

Traitement de la polyphonie pour l'analyse informatique de partitions musicales

Nicolas Guiomard-Kagan

PhD Defense – March 20th 2017


Music can be listened.



And music can also be read.

The image shows a musical score for the song "Le petit cheval" by Georges Brassens. It consists of three staves. The top staff is the vocal line in treble clef with a common time signature (C). The lyrics are written below the notes: "Le p'tit ch'val dans le mau- vais temps Qu'il avait donc du cou-ra - ge". The middle and bottom staves are for piano accompaniment, with the middle staff in treble clef and the bottom staff in bass clef, both in common time. The piano part features a rhythmic accompaniment with chords and moving lines in both hands.

Figure: Extract of the song *Le petit cheval* - Georges Brassens



Music analysis?

Music Analysis: Everyone

- ▶ Extract a melody
- ▶ Recognize patterns, variations, a "covers" of a song
- ▶ Recognize the shape of a piece (ex: Segment a pop song in verse/chorus)

Music Analysis: Music theorists

Expert in music analysis:

- ▶ Context
- ▶ Shape
- ▶ Structure

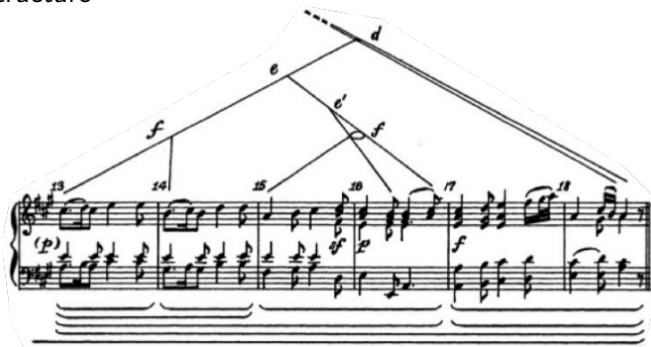


Figure: Arborescent analysis - Lerdhal, Jackendoff - 1983

Music Analysis: Computers

An idea that goes back to the premise of computing.



"[...] the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent"

Ada Lovelace - 1842

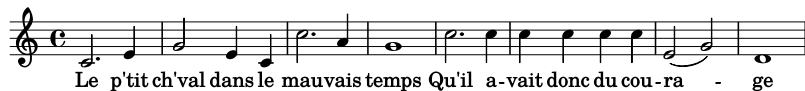
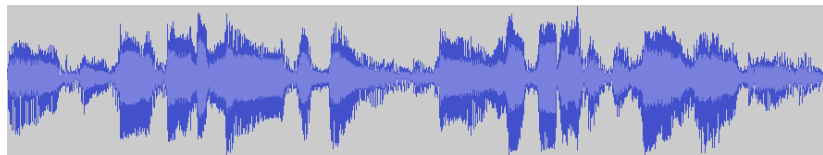
Can computer analyze music?

Music Analysis: Computers

What kind of sources for computer?

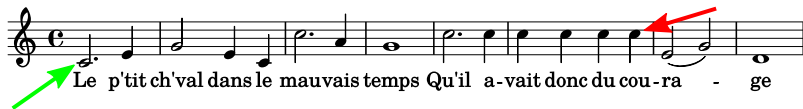
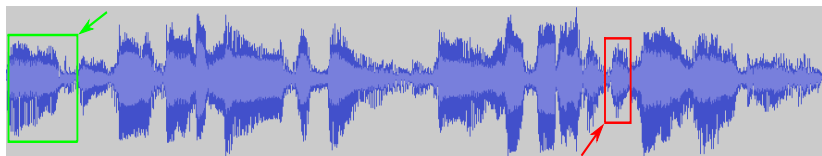
Music Analysis: Computers

What kind of sources for computer?



Music Analysis: Computers

What kind of sources for computer?



Music Analysis: Computers

What kind of files for computer?

```
!!!COM:Brassens, George
!!!CNT:French
!!!OTL:Le petit cheval
**kern      **kern      **kern      **kern
*staff1     *staff2     *staff2     *staff3
*clefG4     *clefG4     *clefG4     *clefF3
*k[]        *k[]        *k[]        *k[]
*a:         *a:         *a:         *a:
*M4/4       *M4/4       *M4/4       *M4/4
*MM120     *MM120     *MM120     *MM120
=7          =7          =7          =7
2.c         2.cc        4r          4C
.           .           4e          4g
.           .           4r          4C
4e         4ee        4e          4g
=8         =8         =8          =8
2g         2gg        4r          4C
.           .           4g          4c
4e         4ee        4r          4C
4c         4r         4g          4c
=9         =9         =9          =9

0, 0, Header, 1, 4, 1024
1, 0, Start_track
1, 0, SMPTE_offset, 0, 0, 0, 0, 0
1, 0, Time_signature, 4, 2, 24, 8
1, 0, Key_signature, 0, "major"
1, 0, Tempo, 499999
1, 0, Tempo, 999998
1, 61440, End_track
2, 0, Start_track
2, 0, Title_t, « Voice »
2, 0, Program_c, 0, 0
2, 0, Control_c, 0, 7, 101
2, 0, Control_c, 0, 10, 64
2, 0, Note_on_c, 0, 60, 90
2, 3072, Note_off_c, 0, 60, 0
2, 3072, Note_on_c, 0, 64, 90
2, 4096, Note_off_c, 0, 64, 0
2, 4096, Note_on_c, 0, 67, 90
2, 6144, Note_off_c, 0, 67, 0
2, 6144, Note_on_c, 0, 64, 90
2, 7168, Note_off_c, 0, 64, 0
```

****Kern**

MIDI

Figure: Extract of the song *Le petit cheval* - Georges Brassens

Music Analysis: Computers

What kind of files for computer?

```
!!!COM:Brassens, George
!!!CNT:French
!!!OTL:Le petit cheval
**kern      **kern      **kern      **kern
*staff1     *staff2     *staff2     *staff3
*clefG4     *clefG4     *clefG4     *clefF3
*k[]        *k[]        *k[]        *k[]
*a:         *a:         *a:         *a:
*M4/4       *M4/4       *M4/4       *M4/4
*MM120      *MM120      *MM120      *MM120
=7          =7          =7          =7
2.c         .           4r          4C
.           .           4e          4g
.           .           4r          4C
4e         4ee        4e          4g
=8         =8         =8          =8
2g         2gg        4r          4C
.           .           4g          4c
4e         4ee        4r          4C
4c         4r         4g          4c
=9         =9         =9          =9

0, 0, Header, 1, 4, 1024
1, 0, Start_track
1, 0, SMPTE_offset, 0, 0, 0, 0, 0
1, 0, Time_signature, 4, 2, 24, 8
1, 0, Key_signature, 0, "major"
1, 0, Tempo, 499999
1, 0, Tempo, 999998
1, 61440, End_track
2, 0, Start_track
2, 0, Title_t, « Voice »
2, 0, Program_c, 0, 0
2, 0, Control_c, 0, 7, 101
2, 0, Control_c, 0, 10, 64
2, 0, Note_on_c, 0, 60, 90
2, 3072, Note_off_c, 0, 60, 0
2, 3072, Note_on_c, 0, 64, 90
2, 4096, Note_off_c, 0, 64, 0
2, 4096, Note_on_c, 0, 67, 90
2, 6144, Note_off_c, 0, 67, 0
2, 6144, Note_on_c, 0, 64, 90
2, 7168, Note_off_c, 0, 64, 0
```

****Kern**

MIDI

Figure: Extract of the song *Le petit cheval* - Georges Brassens

Outline

State of the art:

1. What means to organize polyphony?
2. Voice and stream

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

Monophony and polyphony

Polyphony: Several sounds can be heard at the same time.



The image displays a musical score for the song "Hey Jude" by Paul McCartney and John Lennon. It consists of three staves. The top staff is a vocal line in treble clef with a key signature of one flat (B-flat) and a common time signature (C). The lyrics are: "Hey Jude don't make it bad. Take a sad song and make it better...". The middle and bottom staves are piano accompaniment. The middle staff is in treble clef and features a series of chords, while the bottom staff is in bass clef and features a rhythmic bass line. The overall texture is polyphonic, with multiple sounds heard simultaneously.

Figure: Extract of the song *Hey Jude* - Paul McCartney and John Lennon

Monophony and polyphony

Polyphony: Several sounds can be heard at the same time.



