



If Instruments could talk ...

Vowels and their role as key features for musical instrument timbre recognition

Content

1.1 Early Sources

1.2 Vowel tones and Formants

1.3 Musical Instruments Formants

2.1 Blending and Similarity

2.2 How to calculate Similarity? (Timbre Spaces and MFCCs)

2.3 Formants vs. MFCCs

3 Conclusion



If Instruments could talk ...

Early Sources

In his "Dissertation about the formation of language"
[*"Dissertatio de formatione loquelaе"*, 1781] Christoph Friedrich Hellwag described the **vowel chart** (or triangle) for the very first time.

§ 57.

Princeps vocalium, reliquarum basis, vel in scala positarum centrum est *a*: ex hac duplex ascendit scala, in gradus extremos *i* et *u* terminata: gradibus his extremis et homologis inferioribus termini interjacent intermedii. Graduum et terminorum intermediorum ad basin relatio sub hoc schemate concinno potest repraesentari:

u	ü	i
o	ö	e
	â	ä
	a	

Vocalis *o* medium tenet inter *u* et *â*, *â* inter *o* et *a*; similiter *e* inter *i* et *ä*, *ä* inter *e* et *a*; per *ü* fit transitus ex *u* ad *i*, per *ö* ex *o* ad *e*: exprimi potest terminus, per quem ex *â* ad *ä* transitur. Gradibus hisce scriptione designatis infiniti alii possunt interpolari, quos gentes linguis et linguarum varietatibus differentes inter loquendum constanter exprimunt.

[First reference to the vowel chart in the dissertation "Dissertatio de formatione loquelaе" written by Christoph Friedrich Hellwag \(1781, p. 41 § 57\)](#)



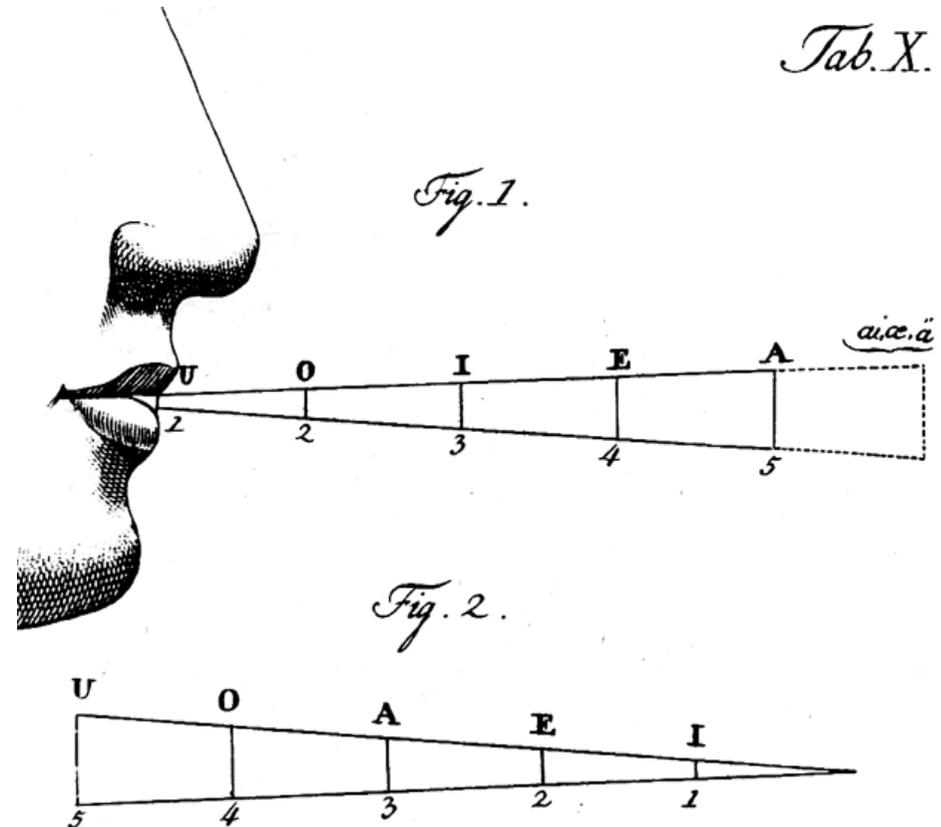
If Instruments could talk ...

