SPEAKER RECOGNITION SYSTEM

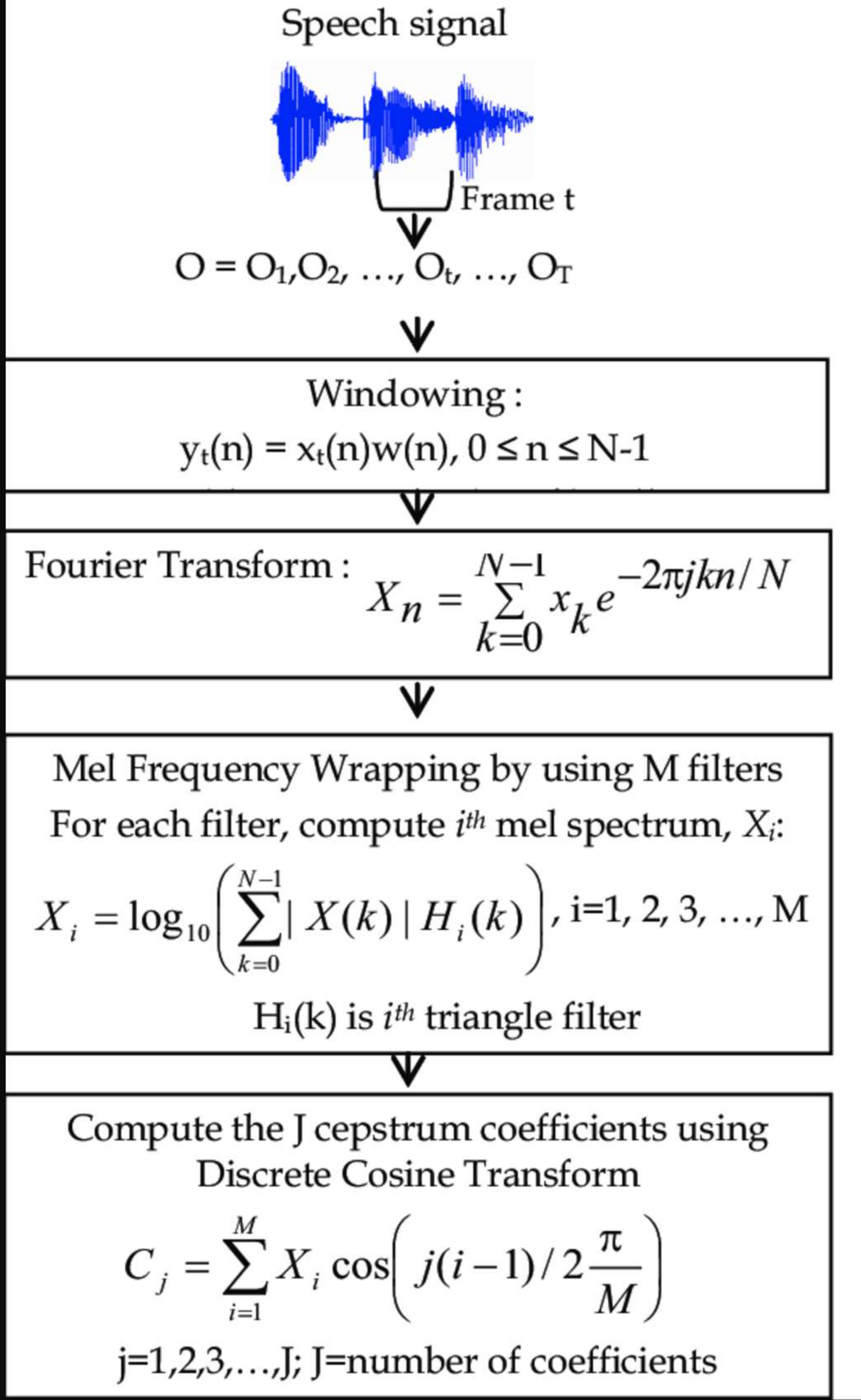
K. ROHITH MANIKANTA 22EC01010
N. CHERISH 22EC01016
Y. HARSHA VARDHAN 22EC01007

Speaker Recognition System Final Implementation

- Developed a Speaker Recognition System using MFCC and Pitch features.
- Implemented using MATLAB with support for different windowing functions.
- Trained 3 models based on:

1. Rectangular Window
2. Hanning Window
3. Hamming Window

- Classification performed using Support Vector Machines (SVM).
- Feature normalization applied to enhance consistency and performance.