The image shows a musical score for the song "Hey Jude" by Paul McCartney and John Lennon. It consists of three staves: a vocal line in treble clef, a piano accompaniment in treble clef, and a piano accompaniment in bass clef. The key signature is one flat (B-flat) and the time signature is common time (C). The lyrics are: "Hey Jude don't make it bad. Take a sad song and make it better...". A red rectangular box highlights the first measure of the music, where the vocal line has a quarter note G4, the piano treble part has a chord of F4-A4-C5, and the piano bass part has a quarter note F3. This illustrates polyphony as multiple distinct sounds are heard simultaneously.

Figure: Extract of the song *Hey Jude* - Paul McCartney and John Lennon

Outline

State of the art:

1. What means to organize polyphony?
2. Voice and stream

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

How to organize polyphony?

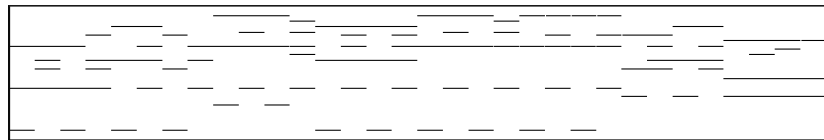


Figure: Extract of the song *Le petit cheval* - Georges Brassens

How to organize polyphony

Le p'tit ch'val dans le mau- vais temps Qu'il avait donc du cou-ra - ge

The image shows a musical score for the song "Le petit cheval" by Georges Brassens. It consists of five staves. The top staff is the vocal line, with lyrics underneath. The second staff is a piano accompaniment. The third and fourth staves are guitar accompaniment, showing a rhythmic pattern of chords. The fifth staff is the bass line. The music is in 4/4 time and G major.

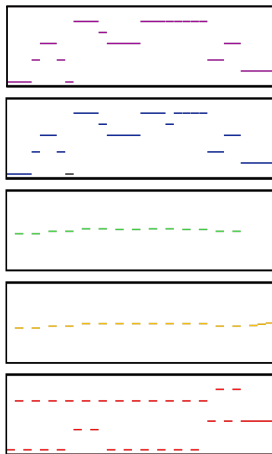


Figure: Extract of the song *Le petit cheval* - Georges Brassens

Outline

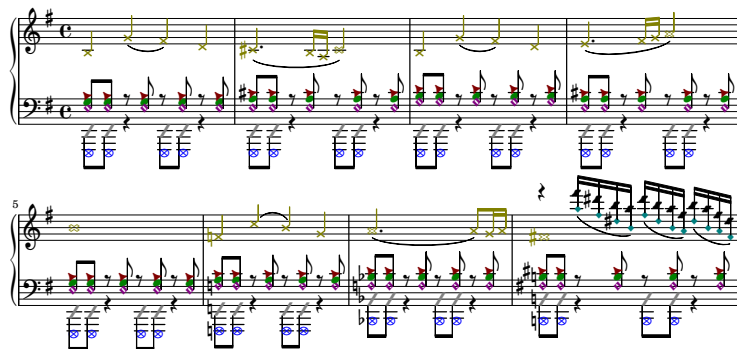
State of the art:

1. What means to organize polyphony?
2. Voice and stream

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

Voice and stream



The image displays a musical score for the 'Theme of Hyrule field' from the game Zelda. The score is presented in two systems, each with a grand staff (treble and bass clefs). The key signature is one sharp (F#) and the time signature is common time (C). The melody is written in the treble clef, and the accompaniment is in the bass clef. The melody consists of a sequence of notes, some of which are marked with yellow 'x' symbols, indicating predicted voices. The accompaniment features a steady rhythm of eighth notes, with some notes marked with red and green symbols. The score is divided into measures, with a measure number '5' indicated at the start of the second system. The overall structure is a simple, melodic theme with a consistent accompaniment.

Figure: Theme of Hyrule field in Zelda

Voice separation:

- ▶ The predicted voices must be monophonic;
- ▶ The predicted number of voices must match the maximum number of notes played simultaneously.

Voice and stream

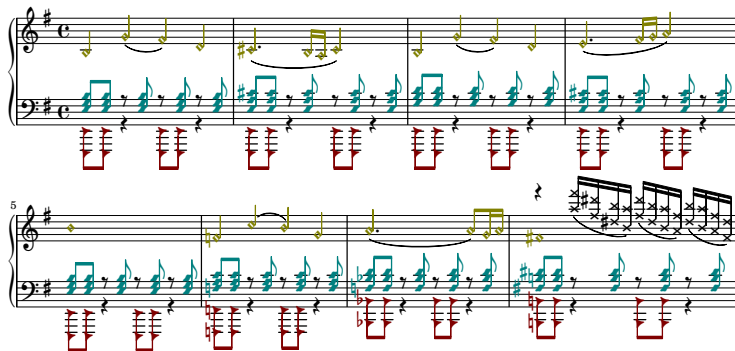


Figure: Theme of Hyrule field in Zelda

Stream separation:

- ▶ The predicted streams can be polyphonic;
- ▶ The size is very variable (few notes to few bars).

Outline

State of the art:

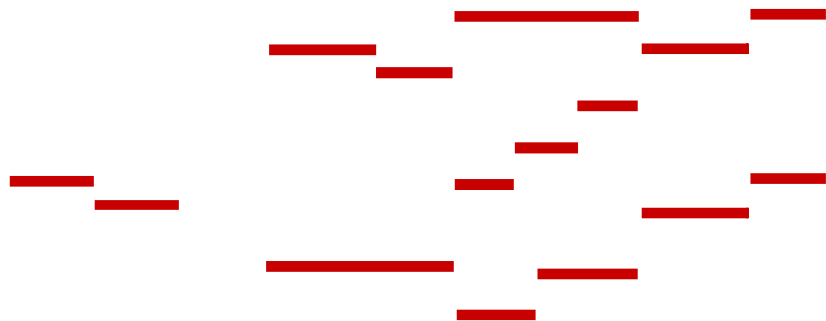
1. What means to organize polyphony?
2. Voice and stream
 - 2.1 Voice separation algorithms
 - 2.2 Stream separation algorithms

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

Voice separation algorithms:

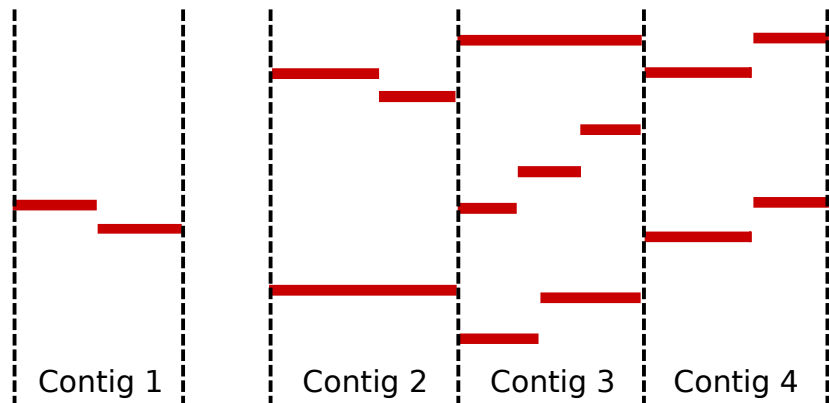
Chew and Wu - 2005



- ▶ Defines **contigs**: The number of notes played simultaneously into a contig is constant.

Voice separation algorithms:

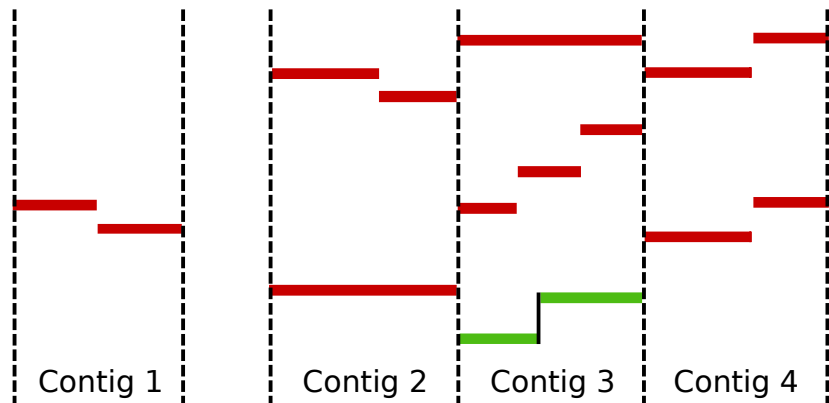
Chew and Wu - 2005



- ▶ Defines **contigs**: The number of notes played simultaneously into a contig is constant.