Early Sources

Wolfgang von Kempelen described the impact of the **lips opening** (Fig. 1) and the **tongue position** (Fig. 2) on the timbres of **vowels** in his book about his speaking machine [„*Mechanismus der menschlichen Sprache*“, 1791] for the first time.

Here he already mentioned that the change between different vowels has an inherent melodic quality.

(von Kempelen 1791, p. 196)



[First reference to the lip and tongue movement and its impact to the vowel quality can be found in the dissertation about the mechanism of the human voice written by Wolfgang von Kempelen \(1791, p. 194\)](#)



If Instruments could talk ...

Vowel tones and Formants

In his work "About vowel tones and tongue pipes" ["Über Vocaltoene und Zungenpfeifen"] Robert Willis found out in 1832 that **fixed emphasized frequencies** in the vowel spectrum are responsible for a certain **vowel quality**.

He found these frequencies with the help of small stopped organ pipes.

Tafel I.

I	See	0,38	g^v
E	Pet	0,6	c^v
	Pay	1,0	d^{iv}
A	Paa	1,8	f^m
	Part	2,2	$a^r b$
A ^o	Paw	3,05	g^u
	Nought	3,8	$e^r b$
O	No	4,7	c^n
	But	unbestimmt	
U	Boot	- -	

Vowels (Column 1; used as in the words in column 2) built by vowel tones in the frequency of the pitches in column 4. These pitches correspond in their frequency to the length of small stopped flute pipes (column 3, length in inch).
(Willis 1832, p. 410, Table 1)



If Instruments could talk ...

Vowel tones and Formants

Similar „*vowel tones*“ have been found as resonances of the oral cavity by Hermann von Helmholtz in 1863.



[Vowel tones as resonance frequencies of the oral cavity
\(Helmholtz 1863, p. 173\)](#)



If Instruments could talk ...

Vowel tones and Formants

Similar „*vowel tones*“ have been found as resonances of the oral cavity by Hermann von Helmholtz in 1863.

He measured the pitch of the vowel tones with the help of a set of tuning forks, striking them close to the open mouth:

The louder the tuning fork sounds, the stronger is the **self-resonance of the oral cavity**.



[Vowel tones as resonance frequencies of the oral cavity](#)
(Helmholtz 1863, p. 173)



If Instruments could talk ...

Vowel tones and Formants

Ludimar Hermann introduced the term "**formant**" in his "Phonophotographical Studies" (of the voice) in 1894.

Wiederum habe ich nach der Schwerpunktsmethode (über die Auswahl der Amplituden s. dies Archiv Bd. 53, S. 50; vgl. ferner den Anhang zu dieser Abtheilung) die Lage des charakteristischen Tones oder Formanten genauer zu ermitteln versucht. Ich erlaube mir den Vorschlag, die letztere Bezeichnung statt der schleppenden erstgenannten generell einzuführen.

Schwerpunktsbestimmung der Vocalformanten.

Note (und Nr. des Blattes)	Ordnungszahlen der benutzten Amplituden	Resultirende Ordnungszahl des Formanten	Note des Formanten
Vocal A.			
G (202)	6 bis 10	7,8	$\langle g^2$
H (202)	7	7,0	$\langle a^2$
c (182)	5 bis 7	6,2	$\langle gis^2$
d (202)	4 " 6	4,6	$\cdot e^2 - f^2$
e (182)	4 " 6	5,1	$\langle as^2$
g (182)	3 " 5	4,3	as^2
" (202)	3 " 5	4,2	gis^2
c ¹ (182)	3 " 5	4,4	$dis^2 - e^2$

Ludimar Hermann introduced the term „formant“ in 1894 (Hermann 1894, p. 267).



If Instruments could talk ...

Vowel tones and Formants

Ludimar Hermann introduced the term "**formant**" in his "Phonophotographical Studies" (of the voice) in 1894.

"**Formant**" describes a single emphasized and **dominant partial** in the vowel spectrum, which is **characteristic** for the vowel and which is **independent** from the fundamentals' **pitch**.

Wiederum habe ich nach der Schwerpunktsmethode (über die Auswahl der Amplituden s. dies Archiv Bd. 53, S. 50; vgl. ferner den Anhang zu dieser Abtheilung) die Lage des charakteristischen Tones oder Formanten genauer zu ermitteln versucht. Ich erlaube mir den Vorschlag, die letztere Bezeichnung statt der schleppenden erstgenannten generell einzuführen.

Schwerpunktsbestimmung der Vocalformanten.

