SFU SCHOOL OF COMPUTING SCIENCE

Dynamic Signal Processing

CMPT 419/983, Summer 2020

Dr. Angelica Lim

This lecture will be recorded and linked in Canvas.
You will be able to download it, but please don't post it anywhere. Thanks!

Assignment 2 Updates

The Phoebe dataset was updated (as per my email) to remove timestamps in favour of frame numbers.

Thanks Tien, Ye and Justin for discussions in Piazza and Bita for regenerating our .csv files!

Activity Debrief

FACS, gaze and head orientation can be used as distinctive, measurable characteristics called **features**.

"The context plays a very important role in this clip. The shift in gaze directly to the object of interest followed by head-tilt highlights the object of curiosity. The lack of activations followed by scanning the environment implies the subject is trying to maintain a neutral face for the fear of drawing attention."

- Sachini Herath

Activity Debrief

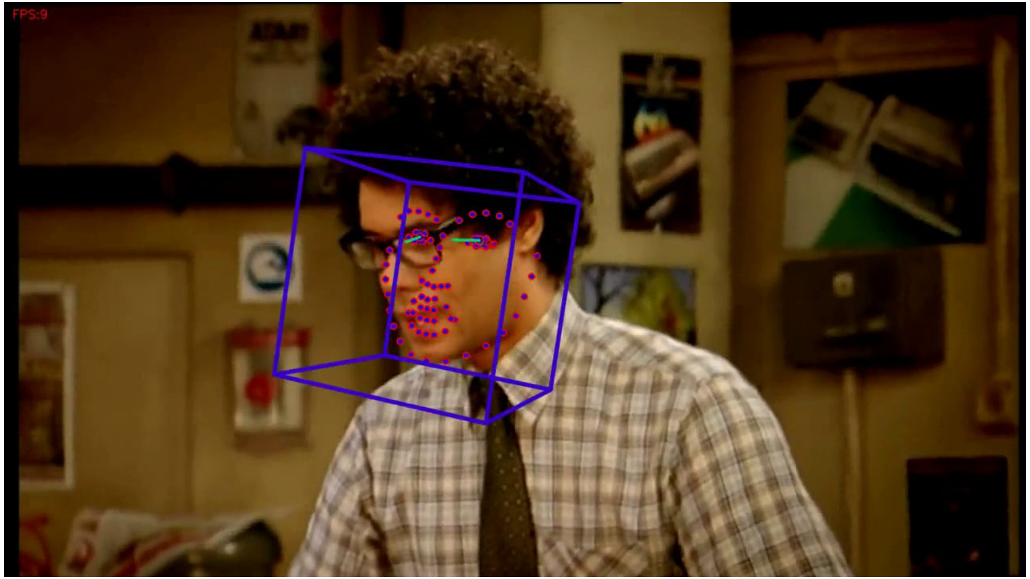
Other types of features like body pose may be useful.

"Context definitely matters, as Devon is staring down his fellow armwrestler, in a loud cheering place." - Tien Vu

→ **Body pose as a feature, relationship to another person, dominance and arousal dimensions**

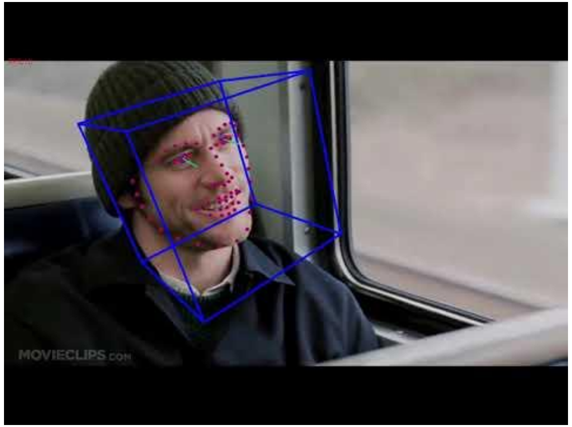
Activity Debrief

Automatic FACS detection is not yet perfect.



Activity Debrief

Multiple modalities may be necessary (Week 7):



The context is incredibly important, with the tone of their speaking really coloring the emotional impact of their facial expressions. - Dylan

"Lots of lip movement in saying a sentence, so there are lip movement related AU's but don't contribute to the actual mood of the clip [...] The overall emotion that this segment is conveying is confusion, the AU1 kind of validates this. However, it is more telltale to hear Maria's tone of speech (clips here are silent) to know that she really IS confused. - Jack Zhao

Activity Debrief: Context

Context plays a decent role because it makes it more clear which emotion the person should be feeling. This can help when there are AUs that can apply to multiple emotions and it is not certain which one is right. - Andrew Liu

It matters what a person is looking at when one tries to determine that person's emotion. For example, if a person expresses social signals associated with happiness when the person is looking at his worst enemy losing in a battle, the person may feel happy and contempt on the inside; if the person is looking at his best friend winning a contest, that person is likely to have happy emotion only. The context a person is in can provide clues to the person's emotion. - Alina Li

However, understanding the context may provide additional details; observers may infer a component of relief in the happiness/joy the character is showing. - Sasha Yao

One thing to note was that part 00:01~00:03 is a 'mocking' laugh. In other words, despite the physical appearance that can be detected through the detector, it was the context that determines the actual emotion delivered through the signal. - Jihoon Sung

Kuleshov Effect

A word about final projects

- Keep track of any interesting topics you'd like to explore. A lot of what we discuss (especially about context) are open problems!
- Feel free to get feedback anytime, via Piazza DM to instructors or Office Hours

- Week 7's Activity (**June 24**) will be to share your project idea on the Canvas Discussion thread, and read what others are considering
- Team project proposals will be due on **July 7th**

Project details here: https://canvas.sfu.ca/courses/53138/pages/project

Last week...

