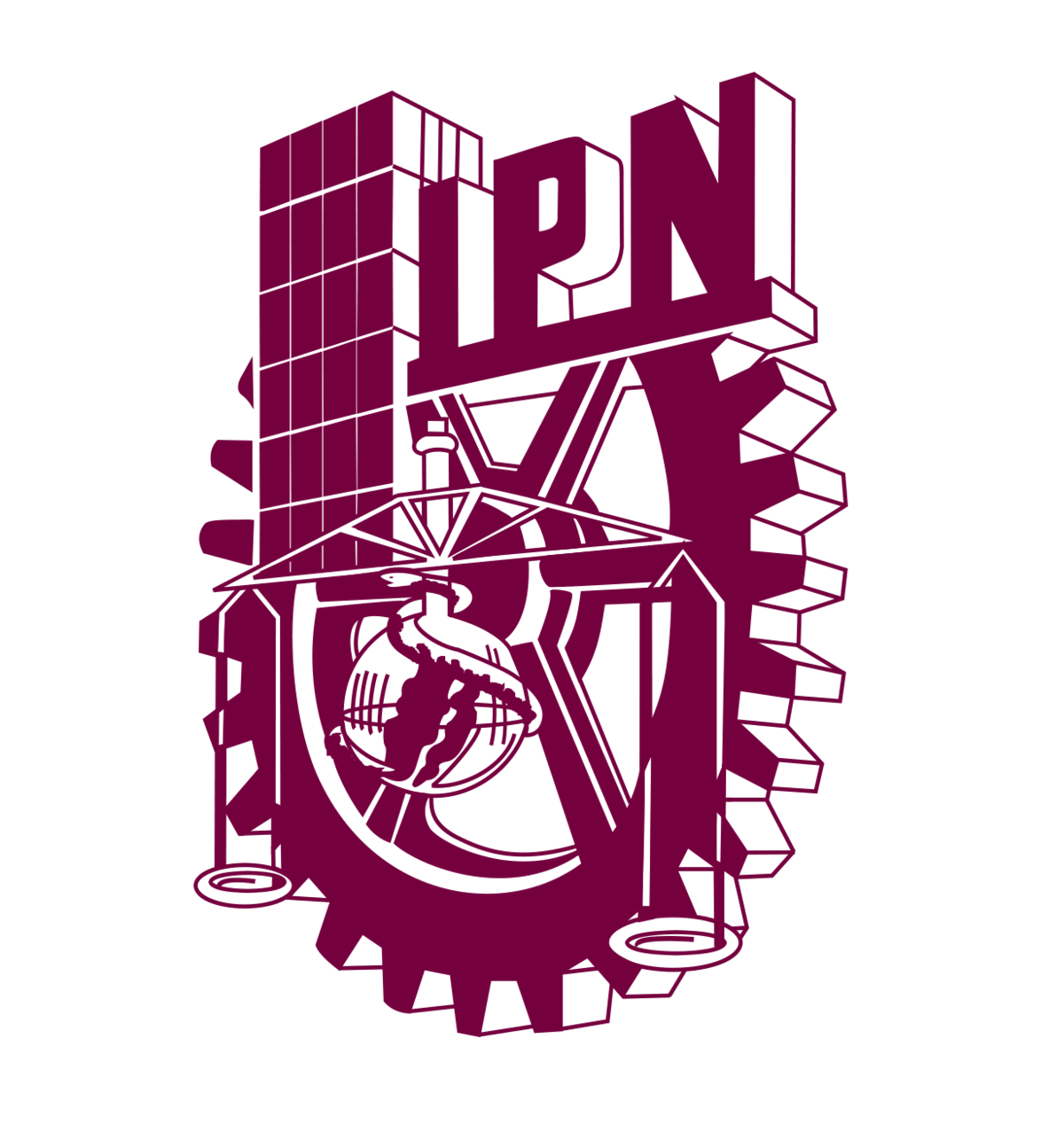
2024-02-27

**Instituto Politécnico Nacional**  
**Escuela Superior de Cómputo**  
Unidad Zacatenco  
Ingeniería en Inteligencia Artificial

**Reconocimiento de voz**

|  |  |
| --- | --- |
| **Tarea 2** | Métodos de extracción de características |
| **Fecha** | 2024-02-27 |
| **Grupo** | 7BM1 |
| **Profesor** | Carmona García Enrique Alfonso |
| **Alumno** | Juárez Botello Josué Adalid |
|  | Hola |

# Resumen

El presente documento aborda una exploración detallada de diversas técnicas avanzadas de extracción de características, fundamentales para el procesamiento de señales y el reconocimiento de patrones. En particular, se analizan el Análisis Cepstral, el Análisis en Escala de Mel, el Análisis de Bancos de Filtros, los Coeficientes Cepstrales en Escala de Mel (MFCC), y la Extracción de Características Basada en Kernel. Este estudio destaca la importancia de seleccionar la metodología adecuada para mejorar la precisión y eficiencia en la identificación y clasificación de señales en diversas aplicaciones, en este caso, el procesamiento de voz.

