MUSIC TIMBRE ANALYSIS AND SYNTHESIS

HAN ZHANG'S MASTER THESIS

HanZhang2020@u.northwestern.edu

PURPOSE

- Design a framework for the extraction and modification of harmonics morphological features for musical sounds and develop a synthesis method that allows the sound reconstruction, design, and morphing based on the features.

Analysis

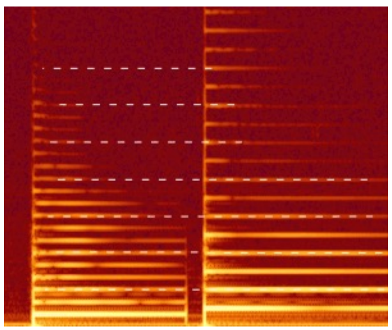
- Extract morphological features of harmonics
- Relate the features to timbre descriptors
- Distinguish musical instruments

→

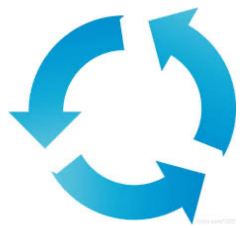
Synthesis

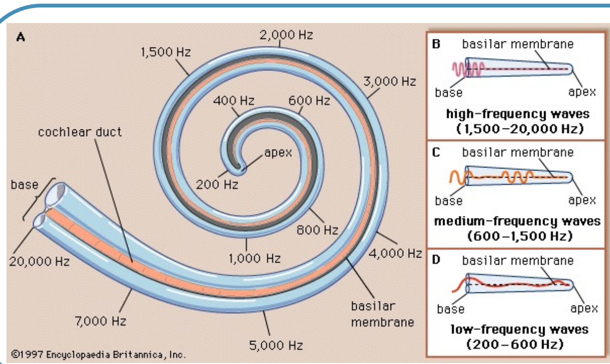
- Reconstruct sounds from the features
- Modify some parameters for new musical timbre
- Morph sounds

MOTIVATION

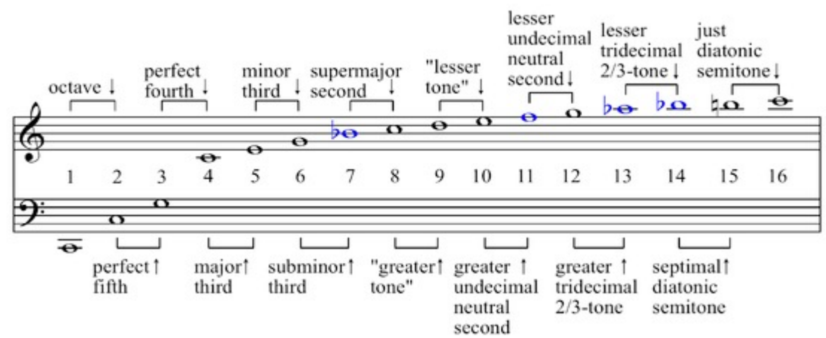


Fact of sound:
Additive model



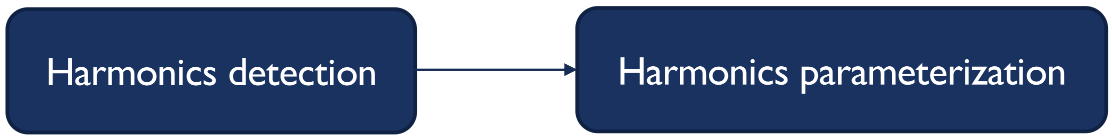


Fact of human perception:
Cochlea structure



Development of music:
Spectral music

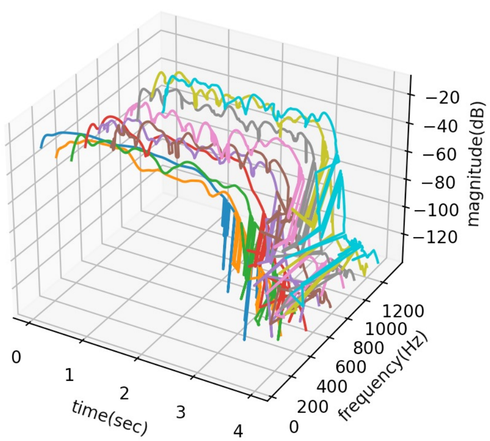
FEATURE EXTRACTION MODEL



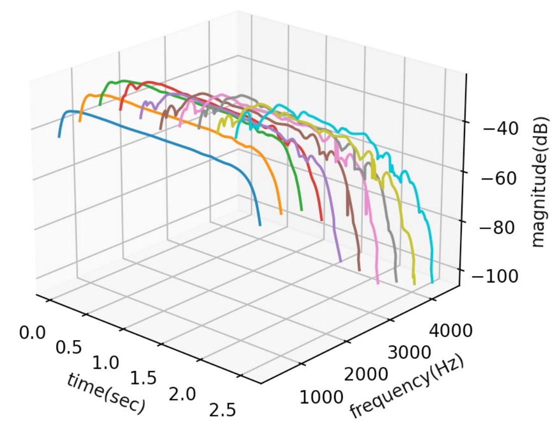
Step1: Harmonics detection

- SMS Tools: Harmonics plus Stochastic model (Xavier Serra, Julius Smith, 1990)