Note (und Nr. des Blattes)	Ordnungszahlen der benutzten Amplituden	Resultirende Ordnungszahl des Formanten	Note des Formanten
Vocal A.			
G (202)	6 bis 10	7,8	$\langle g^2$
H (202)	7	7,0	$\langle a^2$
c (182)	5 bis 7	6,2	$\langle gis^2$
d (202)	4 " 6	4,6	$\langle e^2 - f^2$
e (182)	4 " 6	5,1	$\langle as^2$
g (182)	3 " 5	4,3	as^2
" (202)	3 " 5	4,2	gis^2
c ¹ (182)	3 " 5	4,4	$dis^2 - e^2$

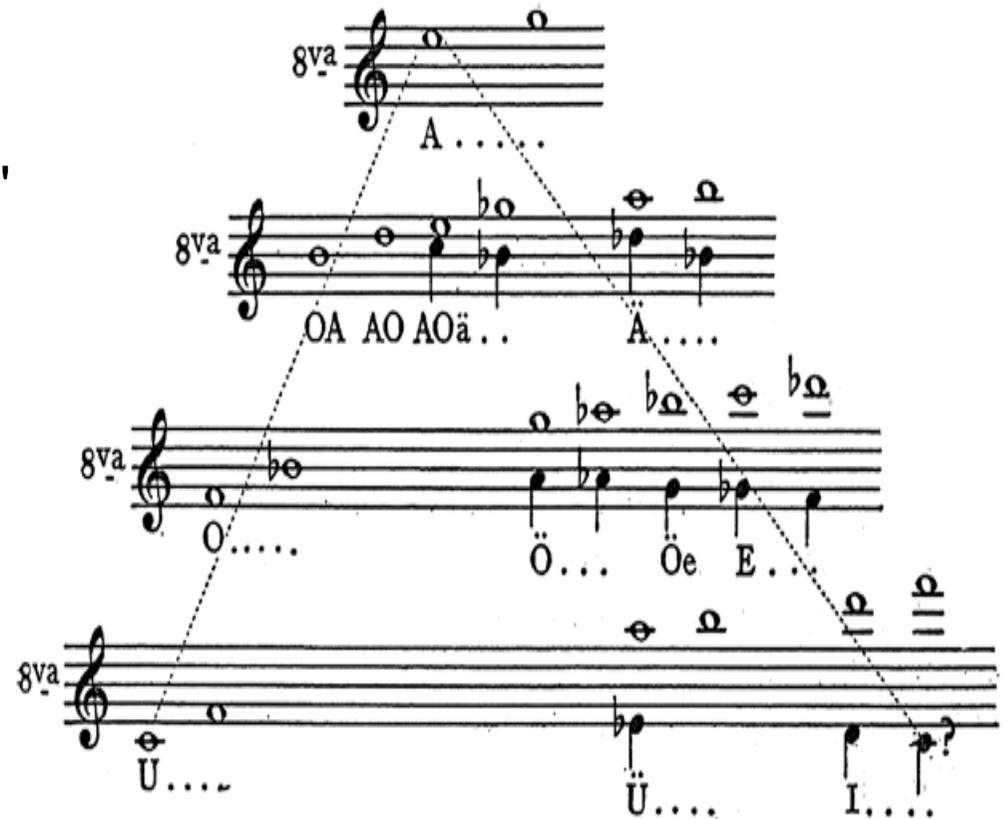
Ludimar Hermann introduced the term „formant“ in 1894 (Hermann 1894, p. 267).



If Instruments could talk ...

Vowel tones and Formants

In his book "The Sounds of Speech" [*Die Sprachlaute*] in 1926] Carl Stumpf assigned the **vowel formants** to the **vowel chart** of Hellwag.



Pitches of the vowel formants in the whispering voice
(one octave above the notation)
(Stumpf 1926, p. 145)