# Introducción

El análisis y procesamiento de señales se desempeñan en la interpretación y comprensión de datos en múltiples disciplinas. La extracción de características es la base para lograr altos niveles de precisión en tareas de reconocimiento de patrones y clasificación. Este documento se centra en cinco metodologías prominentes en el ámbito de la extracción de características: Análisis Cepstral, Análisis en Escala de Mel, Análisis de Bancos de Filtros, Coeficientes Cepstrales en Escala de Mel (MFCC), y Extracción de Características Basada en Kernel. Cada técnica ofrece una perspectiva única y ventajas específicas para abordar los desafíos inherentes al procesamiento de señales complejas, subrayando la necesidad de una elección metodológica informada basada en las características particulares de los datos en estudio.

# Análisis cepstral

## Conceptos del análisis cepstral

El análisis cepstral se fundamenta en varios conceptos clave. La Transformada de Fourier descompone una señal en sus frecuencias, mientras que el espectro de potencia, derivado de esta transformada, resalta la distribución de potencia de la señal a lo largo de las frecuencias. Aplicar el logaritmo al espectro de potencia comprime su rango dinámico, facilitando el análisis al mejorar la visibilidad de componentes de baja intensidad. Además, la aplicación de la Transformada de Fourier Inversa al espectro de potencia logaritmizado produce el cepstrum, que captura variaciones en los componentes espectrales y es crucial para tareas como la detección de fallas en maquinaria industrial.

## Técnicas del análisis cepstral

El análisis cepstral ha evolucionado con el tiempo, desarrollando técnicas avanzadas para abordar señales complejas. El Análisis Cepstral Homomórfico (HCA), por ejemplo, extiende el análisis cepstral tradicional mediante el filtrado homomórfico, separando componentes de alta y baja frecuencia y mejorando el análisis de señales complejas. Esta técnica se destaca en diagnósticos de maquinaria industrial, donde las señales no estacionarias son comunes. Otro enfoque es el Análisis de Quefrenia, que examina los coeficientes cepstrales en el dominio de la quefrenia, ofreciendo perspectivas sobre patrones temporales ocultos y beneficiando especialmente el análisis de señales no estacionarias en diagnósticos de maquinaria.

## Extracción de características

La extracción de características es una parte importante del análisis cepstral. Las representaciones tiempo-frecuencia, obtenidas mediante la combinación de análisis de espectrogramas y técnicas cepstrales, mejoran la extracción de características de señales no estacionarias al capturar tanto características espectrales como temporales. Explorar coeficientes cepstrales de orden superior permite una representación más detallada de la señal, capturando variaciones sutiles útiles para la detección de fallas.

# Introducción a la escala de mel

La Escala de Mel, derivada de la palabra "melodía", es una escala perceptual diseñada para reflejar la manera en que los humanos perciben el sonido. A diferencia de la escala de frecuencias en Hercios (Hz), que mide las frecuencias objetivamente, la Escala de Mel está diseñada para imitar la percepción auditiva humana, donde los cambios en tonos a diferentes frecuencias no son percibidos de manera uniforme. La relación entre la Escala de Mel y la frecuencia en Hz se establece mediante la asignación de 1000 mels a un tono de 1000 Hz, lo que facilita la comparación entre la percepción humana y las mediciones acústicas estándar.

Las conversiones entre la frecuencia en Hercios y la Escala de Mel se realizan mediante fórmulas matemáticas que reflejan cómo el oído humano discrimina entre tonos a diferentes frecuencias. Una de las fórmulas más utilizadas es:

Esta fórmula y sus variantes permiten una aproximación matemática a cómo los humanos perciben las diferencias de tono, especialmente útil en el procesamiento de señales de audio para aplicaciones como el reconocimiento de voz y la síntesis de música.

A pesar de su amplia adopción, la Escala de Mel ha sido objeto de críticas, principalmente por posibles sesgos introducidos durante los experimentos originales. Estas críticas subrayan la importancia de realizar investigaciones adicionales y refinamientos para asegurar que la escala refleje de manera precisa la percepción auditiva humana sin sesgos.