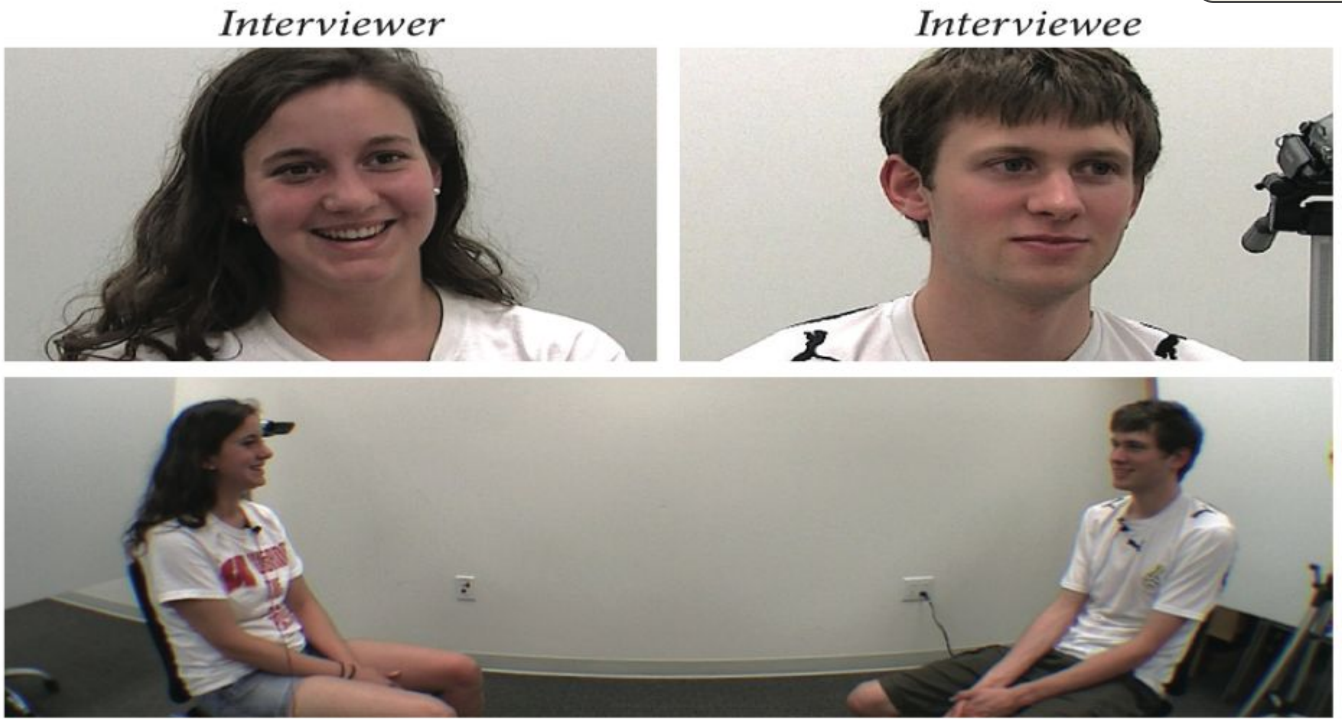
We talked about clustering and creating models of data (K-Means and GMM).

This data was generally **static** in that we worked with image-based features such as AU intensity and presence.

What are some other features?

Gaze and Head Orientation

These 2 are different!



Conversational gaze aversion for humanlike robots, Andrist (2014)

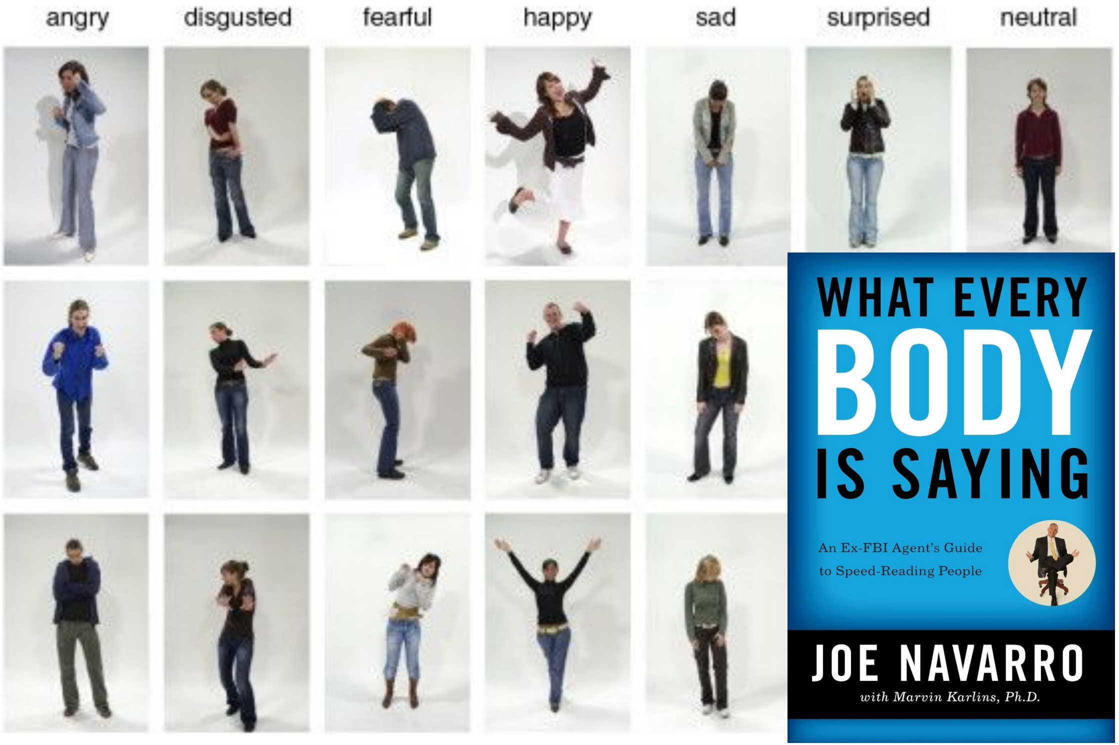
Body Posture

Beatrice De Gelder, Neuropsychologist studies emotion in bodies

In this study, she found:

- 2D features were enough (we don't need to infer 3D bodies first)
- Disgusted and fearful poses were confused

From the Tilburg University stimulus set



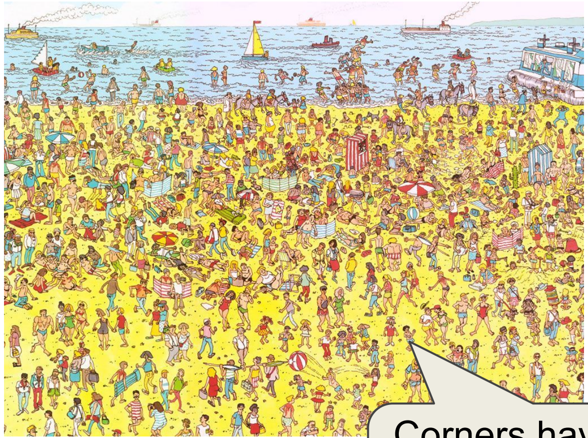
Recognizing emotions expressed by body pose: A biologically inspired neural model

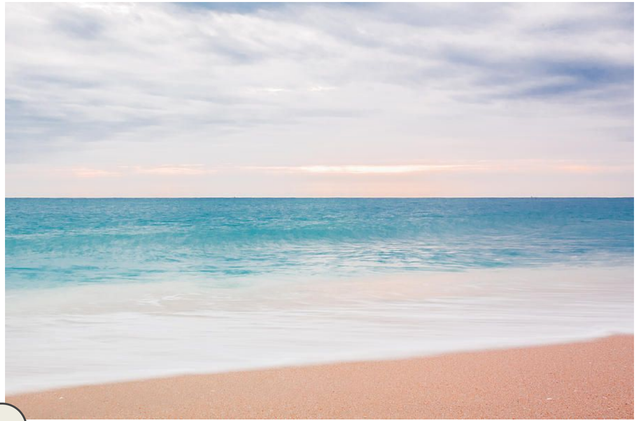
Arousal and Distance

https://learning.oreilly.com/library/view/design-for-emotion/9780123865311/xhtml/CHP005.html#CHP005tit1

What are the features here?

