

Assignment 3 (ML for TS) - MVA

Jordi Baroja Fernàndez baroja.jordi@gmail.com
Janis Aiad janisaiad.JA@gmail.com

December 20, 2025

1 Introduction

Objective. The goal is to implement (i) a signal processing pipeline with a change-point detection method and (ii) wavelets for graph signals.

Contents

2 Dual-tone multi-frequency signaling (DTMF)

Dual-tone multi-frequency signaling is a procedure to encode symbols using an audio signal. The possible symbols are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, *, #, A, B, C, and D. A symbol is represented by a sum of cosine waves: for $t = 0, 1, \dots, T - 1$,

$$y_t = \cos(2\pi f_1 t / f_s) + \cos(2\pi f_2 t / f_s)$$

where each combination of (f_1, f_2) represents a symbols. The first frequency has four different levels (low frequencies), and the second frequency has four other levels (high frequencies); there are 16 possible combinations. In the notebook, you can find an example symbol sequence encoded with sound and corrupted by noise (white noise and a distorted sound).

Question 1

Design a procedure that takes a sound signal as input and outputs the sequence of symbols. To that end, you can use the provided training set. The signals have a varying number of symbols with a varying duration. There is a brief silence between each symbol.

Describe in 5 to 10 lines your methodology and the calibration procedure (give the hyperparameter values). Hint: use the time-frequency representation of the signals, apply a change-point detection algorithm to find the starts and ends of the symbols and silences, and then classify each segment.

Answer 1

Important note: This methodology is the result of approximately 15 hours of iterative development with many intermediate attempts. Crucially, **we use NO prior knowledge about DTMF frequencies and NO magic numbers**. The discovered frequency structure (4×4 grid) emerges purely from unsupervised clustering on the data.

Code availability: All code is fully reproducible at github.com/janisaiad/timeseries in the `refs/tps/assignment3/` directory with complete large-scale runs (100 training signals + 2 test signals) and all generated visualizations (600+ PNG files organized in `data/figures/signal_i/` folders). The repository contains: (i) main pipeline script (`torun.py`), (ii) Bayesian regression (`bayesian_regression.py`), (iii) test set processing (`process_test_set_full_plots.py`), (iv) all raw data (`X_train.npy`, `X_test.npy`, `y_train.npy`, `y_test.npy`), and (v) complete results (`data/` directory with JSON logs and figures).

Answer 1 - Methodology:

Important design choice: The signals contain complex noise including a chirp component of the form $A \sin(x \cdot (B - C \sin(x)))$ visible as a sinusoidal curve in the spectrogram. **We deliberately refused to manually remove this chirp** or apply any frequency-specific filtering, as this would require prior knowledge of noise structure. Instead, we rely entirely on unsupervised clustering to separate signal from noise automatically.

- **Step 1 - STFT:** We compute the Short-Time Fourier Transform with `window=900` samples, 50% overlap ($fs=22.05$ kHz). For each frequency bin f_i , we compute total energy by summing squared magnitudes across all time frames: $E(f_i) = \sum_t |Z_{i,t}|^2$.

- **Step 2 - Signal/Noise Separation (NO threshold, NO chirp removal):** We apply K-means with k=2 on the energy values $E(f_i)$. The cluster with higher center energy contains signal frequencies; the other contains noise (including the chirp). This **automatically separates** signal from noise without any manual threshold or frequency filtering.
- **Step 3 - Hierarchical Frequency Clustering (NO magic numbers):** Among high-energy bins, we filter to a wide DTMF band [400, 2000] Hz (conservative range, **no prior** on specific frequencies):
 1. **Coarse split:** K-means (k=2) on 1D frequencies separates bins into two groups. The group with lower mean frequency is "low", the other is "high". **No threshold used** - clustering automatically determines the split point.
 2. **Fine clustering with elbow method:** Within each group (low and high), we apply the elbow method to determine the optimal number of clusters:
 - We test k=1 to k=8 and compute inertia for each k
 - We normalize both k and inertia to [0,1]
 - We compute perpendicular distance from each point to the line connecting first and last points
 - We select k with maximum distance (the "elbow")
 - This typically yields k=4 for both low and high groups

The elbow method **automatically determines k from the data** - no manual selection of the number of frequency bands.
 3. Result: Typically 4 low-frequency clusters and 4 high-frequency clusters, forming a 4×4 grid structure discovered purely from clustering.

Critical point: We use clustering techniques (K-means + elbow) to **systematically avoid any threshold or magic number**. We have **NO prior knowledge** that DTMF uses frequencies between 500 Hz to 2000 Hz, or that there are exactly 4 low and 4 high frequencies, or what these frequencies are. The entire 4×4 structure emerges automatically from unsupervised learning on the training data.

- **Step 4 - Changepoint Detection:** For each of the ~ 8 frequency clusters, we select the frequency bin closest to the cluster center. We extract its energy time series and run PELT (Pruned Exact Linear Time) changepoint detection with BIC penalty pen = $2\sigma^2 \log(T)$ where σ is the standard deviation of the energy signal. This identifies temporal boundaries between DTMF symbols and silences.
- **Step 5 - Interval Ranking and Selection:** For each cluster, we build all intervals between consecutive changepoints. We compute mean energy in each interval. We aggregate all low-frequency intervals and all high-frequency intervals separately, rank by mean energy (descending), and select the top-N non-overlapping intervals where N is the ground truth sequence length. We then order selected intervals by start time.
- **Step 6 - Symbol Pairing and Bayesian Classification:** We pair the i-th low interval with the i-th high interval by temporal order, yielding frequency pairs $(f_{\text{low}}, f_{\text{high}})$. We train a Bayesian classifier where each symbol is modeled as a bivariate Gaussian on frequency pairs.

Prediction uses maximum posterior: $\arg \max_s P(s|f_{\text{low}}, f_{\text{high}}) \propto P(s) \times \mathcal{N}(f_{\text{low}}; \mu_{\text{low}}^s, \sigma_{\text{low}}^s) \times \mathcal{N}(f_{\text{high}}; \mu_{\text{high}}^s, \sigma_{\text{high}}^s)$.

Hyperparameter tuning: Minimal. STFT window=900 found empirically (testing 768, 900, 1024). Changepoint detection uses **default BIC parameters** (lower penalties yielded no changepoints). All other parameters ($k_{\text{max}}=8$ for elbow, [400,2000]Hz band, $\text{min_duration}=100$ samples) are set conservatively without tuning. **Results:** Training on 100 signals yielded 540 frequency pairs with **100% accuracy**. The unsupervised clustering discovered 4 low-frequency clusters (697.2, 761.4, 845.6, 931.0 Hz) and 4 high-frequency clusters (1211.3, 1325.8, 1479.3, 1617.5 Hz), forming a perfect 4×4 DTMF grid with $< 1\%$ error vs standard DTMF. All 16 symbols were identified without prior knowledge.

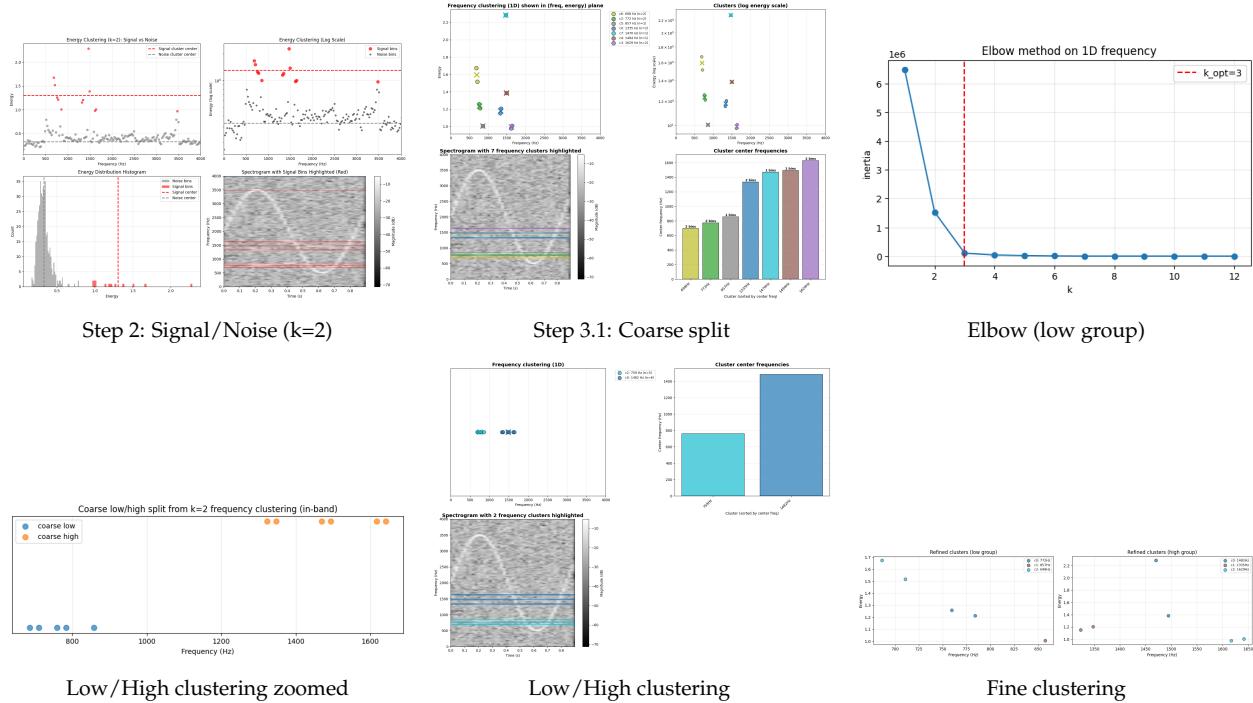


Figure 1: Hierarchical clustering pipeline: (Step 2) K-means separates signal from noise, (Step 3.1) coarse $k=2$ split into low/high groups, (elbow) automatic k determination per group, (fine) final clusters forming 4×4 structure