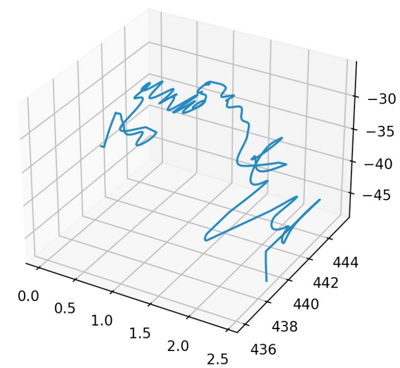
Trumpet in A4 original



Trumpet in A4 HpS

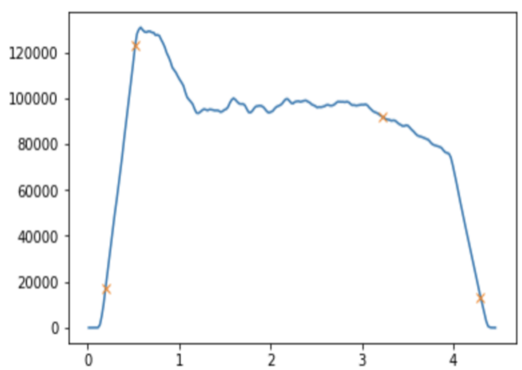


FEATURE EXTRACTION MODEL

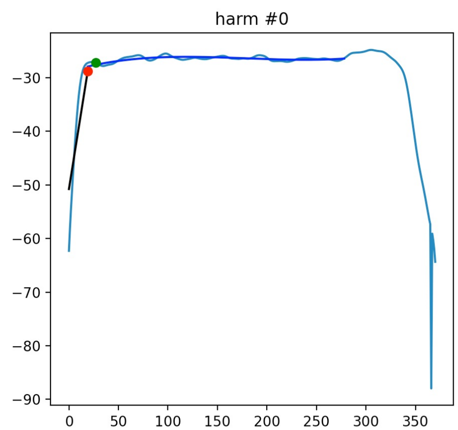


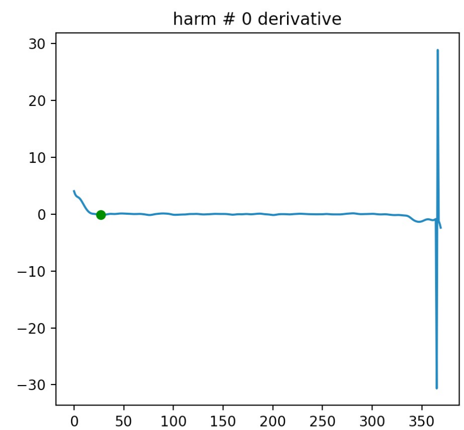
Step2: Harmonics parameterization

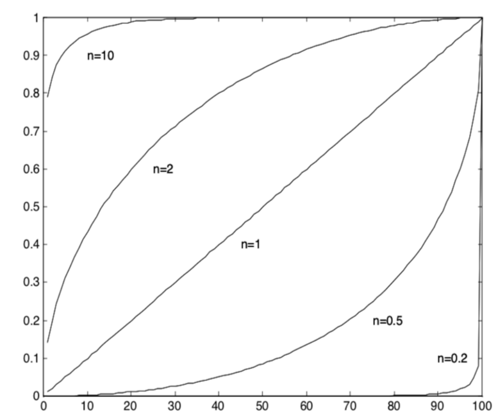
- Frequency: Control with the first and the second moment.
- Magnitude: Segment each harmonic into attack, steady and release parts. Fit each segment with a single-degree-of-freedom curve. (Kristoffer, 2006)



SOA: start of attack (10% of max)
EOA: special detecting model
SOR: start of release (70% of max)
EOR: end of release (10% of max)







$$Curve_s = v_0 + (v_1 - v_0)(1 - (1 - x)^n)^{\frac{1}{n}}$$

FEATURE EXTRACTION MODEL

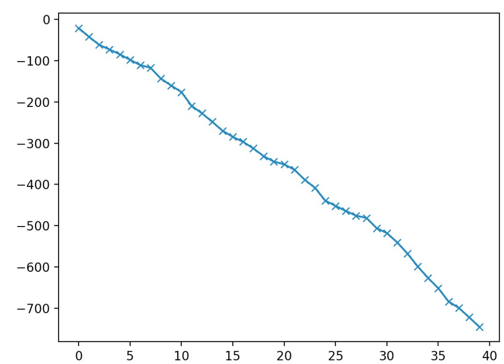
Step2: Harmonics parameterization

- Phase: Phase propagation through time.

$$\varphi_{i+1} = \varphi_i + \frac{\pi*(f_i+f_{i+1})*H}{f_s}.$$

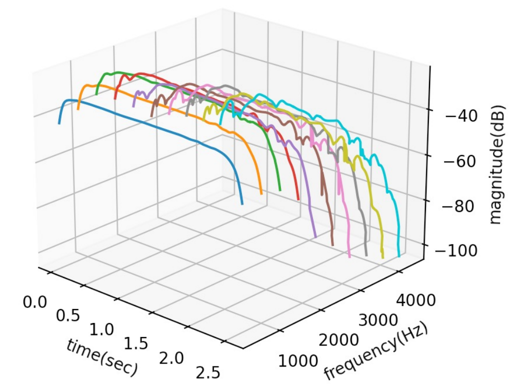
H: hop size; φ_i: the phase of frame I; f_i: the frequency of frame I; f_s: the sample frequency.

- Linear first frame phases: Apply a linear regression.

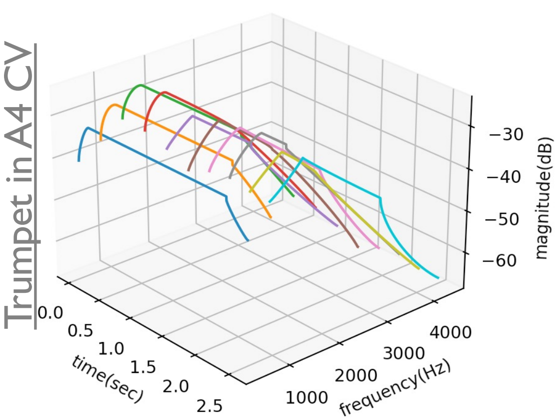


First frame phases of flute on A4.
Horizontal: harmonic number, vertical: phase

Trumpet in A4 HpS



Trumpet in A4 CV

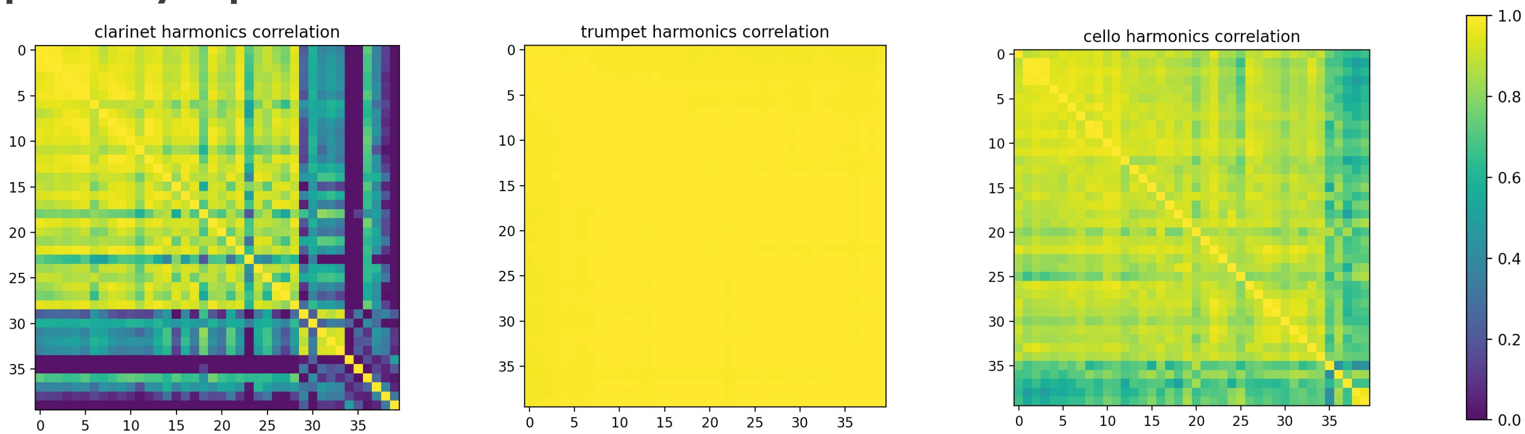


FEATURE EXTRACTION MODEL

- Performance Evaluation
  - Details degraded.
  - Good reconstruction for continuous sounds, bad for impulsive sounds.
- Pros:
  - Small number of parameters. (O(nH))
  - Interpretability
- Cons:
  - Limited possibility for timbre. → Transformations.
  - Lack of numerical sound quality estimation. → Listening tests.
- Comparing with sub-band analysis models:
  - SMS is more adaptive to pitch changes.
  - More detailed parameters comparing to coefficients.

