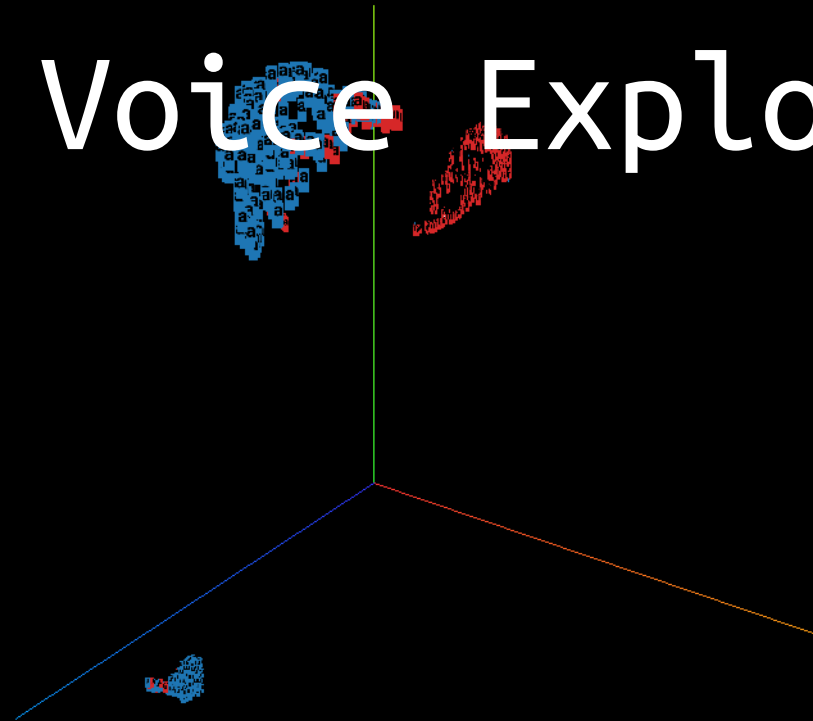




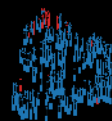
VoxRimor

—

The Voice Explorer



Alessandro De Luca



LIRI - CL UZH

Schedule

1. Motivation
2. What is *VoxRimor* and where to find it
3. How to install *VoxRimor*
4. DEMO 1: Visualizer mode (Saarbrücken data)
5. DEMO 2: Speaker embeddings mode (CANDOR)
6. Feature extraction mode: parameter settings
7. DEMO 3: Feature extraction mode (CANDOR)
8. DEMO 4: Filtering mode
9. Future plans, goals, and Q&A

Motivation

- With the recent and fast improvements in computational power and storage capacity, driven by the need of more and better descriptive data that covers a large study population, speech corpora have started to become increasingly large and complex.
- E.g. CANDOR:
 - Multi-modal data: archived (.zip) = 1TB
 - Only full-length WAV files \cong 440 GB
 - Processed (VAD + filtering and snipping) 28480 files [942 speakers]

Motivation

**Better access to computational resources
for all!**



What is *VoxRimor*?

- *VoxRimor* started as an interactive visualization tool, BUT...
- An interactive tool for exploration and subsetting of large corpora
- A powerful automatic feature extraction tool

Where to find *VoxRimor*:

<https://github.com/delale/VoxRimor>

Installation demo...

- Clone the [repository](#)
- Install [miniconda](#) or [Anaconda](#)
- Create the *VoxRimor* conda environment
- Activate the environment
- Have fun!

```
🍏 ~ [base]
→ git clone https://github.com/delale/VoxRimor.git
```

```
🍏 ~ [base]
→ conda env create -f voxrimor_env OSX.yml
```

```
🍏 ~ [base]
→ conda activate voxrimor
```

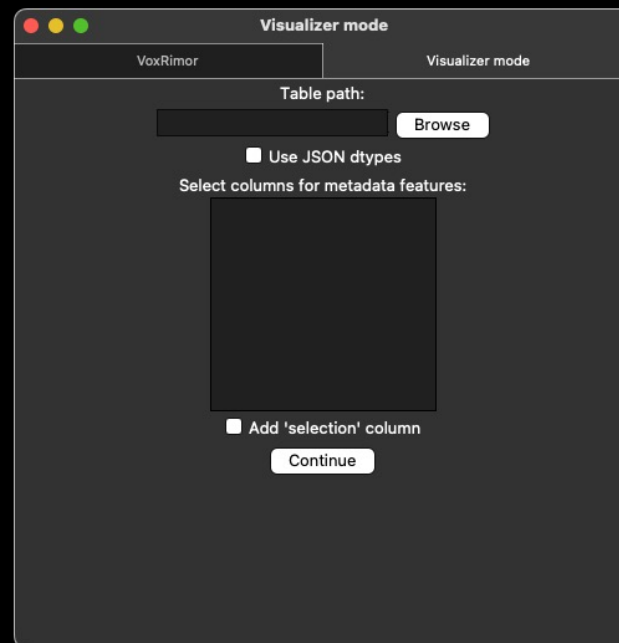
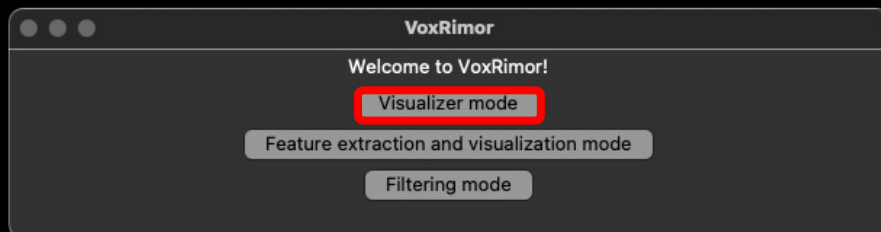
```
🍏 ~ [voxrimor]
→ which python3
/Users/delale/miniconda3/envs/voxrimor/bin/python3
```

```
🍏 ~ [voxrimor]
→ cd VoxRimor
```

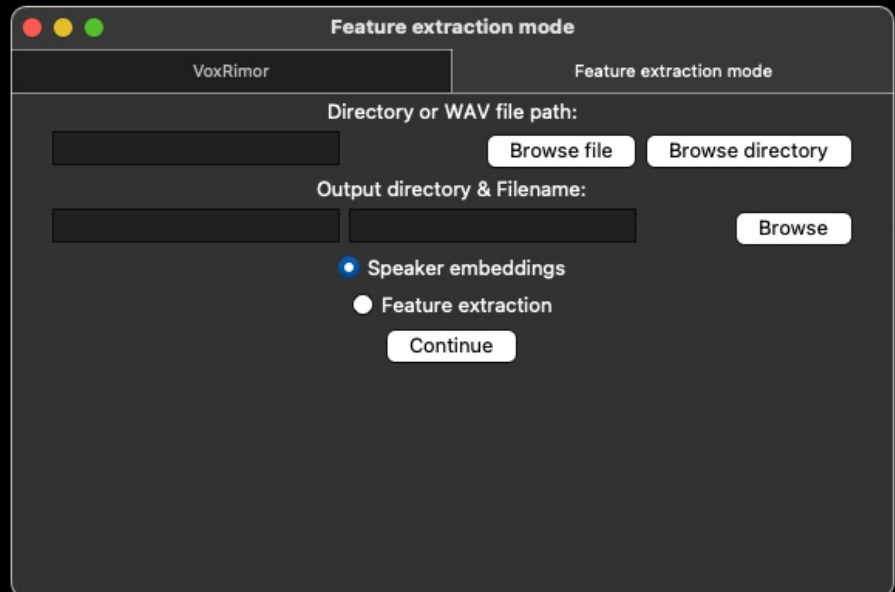
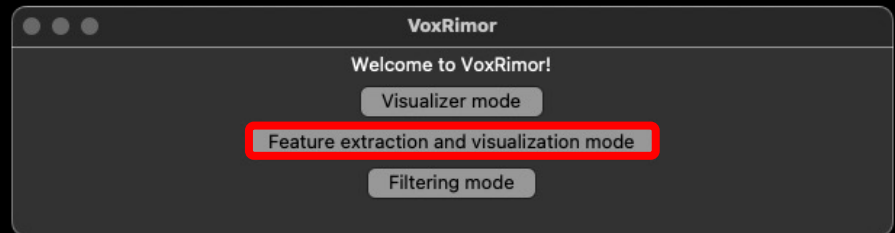
```
🍏 ~/VoxRimor [voxrimor]
→ python3 voxrimor.py
```

DEMO 1: Visualizer mode

	filename	vowel	gender	age	health	F0mean	F0median	...	mfcc14	mfcc15	mfcc16	mfcc17	mfcc18	mfcc19	mfcc20
0	1-a_n_20f-hlth.wav	a	f	20	hlth	274.431673	275.862069	...	0.356511	0.556554	-0.127219	-1.107133	-1.910601	-0.961069	0.149626
1	10-a_n_22f-hlth.wav	a	f	22	hlth	199.024390	200.000000	...	0.353346	-0.046021	0.196454	-0.045817	0.052015	0.299132	0.111662
2	1006-a_n_32f-hlth.wav	a	f	32	hlth	205.128205	205.128205	...	0.328761	0.124032	0.082613	0.151899	0.153524	0.160330	0.650411
3	102-a_n_39f-hlth.wav	a	f	39	hlth	198.482385	200.000000	...	0.102856	0.202720	0.078172	0.003700	0.161462	-0.053153	0.284400
4	1022-a_n_45f-hlth.wav	a	f	45	hlth	216.750083	216.216216	...	0.198137	0.186112	-0.140532	0.402129	0.362395	0.864775	0.845723
...
4126	95-u_n_51f-hlth.wav	u	f	51	hlth	222.222222	222.222222	...	0.252393	0.344417	0.298158	0.500797	0.736506	0.685966	0.417393
4127	962-u_n_30f-hlth.wav	u	f	30	hlth	223.774250	222.222222	...	0.316520	0.208523	0.332267	0.532395	0.701455	0.793471	0.230223
4128	97-u_n_21f-hlth.wav	u	f	21	hlth	242.760943	242.424242	...	0.367020	0.428709	0.891417	0.819015	0.402244	-0.548269	-1.293228
4129	99-u_n_84f-hlth.wav	u	f	84	hlth	239.889087	242.424242	...	0.257677	0.647518	0.500471	0.782175	0.399972	-0.234390	-0.587947
4130	990-u_n_29f-hlth.wav	u	f	29	hlth	233.949580	235.294118	...	0.421929	0.605091	0.535736	0.667520	0.510483	0.192911	-0.264604

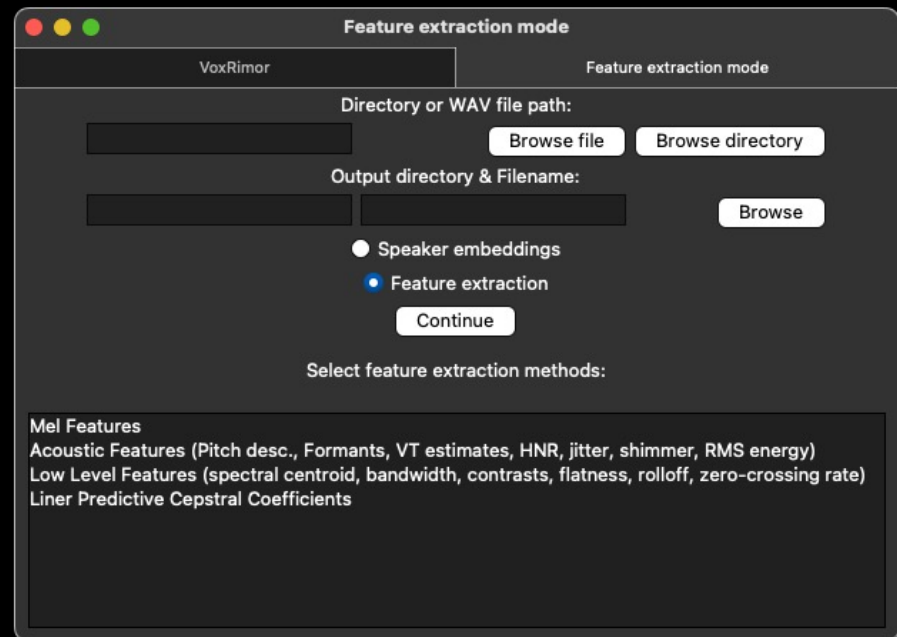
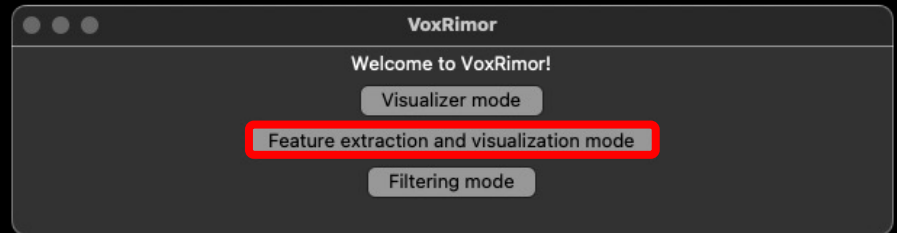


DEMO 2: Speaker embeddings mode



Feature extraction mode

- Mel-features
- Acoustic features
- Low-level features
- Linear predictive cepstral coefficients



Feature extraction mode: Mel-features

- MFCCs: coefficients representing the short-term power spectrum of an audio signal computed using Mel-frequency scaling and often used in automatic speech processing tasks.
- Delta features:
 - delta: first derivative of the MFCCs
 - delta-delta: second derivative of the MFCCs

Feature extraction

VoxRimor Feature extrac... Feature extrac...

Set parameters for Mel Features

n_mfcc	13
n_mels	40
win_length	25.0
overlap	10.0
fmin	150.0
fmax	4000.0
preemphasis	0.95
lifter	22.0
deltas	<input checked="" type="checkbox"/>
summarise	<input checked="" type="checkbox"/>

Continue

Feature extraction mode: Mel-features

- `n_mfcc`: number of coefficients
- `n_mels`: number of Mel bands
- `win_length`: length of analysis frame (ms)
- `overlap`: length of overlap between successive frames (ms)
- `fmin`: lowest frequency (Hz)
- `fmax`: highest frequency (Hz)
- `preemphasis`: pre-emphasis coefficient
- `lifter`: cepstral filtering coefficient
- `deltas`: compute also delta and delta-delta features?
- `summarise`: summarise (μ and σ) of each feature at the utterance (file) level

Feature extraction

VoxRimor Feature extrac... Feature extrac...

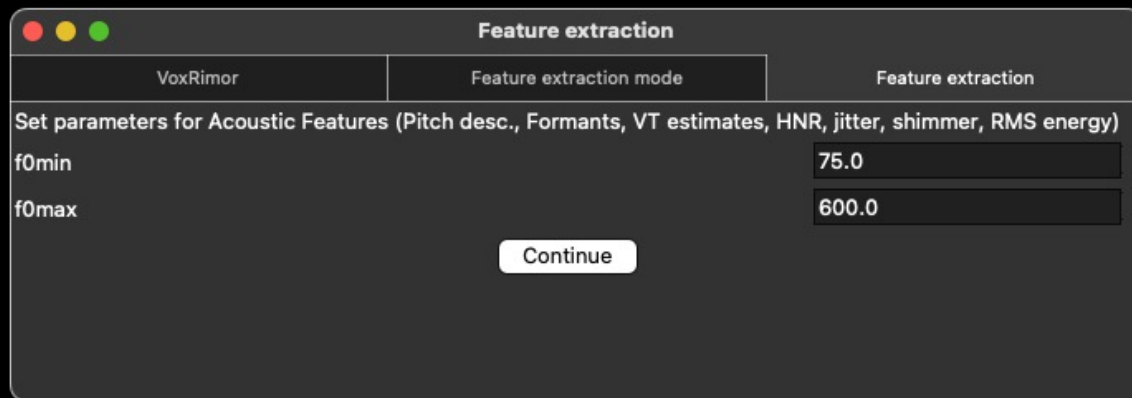
Set parameters for Mel Features

n_mfcc	13
n_mels	40
win_length	25.0
overlap	10.0
fmin	150.0
fmax	4000.0
preemphasis	0.95
lifter	22.0
deltas	<input checked="" type="checkbox"/>
summarise	<input checked="" type="checkbox"/>

Continue

Feature extraction mode: Acoustic features

- Pitch: mean, median, minimum, maximum, std. dev.
- Formants: F1, F2, F3, F4
- VT estimates: formant dispersion, avg. formant, formants' geometric mean, [Fitch VTL](#), [VTL \$\Delta f\$](#)
- HNR
- Jitter & Shimmer
- RMS energy



The screenshot shows a software window titled "Feature extraction" with a dark gray background. At the top, there are three colored window control buttons (red, yellow, green) on the left. Below the title bar, there is a header section with three tabs: "VoxRimor", "Feature extraction mode" (which is selected), and "Feature extraction". The main area of the window contains the text "Set parameters for Acoustic Features (Pitch desc., Formants, VT estimates, HNR, jitter, shimmer, RMS energy)". Below this text, there are two input fields: "f0min" with a value of "75.0" and "f0max" with a value of "600.0". At the bottom center of the window, there is a white button with the text "Continue".

VoxRimor	Feature extraction mode	Feature extraction
Set parameters for Acoustic Features (Pitch desc., Formants, VT estimates, HNR, jitter, shimmer, RMS energy)		
f0min		75.0
f0max		600.0
<div>Continue</div>		

Feature extraction mode: Acoustic features

- f0min: pitch floor (Hz)
- f0max: pitch ceiling (Hz)

Uses Praat (Sound: To Pitch (cc)); Soon more parameters to come...



The screenshot shows a macOS-style dialog box titled "Feature extraction". It has three tabs: "VoxRimor", "Feature extraction mode", and "Feature extraction". The "Feature extraction" tab is selected. Below the tabs, the text "Set parameters for Acoustic Features (Pitch desc., Formants, VT estimates, HNR, jitter, shimmer, RMS energy)" is displayed. There are two input fields: "f0min" with a value of "75.0" and "f0max" with a value of "600.0". A "Continue" button is located at the bottom center.

Feature extraction mode	Feature extraction
Set parameters for Acoustic Features (Pitch desc., Formants, VT estimates, HNR, jitter, shimmer, RMS energy)	
f0min	75.0
f0max	600.0
<button>Continue</button>	

Feature extraction mode: Low-level features

- spectral centroid: “centre of mass” of avg. frequency in frame
- spectral bandwidth
- spectral contrast: mean energy contrast between peaks and valleys in each frequency sub-band
- spectral flatness: how “noise-like” vs. “tone-like” is the sound?
- spectral roll-off: roll-off frequency (freq. below which 85% of the energy is contained)
- zero-crossing rate

The screenshot shows a software window titled "Feature extraction" with a sub-header "VoxRimor". The main content area is titled "Feature extraction mode" and "Feature extraction". It contains a section "Set parameters for Low Level Features (spectral centroid, bandwidth, contrasts, flatness, rolloff, zero-crossing rate)". Below this, there are several input fields and checkboxes:

Parameter	Value	Checkbox
win_length	25.0	
overlap	10.0	
preemphasis	0.95	
n_bands_contrasts	6	
use_mean_contrasts		<input checked="" type="checkbox"/>
summarise		<input checked="" type="checkbox"/>

At the bottom of the window, there is a "Continue" button.

Feature extraction mode: Low-level features

- `win_length`: length of analysis frame (ms)
- `overlap`: length of overlap between successive frames (ms)
- `preemphasis`: pre-emphasis coefficient
- `n_bands_contrasts`: number of frequency bands by which to divide the spectrum to calculate contrasts (num. contrasts = `n_bands`+1)
- `use_mean_contrasts`: calculates the average contrast
- `summarise`: summarise (μ and σ) of each feature at the utterance (file) level

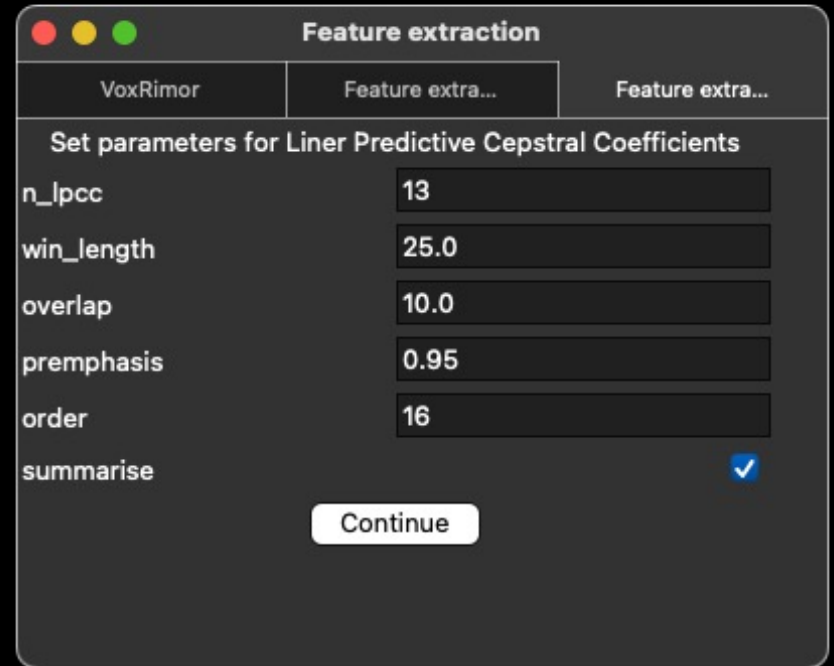
The screenshot shows a macOS-style dialog box titled "Feature extraction". It has three tabs: "VoxRimor", "Feature extraction mode" (which is selected), and "Feature extraction". Below the tabs, the text "Set parameters for Low Level Features (spectral centroid, bandwidth, contrasts, flatness, rolloff, zero-crossing rate)" is displayed. The parameters are listed on the left, and their values are in input fields on the right:

Parameter	Value
<code>win_length</code>	25.0
<code>overlap</code>	10.0
<code>preemphasis</code>	0.95
<code>n_bands_contrasts</code>	6
<code>use_mean_contrasts</code>	<input checked="" type="checkbox"/>
<code>summarise</code>	<input checked="" type="checkbox"/>

At the bottom center of the dialog is a "Continue" button.

Feature extraction mode: LPC features

- LPC: linear prediction coefficients (Burg); a representation of the spectral envelope after applying a linear filter
- LPCCs: linear predictive cepstral coefficients; coefficients representing the cepstrum (log-power spectrum) of the linear prediction coefficients



The screenshot shows a macOS-style dialog box titled "Feature extraction". It has three tabs: "VoxRimor", "Feature extra...", and "Feature extra...". The "Feature extra..." tab is selected. Below the tabs, the text "Set parameters for Liner Predictive Cepstral Coefficients" is displayed. There are six parameters listed on the left, each with a corresponding input field on the right:

Parameter	Value
n_lpcc	13
win_length	25.0
overlap	10.0
preemphasis	0.95
order	16
summarise	<input checked="" type="checkbox"/>

At the bottom center of the dialog is a "Continue" button.

Feature extraction mode: LPC features

- `n_lpcc`: number of coefficients
- `win_length`: length of analysis frame (ms)
- `overlap`: length of overlap between successive frames (ms)
- `preemphasis`: pre-emphasis coefficient
- `order`: order of the linear filter
- `summarise`: summarise (μ and σ) of each feature at the utterance (file) level

Feature extraction

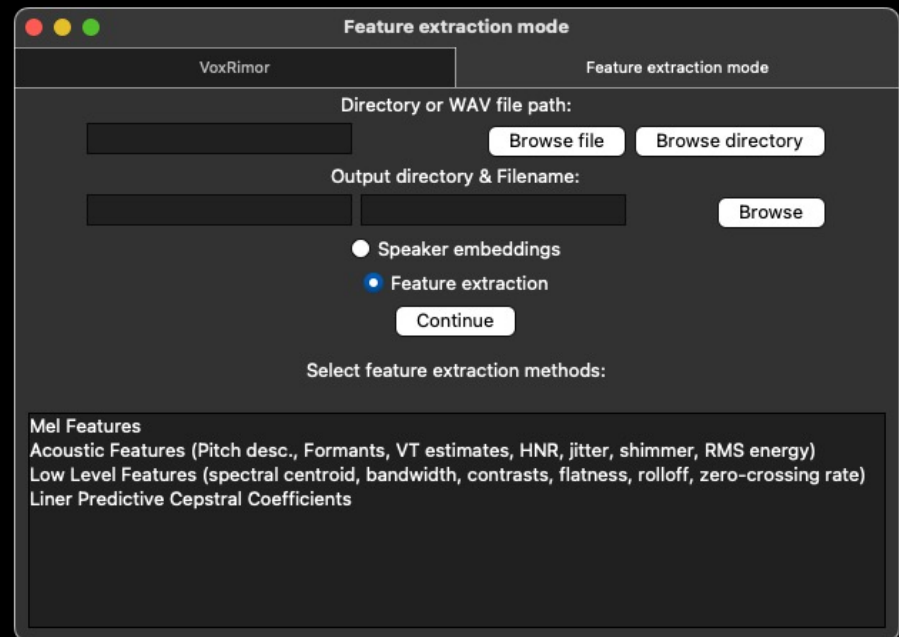
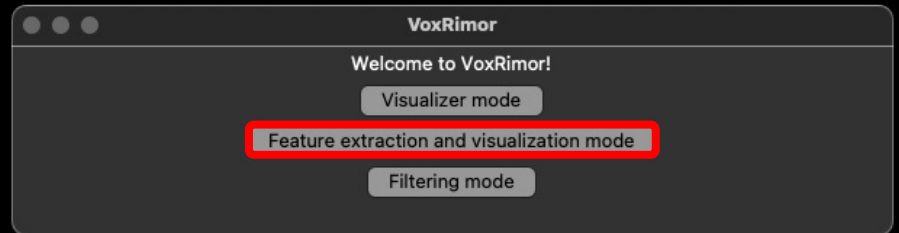
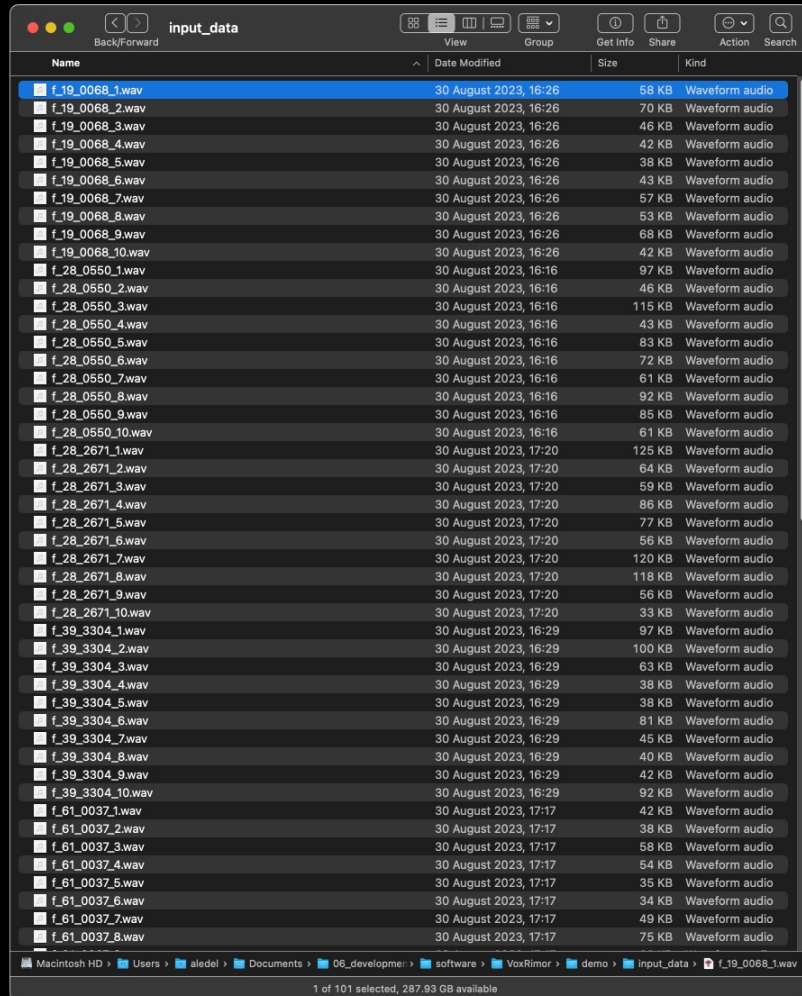
VoxRimor Feature extra... Feature extra...