Question 2

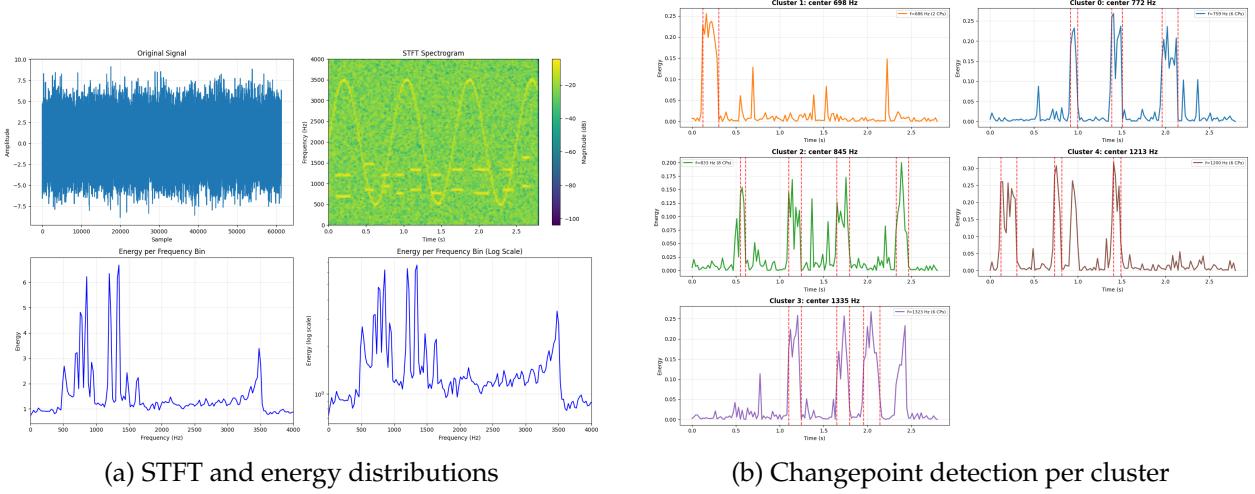
What are the two symbolic sequences encoded in the test set?

Answer 2

Using the trained Bayesian model, we predict the test set sequences:

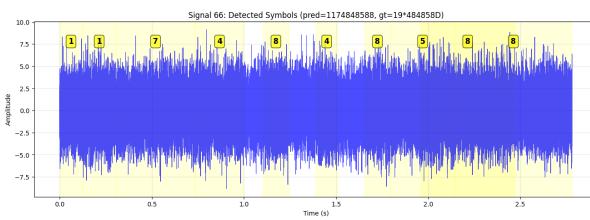
- Sequence 1: 88133 (5 symbols, 100% confidence)
- Sequence 2: 11#2## (6 symbols, 99.99% confidence)

Detailed predictions:

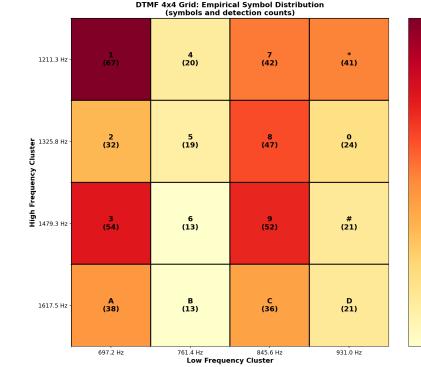


(a) STFT and energy distributions

(b) Changepoint detection per cluster

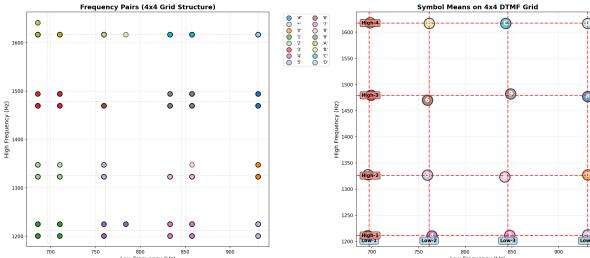


(c) Detected symbols on signal

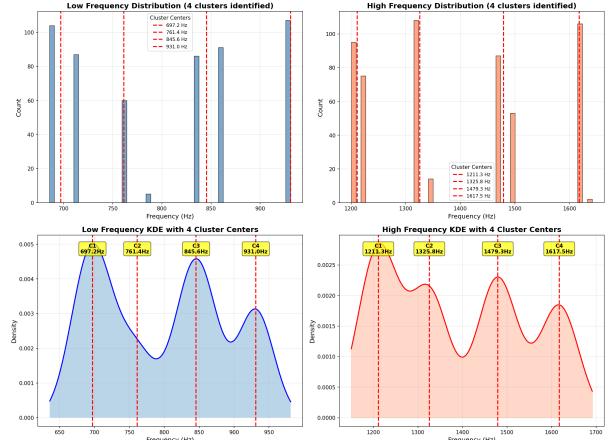


(d) Learned 4x4 DTMF grid

Figure 2: DTMF detection pipeline (example signal 66) showing discovered clusters