Voice separation algorithms:

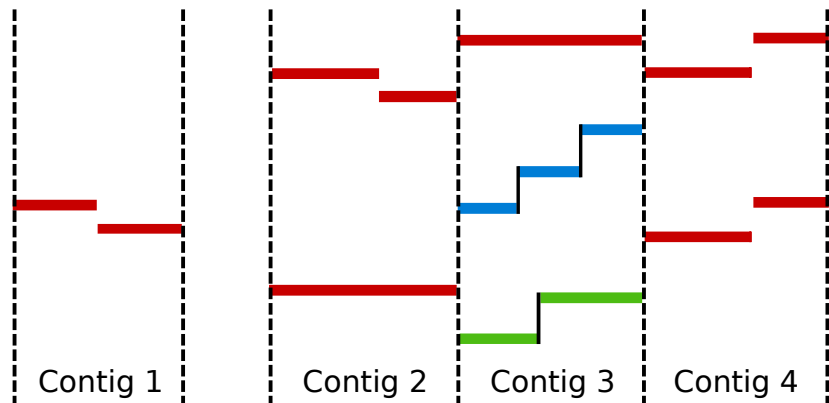
Chew and Wu - 2005



- ▶ Defines **fragments**: The notes are grouped from the highest pitch to the lowest pitch in fragments of voice.

Voice separation algorithms:

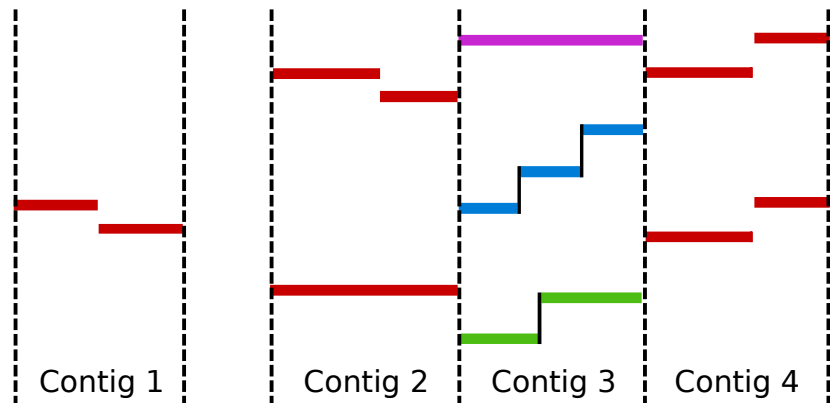
Chew and Wu - 2005



- ▶ Defines **fragments**: The notes are grouped from the highest pitch to the lowest pitch in fragments of voice.

Voice separation algorithms:

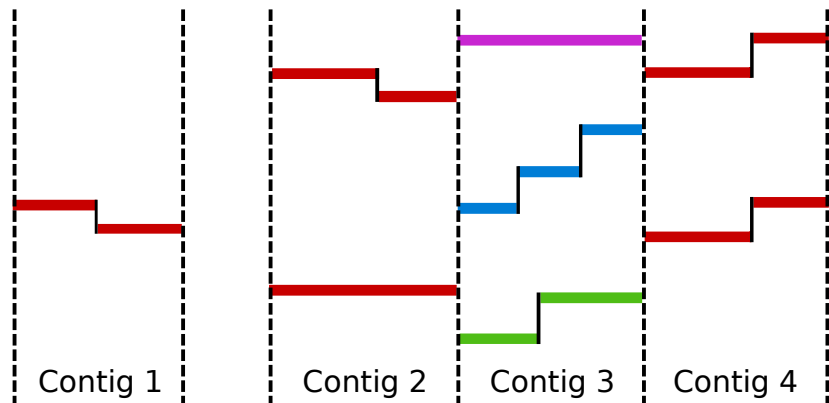
Chew and Wu - 2005



- ▶ Defines **fragments**: The notes are grouped from the highest pitch to the lowest pitch in fragments of voice.

Voice separation algorithms:

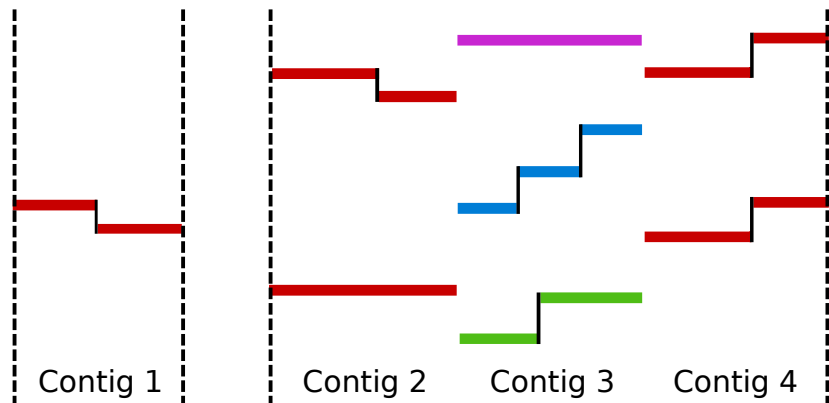
Chew and Wu - 2005



- ▶ Defines fragments links, starts with **maximal contigs**.

Voice separation algorithms:

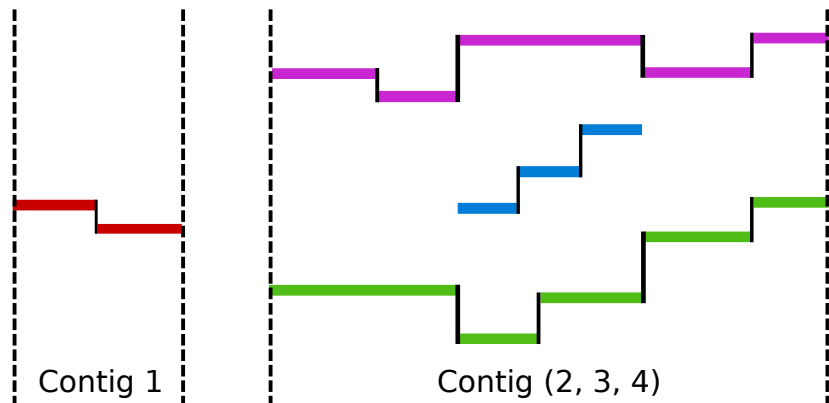
Chew and Wu - 2005



- ▶ Defines fragments links, starts with **maximal contigs**.

Voice separation algorithms:

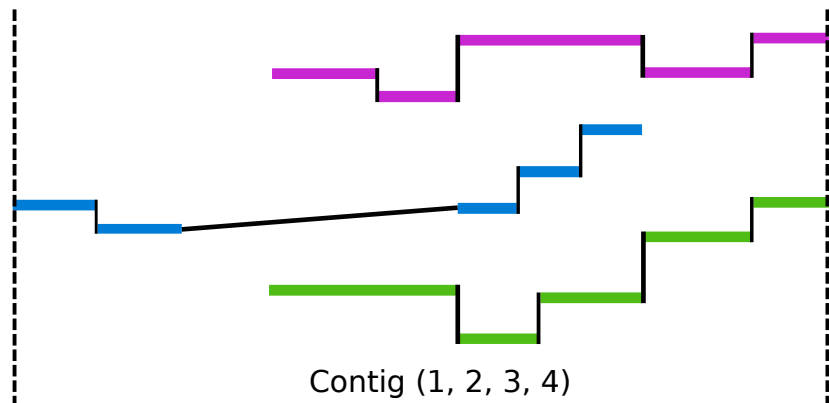
Chew and Wu - 2005



- ▶ Defines fragments links, starts with **maximal contigs**.

Voice separation algorithms:

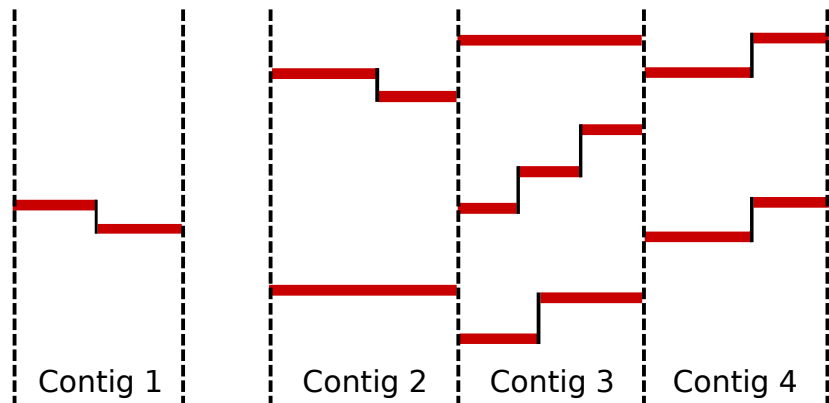
Chew and Wu - 2005



- ▶ Defines fragments links, starts with **maximal contigs**.

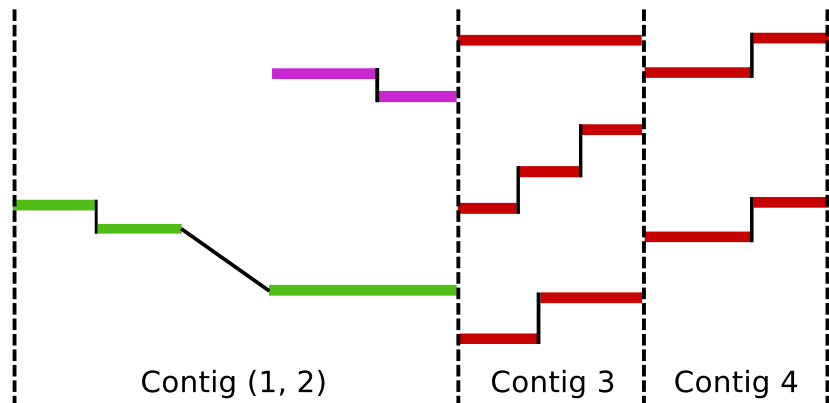
Voice separation algorithms:

Ishigaki, Matsubara and Saito - 2011



- ▶ Starts to merge only adjacent contigs with an increasing number of voices.

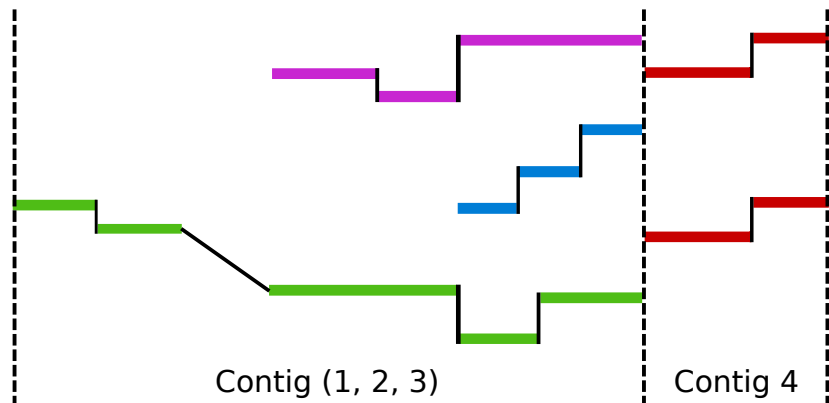
Voice separation algorithms: Ishigaki, Matsubara and Saito - 2011



- ▶ Starts to merge only adjacent contigs with an increasing number of voices.

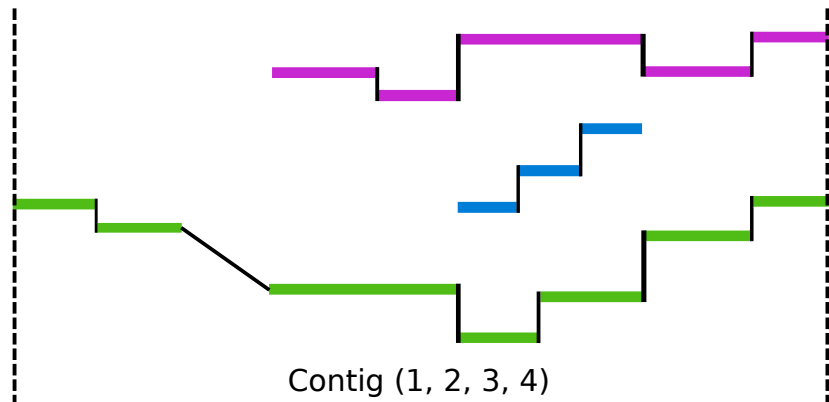
Voice separation algorithms:

Ishigaki, Matsubara and Saito - 2011



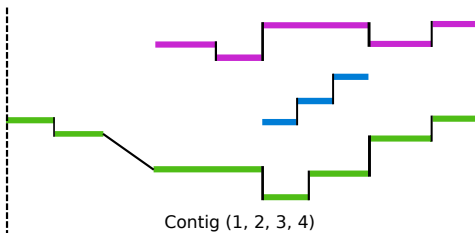
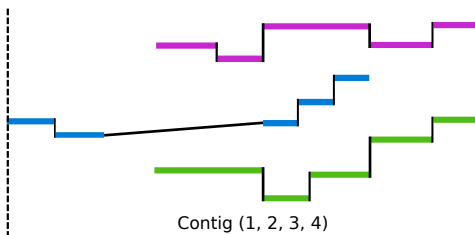
- ▶ Starts to merge only adjacent contigs with an increasing number of voices.

Voice separation algorithms: Ishigaki, Matsubara and Saito - 2011



- ▶ Merge the remaining contigs

Voice separation algorithms



Two different separations only by changing the connection order of the contigs.

Outline

State of the art:

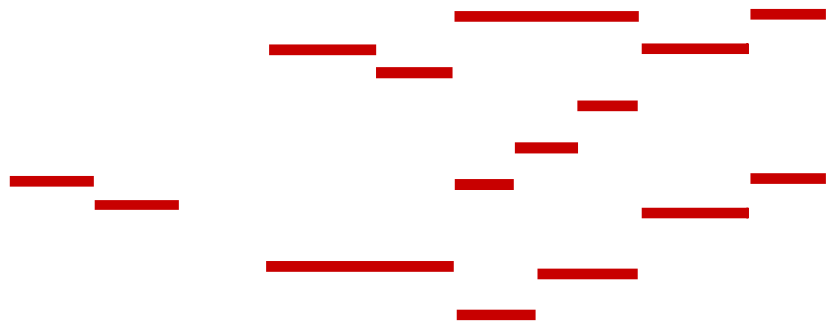
1. What means to organize polyphony?
2. Voice and stream
 - 2.1 Voice separation algorithms
 - 2.2 Stream separation algorithms

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

Stream separation algorithms:

Temperley - 2001

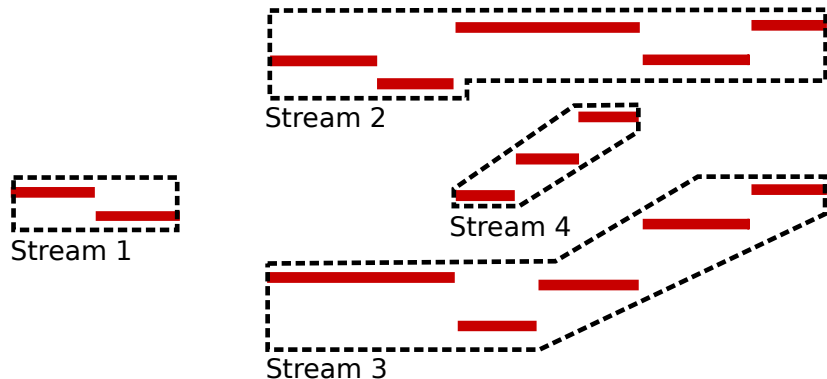


Principles:

- ▶ Monophonic streams
- ▶ Temporal proximity
- ▶ Pitch proximity

Stream separation algorithms:

Temperley - 2001

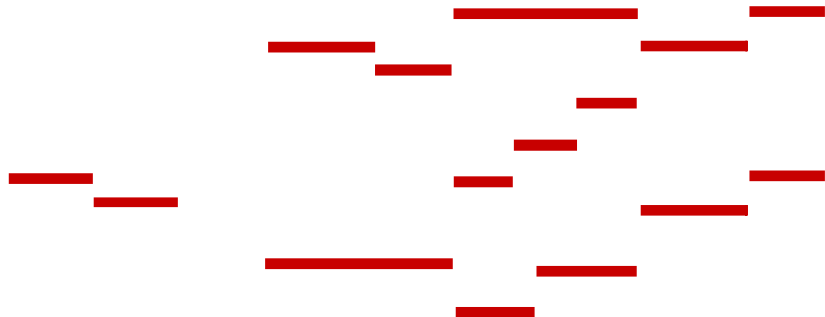


Principles:

- ▶ Monophonic streams
- ▶ Temporal proximity
- ▶ Pitch proximity

Stream separation algorithms:

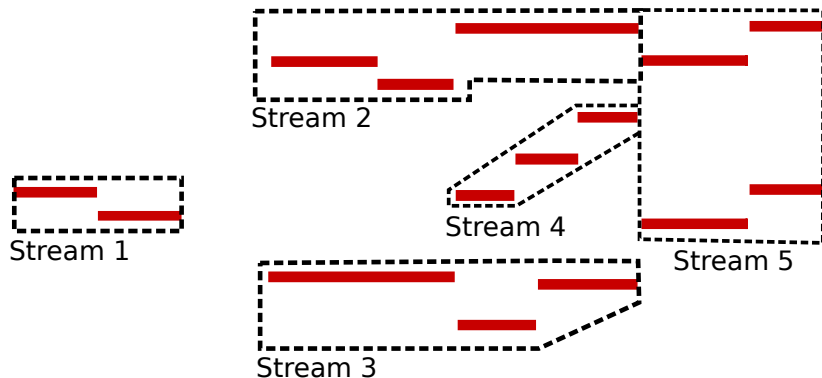
Rafailidis *et al.* - 2008



- ▶ The distance between two notes is computed according to their synchronicity, pitch and onset proximity.
- ▶ Streams can be polyphonic

Stream separation algorithms:

Rafailidis *et al.* - 2008



- ▶ The distance between two notes is computed according to their synchronicity, pitch and onset proximity.
- ▶ Streams can be polyphonic

Outline

State of the art:

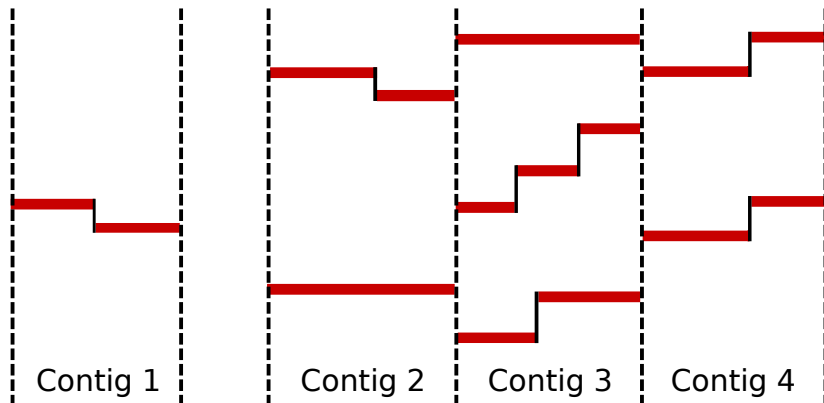
1. What means to organize polyphony?
2. Voice and stream

My contributions (ISMIR 2015, ISMIR 2016):

- 3. Voice separation can be viewed as stream separation**
4. How to compare voice and stream separation?
5. Improving voice and stream separation

Stream separation algorithms:

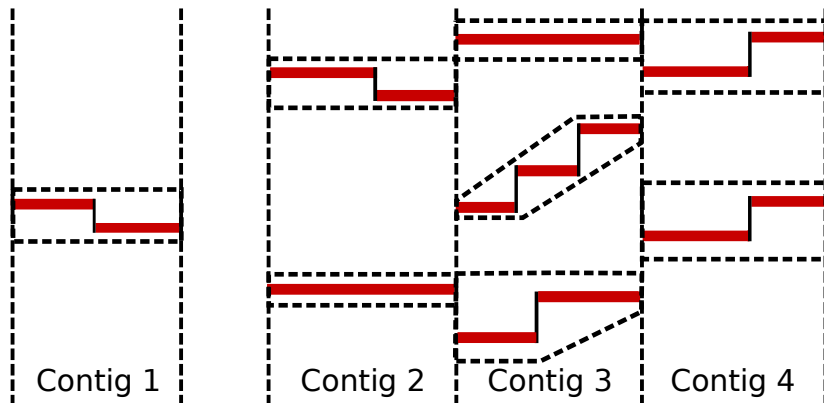
Chew and Wu Contigs



The first step of the Chew and Wu algorithm (contig and fragment creation) can be considered as a stream segmentation algorithm.

Stream separation algorithms:

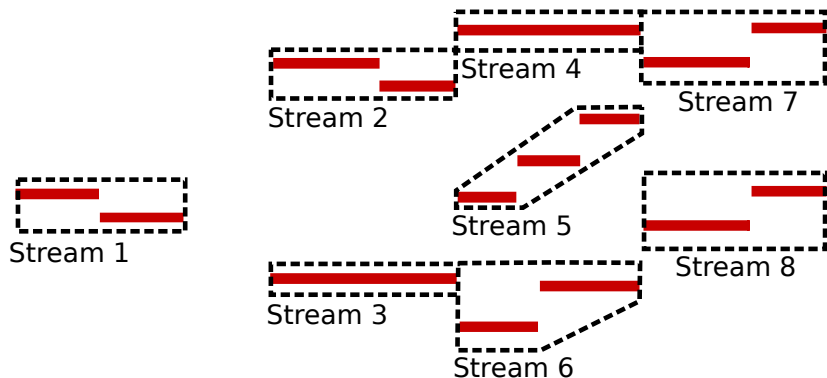
Chew and Wu Contigs



The first step of the Chew and Wu algorithm (contig and fragment creation) can be considered as a stream segmentation algorithm.

Stream separation algorithms:

Chew and Wu Contigs



The first step of the Chew and Wu algorithm (contig and fragment creation) can be considered as a stream segmentation algorithm.

Outline

State of the art:

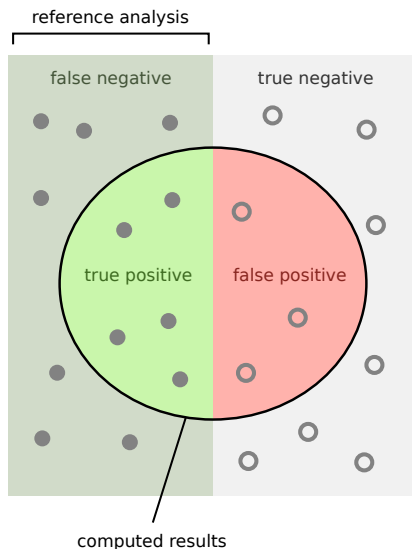
1. What means to organize polyphony?
2. Voice and stream

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

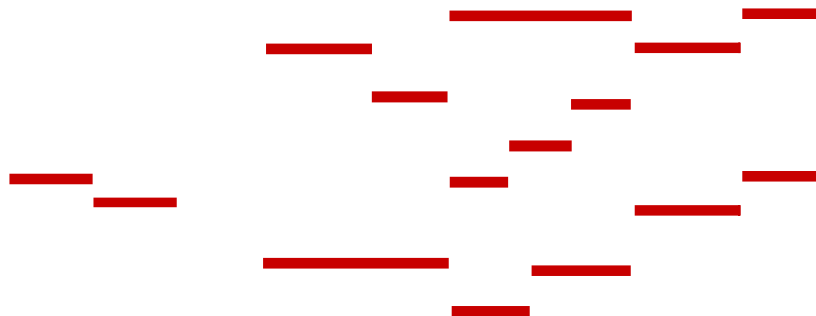
Information retrieval context

Comparison to reference analysis



Evaluation measures

Are notes in the right voice?



Evaluation measures

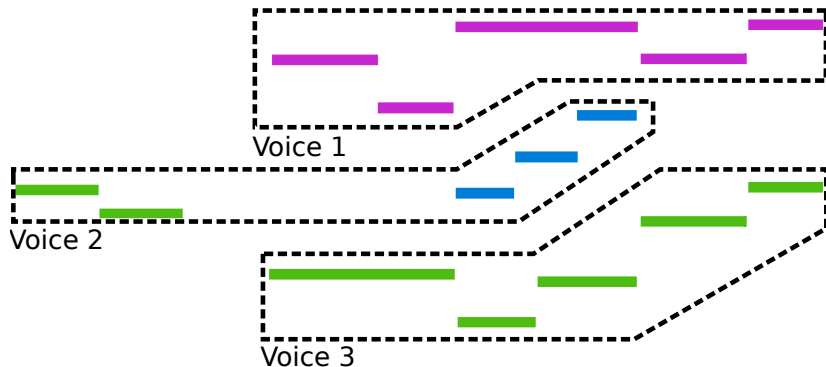
Are notes in the right voice?



$$\text{Note Precision} = \frac{\text{\#True Positive}}{\text{\#True Positive} + \text{\#False Negative}}$$

Evaluation measures

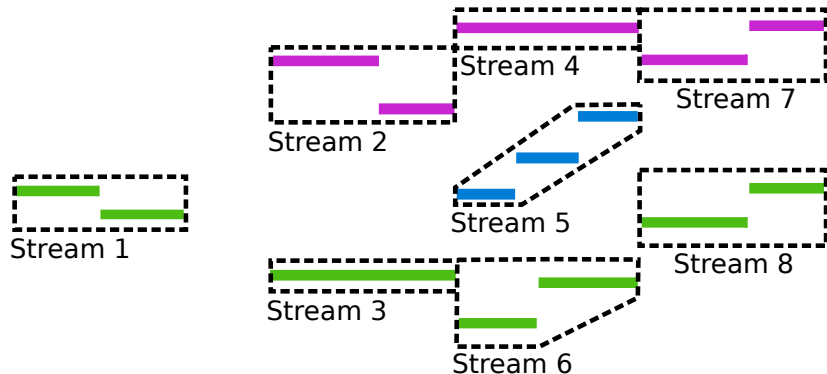
Are notes in the right voice?



$$\begin{aligned}\text{Note Precision} &= (\# \text{True Positive}) / (\# \text{True Positive} + \# \text{False Negative}) \\ &= (5 + 3 + 5) / 15 = 13 / 15 = 0.87\end{aligned}$$

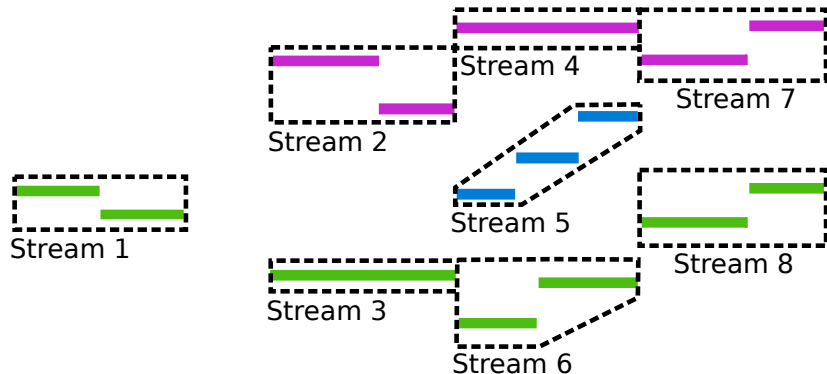
Evaluation measures

Are notes in the right voice?



Evaluation measures

Are notes in the right voice?



$$\begin{aligned}\text{Note Precision} &= (\# \text{True Positive}) / (\# \text{True Positive} + \# \text{False Negative}) \\ &= (2 + 3 + 2) / 15 = 7 / 15 = 0.47\end{aligned}$$

Evaluation measures

Are notes well grouped together?



Evaluation measures

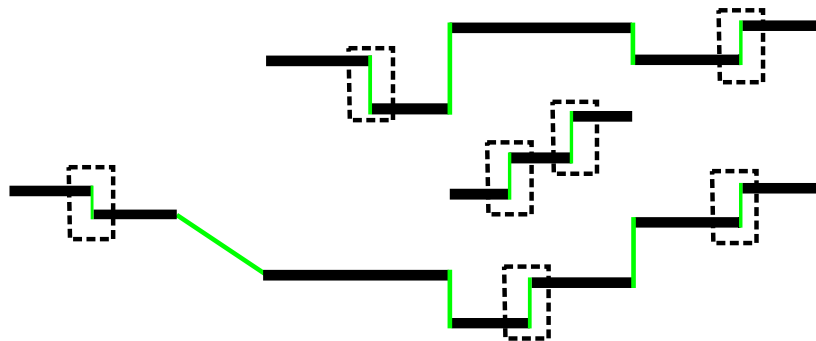
Are notes well grouped together?



Transition Recall = $(\# \text{True Positive}) / (\# \text{True Positive} + \# \text{False Negative})$
(sensitivity)

Evaluation measures

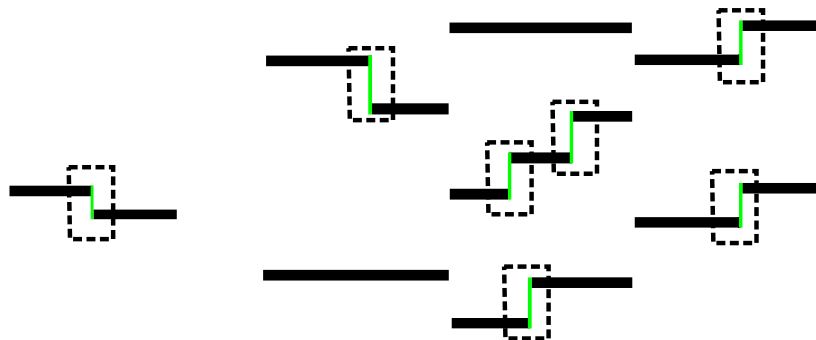
Are notes well grouped together?



$$\begin{aligned}\text{Transition Recall (sensitivity)} &= (\# \text{ True Positive}) / (\# \text{ True Positive} + \# \text{ False Negative}) \\ &= 7/12 = 0.58\end{aligned}$$

Evaluation measures

Are notes well grouped together?

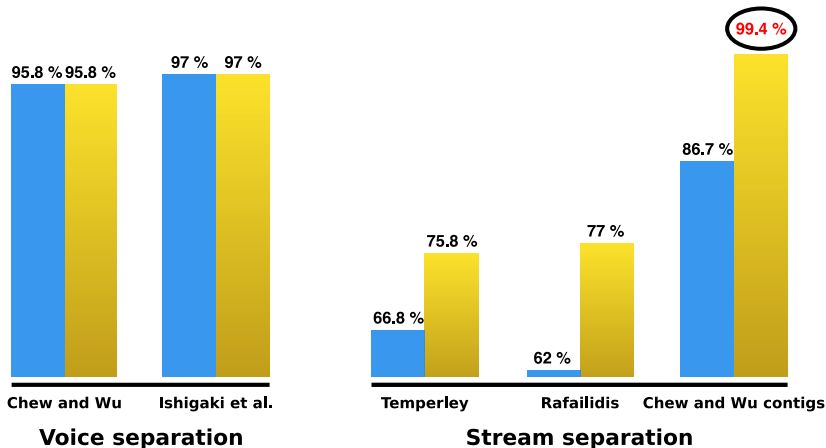


$$\begin{aligned}\text{Transition Precision (specificity)} &= (\# \text{ True Positive}) / (\# \text{ True Positive} + \# \text{ False Positive}) \\ &= 7/7 = 1.0\end{aligned}$$

Transitions Recall/Precision

- TR-rec: Transition Recall
- TR-prec: Transition Precision

		Files	Voices	notes
Corpus: Bach WTC	Book I	24	3.5	1041
	Book II	24	3.4	1071



Outline

State of the art:

1. What means to organize polyphony?
2. Voice and stream

My contributions (ISMIR 2015, ISMIR 2016):

3. Voice separation can be viewed as stream separation
4. How to compare voice and stream separation?
5. Improving voice and stream separation

A new view on contig-based voice separation

- ▶ Which contigs should be connected?
- ▶ How to connect them?

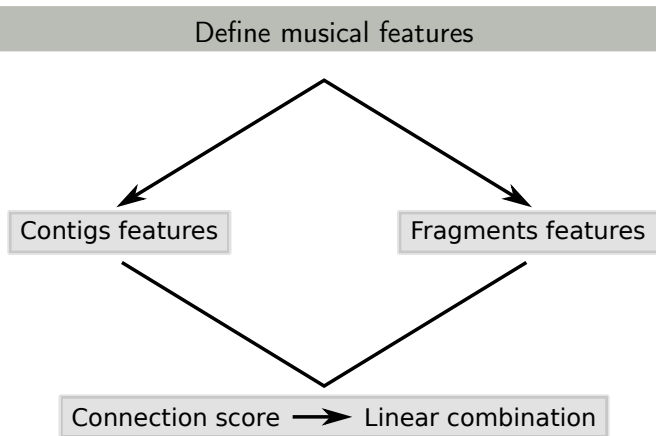
A new view on contig-based voice separation

- ▶ Which contigs should be connected?
- ▶ How to connect them?

Define musical features

A new view on contig-based voice separation

- ▶ Which contigs should be connected?
- ▶ How to connect them?



List of features

	Features	SimCW	SimIMS
Contigs features	<i>increase()</i>	0	0
	<i>increase_one()</i>	0	0
	<i>increase_equal()</i>	0	0.250
	<i>decrease()</i>	0	0
	<i>decrease_one()</i>	0	0
	<i>decrease_equal()</i>	0	0
	<i>difference_nb_voices()</i>	0	0
	<i>maximal_voices()</i>	0.500	0
	<i>minimal_voices()</i>	0	0.250
	<i>maximal_sim_notes()</i>	0	0
Fragments features	<i>crossed_voices()</i>	0	0
	<i>no_crossed_voices()</i>	0.250	0.250
	<i>extreme_pitch()</i>	0.250	0.250
	<i>avg_pitch_right()</i>	0	0
	<i>avg_pitch_left()</i>	0	0
	<i>avg_pitch()</i>	0	0
	<i>extreme_dur()</i>	0	0
	<i>avg_dur_right()</i>	0	0
	<i>avg_dur_left()</i>	0	0
	<i>avg_dur()</i>	0	0

Table: Weighting coefficients of musical features used to measure the quality of the connection.

List of features

	Features	SimCW	SimIMS
Contigs features	<i>increase()</i>	0	0
	<i>increase_one()</i>	0	0
	<i>increase_equal()</i>	0	0.250
	<i>decrease()</i>	0	0
	<i>decrease_one()</i>	0	0
	<i>decrease_equal()</i>	0	0
	<i>difference_nb_voices()</i>	0	0
	<i>maximal_voices()</i>	0.500	0
	<i>minimal_voices()</i>	0	0.250
	<i>maximal_sim_notes()</i>	0	0
Fragments features	<i>crossed_voices()</i>	0	0
	<i>no_crossed_voices()</i>	0.250	0.250
	<i>extreme_pitch()</i>	0.250	0.250
	<i>avg_pitch_right()</i>	0	0
	<i>avg_pitch_left()</i>	0	0
	<i>avg_pitch()</i>	0	0
	<i>extreme_dur()</i>	0	0
	<i>avg_dur_right()</i>	0	0
	<i>avg_dur_left()</i>	0	0
	<i>avg_dur()</i>	0	0

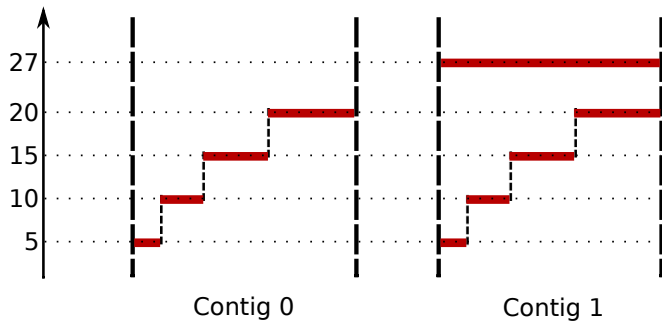
Table: Weighting coefficients of musical features used to measure the quality of the connection.

List of features

	Features	SimCW	SimIMS
Contigs features	<i>increase()</i>	0	0
	<i>increase_one()</i>	0	0
	<i>increase_equal()</i>	0	0.250
	<i>decrease()</i>	0	0
	<i>decrease_one()</i>	0	0
	<i>decrease_equal()</i>	0	0
	<i>difference_nb_voices()</i>	0	0
Fragments features	<i>maximal_voices()</i>	0.500	0
	<i>minimal_voices()</i>	0	0.250
	<i>maximal_sim_notes()</i>	0	0
	<i>crossed_voices()</i>	0	0
	<i>no_crossed_voices()</i>	0.250	0.250
	<i>extreme_pitch()</i>	0.250	0.250
	<i>avg_pitch_right()</i>	0	0
Fragments features	<i>avg_pitch_left()</i>	0	0
	<i>avg_pitch()</i>	0	0
	<i>extreme_dur()</i>	0	0
	<i>avg_dur_right()</i>	0	0
	<i>avg_dur_left()</i>	0	0
	<i>avg_dur()</i>	0	0

Table: Weighting coefficients of musical features used to measure the quality of the connection.

Features on notes pitch



Features on notes pitch

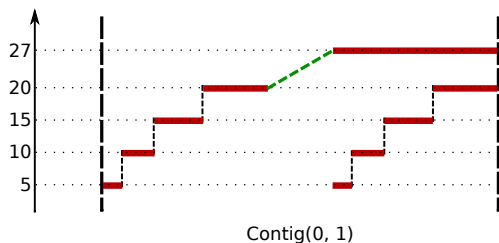


Figure: With the pitches of the notes at the ends of the fragments \Rightarrow Bad merge

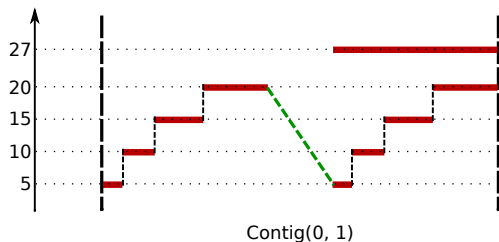
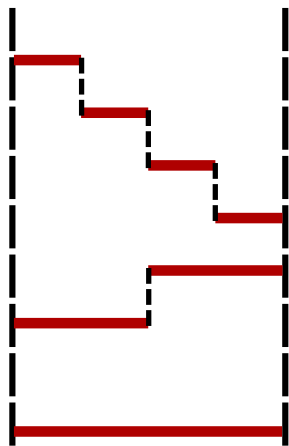
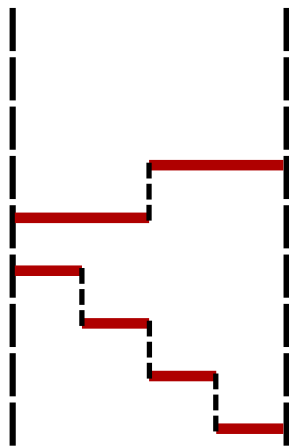


Figure: Taking average pitches of fragment notes \Rightarrow Correct merge 39 / 51

Features on notes length



Contig 0



Contig 1

Features on notes length

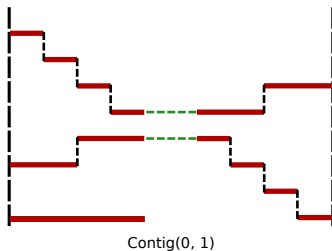


Figure: Without length of notes \Rightarrow Bad merge

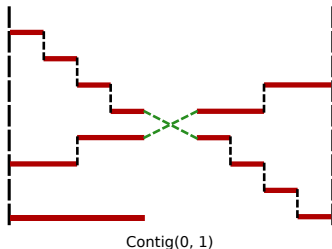
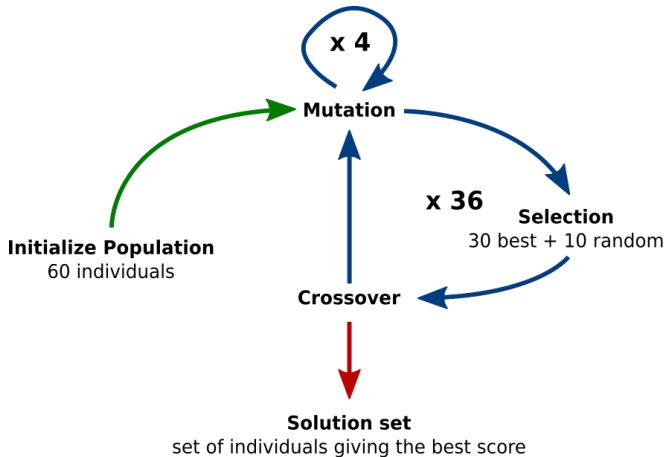


Figure: Taking length of notes \Rightarrow Correct merge

Find coefficients features values

Genetic algorithm



Using MeCS calculator - 288 CPU cores and stocking harbor 100 To

Coefficients features values

	Features	SimCW	SimIMS	GA1
Contigs features	<i>increase()</i>	0	0	0.004
	<i>increase_one()</i>	0	0	0.004
	<i>increase_equal()</i>	0	0.250	0.137
	<i>decrease()</i>	0	0	0.013
	<i>decrease_one()</i>	0	0	0.019
	<i>decrease_equal()</i>	0	0	0.112
	<i>difference_nb_voices()</i>	0	0	0.009
	<i>maximal_voices()</i>	0.500	0	0.026
	<i>minimal_voices()</i>	0	0.250	0.007
	<i>maximal_sim_notes()</i>	0	0	0.007
Fragments features	<i>crossed_voices()</i>	0	0	0.009
	<i>no_crossed_voices()</i>	0.250	0.250	0.248
	<i>extreme_pitch()</i>	0.250	0.250	0.090
	<i>avg_pitch_right()</i>	0	0	0.117
	<i>avg_pitch_left()</i>	0	0	0.023
	<i>avg_pitch()</i>	0	0	0.041
	<i>extreme_dur()</i>	0	0	0.007
	<i>avg_dur_right()</i>	0	0	0.048
	<i>avg_dur_left()</i>	0	0	0.006
	<i>avg_dur()</i>	0	0	0.073