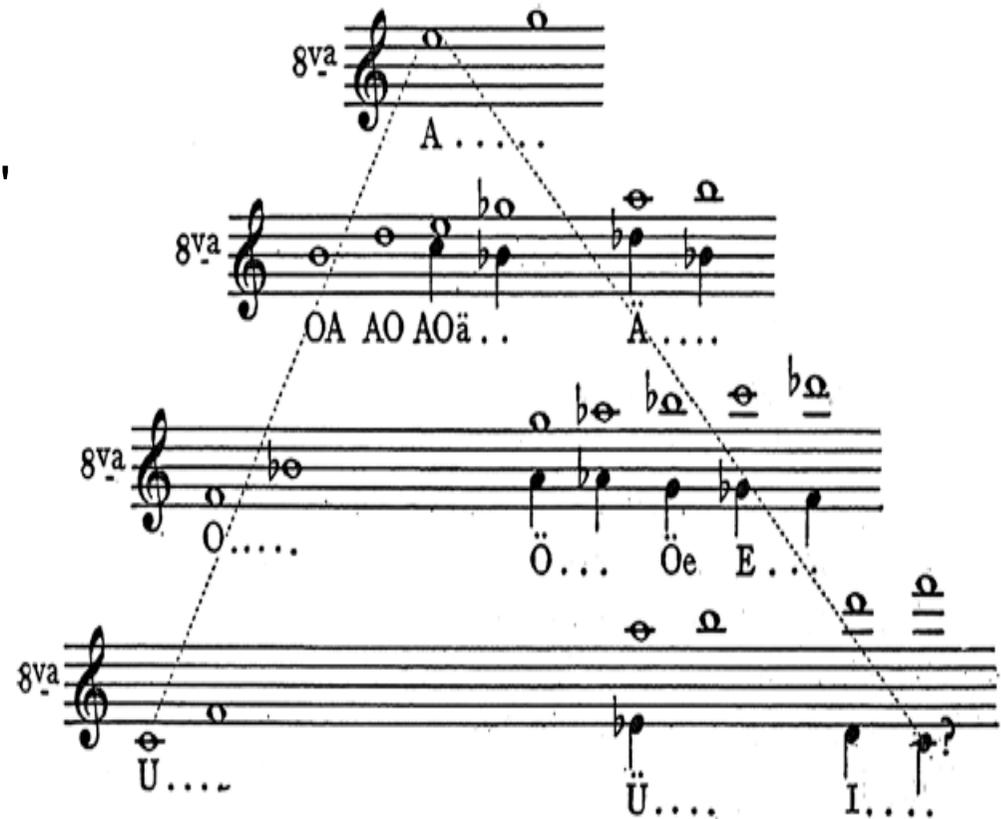


If Instruments could talk ...

Vowel tones and Formants

In his book "The Sounds of Speech" [*Die Sprachlaute* in 1926] Carl Stumpf assigned the **vowel formants** to the **vowel chart** of Hellwag.

Furthermore he found out that timbres of musical instruments seem to have formants too. He called them „**secondary formants**“ [*Nebenformanten*],
Here: formant is not a single partial but a certain **frequency band**.



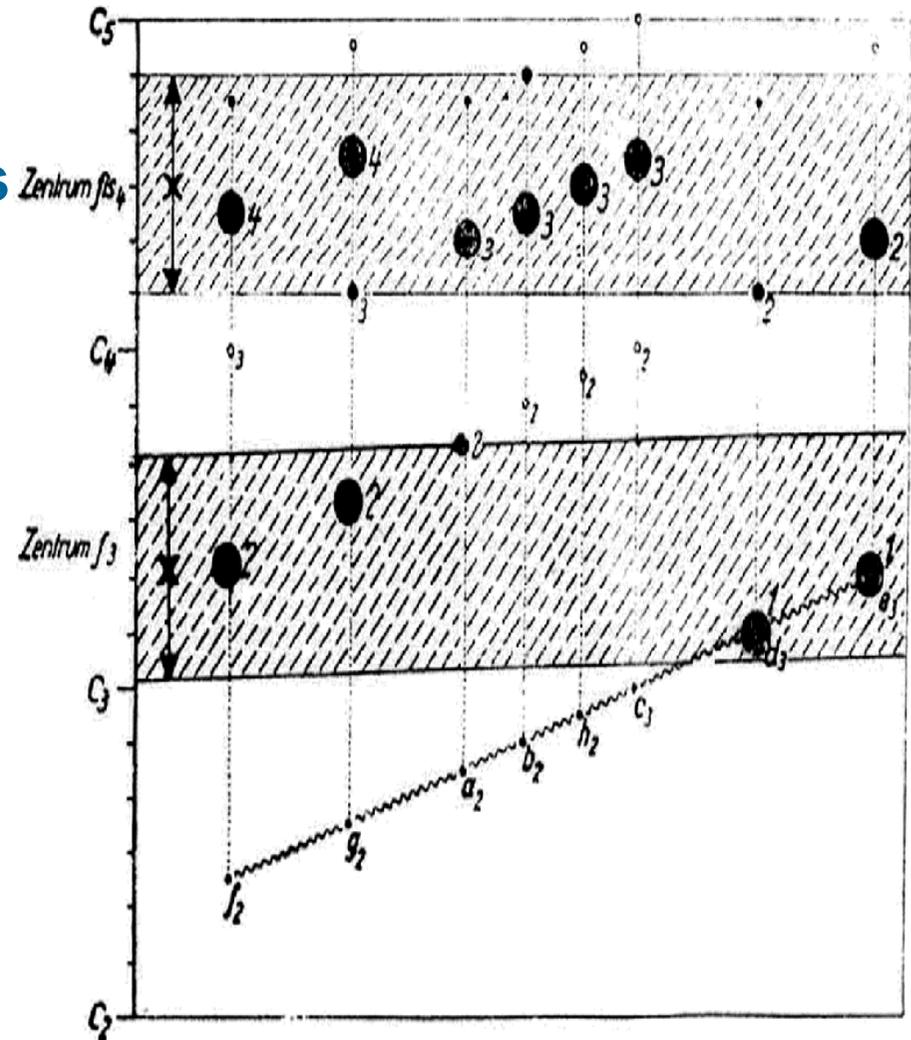
Pitches of the vowel formants in the whispering voice
(one octave above the notation)
(Stumpf 1926, p. 145)