FEATURE EXTRACTION MODEL

Exploratory experiment: Harmonics correlation

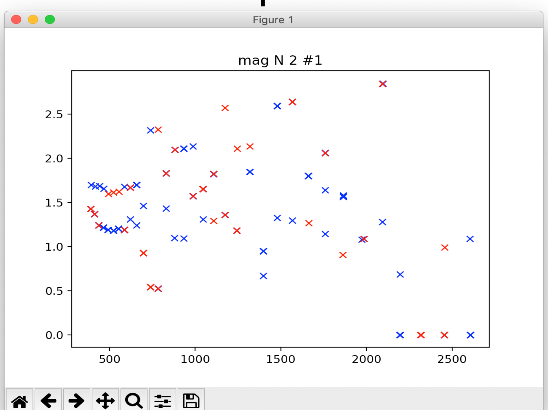


- Relationship between the directionality and the correlation.
- Possibility of further shrink the parameter space: grouping harmonics.

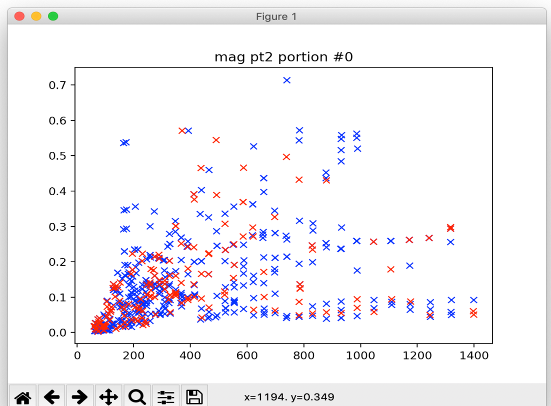
SOUND FEATURE ANALYSIS

Relationship between features and fundamental frequency

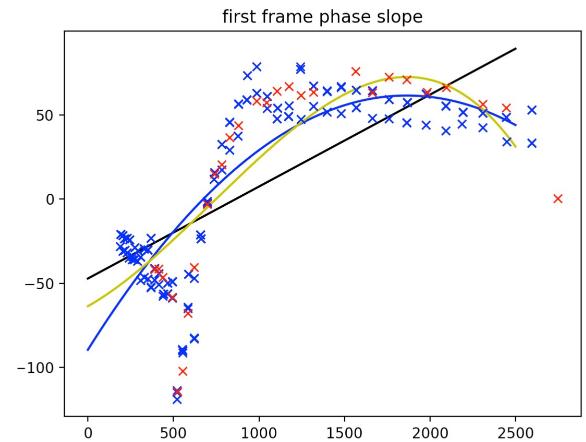
I: No pattern



II: Distribution



III: Curve

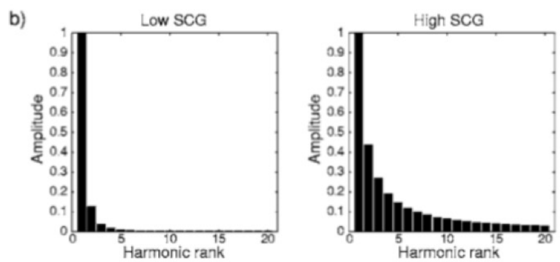


- Inconsistency among different instruments for the same feature
- Still debatable:
  - Spectral envelopes of sustained orchestral instrument sounds are invariant to variation in F0. (Patterson *et al.* 2010)
  - Dynamics should be considered. Some instruments show linearity in spectral centroid. (Siedenburg *et al.* 2021)

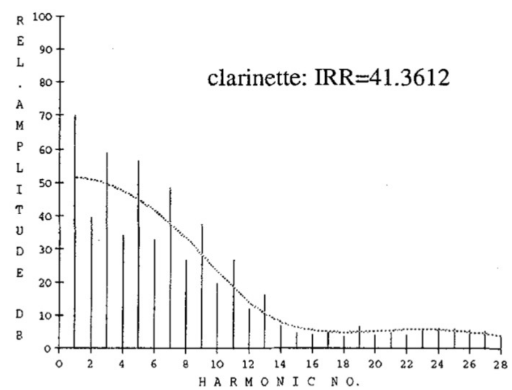
SOUND FEATURE ANALYSIS

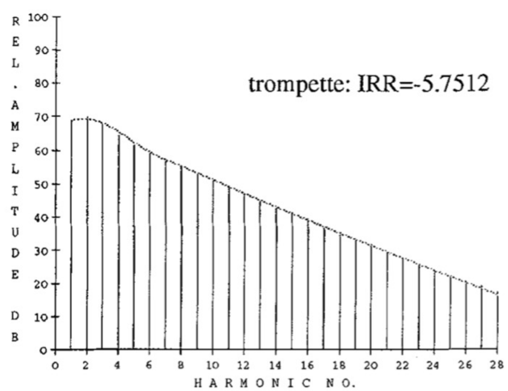
Semantic descriptions for the parameters

