(13) Dm7 G7(13) C / D
F#m7 Em6/G G#m7(11) / C#7(9)	2X	/ D / C#7 / D / Bm7 /
/ G#m7 / C#7 /		Em7 / A7 A6(9) D6(9) / / /

Figure 7. Progressions based on the harmony learned from Jobim's database.

One additional way of using this database is to make a comparative analysis based on the information contained in the songs' headers. In other words, we can divide the songs according to some characteristic, and compare each sub dataset. In our study, we used the volume information (which indicates the period when the song was composed) to split the database into 5 datasets in order to see the evolution of Jobim's harmonies over the course of his career. This result, as well expanded color graphics and other results from the project is available at <http://willshare.com/jobim>.

5. Conclusion

The availability of a database containing Jobim's chord progressions has created an opportunity to perform statistical analysis on his music, which we have begun to do from a variety of perspectives. Our progress in this area is described in another paper submitted to the conference.

6. References

- Cabral, Giordano, and Willey, Robert. "Analyzing Harmonic Progressions with HarmIn: the Music of Antônio Carlos Jobim". Submitted to the *SBCM 2007*, São Paulo, 2007.
- Cabral, Giordano. "Comping by Humming : Adaptation x Continuité dans l'Interaction Musicale". Rapport de Pré-Soutenance de Thèse, Paris, 2005.
- Cardim, Alfredo. "Jobim." Internet Communication, 9 August, 2006.
- J. Stephen Downie, Kris West, Andreas Ehmann & Emmanuel Vincent. "The 2005 Music Information Retrieval Evaluation Exchange (MIREX 2005): Preliminary Overview". 6th International Conference on Music Information Retrieval online proceedings, pp. 320-323. <http://ismir2005.ismir.net/proceedings/xxxx.pdf>
- Knuth, Donald E. "Backus Normal Form vs. Backus Naur Form," Communications of the ACM 7(12):735-736, 1964.
- Fujita, Tesuya. The Beatles: Complete Scores, Milwaukee: Hal Leonard Publishing Corporation, 1993.
- Jobim, Antônio Carlos. Cancioneiro Jobim: complete works, arrangements for piano, 5 volumes, Paulo Jobim, coordination. Rio de Janeiro: Jobim Music, 2001.
- Johansson, KG. Beatles harmoniska språk. En beskrivande undersökning av harmoniken i Beatles musik. Musikhögskolan i Piteå, rapportserie, Luleå Tekniska Universitet: 1998:01
- Johansson, KG. "The Harmonic Language of the Beatles", STM-Online Vol. 2, 1999. http://www.musik.uu.se/ssm/stmonline/vol_2_1/KGJO/index.html
- Mauro, Jr., Haroldo. "Cancioneiro Jobim." Internet Communication, 12 August, 2006.
- Willey, Robert. "Antônio Carlos ('Tom') Jobim", Encyclopedia of Recorded Sound in the United States, edited by Frank Hoffmann. New York: Routledge, 2005.