Nombre de la materia:	Síntesis y reconocimiento de voz
Clave:	IA7710-T
No. de horas/semana:	3
Total de horas:	48
No. de créditos:	6
Prerrequisitos:	Cálculo IV (CB0003-T), Programación de Computadoras (CI0000-T)

Objetivo general: Que el alumno comprenda y domine las técnicas que se utilizan en los diferentes componentes de un sistema de reconocimiento de voz así como de un sintetizador de voz de manera que pueda aplicarlos al desarrollo de interfaces hombre-máquina.

Contribución a los atributos de egreso y su nivel de aportación

- **AE1.** Aplicar los conocimientos de ingeniería adquiridos durante sus estudios para elaborar proyectos de ingeniería que resuelvan problemas específicos. **(Avanzado)**
- **AE2.** Identificar, formular y resolver problemas de ingeniería mediante un pensamiento crítico y asertivo, basados en los principios de ciencias básicas e ingeniería. **(Avanzado)**
- **AE3.** Presentar y defender su trabajo en diversos foros, tanto académicos como profesionales. **(Medio)**
- **AE4.** Intercambiar su conocimiento y puntos de vista con profesionales del área e integrarse en equipos de trabajo multidisciplinarios. **(Inicial)**
- **AE5.** Respetar su entorno social y disciplinar, enmarcado siempre por valores humanos y de ética profesional, con una actitud creativa y positiva para enfrentar nuevos retos. **(Inicial)**

Programa sintético

1. Fundamentos de Procesamiento Digital de Señales ..... 10 hrs.
2. Modelos digitales para la señal de voz y teoría acústica ..... 8 hrs.
3. Procesamiento de la señal de voz en el dominio del tiempo ..... 6 hrs.
4. Procesamiento de la señal de voz en el dominio de la frecuencia ..... 8 hrs.
5. Procesamiento Homomórfico de la señal de voz ..... 6 hrs.
6. Codificación Lineal Predictiva ..... 8 hrs.
7. Revisión Final de Trabajos ..... 2 hrs.

Total: 48 hrs.

Programa desarrollado

1. Fundamentos de Procesamiento Digital de Señales ..... 10 hrs.
   - 1.1 Introducción
   - 1.2 La transformada Z. Definición, Propiedades de la transformada Z
   - 1.3 La transformada de Fourier. Definición a partir de la Transformada Z. La transformada Discreta de Fourier (DFT) y sus propiedades y limitaciones.

- 1.3.1 Diseño de un filtro usando la DFT y la DFT inversa (1er Trabajo)
- 1.4 Filtros digitales, causalidad, linealidad, sistemas invariantes en tiempo
  - 1.4.1 Filtros FIR
  - 1.4.2 Diseño de un filtro FIR (determinación de los coeficientes del filtro mediante la técnica de “Window design” (2º trabajo)
- 1.5 Filtros IIR
  - 1.5.1 Implementación de filtros IIR con requerimientos mínimos de almacenamiento
  - 1.5.2 Filtros Butterworth
- 1.6 Muestreo. Regla de Nyquist. El efecto Aliasing. Filtro antialiasing
- 1.7 Sub-muestreo crítico.

2. Modelos digitales para la señal de voz y teoría acústica ..... 8 hrs.
   - 2.1 Anatomía y Fisiología de la producción de la voz
   - 2.2 Fonética: Sonidos Nasales, Vocalizados, fricativos y plosivos
   - 2.3 Teoría Acústica
   - 2.4 El concepto de formantes
     - 2.4.1 Mapa de formantes para los sonidos vocalizados mas comunes (3er trabajo)
     - 2.4.2 Técnica de síntesis de sonidos vocalizados usando los primeros tres formantes (4º
   - 2.5 Propagación del sonido en un tubo sin pérdidas. Modelado a partir de ecuaciones diferenciales
   - 2.6 Modelo del tracto vocal basado en tubos sin pérdidas de sección uniforme concatenados
     - 2.6.1 Determinación de los coeficiente de reflexión de la onda de sonido a partir de las condiciones de frontera
     - 2.6.2 Sintesis de voz usando los coeficientes reflejantes
   - 2.7 Pulso glotal
   - 2.8 Modelo completo del tracto vocal

3. Procesamiento de la señal de voz en el dominio del tiempo ..... 6 hrs.
   - 3.1 Energía de tiempo corto
   - 3.2 Régimen de cruces por cero de tiempo corto
   - 3.3 Estimación del tono
   - 3.4 Entropía de la señal de voz
   - 3.5 Segmentación de palabras aisladas (5º trabajo)
   - 3.6 Autocorrelación de tiempo corto.
     - 3.6.1 Algoritmo de discriminación silencio/voz vocalizado/no-vocalizado y estimación del tono mediante autocorrelación y “center clipping” (6º trabajo)
     - 3.6.2 Autocorrelación Modificada de tiempo corto (Correlación cruzada entre marcos consecutivos)
     - 3.6.3 Algoritmo de Blankenship para el cálculo eficiente de la autocorrelación

4. Procesamiento de la señal de voz en el dominio de la frecuencia ..... 8 hrs.
   - 4.1 Transformada de Fourier de tiempo corto
   - 4.2 Aplicación de ventanas (Hann, Hamming, Parzen, etc) y el efecto “leakage” (escurrimiento)
   - 4.3 Determinación del espectrograma mediante la Transformada de Fourier de tiempo corto (7º trabajo)
   - 4.4 Determinación del espectrograma de la señal de voz mediante Bancos de Filtros

4.5 Escala de Bark
4.6 Doblado Dinámico en Tiempo
4.7 Distancia de Manhatan, Distancia Euclidiana, Distancias LP y Distancia coseno
4.8 Criterio del vecino más cercano y criterio de los K-vecinos
4.9 Implementación de un sistema de reconocimiento de palabras aisladas mediante espectrogramas (8º trabajo)
4.10 Evaluación de la transformada continua de Fourier en los ceros de los polinomios de Hermite

5. Procesamiento Homomórfico de la señal de voz ..... 6 hrs.
5.1 El Cepstrum y el Cepstrum complejo
5.2 Aplicación a la estimación
5.3 Aplicación a la estimación de los formantes
5.4 Determinación de los coeficientes MFCC (Mel-frequency Cepstral coefficients)
5.4.1 La escala de Mel
5.4.2 Filtros triangulares de Mel
5.4.3 Transformada Coseno de la salida de los filtros de Mel
5.5 Sistema de reconocimiento de voz usando los MFCC (9º trabajo)

6. Codificación Lineal Predictiva ..... 8 hrs.
6.1 Principios del análisis lineal predictivo
6.1.1 Método de la Autocorrelación
6.2 Solución de las ecuaciones LPC
6.2.1 Método recursivo de Durbin para solucionar las ecuaciones de autocorrelación
6.3 La señal de error y su relación con la estimación del pulso glotal y aplicación a identificación de individuos por su voz
6.4 Interpretación en el dominio de la frecuencia del análisis lineal predictivo
6.5 Interpretación en el dominio de la frecuencia del error de predicción
6.6 Relación entre el análisis predictivo el los modelos de tubos sin pérdidas
6.7 Los coeficientes PARCOR
6.8 Síntesis de voz mediante parámetros LPC
6.9 Determinación del tono usando los coeficientes LPC
6.10 Análisis de formantes usando coeficientes LPC
6.11 Reconocedor de palabras aisladas basado en coeficientes LPC
6.11.1 Distancia de Itakura
6.12 Síntesis de voz basado en coeficientes LPC (10º trabajo)

7. Revisión Final de Trabajos ..... 2 hrs.

Bibliografía básica:

- Lawrence R. Rabiner & Ronald W. Schafer. Digital Processing of Speech Signals. Prentice Hall, 1978.
- Philip Lieberman & Sheila E. Blumstein. Speech physiology, speech perception and acoustic phonetics Cambridge University Press. 1988.
-John G. Proakis & Dimitris G. Manolakis. Digital Signal Processing 4 th Ed. Prentice Hall. 2006.

- Gilbert Strang & Truong Nguyen. Wavelets and Filter Banks. Wellesley-Cambridge Press, 1996.

Metodologías de enseñanza-aprendizaje:

- Revisión de conceptos, análisis y solución de problemas en clase (X)
- Lectura de material fuera de clase (X)
- Ejercicios fuera de clase (tareas) (X)
- Elaboración de reportes técnicos o proyectos (X)

Metodologías de evaluación:

- Asistencia (X)
- Tareas (X)
- Elaboracion de reportes técnicos o proyectos (X)

Revisores:

Programa propuesto por Dr. José Antonio Camarena Ibarrola para el semestre Agosto 2011/ Febrero de 2012