International Conference on Speech and Computer, SPECOM
2025, Szeged, Hungary

[1] Pattern Recognition and Artificial Intelligence Group, University of Alicante, Spain
[2] Université Grenoble Alpes, CNRS, GIPSA-Lab, Grenoble, France

Whistler Identification in Whistled Spanish (Silbo): A Case Study

Alejandro López-García[1], María Alfaro-Contreras[1], Julien Meyer[2] and Jose J. Valero-Mas[1]
alg166@gcloud.ua.es - malfaro@dlsi.ua.es - julien.meyer@cnrs.fr - jjvalero@dlsi.ua.es

1) Silbo

- **Whistled Spanish** from **La Gomera** (Canary Islands).
- Most **representative** language of its kind, with an active community of **22000 whistlers**.
- Computational exploitation has been hindered due to **data scarcity**.

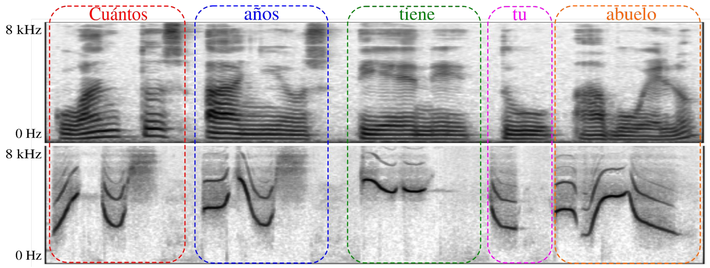


Fig. 1: Spectrogram comparison in spoken Spanish (top) versus Silbo (bottom).

2) Objectives

1. Present the **first approach** to Speaker Identification for **Silbo** speech.
2. Find the most **suitable** combination of **parameters** for the task.

3) Scheme proposed

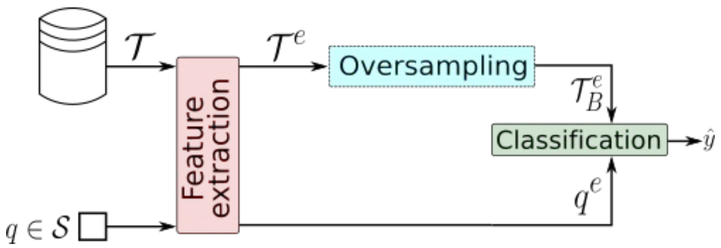


Fig. 2: Graphical description of the proposed methodology for the task.

- $\mathcal{T}$ **:** initial audio recordings.
- $\mathcal{T}^e$**:** embedded audios after undergoing the *Feature extraction* phase.
- $\mathcal{T}_B^e$**:** embedded representations artificially balanced after the *Oversampling* stage.
- q^e **:** embedded audios from a disjoint test partition.
- $\hat{y}$ **:** final speaker prediction by the classifier.

4) Experimental set-up

- **Data:** Only existing Silbo recordings (Jakubiak).

Number of Samples	Total Duration	Average Duration
529	1h 2min 3.3s	7.0s ± 2.9s

Table 1: Details of the Silbo dataset compiled by Jakubiak.

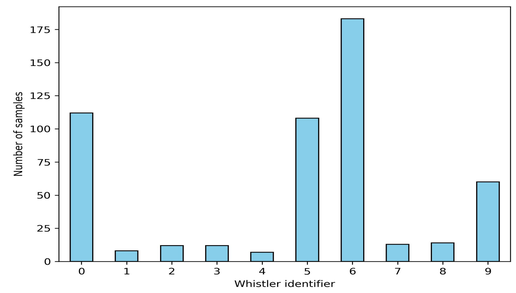


Fig. 3: Speaker identifier distribution of the Jakubiak Silbo dataset.

- ***Feature extraction*:** MFCC, Wav2vec, Whisper
- ***Oversampling*:** SMOTE, B-SMOTE, ADASYN
- ***Classification*:** GMM, *k*NN, MLP, RaF, SVM
- **Metrics**: 5-fold cross-validation + macro average F_1-score.

5) Results

Feature extraction + classifier

	GMM	*k*NN	MLP	RaF	SVM
MFCC					
Base (20)	72.7	75.2	80.5	**68.8**	85.8
Base +Δ (40)	**79.3**	**76.2**	**82.5**	66.1	**87.9**
Base +Δ + Δ² (60)	78.0	75.6	78.8	68.3	86.3
Wav2vec					
64	9.6	**14.0**	12.6	10.3	12.5
256	**11.3**	13.9	12.0	11.9	**13.8**
1024	6.5	11.3	12.3	**13.7**	**13.8**
4096	6.8	12.4	**16.0**	12.0	13.0
Whisper					
Tiny (384)	49.5	39.6	**57.1**	26.3	83.7
Base (512)	**65.4**	**47.4**	45.2	**36.2**	**83.8**
Small (768)	23.0	26.3	31.2	21.7	71.3
Medium (1024)	59.0	35.4	34.4	32.2	74.5

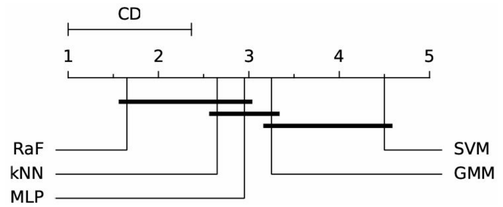
Table 2: Average test results for the classifiers evaluated with respect to the embedding strategy, without oversampling.

Oversampling

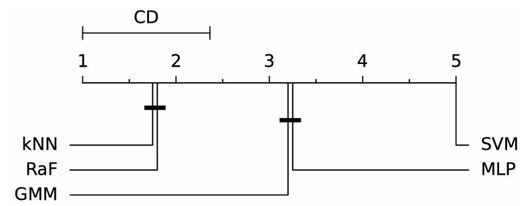
	Oversampling method			
	None	SMOTE	B-SMOTE	ADASYN
MFCC				
GMM	79.3	78.5	79.3	81.7
*k*NN	76.2	79.6	78.2	79.0
MLP	82.5	72.1	81.3	69.0
RaF	68.8	73.2	73.3	73.5
SVM	**87.9**	**86.2**	**85.8**	**86.2**

Table 3: Average test results for the best configurations in Table 2, applying oversampling to the training data.

Nemenyi test



(a) Feature representation using MFCC.



(b) Feature representation using the Whisper encoder.

Fig. 4: Results of the Nemenyi post-hoc test across the different classification strategies for the best MFCC and Whisper representations.

6) Conclusions

- **Automatic Speaker Identification** on Silbo can successfully be carried out, with **competitive performance** under **data-scarcity** conditions.
- **Best** results are obtained using **MFCC** with **SVM**, achieving **87.9%** F_1-score.
- **Oversampling** methods yield general **improvements** in classification performance, highly depending on the specific parameter configuration.
- **Future work** will focus on **expanding** the existing **data** collection.

Universitat d'Alacant
Universidad de Alicante

Funded by the European Union
NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia