Audio Files Realignment by Dynamic Time Warping (DTW)

Florian Picard, Florian Tilquin

June 27, 2013



Introduction	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion
●0	0000	00		O
What is DTW ?				

- Matching algorithm estimating similarities and establishing pointwise correspondance between two time series *U* and *V*;
- In sound processing:
 - Realignment of audio files;
 - Speech recognition
- Robust to time fluctuations (interpretation, rhythm, noise...)







Signal representation: spectrograms







Figure : Example of path





Figure : Example of path





Figure : Example of path





Figure : Example of path

 Introduction
 Dynamic Time Warping
 Audio files realignment
 Speech recognition
 Conclusion

 Cost function and set of path P

Cost function: $W(P) = \sum_{k=1}^{K} \|U_{i(k)} - V_{j(k)}\|_2$ Definition of \mathcal{P} : path in \mathcal{P} are continuous monotonic and bounded.

> (i, j - 1) (i, j) (i - 1, j - 1) (i - 1, j) U U V (a) Continuity, mono- (b) Example of path in \mathcal{P}

We want to minimize W over \mathcal{P} : $P^* = \arg \min_{P \in \mathscr{P}} W(P)$

tonicity



We note D the matrix of distances: $D = (||U_i - V_j||_2)$. We define a cost matrix C, defined by recursion: C(i,j) is the minimal cost of a path finishing at (i,j).



Figure : Construction of the cost matrix

Knowing C we easily can deduce P^* .



- α , β , γ put weight on transitions for the path;
- p percentage of coefficients actually computed in C;
- δ controls the location of the end of the path¹.



Figure : Parameters: α , β , γ , p, δ

¹Considerations on DTW algorithm for spoken word recognition, L. Rabiner, A. Rosenberg, S. Levinson

Introduction 00	Dynamic Time Warping 0000	Audio files realignment	Speech recognition	Conclusion O
Audio file	realignment			

- We recorded 18 vyniles (extracts of 30 seconds), some of records alterated (jump, acceleration...) into wav;
- U = reference signal, V = tested signal;
- We tested our algorithm on the normalized spectrograms.

Objectives

Use the path given by the algorithm in order to:

- output a pointwise time correspondance between the two signals;
- reconstruct the alterated signal to play both signals together.

Introduction	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion
00	0000	⊙●		O
Influence	of parameters			



(a) Influence of the di- (b) Influence of the (c) Influence of the agonal transition pa- truncature parameter p, boundary parameter δ , rameter β , here set to 2 here set to 0.5 here set to 0.2

Introduction	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion
00		00	●00	O
Isolated	word recognitio	'n		

- DTW gives the minimal cost of a realignement between two audio files, which is independant of time fluctuations.
- Provide a distance measure adapted to word recognition

Data

- 13 subjects (3 females and 10 males);
- 31 words, some of them similar ("mène-mère", "irruption-éruption", "rateau-bateau"...);









Introduction 00	Dynamic Time Warping 0000	Audio files realignment 00	Speech recognition	Conclusion
Conclusion				

What we were able to do during this internship:

- Do an extended bibliographical research on the subject and a detailed state of the art;
- Provide a code for the algorithm studied in matlab;
- Put it to the test with a lot of data that we collected;
- Explore two applications of the algorithm in order to test its precision and its limits;
- Submit a research article on the subject containing our results on observations to the journal IPOL/SPOL.

Introduction	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion
00	0000	00		O
Sets of r	ecords			

- 3 sets of records: recorded in wav (sample rate: 44100 Hz)
- Various modifications (jumps, accelerations, decelerations, offset...)
- Various origins (speech, classical music, instrumental music)

For each of those record, we manage to realign the alterated records with the reference one.

Introduction 00	Dynamic Time Warping 0000	Audio files realignment	Speech recognition	Conclusion O
Le concei	rto de la mer			

Track	Modifications
1	Our reference
2	Acceleration at 10", deceleration at 20"
3	A jump in the record
4	With another vinyle disc of the same music
5	Played faster
6	Played louder
7	Played backward
8	Everything (except playing backward)
Та	ble : Modifications on <i>Le concerto de la mer</i>

Introduction 00	Dynamic Time Warping 0000	Audio files realignment	Speech recognition	Conclusion O
Mozart K	550-1			

Track	Modifications
1	Our reference
2	Deceleration at 15"
3	Acceleration at 15"
4	Played faster
5&6	With another vinyle disc of the same music

Table : Modifications on K.550 -1.

Le petit p	orince			
Introduction 00	Dynamic Time Warping 0000	Audio files realignment	Speech recognition	Conclusion O

Track	Modifications
1	Our reference
2&3	With another vinyle disc of the same music (huge offset)
4	Speed variations

Table : Modifications on Le petit prince

Introduction	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion
00	0000	00		O
Effect of	acceleration			

We show here the result of a DTW between U = reference file (Mozart), and V = tested file (Mozart accelerated).



V

Figure : Change in speed (acceleration)

Speakers	caracteristics			
Introduction 00	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion O

Speakers were male and female, and some subjects are speaking with accent, heavy or slight.

Sp1	Male, no accent	Sp8	Male, no accent
Sp2	Male, no accent	Sp9	Male, heavy accent
Sp3	Male, no accent	Sp10	Male, slight accent
Sp4	Female, no accent	Sp11	Female, slight accent
Sp5	Male, no accent	Sp12	Female, no accent
Sp6	Male, no accent	Sp13	Male, no accent
Sp7	Male, no accent		

Introduction	Dynamic Time Warping	Audio files realignment	Speech recognition	Conclusion
00	0000	00		O
Word list				

1	Mène	9	Soleil	17	Mêler	25	Poteau
2	Usuel	10	Forme	18	Limonade	26	Vision
3	Sommeil	11	Mère	19	Tomate	27	Oseille
4	Camion	12	Canon	20	Oreiller	28	Même
5	Tome	13	Mireille	21	Groseille	29	Éruption
6	Mémoire	14	Bateau	22	Passoire	30	Télévision
7	Pareil	15	Homme	23	Rome	31	Rateau
8	Abricot	16	Irruption	24	Bravo		

Table : List of the 31 words pronunced, in the order of pronunciation respected by all speakers.



Speaker to speaker matches





