

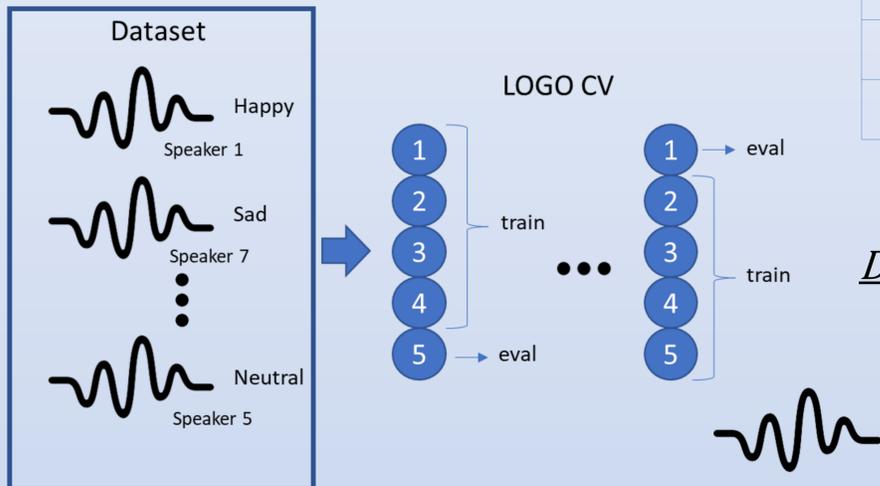
Detección automática de emociones a partir de la voz

Combinación de bases de datos para aumentar el entrenamiento

Miguel Ángel Pastor, Dayana Ribas, Alfonso Ortega
ViVoLab, I3A, Universidad de Zaragoza

SISTEMA

Pre-procesamiento



DATASETS

EmoDB	RAVDESS	IEMOCAP
Alemán	Inglés	Inglés
45 min	3 horas	12 horas
10 actores	24 actores	5 sesiones con 10 actores

Detección

Representación

- GeMAPS
- eGeMAPS
- Compare

Normalización

- Zscore

Clasificación

- SVM

REPRESENTACIONES

GeMAPS (62) y eGeMAPS (88)

Parámetros espectrales

- Alfa Ratio, Spectral Slope, Hammarberg Index
- Diferencia armónica H1-H2 y H1-A3
- Energía formantes 1, 2 y 3
- Flujo espectral y MFCC 1-4 (eGeMAPS)

Parámetros frecuenciales

- Pitch, Jitter, Formantes 1,2 y 3
- BW formantes 1, 2 y 3 (en GeMAPS, solo el 1º)

Parámetros de energía

- Shimmer, Loudness, HNR

ComPARE (6373)

Parámetros espectrales

- Espectro RASTA: bandas 1-26, MFCC 1-14
- Roll off espectral (0.25, 0.5, 0.75, 0,9)
- Varianza, skewness y kurtosis espectral
- Flujo, centroide, entropía y pendiente espectral

Parámetros de voz

- HNR, Jitter y Shimmer, Pitch, Probabilidad de voz

Parámetros de energía

- Energía RMS, Ratio de Cruces por Cero
- Loudness, Suma del Espectro RASTA

RESULTADOS DE CLASIFICACION

- Mejora del rendimiento del 1%, incluso en IEMOCAP, que es mucho mayor que los otros dos datasets combinados.
- EmoDb, aún estando en alemán, mejora sus resultados y tiene buenos resultados entrenando sólo con IEMOCAP