If Instruments could talk ...

Musical Instruments Formants

In a systematic approach (measuring all pitches in different dynamics)
Karl Erich Schumann, a pupil of Carl Stumpf, found out the “**Principles of Timbre**” [“*Klangfarbengesetze*”] in his habilitation thesis “The Physics of Timbre” [“*Physik der Klangfarben*”] in 1929.



Formant areas in timbre of an oboe
(Schumann 1929, p. 89)



If Instruments could talk ...

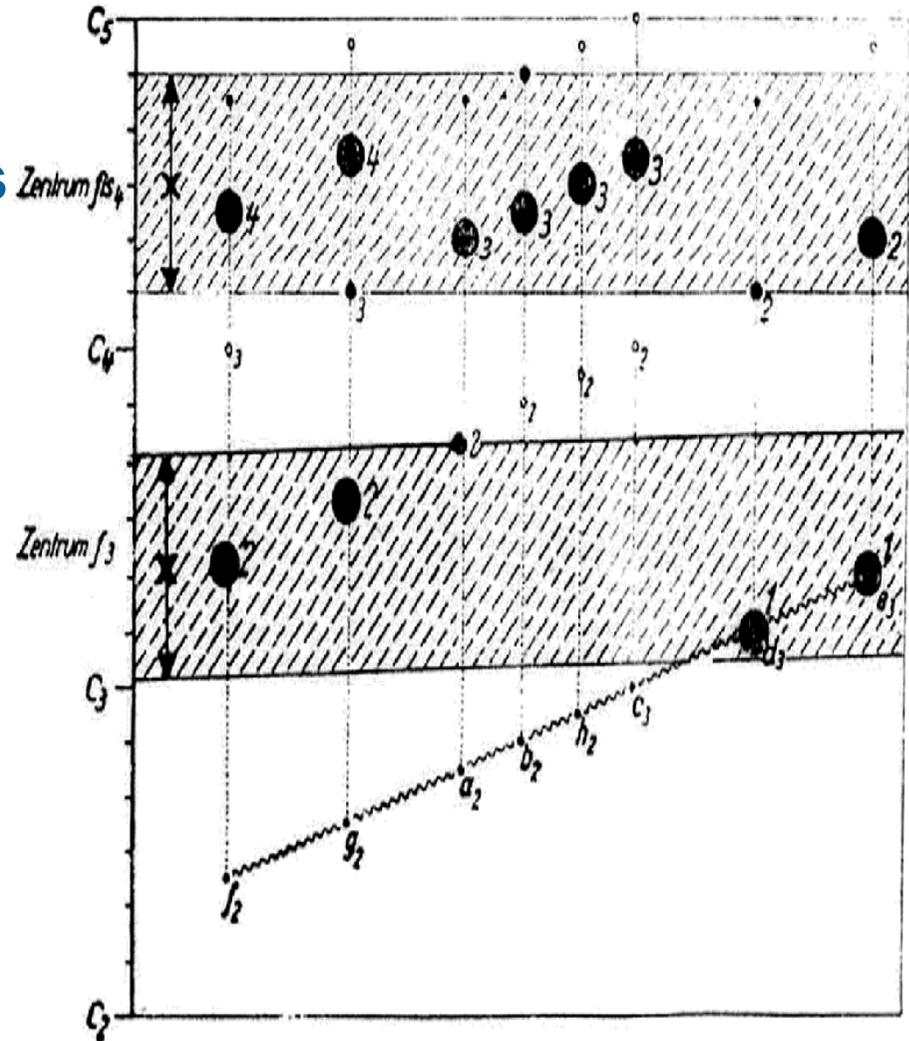
Musical Instruments Formants

- **Principle of Formant Areas**
[“*Formantstreckengesetz*”]