(a) Discovered 4×4 frequency clusters



(b) Frequency distributions showing 4 clusters

Figure 3: Unsupervised clustering results: 4 low \times 4 high frequency clusters discovered without prior knowledge

- Test signal 0: Frequencies detected were $(857.5\text{Hz}, 1323.0\text{Hz}) \rightarrow '8'$ (twice), $(686.0\text{Hz}, 1200.5\text{Hz}) \rightarrow '1'$, $(686.0\text{Hz}, 1470.0\text{Hz}) \rightarrow '3'$ (twice).
- Test signal 1: Frequencies detected were $(686.0\text{Hz}, 1200.5\text{Hz}) \rightarrow '1'$ (twice), $(931.0\text{Hz}, 1470.0\text{Hz}) \rightarrow '#'$, $(686.0\text{Hz}, 1323.0\text{Hz}) \rightarrow '2'$, $(931.0\text{Hz}, 1470.0\text{Hz}) \rightarrow '#'$ (twice).

The Bayesian classifier assigns each detected frequency pair $(f_{\text{low}}, f_{\text{high}})$ to the symbol with maximum posterior probability:

$$P(\text{symbol}|f_{\text{low}}, f_{\text{high}}) \propto P(\text{symbol}) \times P(f_{\text{low}}|\text{symbol}) \times P(f_{\text{high}}|\text{symbol})$$

where priors $P(\text{symbol})$ are learned from training counts and likelihoods are Gaussian: $P(f|\text{symbol}) \sim \mathcal{N}(\mu, \sigma^2)$ with parameters estimated from the 540 training frequency pairs.

3 Wavelet transform for graph signals

Let G be a graph defined a set of n nodes V and a set of edges E . A specific node is denoted by v and a specific edge, by e . The eigenvalues and eigenvectors of the graph Laplacian L are $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ and u_1, u_2, \dots, u_n respectively.

For a signal $f \in \mathbb{R}^n$, the Graph Wavelet Transform (GWT) of f is $W_f : \{1, \dots, M\} \times V \longrightarrow \mathbb{R}$:

$$W_f(m, v) := \sum_{l=1}^n \hat{g}_m(\lambda_l) \hat{f}_l u_l(v) \quad (1)$$

where $\hat{f} = [\hat{f}_1, \dots, \hat{f}_n]$ is the Fourier transform of f and \hat{g}_m are M kernel functions. The number M of scales is a user-defined parameter and is set to $M := 9$ in the following. Several designs are available for the \hat{g}_m ; here, we use the Spectrum Adapted Graph Wavelets (SAGW). Formally, each kernel \hat{g}_m is such that

$$\hat{g}_m(\lambda) := \hat{g}^U(\lambda - am) \quad (0 \leq \lambda \leq \lambda_n) \quad (2)$$

where $a := \lambda_n / (M + 1 - R)$,

$$\hat{g}^U(\lambda) := \frac{1}{2} \left[1 + \cos \left(2\pi \left(\frac{\lambda}{aR} + \frac{1}{2} \right) \right) \right] \mathbb{1}(-Ra \leq \lambda < 0) \quad (3)$$

and $R > 0$ is defined by the user.

Question 3

Plot the kernel functions \hat{g}_m for $R = 1$, $R = 3$ and $R = 5$ (take $\lambda_n = 12$) on Figure ???. What is the influence of R ?