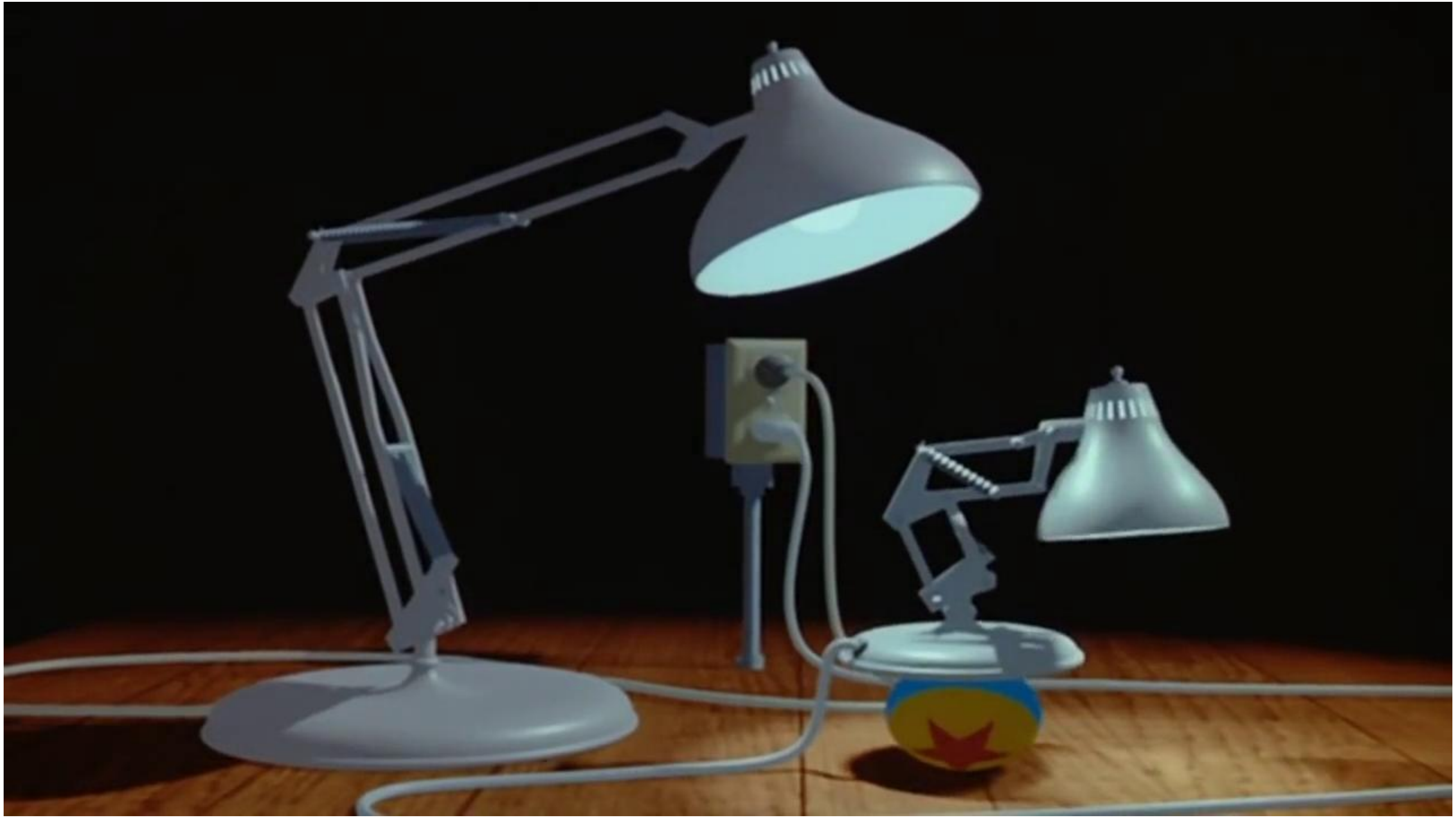
Two pictures of the beach:





Corners have been a typical feature in computer vision

Let's Talk about Dynamics



Dynamic Data

Temporal or dynamic data, also known as "time series" data, can be processed using:

- **Distance**-based methods ⇒ Compute the distance between pairs of time series (e.g. Dynamic Time Warping) *// Today*

- **Feature**-based methods ⇒ Transform data into lower-dimensional feature vectors, before applying traditional classification techniques *// Thurs*

- **Model**-based methods ⇒ Use a model such as Hidden Markov Model (HMM), Recurrent Neural Network (RNN), etc. *// Later*

Working with temporal data

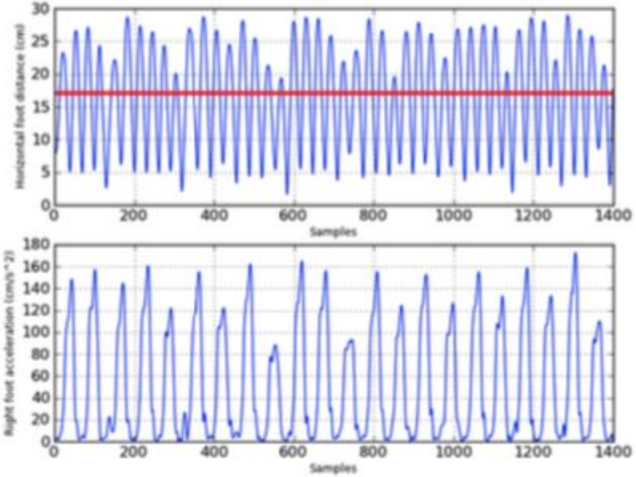
Temporal human signals?

- Music: Tempo, beat detection, pitch…
- Physiological: Heart rate, heart rate variability, skin conductance
- Gesture
- Gait
- Facial or gaze dynamics
- Dance
- Voice

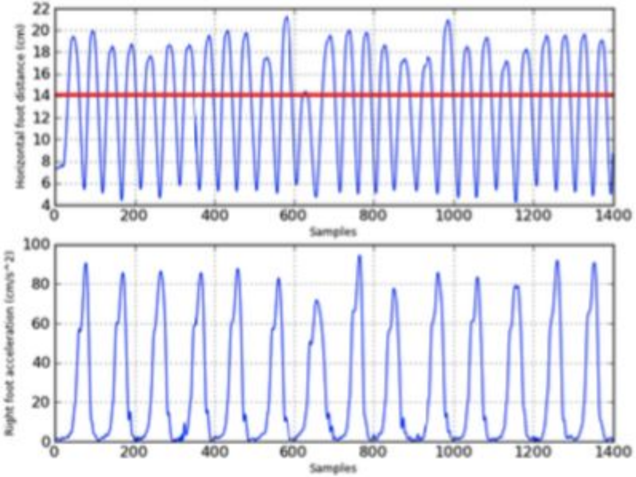
https://arxiv.org/abs/1602.01711

Gait

Fig. 1: Examples of gait analysis. The red line indicates the mean value (threshold for peak-picking)



(a) Example of angry gait of user "ale"



(b) Example of sad gait of user "ale"

Using Speech Data to Recognize Emotion in Human Gait (Lim, 2012)

The Great Time Series Bake-Off

- Multiple rigorous studies show that for time series classification, **Nearest Neighbour DTW** is very hard to beat

- Where NN-DTW can be beaten, it's typically by a small margin at a large cost in code complexity, time and space overhead.

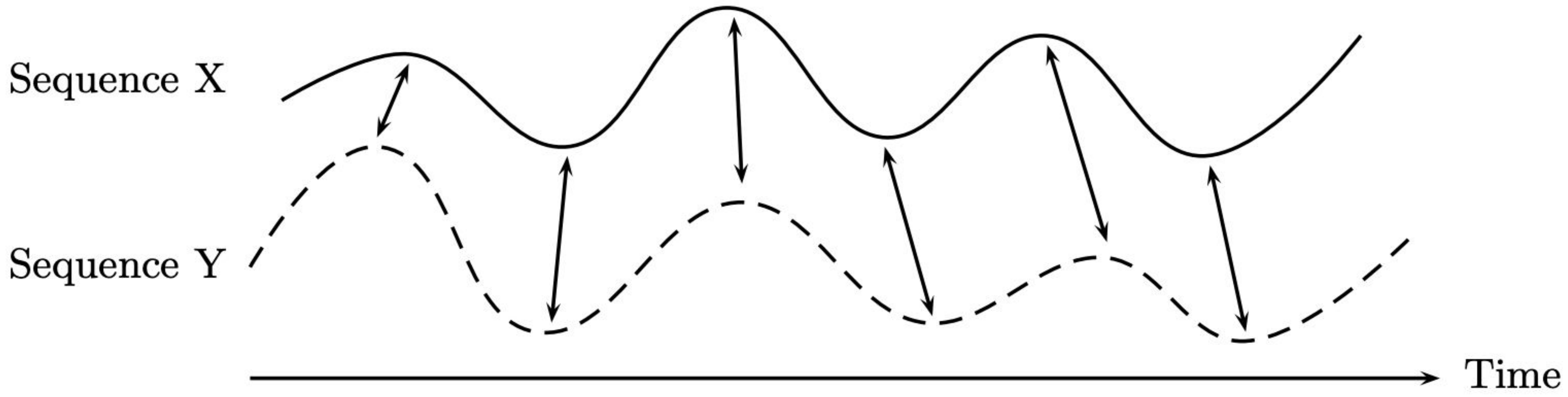
*This paper conducts 35 million experiments, on 85 datasets, with dozens of rival methods
The Great Time Series Classification Bake Off: An Experimental Evaluation of Recently Proposed Algorithms (Bagnall, 2016).

Dynamic Time Warping

Dynamic Time Warping (DTW)

- Finds optimal alignment between two temporal sequences
- Warps the sequences non-linearly to match each other
- Originally used to compare speech patterns in automatic speech recognition, e.g. "Hello" vs. "Helllllllooooo"
- Also used widely in music information retrieval, gesture recognition, robotics, bioinformatics, finance, physics, and other fields requiring data analytics...

Dynamic Time Warping



How do we find the aligned points, indicated by arrows?

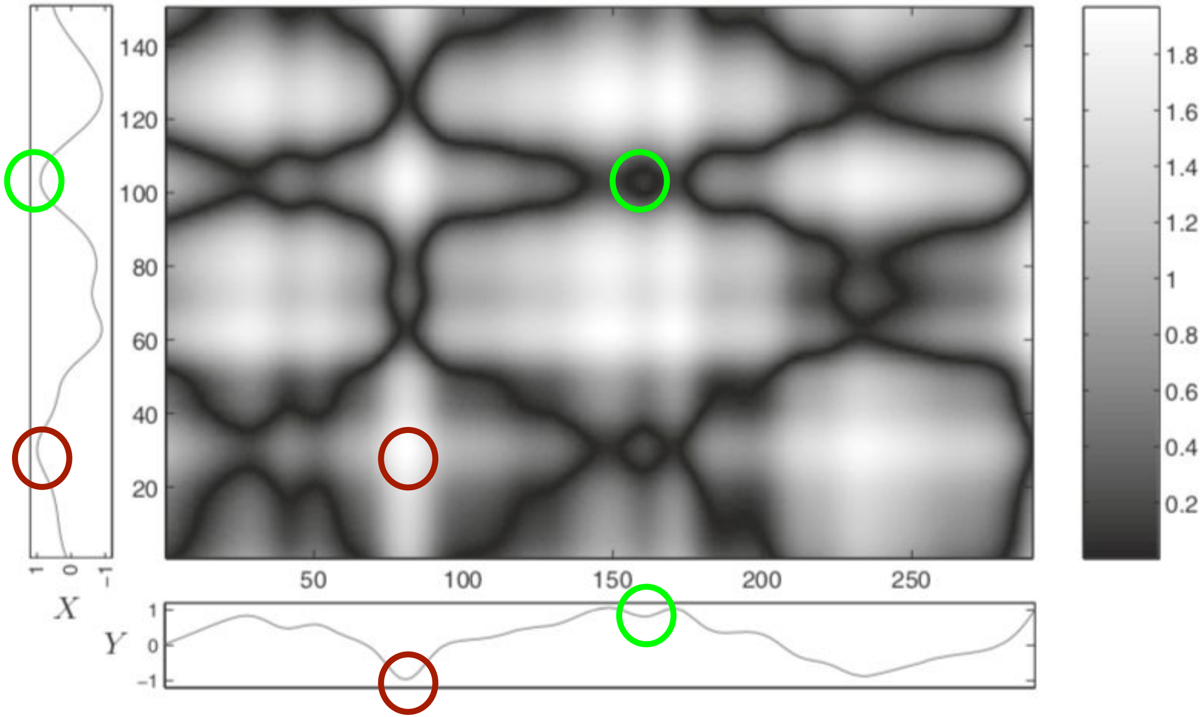
Information Retrieval for Music and Motion, pg. 70

Dynamic Time Warping

Cost matrix of X and Y using absolute value of difference as cost measure.

The goal is to find an alignment between X and Y with **minimal overall cost**.

Intuitively, the **optimal alignment** runs along a "valley" of low cost, i.e. darker coloured areas here.

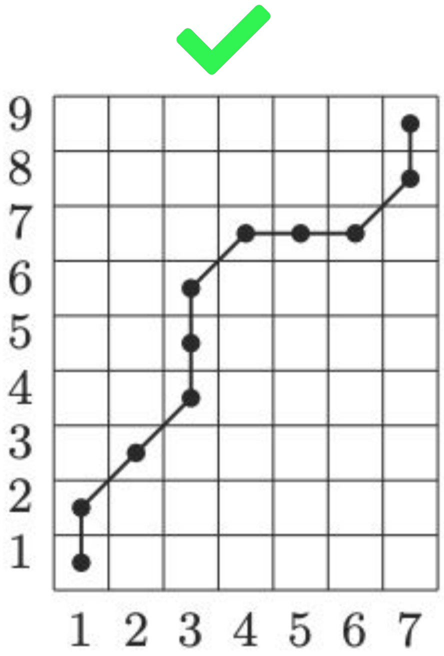


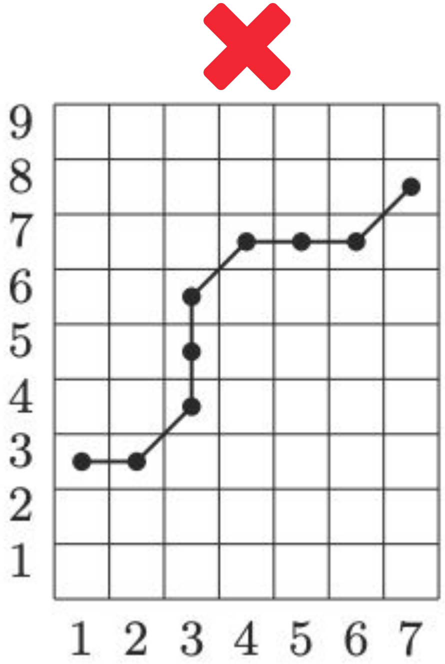
Information Retrieval for Music and Motion, pg. 70

Dynamic Time Warping

Three constraints:

- **Boundary condition**
- Monotonicity condition
- Step size condition



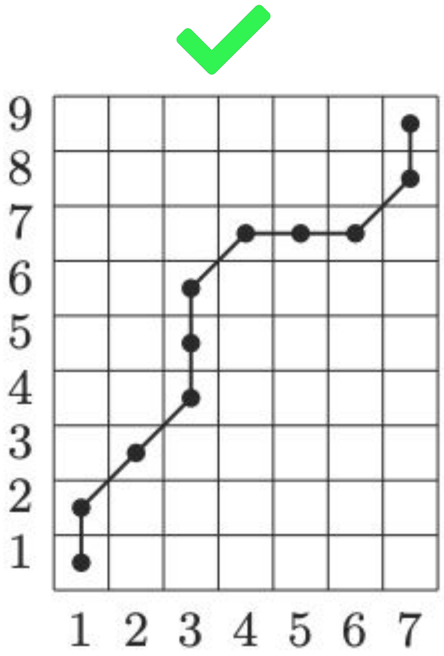


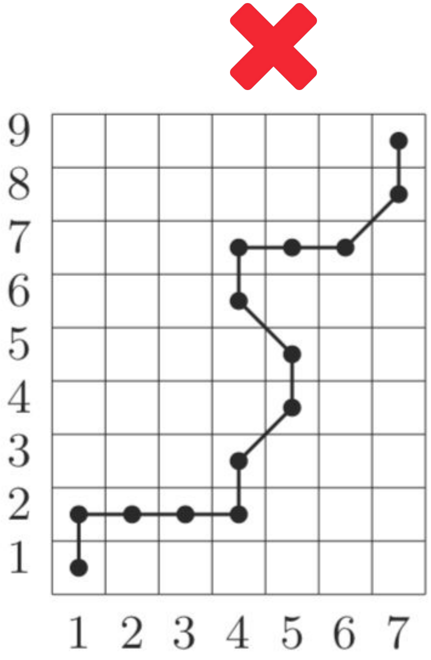
Information Retrieval for Music and Motion, pg. 70

Dynamic Time Warping

Three constraints:

- Boundary condition
- **Monotonicity condition**
- Step size condition



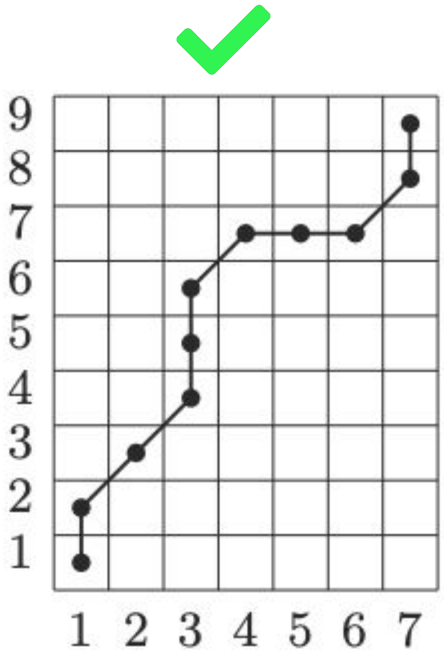


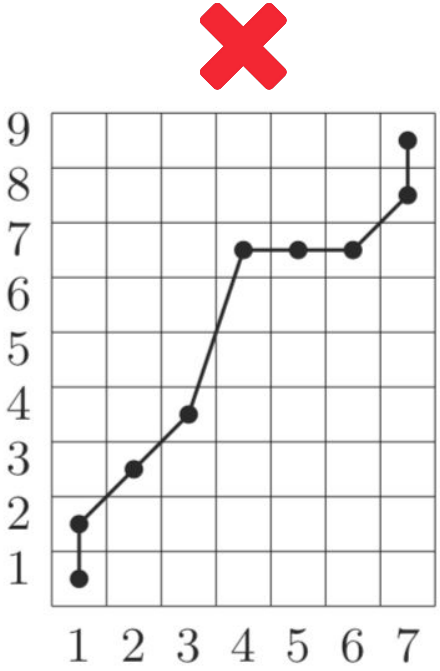
Information Retrieval for Music and Motion, pg. 70

Dynamic Time Warping

Three constraints:

- Boundary condition
- Monotonicity condition
- **Step size condition**





Information Retrieval for Music and Motion, pg. 70

DTW Algorithm

- To determine the optimal path by testing all possible paths between X (length N) and Y (length M) is exponential
- Dynamic programming can find a path in O(NM)
  - Break down the problem into simpler subproblems
  - Derive the optimal warping path for subsequences, apply recursively

DTW Algorithm

Table 3.2 from [Müller, FMP, Springer 2015]

Algorithm: DTW

Input: Cost matrix $\mathbf{C}$ of size $N \times M$
Output: Accumulated cost matrix $\mathbf{D}$
Optimal warping path P^*