Formants of musical instruments are **fixed** and **pitch-independent areas** of the spectrum, wherein partials have exceptionally strong amplitudes, so that the timbre impression is influenced mainly by partials located in these areas



(Schumann, 1929, p. 89).



Formant areas in timbre of an oboe
(Schumann 1929, p. 89)

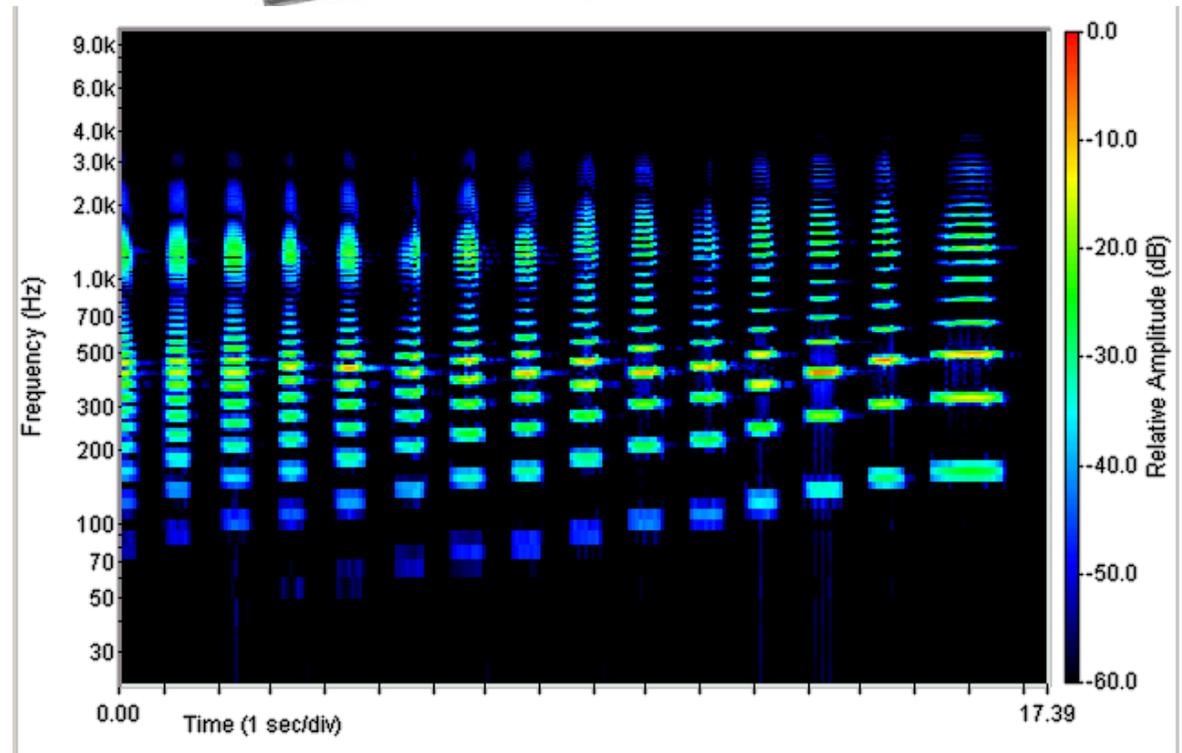


If Instruments could talk ...

Musical Instruments Formants

With increasing pitch of the fundamental, the maximum in the formant area rises too, until it reaches the end of the formant area.

Then the **amplitude maximum swaps** to the **nearest lower partial**, which again rises to the formant areas limit etc.



Sonogramm of the fixed formant areas in the case of the bassoon at different pitches



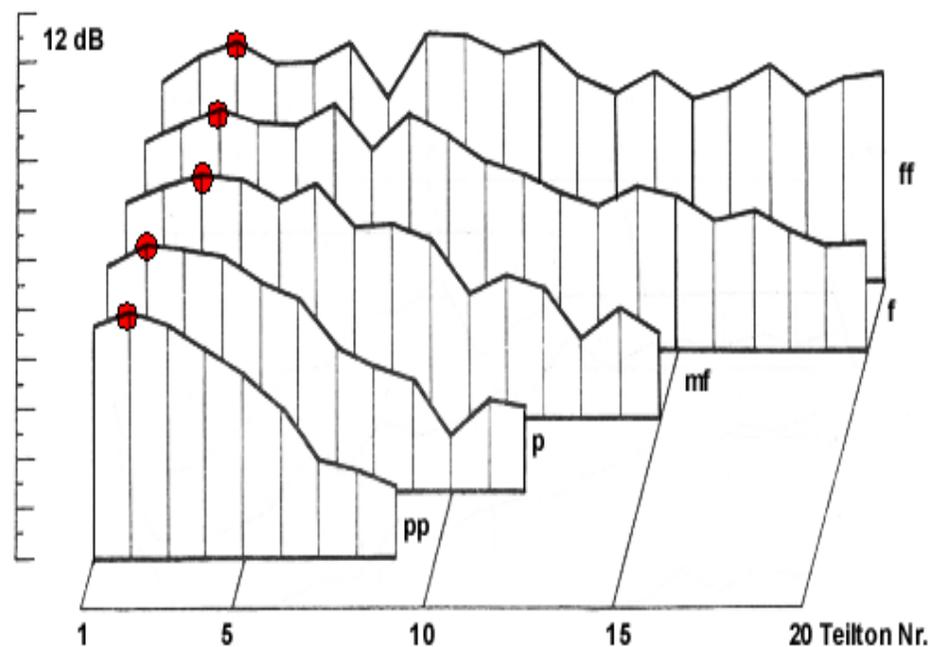
If Instruments could talk ...

Musical Instruments Formants

- **Principle of Formant Shifting**
[“*Formantverschiebungsgesetz*”]

With increasing musical dynamics, the strongest amplitude of the partial in the formant area shifts to a partial of higher order in the same formant area

(Schumann, 1929, p. 15–18, 98 and 100).



Formant shifting in the spectra of the trumpet
at different dynamics
(Müller 1971, p. 95)



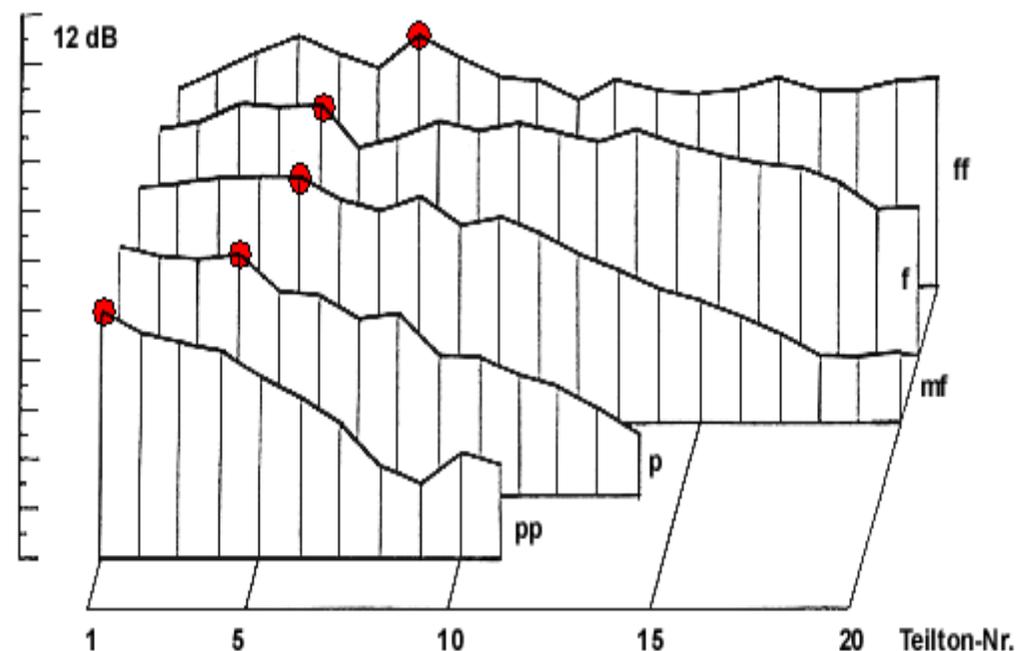
If Instruments could talk ...