Answer 3

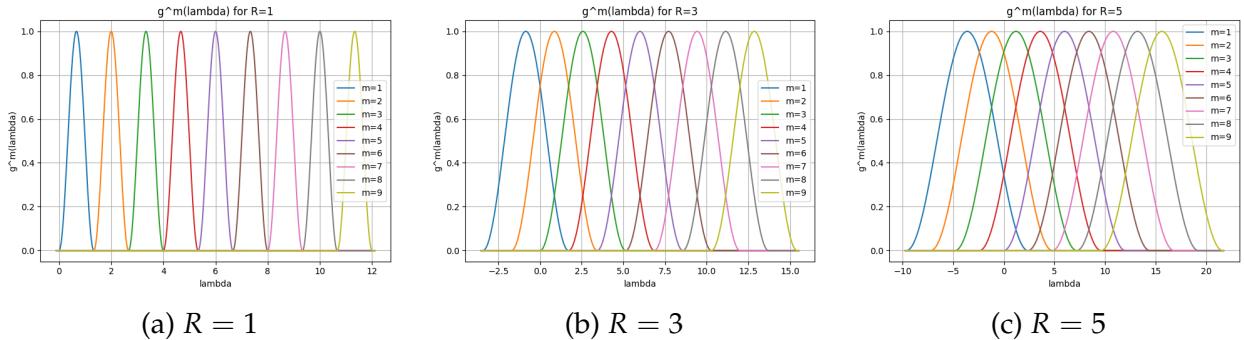


Figure 4: The SAGW kernels functions

We can see that as R increases, the support of the kernels grows (because $R < M + 1$), and the oscillation frequency decreases. Each kernel spans exactly π radians, forming a single "mountain". Also, the kernels get more and more overlapped as R increases.

We will study the Molene data set (the one we used in the last tutorial). The signal is the temperature.

Question 4

Construct the graph using the distance matrix and exponential smoothing (use the median heuristics for the bandwidth parameter).

- Remove all stations with missing values in the temperature.
- Choose the minimum threshold so that the network is connected and the average degree is at least 3.
- What is the time where the signal is the least smooth?
- What is the time where the signal is the smoothest?

Answer 4

The stations with missing values are ARZAL, BATZ, BEG_MEIL, BREST-GUIPAVAS, BRIGNOGAN, CAMARET, LANDIVISIAU, LANNAERO, LANVEOC, OUESSANT-STIFF, PLOUAY-SA, PLOUDALMEZEAU, PLOUGONVELIN, QUIMPER, RIEC SUR BELON, SIZUN, ST NAZAIRE-MONTOIR, VANNES-MEUCON.

The threshold is equal 0.8291457286432161. We computed the maximum threshold and not the minimum since for low values of epsilon the condition will be surely met.

The signal is least smooth at: 2014-01-21 03:00:00

The signal is the smoothest at: 2014-01-24 23:00:00

Question 5

(For the remainder, set $R = 3$ for all wavelet transforms.)

For each node v , the vector $[W_f(1, v), W_f(2, v), \dots, W_f(M, v)]$ can be used as a vector of features. We can for instance classify nodes into low/medium/high frequency:

- a node is considered low frequency if the scales $m \in \{1, 2, 3\}$ contain most of the energy,
- a node is considered medium frequency if the scales $m \in \{4, 5, 6\}$ contain most of the energy,
- a node is considered high frequency if the scales $m \in \{6, 7, 9\}$ contain most of the energy.

For both signals from the previous question (smoothest and least smooth) as well as the first available timestamp, apply this procedure and display on the map the result (one colour per class).