Windowing in Feature Extraction

- Frame blocking followed by windowing improves feature stability.
- Window functions applied:

Rectangular

Simple shape, wider main lobe, higher side lobes cause more spectral leakage.

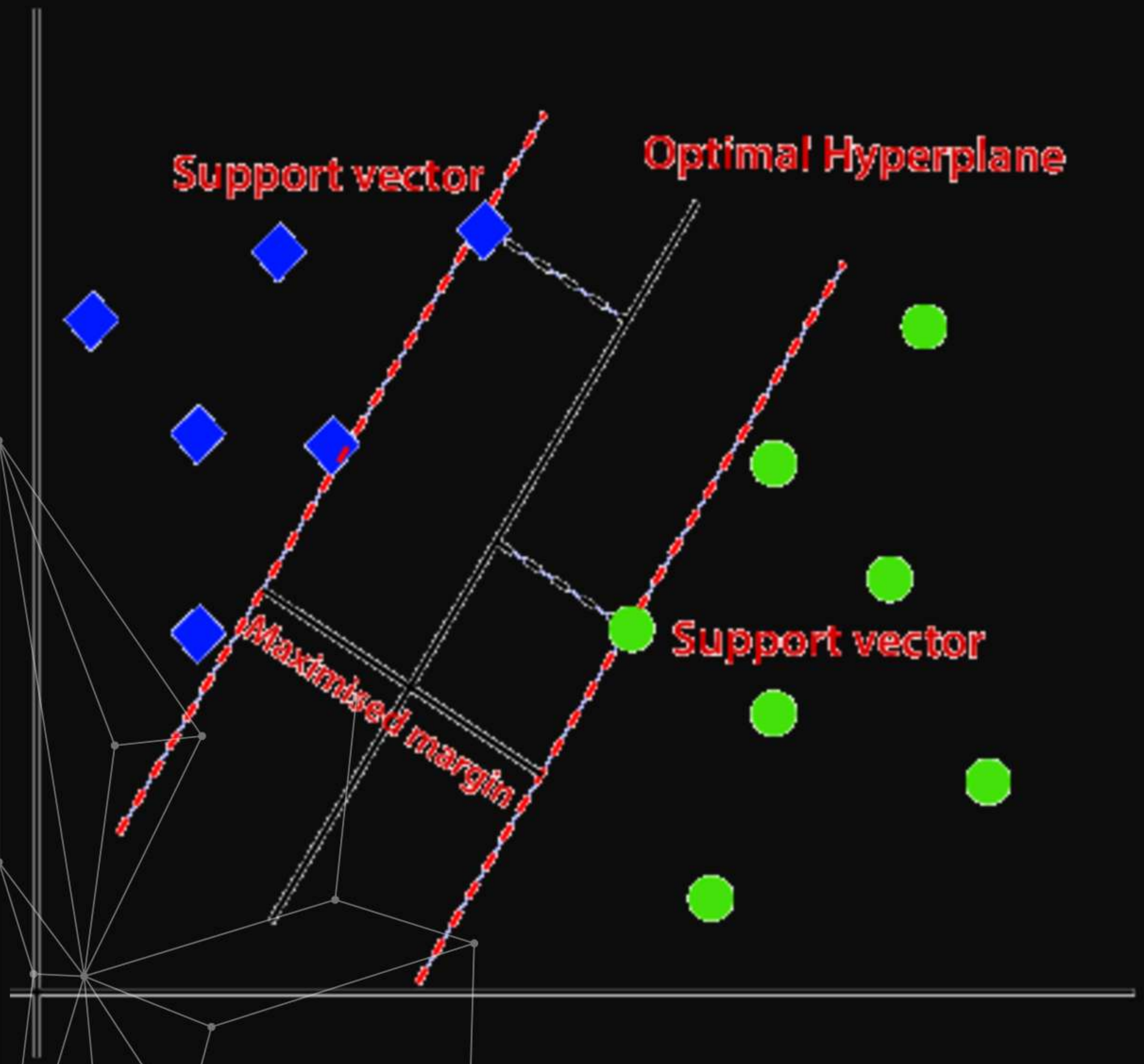
Hamming

Smoother taper reduces side lobes, improving signal-to-noise ratio by ~5% over rectangular.

Hanning

Even smoother taper, better side lobe attenuation, lowers spectral leakage for cleaner features.

- Speech signals are non-stationary.So instead of analyzing the whole signal at once, we break it into short segments (30 ms).
- Applying a window helps focus on that segment, reducing the effect of the rest.



Machine Learning: Classification Using SVM

- Trained a Multi-class SVM classifier using fitcecoc in MATLAB.
- Used linear kernel for fast, effective classification.
- Dataset split: 80% training, 20% testing.
- Normalized feature vectors ensure balanced scaling across speakers.
- SVM distinguishes speakers by finding optimal decision boundaries.

Recognition Accuracy & Results

- Tested system with 3 different windowing functions.
- Accuracy Results:

1. Rectangular: ~ *73.26%*
2. Hamming: ~ *78.36%*
3. Hanning: ~ *78.19%*

- Hanning window gave highest accuracy
- Evaluated with confusion matrices showing correct and misclassified predictions.

CONFUSION TABLES

Rectangular Window

True Class \ Predicted Class	Aasritha	Cherish	Hari	Harsha	Mohith	Rohith	keerthi
Aasritha	3167	75	61	110	23	25	132
Cherish	66	1897	108	244	31	145	178
Hari	90	78	2572	91	266	106	18
Harsha	103	304	104	1465	185	316	134
Mohith	201	26	262	245	1654	214	20
Rohith	53	237	46	295	137	1666	343
keerthi	57	143	23	45	1	181	2510

Hamming Window

True Class \ Predicted Class	Aasritha	Cherish	Hari	Harsha	Mohith	Rohith	keerthi
Aasritha	3386	36	46	43	48	16	18
Cherish	31	1963	103	214	17	215	127
Hari	42	73	2765	63	160	94	24
Harsha	41	243	101	1663	170	334	59
Mohith	131	17	218	243	1829	181	3
Rohith	48	247	32	303	146	1674	326
keerthi	2	109	19	6	2	181	2641

Hanning Window

True Class \ Predicted Class	Aasritha	Cherish	Hari	Harsha	Mohith	Rohith	keerthi
Aasritha	3361	74	46	57	19	22	15
Cherish	46	1991	108	167	15	179	163
Hari	40	87	2815	33	148	60	38
Harsha	50	273	103	1575	218	338	54
Mohith	116	15	249	233	1866	136	7
Rohith	38	235	23	252	185	1657	386
keerthi	4	100	12	7		90	2747

Limitations & Scope for Improvement

Current Limitations:

- Accuracy drops in noisy environments.
- GUI-Based Prediction is Inconsistent.
- Linear SVM may miss nonlinear patterns.
- No post-processing like speaker smoothing or Voice Activity detection (VAD).

Future Directions

- Integrate Deep Learning (CNNs, RNNs) for complex patterns.
- Add noise filtering and voice activity detection (VAD).
- Expand to multilingual or text-independent recognition.
- Port system to mobile platforms for real-world deployment.

THANK YOU