22.47 Procesamiento de voz

Proyecto 2

Problema avanzado de procesamiento de voz.

Los proyectos son grupales (grupos de dos). Deben presentarlos en la fecha estipulada.

Una semana después de tal fecha deben presentar un informe en formato IEEE. Diagramen el informe en la forma clásica de un trabajo de investigación: introducción (contexto, antecedentes, presentación de la proposición), desarrollo (actividades realizadas), resultados obtenidos, conclusiones (análisis de resultados y trabajo a futuro), y bibliografía.

Es importante que en el trabajo **justifiquen** todas las decisiones de investigación y desarrollo tomadas: enfoques, procedimientos, parámetros.

Recomendamos los siguientes temas:

- **Reconocimiento de idioma**. Deberán segmentar una grabación en: idioma 1, idioma 2 o indeterminado (silencio, por ejemplo). Una lista con bases de datos: https://github.com/coqui-ai/open-speech-corpora.
- **Reconocimiento de dígitos**. Deberán reconocer secuencias de dígitos (0 a 9) de una grabación. Bases de datos: https://github.com/soerenab/AudioMNIST.
- **Reconocimiento de estado de ánimo**. Deberán segmentar una grabación en estados de ánimo: neutro, alegre, enojado, disgustado, dormido e indeterminado (silencio, por ejemplo). Bases de datos: https://github.com/numediart/EmoV-DB, https://www.nature.com/articles/s41562-019-0533-6, https://github.com/SenticNet/MELD.
- **Verificación de locutores**. Deberán realizar un control de acceso por voz, determinando si la voz es de quien dice ser. Para determinar la calidad del sistema deben representar las curvas DET y compararlas con sistemas state of the art. Bases de datos: https://github.com/JRMeyer/open-speech-corpora.
- **Speaker diarisation**. Deberán segmentar una grabación en: locutor 1, locutor 2, o locutor desconocido (silencio por ejemplo). El sistema no debe ser entrenado con las voces de los locutores sino determinarlos sobre la marcha. Bases de datos: NIST 2008 (pedir a la cátedra).
- **Neural speech coding**. Deberán codificar la voz en forma inteligible a ultra baja tasa de bits. Objetivo: 150 bps. Keywords: autoencoder, LPCNet.
- **Voice style transfer**. Deberán transferir el estilo del habla de un locutor destino en una grabación de un locutor fuente. Paper de referencia: https://ebadawy.github.io/post/speech_style_transfer/.