- Timbre Space (*S. McAdam, 1995*)
  - I: Spectral Gravity Centroid →
  - II: Logarithm of rise time
  - III: Spectral Flux

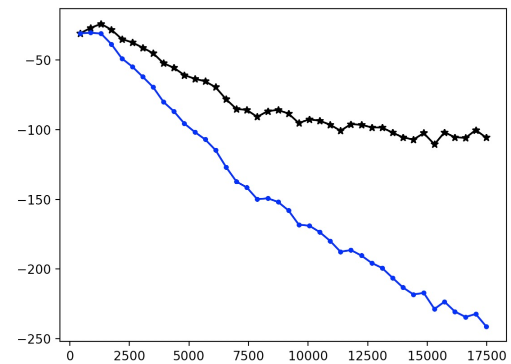


$$IRR = \log_{10}\left(\sum_{k=2}^{n-1}\left|20\log_{10}(A_k) - \frac{20\log_{10}(A_{k+1}) + 20\log_{10}(A_k) + 20\log_{10}(A_{k-1})}{3}\right|\right)$$





[Trumpet in A4 original]

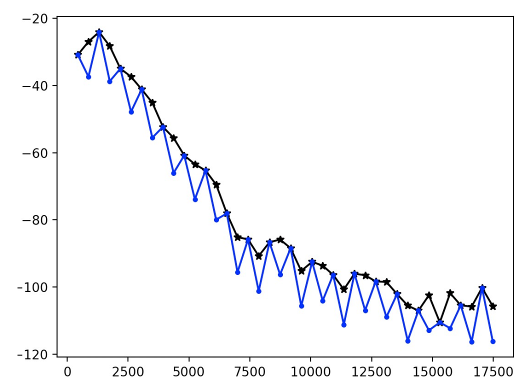


#1 harmonic magnitude

Black: original

Blue: modified

[Trumpet in A4, exponential attenuation alpha=0.4]

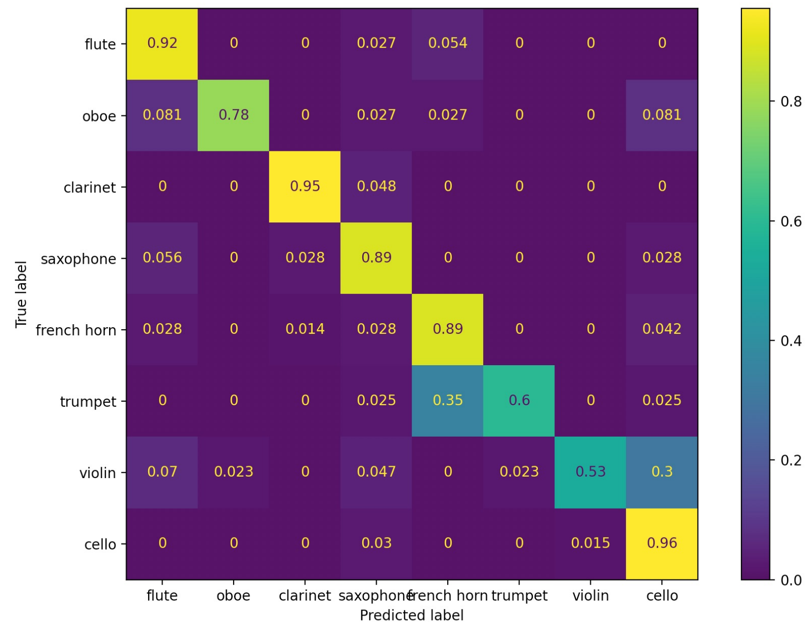


[Trumpet in A4, Even harmonics attenuation, alpha=0.3]