Set parameters for Liner Predictive Cepstral Coefficients

n_lpcc	13
win_length	25.0
overlap	10.0
preemphasis	0.95
order	16
summarise	<input checked="" type="checkbox"/>

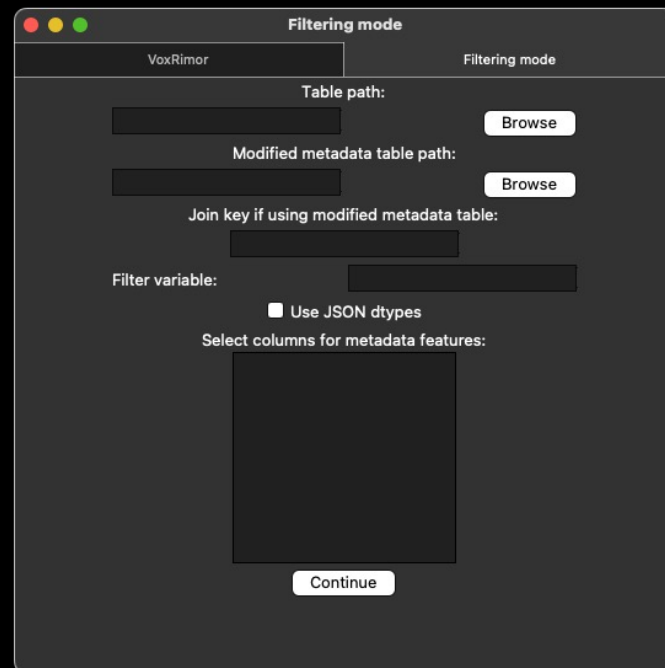
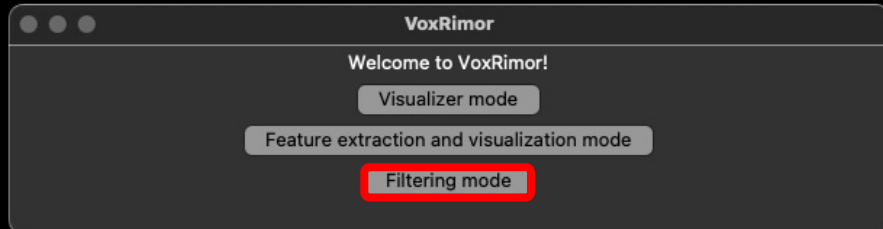
Continue