## Evolución histórica

La Escala de Mel ha evolucionado a través del tiempo, con varias fórmulas propuestas desde la década de 1930, reflejando diferentes aproximaciones para mapear las percepciones auditivas humanas a una escala mensurable. Desde las primeras curvas psicofísicas hasta las fórmulas más refinadas propuestas por Fant y posteriormente por Makhoul y Cosell, la Escala de Mel ha sido ajustada para ofrecer una representación más precisa de la percepción del tono, particularmente para frecuencias por debajo de 1000 Hz.

## Aplicaciones en el análisis de audio

La Escala de Mel es fundamental en la extracción de características de audio, donde se utilizan técnicas como los Coeficientes Cepstrales en Escala de Mel (MFCC) para el reconocimiento de voz, identificación de música, y otras aplicaciones de procesamiento de señales de audio. Al alinear las características extraídas con la percepción auditiva humana, la Escala de Mel permite que los sistemas de procesamiento de audio sean más efectivos y naturales en su interpretación de sonidos.

# Bancos de filtros

Un Banco de filtros es una herramienta que se utiliza para descomponer una señal en diferentes componentes de frecuencia. Es decir, la señal se divide en diferentes bandas de frecuencia y luego se procesa cada componente de manera individual.

Existen diferentes tipos de Banco de filtros, como DCT, Polyphase, Gabor, Mel y FBMC. Cada uno de ellos tiene su propia forma de dividir la señal en componentes de frecuencia.

## Bancos de filtros digitales

Durante las últimas dos décadas, los bancos de filtros han emergido como herramientas fundamentales en el procesamiento de señales, encontrando aplicación en campos tan diversos como la codificación de voz, el scrambling, la compresión de imágenes, y la transmisión de múltiples señales a través de un mismo canal. Su principal función radica en la capacidad de separar una señal en distintas bandas de frecuencia o, inversamente, componer varias señales en una única señal.

Los bancos de filtros se clasifican en dos categorías principales: los bancos de filtros de análisis, que dividen la señal en sub-bandas, y los bancos de filtros de síntesis, que combinan estas sub-bandas en una señal unificada. Esta dualidad facilita tanto el análisis como la síntesis espectral de señales, siendo particularmente útiles en la implementación eficiente de la Transformada Discreta de Fourier (DFT) mediante técnicas multifrecuencia.

## Codificación en sub-banda de señales de voz e imágenes

La codificación en sub-banda se basa en la premisa de que muchas señales tienen energía concentrada en regiones específicas del espectro. Por tanto, dividir la señal en sub-bandas permite asignar bits de manera más eficiente, basándose en la energía contenida en cada sub-banda. Este enfoque se alinea con las propiedades perceptuales del oído humano, permitiendo optimizar la calidad del sonido codificado mediante un uso eficiente del ancho de banda y la asignación de bits.

## Errores en el Banco QMF y su mitigación

El diseño de bancos de filtros, como el banco QMF (Quadrature Mirror Filter), implica desafíos significativos debido al aliasing y las distorsiones en amplitud y fase introducidas durante el procesamiento de señales. Sin embargo, es posible diseñar filtros que mitiguen estos efectos, logrando lo que se denomina reconstrucción perfecta. Este proceso involucra la cancelación del aliasing y la minimización de las distorsiones, permitiendo que la señal reconstruida mantenga una alta fidelidad respecto a la señal original.

## Simulación y aplicación práctica

La simulación de bancos de filtros, como se ejemplifica a través de Simulink, demuestra la capacidad de estos sistemas para procesar y reconstruir señales con alta precisión. Mediante la selección adecuada de filtros y la configuración óptima de parámetros, se logra una reconstrucción perfecta de la señal de entrada, destacando la eficacia de los bancos de filtros digitales en aplicaciones prácticas.

Los bancos de filtros digitales representan una herramienta invaluable en el procesamiento de señales, permitiendo una manipulación eficiente y efectiva de la información en el dominio de la frecuencia. A través de la codificación en sub-banda y el diseño cuidadoso de filtros, es posible superar desafíos como el aliasing y las distorsiones, logrando una alta calidad en la señal reconstruida. Las simulaciones demuestran la aplicabilidad de estos conceptos en la práctica, subrayando la importancia de los bancos de filtros en la tecnología de procesamiento de señales contemporánea.

# Kernel PCA

Mientras que el PCA tradicional es altamente efectivo para transformaciones lineales de datos, puede no capturar la estructura subyacente de conjuntos de datos complejos y no lineales. Para abordar este problema, introducimos el análisis de componentes principales con kernel (KPCA). KPCA se basa en la intuición de que muchos conjuntos de datos que no son separables linealmente en su dimensión actual, pueden serlo al ser proyectados en un espacio de mayor dimensión.

## Funciones Kernel

La función kernel es una función que usamos para transformar los datos originales a un espacio de mayor dimensión donde se vuelven linealmente separables. Generalmente, los kernels se emplean en varios modelos de aprendizaje automático como las máquinas de vectores de soporte (SVM), la regresión ridge con kernel (KRR), el k-means kernelizado, entre otros. s. Aquí mencionaremos algunas de ellas.

Algunas de las ventajas del kernel PCA incluyen la transformación a dimensiones superiores, la capacidad de capturar relaciones complejas y no lineales, y la flexibilidad, lo que lo hace adaptable a diversos tipos de datos.

El kernel PCA encuentra aplicaciones en varios campos donde se deben capturar relaciones no lineales en los datos. Algunas aplicaciones notables incluyen el reconocimiento de imágenes, el procesamiento del lenguaje natural (NLP), la genómica y bioinformática, y el modelado financiero.

# Cepstrum de Frecuencia Mel y sus Coeficientes

En el procesamiento de sonido, el Cepstrum de Frecuencia Mel (MFC) es una representación del espectro de potencia a corto plazo de un sonido. Se basa en una transformada coseno lineal de un espectro de potencia logarítmico en una escala de frecuencia mel no lineal. Los Coeficientes Cepstrales de Frecuencia Mel (MFCC, por sus siglas en inglés) son coeficientes que conforman un MFC, derivados de una representación cepstral del clip de audio, ofreciendo una "espectro-de-espectro" no lineal. La principal diferencia entre el cepstrum y el MFC radica en que, en el MFC, las bandas de frecuencia están espaciadas igualmente en la escala mel, aproximando más de cerca la respuesta del sistema auditivo humano que las bandas de frecuencia espaciadas linealmente usadas en el espectro normal.

## Historia

Paul Mermelstein es típicamente acreditado con el desarrollo del MFC. Mermelstein acredita a Bridle y Brown por la idea de usar un conjunto de coeficientes de forma de espectro ponderados por la transformada coseno de las salidas de un conjunto de filtros de banda pasante espaciados de manera no uniforme.

## Proceso de Derivación de los MFCC

El proceso común para derivar los MFCC es el siguiente:

1. Tomar la transformada de Fourier de un extracto de señal con ventana.
2. Mapear las potencias del espectro obtenido a la escala mel, utilizando ventanas triangulares superpuestas o ventanas coseno superpuestas.
3. Tomar los logaritmos de las potencias en cada una de las frecuencias mel.
4. Realizar la transformada coseno discreta de la lista de potencias logarítmicas mel, como si fuera una señal.
5. Los MFCC son las amplitudes del espectro resultante.

## Aplicaciones de los MFCC

Los MFCC se utilizan comúnmente como características en sistemas de reconocimiento de voz, tales como sistemas automáticos que reconocen números hablados en un teléfono. Además, los MFCC están encontrando cada vez más aplicaciones en la recuperación de información musical, como la clasificación de géneros, medidas de similitud de audio, entre otros.

## MFCC para el Reconocimiento de Hablantes

Dado que las bandas de frecuencia Mel se distribuyen uniformemente en el MFCC y son muy similares al sistema vocal humano, el MFCC puede usarse eficientemente para caracterizar hablantes. Por ejemplo, puede utilizarse para reconocer los detalles del modelo del teléfono celular del hablante y, además, los detalles del propio hablante.

## Inversión y Sensibilidad al Ruido

Es posible invertir aproximadamente un MFCC a audio en cuatro pasos, los cuales incluyen la transformada coseno discreta inversa para obtener un espectrograma log-potencia mel, y la reconstrucción de la fase y síntesis de audio usando Griffin-Lim. Los valores MFCC no son muy robustos en presencia de ruido aditivo, por lo que es común normalizar sus valores en sistemas de reconocimiento de voz para disminuir la influencia del ruido.

# Conclusiones

La investigación de las distintas técnicas de extracción de características revela la complejidad y la profundidad del campo del procesamiento de señales. Cada método examinado, desde el Análisis Cepstral hasta la Extracción de Características Basada en Kernel, contribuye de manera significativa a la optimización del reconocimiento y clasificación de patrones. Los MFCC, en particular, demuestran ser excepcionalmente eficaces para el procesamiento de señales de audio, mientras que los métodos basados en Kernel ofrecen una solución robusta para capturar la no linealidad en conjuntos de datos complejos. Esta diversidad de enfoques subraya la importancia de una selección cuidadosa de técnicas de extracción de características, adecuadas a las necesidades específicas de cada aplicación.