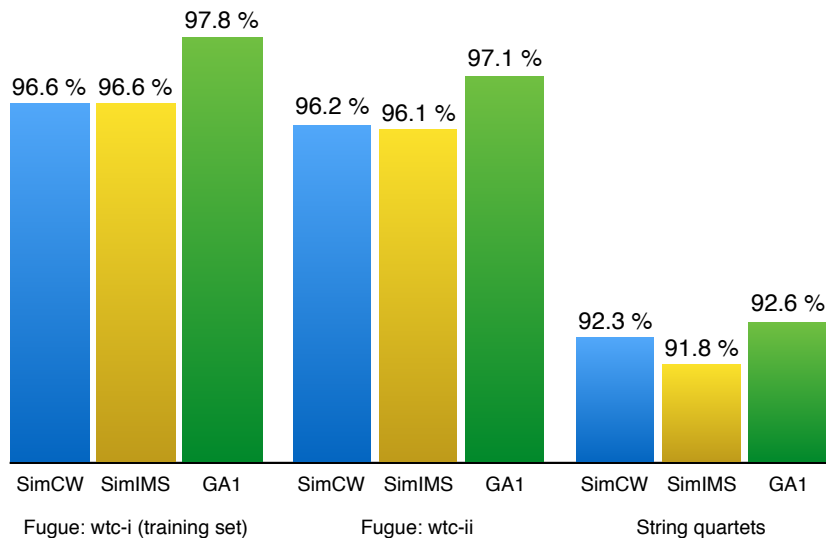
Table: Weighting coefficients of musical features used to measure the quality of the connection.

Coefficients features values

	Features	SimCW	SimIMS	GA1
Contigs features	<i>increase()</i>	0	0	0.004
	<i>increase_one()</i>	0	0	0.004
	<i>increase_equal()</i>	0	0.250	0.137
	<i>decrease()</i>	0	0	0.013
	<i>decrease_one()</i>	0	0	0.019
	<i>decrease_equal()</i>	0	0	0.112
	<i>difference_nb_voices()</i>	0	0	0.009
	<i>maximal_voices()</i>	0.500	0	0.026
	<i>minimal_voices()</i>	0	0.250	0.007
	<i>maximal_sim_notes()</i>	0	0	0.007
Fragments features	<i>crossed_voices()</i>	0	0	0.009
	<i>no_crossed_voices()</i>	0.250	0.250	0.248
	<i>extreme_pitch()</i>	0.250	0.250	0.090
	<i>avg_pitch_right()</i>	0	0	0.117
	<i>avg_pitch_left()</i>	0	0	0.023
	<i>avg_pitch()</i>	0	0	0.041
	<i>extreme_dur()</i>	0	0	0.007
	<i>avg_dur_right()</i>	0	0	0.048
	<i>avg_dur_left()</i>	0	0	0.006
	<i>avg_dur()</i>	0	0	0.073

Table: Weighting coefficients of musical features used to measure the quality of the connection.

Results – Notes transitions



Quality of connection

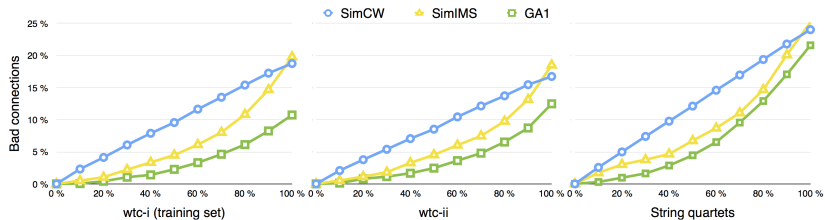


Figure: Errors done during the successive connection steps



Conclusion and future works

Conclusion

- ▶ Voice separation can be viewed as Stream separation,
- ▶ Contigs approach provides the best results,
- ▶ More musical features to improve voice separation.

Publications:

- ▶ Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, and Florence Levé. Comparing voice and stream segmentation algorithms. In *International Society for Music Information Retrieval Conference (ISMIR 2015)*, pages 493–499, 2015.
- ▶ Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, and Florence Levé. Improving voice separation by better connecting contigs. In *International Society for Music Information Retrieval Conference (ISMIR 2016)*, pages 192–198, 2016.
- ▶ Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, and Florence Levé. Séparation de voix en musique classique et en musique populaire. In *Journées d'Analyse Musicale de la Sfam (JAM 2014)*, 2014.

Improve polyphonic separation by searching patterns

Bach, book I, Fugue #16 – measures 12 to 15

The image displays a musical score for measures 12 to 15 of J.S. Bach's Fugue #16, Book I. The score is written in G minor (one flat) and common time (C). It consists of two staves: a treble clef staff on top and a bass clef staff on the bottom. The key signature is one flat (B-flat), and the time signature is common time (C). The score is divided into four measures, labeled 12, 13, 14, and 15. Measure 12 shows the beginning of the fugue with a treble staff starting on a whole note G and a bass staff starting on a whole note B-flat. Measure 13 features a complex polyphonic texture with multiple voices. Measure 14 continues the intricate counterpoint. Measure 15 concludes the section with a final cadence. The notation includes various rhythmic values such as eighth and sixteenth notes, as well as rests and accidentals.

Improve polyphonic separation by searching patterns

Bach, book I, Fugue #16 – measures 12 to 15

The image displays a musical score for measures 12 through 15 of Fugue #16 from Bach's Notebook for Anna Bach. The score is written in G minor and common time (C). It features two staves: a treble clef staff (soprano) and a bass clef staff (alto). The music is polyphonic, with multiple voices. In measure 15, a specific voice crossing is highlighted in orange. This crossing occurs when the soprano voice (treble clef) moves down to a lower pitch than the alto voice (bass clef), which then moves up to a higher pitch, effectively swapping their relative positions. The orange markings include stems and beams for the notes involved in this exchange.

Figure: Crossing voice between alto and soprano

Improve polyphonic separation by searching patterns

Bach, book I, Fugue #16 – measures 12 to 15



The image shows a musical score for three voices: Soprano (top staff), Alto (middle staff), and Bass (bottom staff). The key signature is one flat (B-flat) and the time signature is common time (C). The score covers measures 12, 13, 14, and 15. In measure 15, the Soprano and Alto parts cross, with the Soprano moving to a lower pitch and the Alto moving to a higher pitch, which is highlighted with orange and green annotations.

Figure: Crossing voice between alto and soprano

No algorithm was able to separate this passage correctly.

Make a musical analysis based on voice separation

Algorithm	Corpus	True Positive	False Positive	False Negative	Sensitivity	Precision
SimCW	wtc-i	213	46	78	73.2%	82.2%
SimIMS		217	44	74	74.57%	83.14%
GA1		230	38	61	79.04%	85.82%
Monophonic		266	29	25	91.41%	90.17%
SimCW	Shosta	69	4	105	39.66%	94.52%
SimIMS		70	2	104	40.21%	97.22%
GA1		77	4	97	44.25%	95.06%
Monophonic		94	12	80	54.02%	88.68%

Table: Results of fugue analysis algorithm after voice separation.

Make a musical analysis based on voice separation

Algorithm	Corpus	True Positive	False Positive	False Negative	Sensitivity	Precision
SimCW	wtc-i	213	46	78	73.2%	82.2%
SimIMS		217	44	74	74.57%	83.14%
GA1		230	38	61	79.04%	85.82%
Monophonic		266	29	25	91.41%	90.17%
SimCW	Shosta	69	4	105	39.66%	94.52%
SimIMS		70	2	104	40.21%	97.22%
GA1		77	4	97	44.25%	95.06%
Monophonic		94	12	80	54.02%	88.68%

Table: Results of fugue analysis algorithm after voice separation.

Opening

- ▶ Support for automatic partition analysis,
- ▶ Introduction of new features:
 - ▶ Helping to find other elements;
 - ▶ Real-time interpretation of a piece for voice detection;
 - ▶ Calculation of the proximity between two pieces.

Thank you for your attention