DEMO 3: Feature extraction mode



DEMO 4: Filtering mode

	filename	vowel	gender	age	health	F0mean	F0median	...	mfcc14	mfcc15	mfcc16	mfcc17	mfcc18	mfcc19	mfcc20
0	1-a_n_20f-hlth.wav	a	f	20	hlth	274.431673	275.862069	...	0.356511	0.556554	-0.127219	-1.107133	-1.910601	-0.961069	0.149626
1	10-a_n_22f-hlth.wav	a	f	22	hlth	199.024390	200.000000	...	0.353346	-0.046021	0.196454	-0.045817	0.052015	0.299132	0.111662
2	1006-a_n_32f-hlth.wav	a	f	32	hlth	205.128205	205.128205	...	0.328761	0.124032	0.082613	0.151899	0.153524	0.160330	0.650411
3	102-a_n_39f-hlth.wav	a	f	39	hlth	198.482385	200.000000	...	0.102856	0.202720	0.078172	0.003700	0.161462	-0.053153	0.284400
4	1022-a_n_45f-hlth.wav	a	f	45	hlth	216.750083	216.216216	...	0.198137	0.186112	-0.140532	0.402129	0.362395	0.864775	0.845723
...
4126	95-u_n_51f-hlth.wav	u	f	51	hlth	222.222222	222.222222	...	0.252393	0.344417	0.298158	0.500797	0.736506	0.685966	0.417393
4127	962-u_n_30f-hlth.wav	u	f	30	hlth	223.774250	222.222222	...	0.316520	0.208523	0.332267	0.532395	0.701455	0.793471	0.230223
4128	97-u_n_21f-hlth.wav	u	f	21	hlth	242.760943	242.424242	...	0.367020	0.428709	0.891417	0.819015	0.402244	-0.548269	-1.293228
4129	99-u_n_84f-hlth.wav	u	f	84	hlth	239.889087	242.424242	...	0.257677	0.647518	0.500471	0.782175	0.399972	-0.234390	-0.587947
4130	990-u_n_29f-hlth.wav	u	f	29	hlth	233.949580	235.294118	...	0.421929	0.605091	0.535736	0.667520	0.510483	0.192911	-0.264604

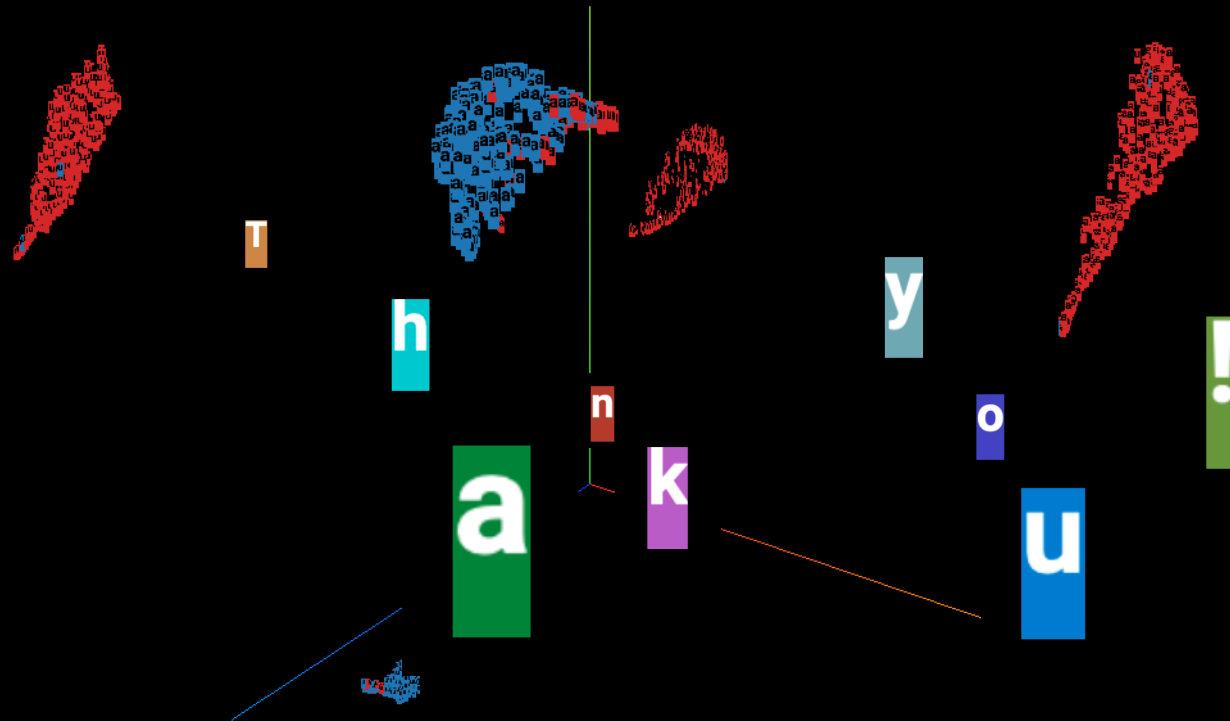


Future plans, goals, and Q&A

- More parameter settings → better feature extraction
- Fully-integrated web application and UI
- *VoxRimor* as an open-source online LiRI service
- Direct selection & downloading of data
- Downloadable reduced dimensions versions of data
- Statistical tools: distances, box-plots, distributions
- Integrated speaker verification system and classification tools



VoxRimor Thanks You!



Any questions?