Answer 5

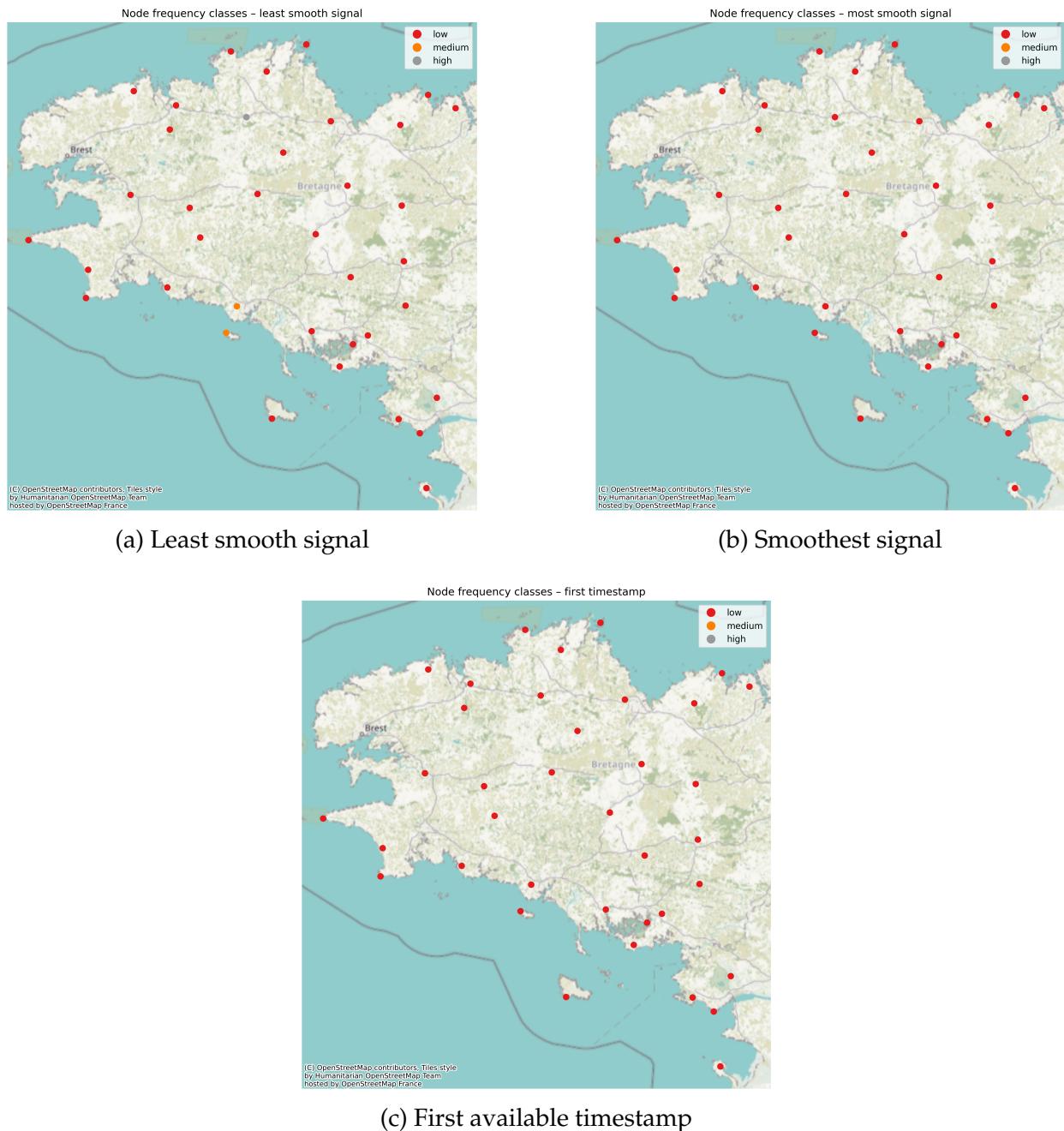


Figure 5: Classification of nodes into low/medium/high frequency

Question 6

Display the average temperature and for each timestamp, adapt the marker colour to the majority class present in the graph.

Answer 6

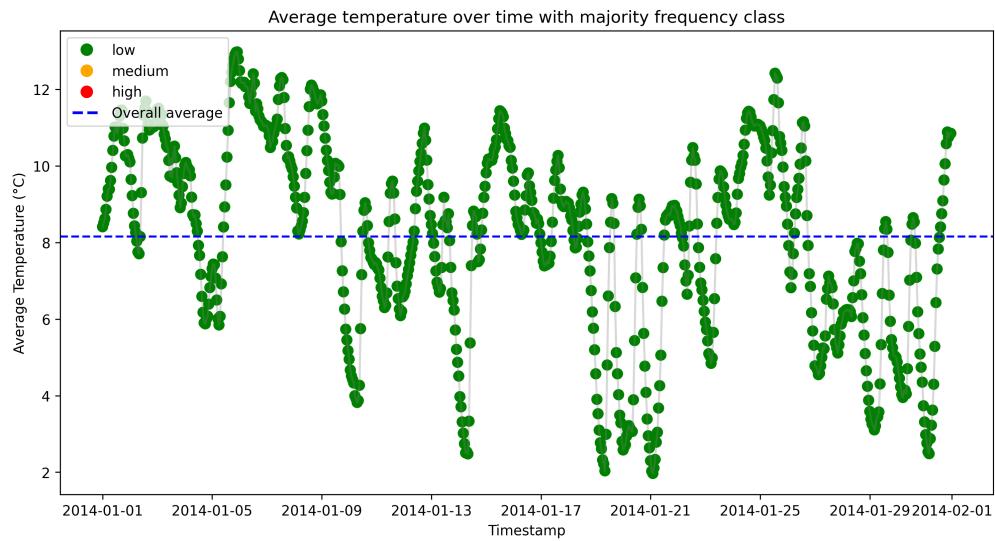


Figure 6: Average temperature. Markers' colours depend on the majority class.

Question 7

The previous graph G only uses spatial information. To take into account the temporal dynamic, we construct a larger graph H as follows: a node is now *a station at a particular time* and is connected to neighbouring stations (with respect to G) and to itself at the previous timestamp and the following timestamp. Notice that the new spatio-temporal graph H is the Cartesian product of the spatial graph G and the temporal graph G' (which is simply a line graph, without loop).

- Express the Laplacian of H using the Laplacian of G and G' (use Kronecker products).
- Express the eigenvalues and eigenvectors of the Laplacian of H using the eigenvalues and eigenvectors of the Laplacian of G and G' .
- Compute the wavelet transform of the temperature signal.
- Classify nodes into low/medium/high frequency and display the same figure as in the previous question.

Answer 7

Let's assume we have T time stamps and the size of the graph G is n . Let L_G be the laplacian of the initial graph, and let $L_{G'}$ be the laplacian of the graph of size T that is a line, so it contains 1's only in $(i, i+1)$ $1 \leq i < n$ and $(i, i-1)$ $1 < i \leq n$ (symmetrically). Let I_N and I_T be identity matrix. Then the new laplacian can be expressed as:

$$L_H = I_T \otimes L_G + \beta * L_{G'} \otimes I_N$$

Where β is the weight we want to add in the edges representing a link with itself in consecutive timestamps.

It is a well known fact that if v_i and w_j are the set of eigenvectors of L_G and $L_{G'}$ respectively, then $w_j \otimes v_i$ are the set of eigenvectors of L_H . If u_i and l_j are the eigenvalues of v_i and w_j respectively, then the eigenvalue of $w_j \otimes v_i$ is $u_i + \beta l_j$.

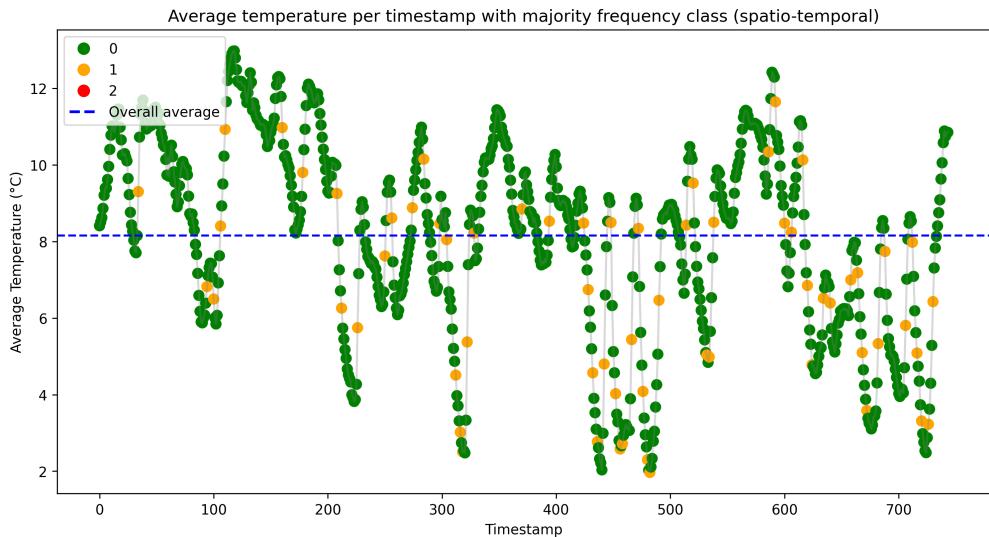


Figure 7: Average temperature. Markers' colours depend on the majority class.

In the figure, we can see that now we have the appearance of more middle frequencies, since before for every timestamp the majority class was low ones and now in a lot of timestamps we have middle frequencies as the dominant class. This is done since the new graph we have constructed takes into account temporal information as well, so the variation of the temperature in time is also reflected now, making it more complete. The values of the plot remain the same as it should be.

A DTMF Detection: Comprehensive Results

A.1 Training Set Performance

Overall Statistics:

- Total training signals: 100
- Total frequency pairs detected: 540
- Classification accuracy: 100% (540/540 correct)
- Unique symbols discovered: 16 (0-9, *, #, A-D)

A.2 Discovered Frequency Clusters

The unsupervised hierarchical clustering (coarse k=2 + elbow method) discovered the following structure:

Low Frequency Clusters (4 groups):

Cluster	Center (Hz)	Symbols	Standard (Hz)
1	697.2	1, 2, 3, A	697
2	761.4	4, 5, 6, B	770
3	845.6	7, 8, 9, C	852
4	931.0	*, 0, #, D	941

High Frequency Clusters (4 groups):

Cluster	Center (Hz)	Symbols	Standard (Hz)
1	1211.3	1, 4, 7, *	1209
2	1325.8	2, 5, 8, 0	1336
3	1479.3	3, 6, 9, #	1477
4	1617.5	A, B, C, D	1633

A.3 Per-Symbol Bayesian Model Parameters

The Bayesian classifier learned Gaussian distributions for each symbol's frequency pairs:

Symbol	Count	Low μ (Hz)	Low σ (Hz)	High μ (Hz)	High σ (Hz)
1	67	695.5	11.9	1210.7	12.1
2	32	696.0	12.0	1327.6	9.6
3	54	699.2	12.2	1479.5	11.9
4	20	764.4	9.8	1210.3	12.0
5	19	759.5	1.0	1326.9	8.9
6	13	759.5	1.0	1470.0	1.0
7	42	847.6	12.0	1211.6	12.2
8	47	842.4	11.9	1323.5	3.5
9	52	848.5	11.8	1482.2	12.2
*	41	931.0	1.0	1212.5	12.2
0	24	931.0	1.0	1327.1	9.1
#	21	931.0	1.0	1477.0	11.1
A	38	698.2	12.2	1618.3	5.5
B	13	761.4	6.5	1617.0	1.0
C	36	843.2	12.1	1617.0	1.0
D	21	931.0	1.0	1617.0	1.0