MUSICAL INSTRUMENT RECOGNITION

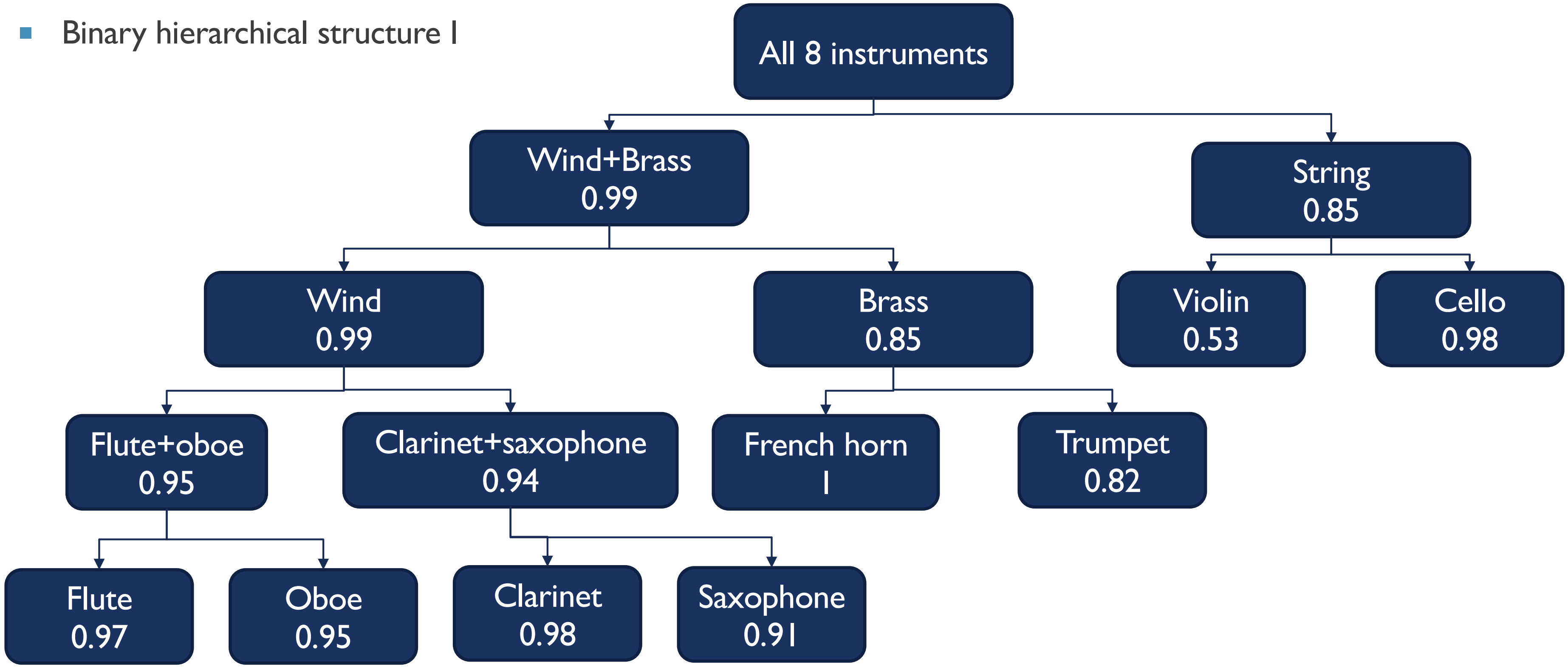
- Database: Phiharmnia Database: records of acoustic instrumental sounds for all orchestra instruments
- Instruments: flute, oboe, clarinet, saxophone, french horn, trumpet, violin, cello
- Classification model: random forest with 80 decision trees
- Result:

  Train 1.0
  Test 0.8492
  OOB score 0.8309



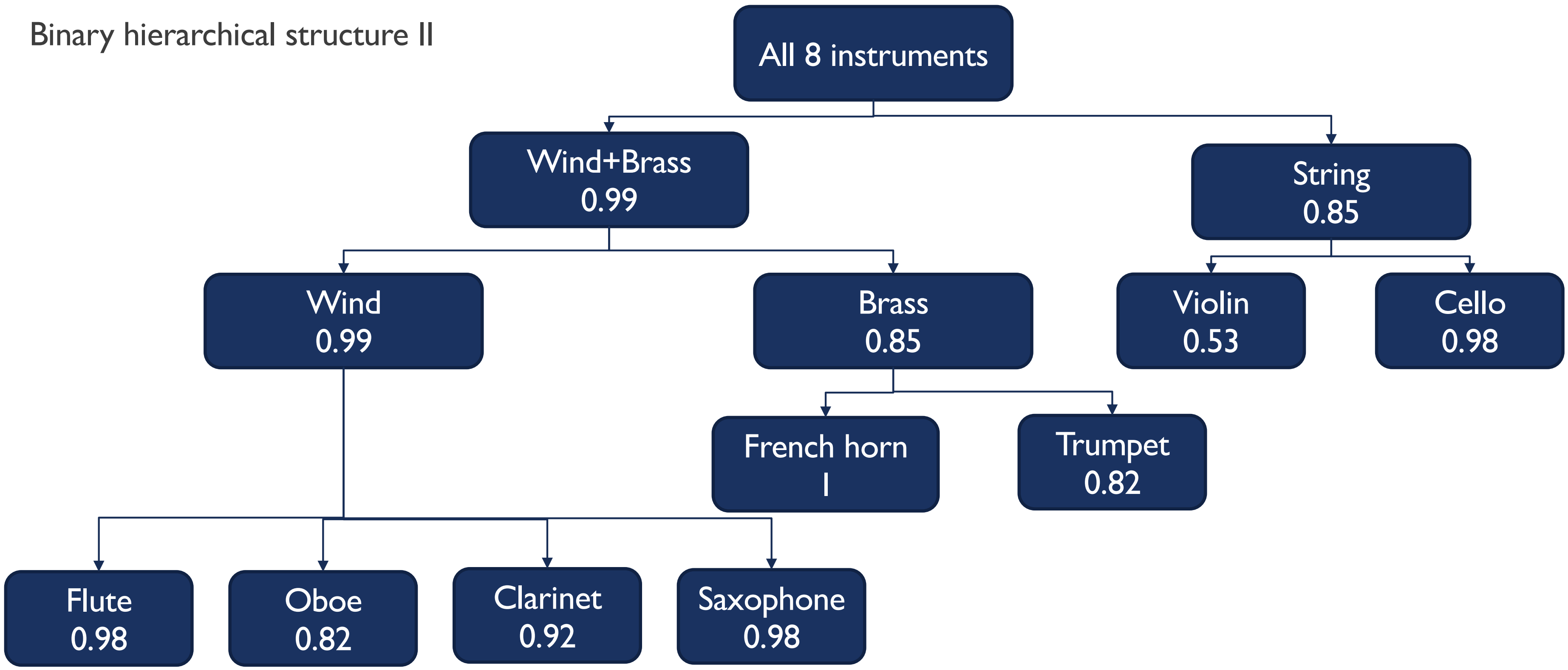
MUSICAL INSTRUMENT RECOGNITION

- Binary hierarchical structure I



MUSICAL INSTRUMENT RECOGNITION

- Binary hierarchical structure II



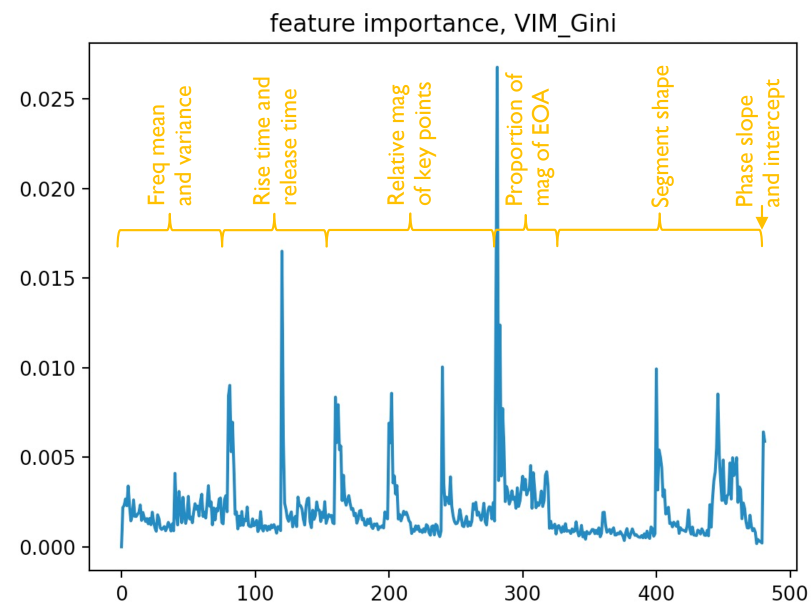
MUSICAL INSTRUMENT RECOGNITION

Paper	Features	Classification method	Accuracy
'MUSICAL INSTRUMENT RECOGNITION USING CEPSTRAL COEFFICIENTS AND TEMPORAL FEATURES' (2000)	Cepstral, spectral and temporal	kNN, hierarchical classification	93% for family, 74.9% for instrument
'COMPARISON OF FEATURES FOR MUSICAL INSTRUMENT RECOGNITION' (2001)	Cepstral, spectral and temporal	kNN	91.7% for family, 45.9% for instrument
'Musical Instrument Recognition by Pairwise Classification Strategies' (2006)	Cepstral, spectral and temporal, OBSI	GMM, SVM, applying pairwise strategy	87% for instrument
'FRAME-LEVEL INSTRUMENT RECOGNITION BY TIMBRE AND PITCH' (2018)	Constant-Q transformation matrix, estimated pitch	CNN	90.0% for instrument
'AN ATTENTION MECHANISM FOR MUSICAL INSTRUMENT RECOGNITION' (2019)	Spectrogram	ATT(attention model), CNN	~85% for instrument
This model	Harmonics morphological features	Random forest	8-class: 79.1% for instrument Hier1: 80.4% for instrument Hier2: 81.2% for instrument

MUSICAL INSTRUMENT RECOGNITION

Feature importance

- Gini index: the effectiveness of reducing impurity
- Observations:
  - Lower harmonics tend to be more informative
  - Magnitude proportion, rise and release time, and some other features show great importance in the classification model.



Strengths of this model

- Less feature numbers, higher efficiency
- Being semantically meaningful, feature importance can be more explicitly explained

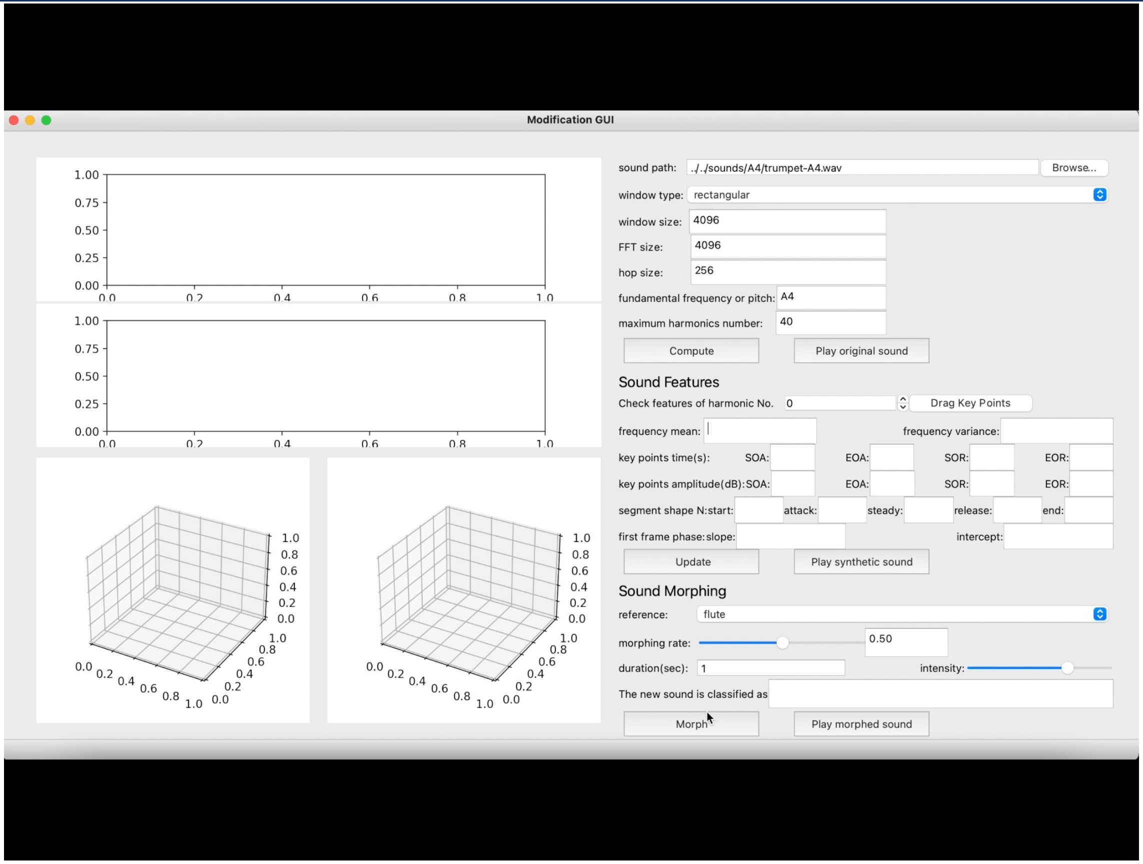
TIMBRAL SOUND MORPHING

- Strategy: Interpolate every feature between two reference sounds, given the morphing rate.
- Example:

instrument1	1+0	0.9+0.1	0.75+0.25	0.5+0.5	0.25+0.75	0.1+0.9	0+1	instrument2
flute	flute	clarinet	flute	horn	flute	oboe	horn	**trumpet**

- Comparing with existing works:
  - Differentiable DSP(Engel *et al.* 2020), Music Translation Network(Mor *et al.* 2019): smaller parameter space; continuous morphing rate.
  - Nsynth sound morphing(Engel *et al.* 2017): smaller parameter space; more flexible in partial morphing

INTERFACE IMPLEMENTATION



INTERFACE IMPLEMENTATION

Comparison with some other sound synthesis/modulation tools

- Synthesizers
- Serum

CONCLUSIONS

- Designed and implemented a complete framework for the extraction and modification of harmonics morphological features for musical sounds. Developed a synthesis method that allows the reconstruction, design, and morphing based on the features.

- Proofed the informativity of the feature space by mapping semantic descriptors to the parameters and testing its capability of timbre recognition. Yielded meaningful conclusions on the importance of the features.

- Built an interface for the demonstration of the model and for further explorations on the features.

FUTURE WORKS AND OUTLOOK

- Feature extraction model refining
  - Attack modeling
  - Noise component modeling
- Mapping semantic descriptions to sound parameters
  - Listening tests
- Transformation
  - Allowing transformations corresponds with classic timbre descriptors
  - Allowing sound design from parameters
- Long term goal
  - Complete a comprehensive model for sound design or music composition based on the idea of shaping the spectrogram.
  - Provide a new perspective for music and an innovative approach for composition.