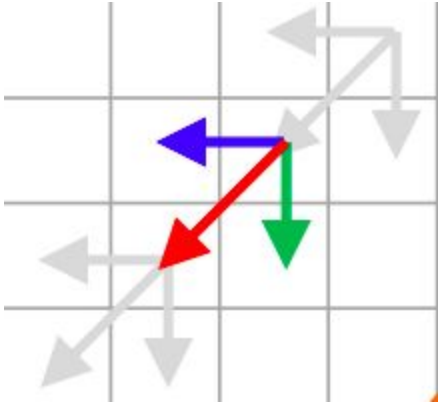
Procedure: Initialize $(N \times M)$ matrix $\mathbf{D}$ by $\mathbf{D}(n,1) = \sum_{k=1}^{n} \mathbf{C}(k,1)$ for $n \in [1:N]$ and $\mathbf{D}(1,m) = \sum_{k=1}^{m} \mathbf{C}(1,k)$ for $m \in [1:M]$. Then compute in a nested loop for $n = 2,\ldots,N$ and $m = 2,\ldots,M$:

$$\mathbf{D}(n,m) = \mathbf{C}(n,m) + \min\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}.$$

Set $\ell = 1$ and $q_\ell = (N,M)$. Then repeat the following steps until $q_\ell = (1,1)$:

Increase ℓ by one and let $(n,m) = q_{\ell-1}$.
If $n = 1$, then $q_\ell = (1, m-1)$,
else if $m = 1$, then $q_\ell = (n-1, m)$,
else $q_\ell = \mathrm{argmin}\{\underline{\mathbf{D}(n-1,m-1)}, \underline{\mathbf{D}(n-1,m)}, \underline{\mathbf{D}(n,m-1)}\}$.
(If 'argmin' is not unique, take lexicographically smallest cell.)

Set $L = \ell$ and return $P^* = (q_L, q_{L-1}, \ldots, q_1)$ as well as $\mathbf{D}$.



https://www.audiolabs-erlangen.de/resources/MIR/FMP/C3/C3S2_DTWbasic.html

Dynamic Time Warping

Here lies the DTW "distance" between two sequences

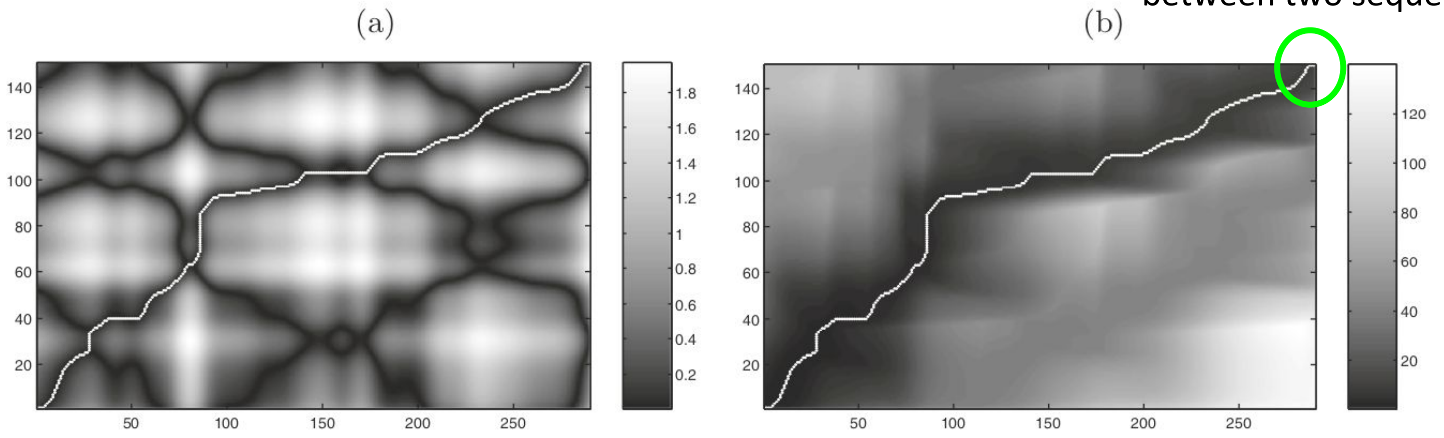
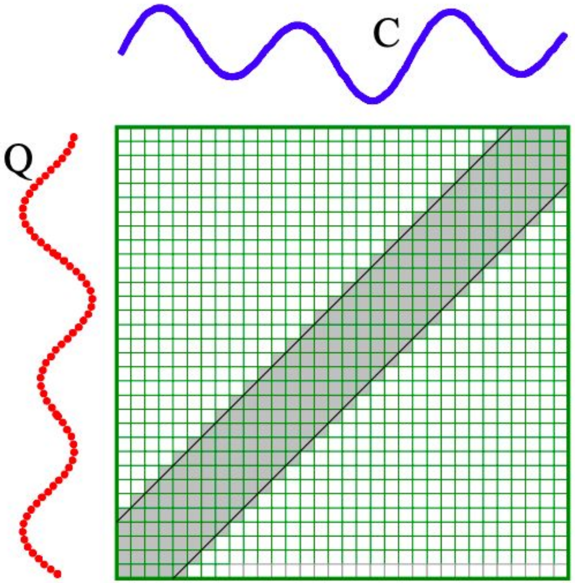


Fig. 4.4. **(a)** Cost matrix C as in Fig. 4.2 and **(b)** accumulated cost matrix D with optimal warping path p^* (*white line*)

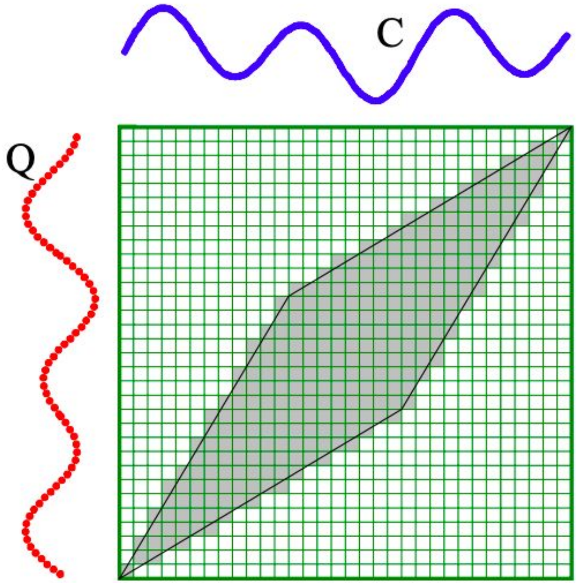
Information Retrieval for Music and Motion, pg. 70

Warping Windows

It is possible to reduce the DTW search space by defining warping constraints.



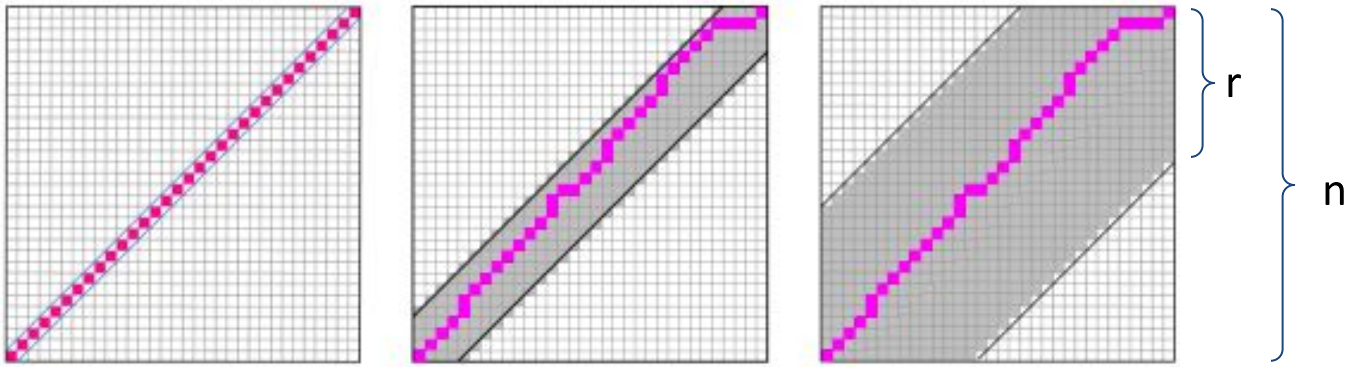
Sakoe-Chiba Band



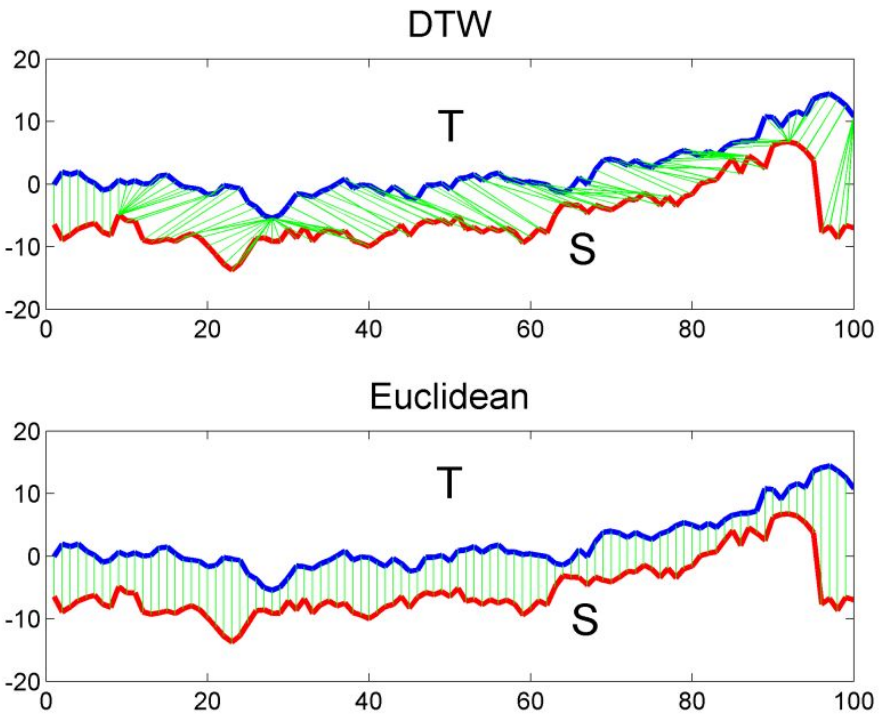
Itakura Parallelogram

w, the warping constraint

w = r/n
allows you to define how much deviation from the diagonal you will allow



DTW vs. Euclidean Distance



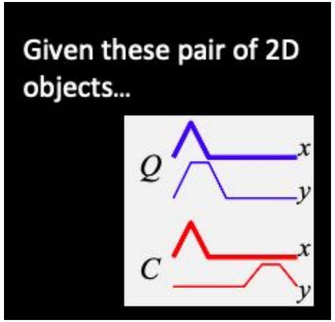
Multi-dimensional DTW

Independent: If the process is coupled loosely, i.e. the physical process generates varying lags. Compute the DTW score for each dimension, and sum the scores

Dependent: If the process is coupled tightly, i.e. the physical process affects the time series simultaneously.

Check out this tutorial from KDD 2016 "Extracting Optimal Performance from Dynamic Time Warping" https://www.cs.unm.edu/~mueen/DTW.pdf

Multi-dimensional DTW



Independent: Compute the DTW score for each dimension independently, and sum up each score.

$$DTW_I(Q,C) = \sum_{m=1}^{M} DTW(Q_m, C_m)$$

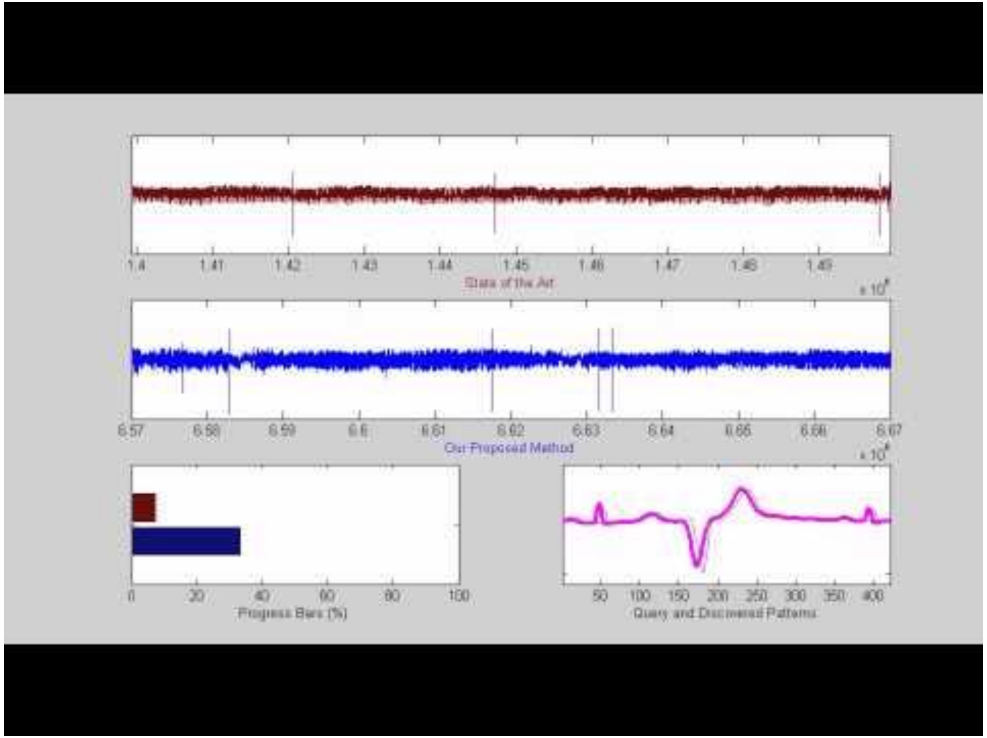
Dependent: Create a single distance matrix that reflect the distance between each corresponding pair of time series, then find the single warping path and distance as per normal.

$$\mathrm{DTW}_I(Q,C) = \mathrm{DTW}(\mathrm{Q}_x,\mathrm{C}_x) + \mathrm{DTW}(\mathrm{Q}_y,\mathrm{C}_y)$$

Check out this tutorial from KDD 2016 "Extracting Optimal Performance from Dynamic Time Warping" https://www.cs.unm.edu/~mueen/DTW.pdf

UCR Suite

- Fast DTW in Python
  https://pypi.org/project/ucrdtw/



http://www.cs.ucr.edu/~eamonn/UCRsuite.html

Synchrony and DTW

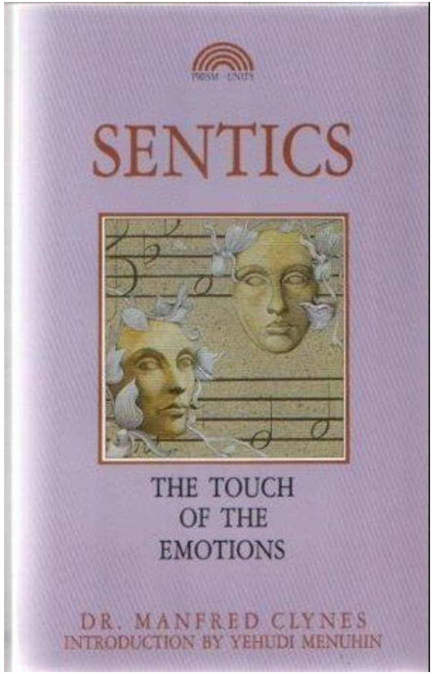
Nonverbal Synchrony as an Adaptation to Social Environments

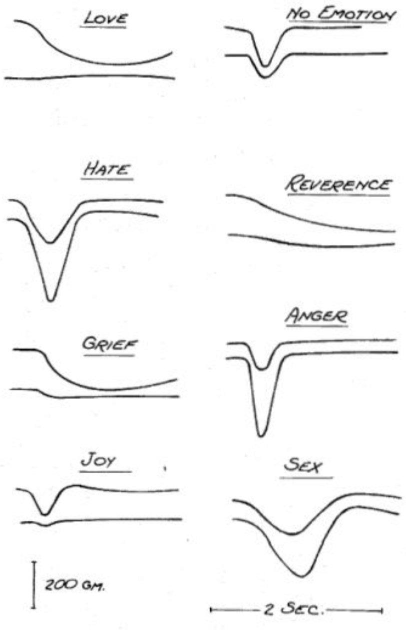
Thinking about traveling to a foreign country, or talking to a person from another place, what would you do to express yourself clearly? How do people use nonverbal coordination to adapt socially in which language may not be able to serve its optimal function? To study this question, I created in lab a set of tasks that elicit emotions, and facilitate cooperation. Using automated facial and bodily analysis, I also developed state-of-the-art methods to measure and automated nonverbal synchrony.

To learn more about this project:
** I will be presenting my work at the SAS 2019 in Boston, MA
Also, check out my poster at SPSP 2019 in Portland, OR [Poster]

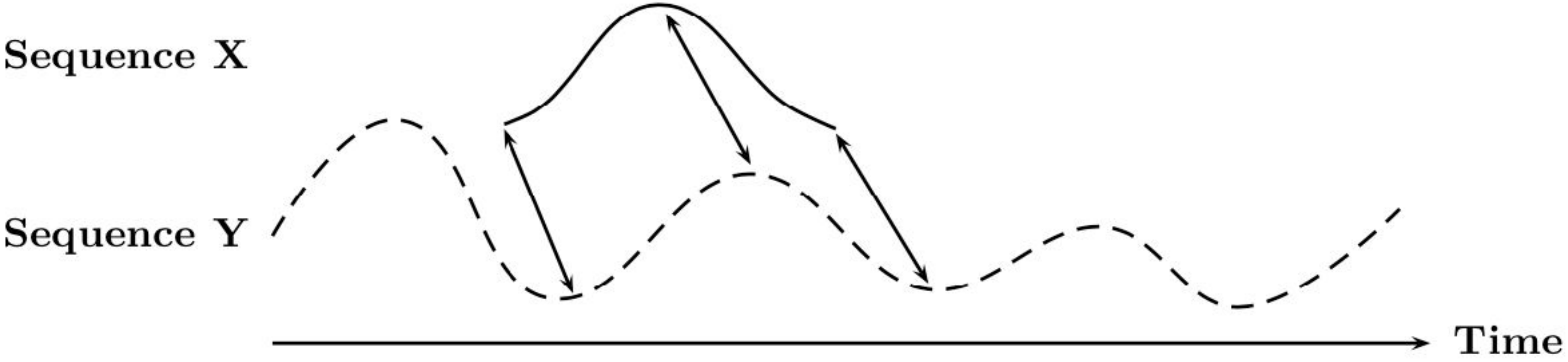
https://www.olivia-zhao.com/

Sentics


SENTICS
THE TOUCH OF THE EMOTIONS
DR. MANFRED CLYNES
INTRODUCTION BY YEHUDI MENUHIN


LOVE
NO EMOTION
HATE
REVERENCE
GRIEF
ANGER
JOY
SEX
200 GM.
2 SEC.

Retrieve Similar Subsequences



Retrieve Similar Subsequences

Algorithm: ComputeSimilarSubsequences

Input: $X = (x_1, \ldots, x_N)$ query sequence
$Y = (y_1, \ldots, y_M)$ database sequence
$\tau \in \mathbb{R}$ cost threshold

Output: Ranked list of all (essential distinct) subsequences of Y that have a DTW distance to X below the threshold τ.

(0) Initialize the ranked list to be the empty list.
(1) Compute the accumulated cost matrix D w.r.t. X and Y,
(2) Determine the distance function Δ as in (4.12).
(3) Determine the minimum $b^* \in [1:M]$ of Δ.
(4) If $\Delta(b^*) > \tau$ then terminate the procedure.
(5) Compute the corresponding DTW-minimizing index $a^* \in [1:M]$.
(6) Extend the ranked list by the subsequence $Y(a^*:b^*)$.
(7) Set $\Delta(b) := \infty$ for all b within a suitable neighborhood of b^*.
(8) Continue with Step (3).

[Online Activity] Week 5: Apply dynamic time warping to gesture data and brainstorm uses

Angelica Lim

Jun 8 at 7:18pm

1 1

In this activity, you will analyze your emotional knocking and waving gesture data from Week 2 using Dynamic Time Warping (DTW).

Tasks

1. Read through this Dynamic Time Warping tutorial for background.

2. Download the class' gesture dataset.

3. Make a copy of the provided Google Notebook, and work through the sections, reading the comments. Use your gesture data (or someone else's data) from Activity 2 to fill in the confusion matrices below, depicting the DTW distances between the various gestures:

	Knock A	Knock B	Knock C	Knock D
Knock A				
Knock B				
Knock C				
Knock D				