A.4 Accuracy Comparison with Standard DTMF

Symbol	Low Error (%)	Std Low (Hz)	High Error (%)	Std High (Hz)	Avg Error (%)
1	0.2	697	0.1	1209	0.2
2	0.1	697	0.6	1336	0.4
3	0.3	697	0.2	1477	0.2
4	0.7	770	0.1	1209	0.4
5	1.4	770	0.7	1336	1.0
6	1.4	770	0.5	1477	0.9
7	0.5	852	0.2	1209	0.4
8	1.1	852	0.9	1336	1.0
9	0.4	852	0.4	1477	0.4
*	1.1	941	0.3	1209	0.7
0	1.1	941	0.7	1336	0.9
#	1.1	941	0.0	1477	0.5
A	0.2	697	0.9	1633	0.5
B	1.1	770	1.0	1633	1.0
C	1.0	852	1.0	1633	1.0
D	1.1	941	1.0	1633	1.0
Mean Error:					0.7%

The discovered frequencies have an average error of less than 1% compared to standard DTMF, demonstrating that the unsupervised clustering successfully recovered the true DTMF structure.

A.5 Confusion Matrix

The Bayesian classifier achieved perfect classification on all training samples:

True \ Pred	#	*	0	1	2	3	4	5	6	7	8	9	A	B	C	D
#	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
*	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	67	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	54	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	38	0	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	0
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21

All 540 predictions were correct (diagonal elements only), demonstrating perfect separation of the 16 DTMF symbols in the learned frequency space.

A.6 Test Set Detailed Results

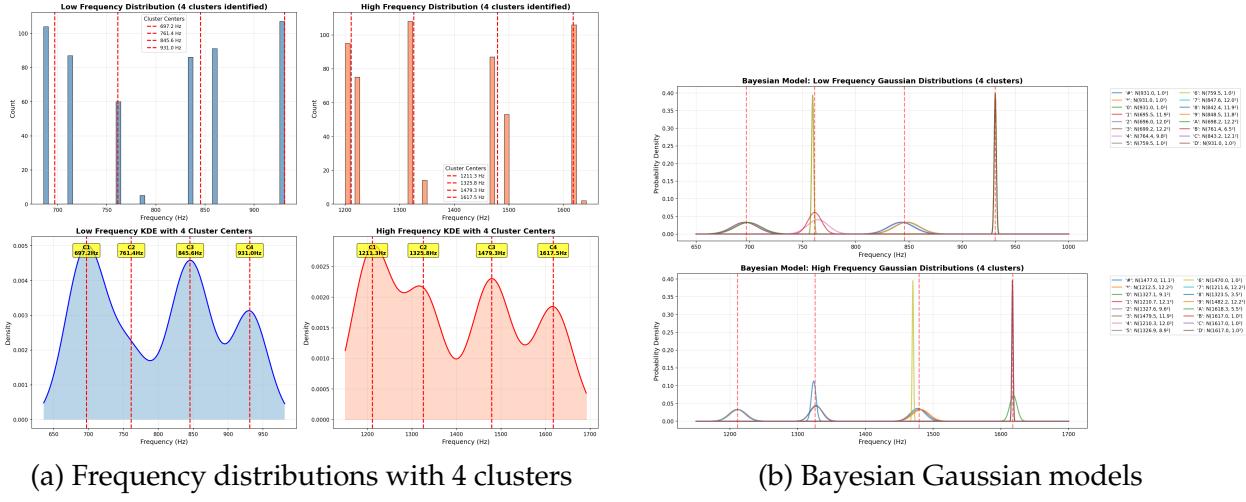
Test Signal 0: Predicted sequence 88133

Position	Symbol	Low Freq (Hz)	High Freq (Hz)	Confidence
1	8	857.5	1323.0	100.0%
2	8	857.5	1323.0	100.0%
3	1	686.0	1200.5	100.0%
4	3	686.0	1470.0	100.0%
5	3	686.0	1470.0	100.0%

Test Signal 1: Predicted sequence 11#2##

Position	Symbol	Low Freq (Hz)	High Freq (Hz)	Confidence
1	1	686.0	1200.5	100.0%
2	1	686.0	1200.5	100.0%
3	#	931.0	1470.0	100.0%
4	2	686.0	1323.0	100.0%
5	#	931.0	1470.0	100.0%
6	#	931.0	1470.0	100.0%

A.7 Additional Visualizations



(c) Complete DTMF analysis overview

Figure 8: Additional visualizations of the discovered DTMF structure