Musical Instruments Formants

- **Principle of Spectral Gap Skipping** [*Sprunggesetz*]

With very intense musical dynamics, the strongest amplitude of the first (or lowest) formant area shifts to a partial in the second (higher) formant area, skipping over the partials between these areas.

(Schumann, 1929, p. 98 and 100).



Spectral gap skipping in the spectra of the trumpet
in case of extreme dynamic changes
(Müller 1971, p. 60)



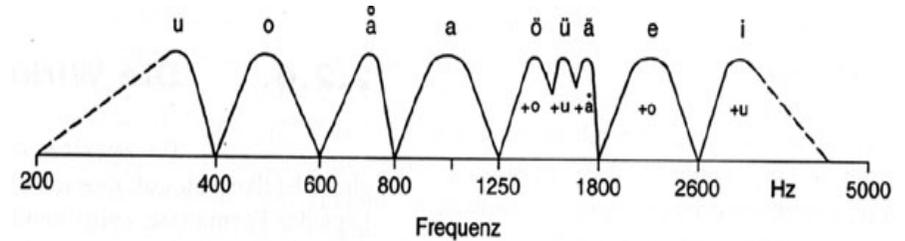
If Instruments could talk ...

Musical Instruments Formants

It is possible to categorize musical instruments sounds in terms of vowel formants.

Without formants, especially the timbres of brass and wind instruments would not sound typical anymore.

One can find more or less the **essence of a timbre** in the formant frequency bands.



Vowel formants (Meyer 2015, p. 33)

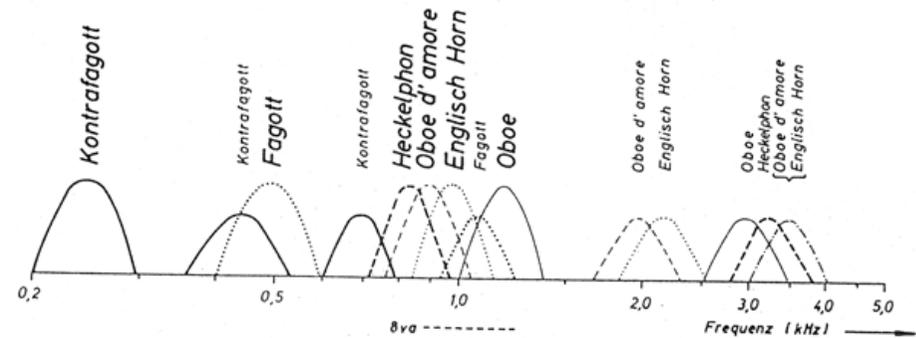


Abb. 6 Frequenzlage der Formanten für die Doppelrohrblattinstrumente, zusammengestellt nach Angaben von E. Meyer und G. Buchmann [3] (Oboen und Englisch Horn) und eigenen Messungen des Verf. (Fagotte [11] und Heckelphon)

Formants of double reed instruments (Meyer 2015, p. 63)



If Instruments could talk ...

Take Home Message

Musical instruments have **characteristic** and **pitch-independent formants** like vowels.

This spectral feature contributes to our ability to **recognize** and **categorize** musical instrument timbres like speech sounds.

The behaviour of musical instruments formants at **changes in pitch and dynamics** can be described by Schumanns „**Principles of Timbre**“

The **formants** of orchestral musical instruments are mostly located in different **frequency bands**. But there is a **common maximum** from **250 to 500 Hz** and **from 1000 to 1500 Hz**.



If Instruments could talk ...

Blending and Similarity

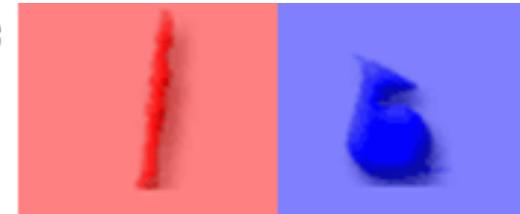
The role of formants in timbre blending while playing simultaneously in unison

Fragmentary masking by non-overlapping main formant areas:

Musical instrument timbres with **non-overlapping** main formant areas playing in unison can be **separated easily**, i.e. they can be distinguished very well from the total sound mixture.

(Fricke 1976 u. 1986)

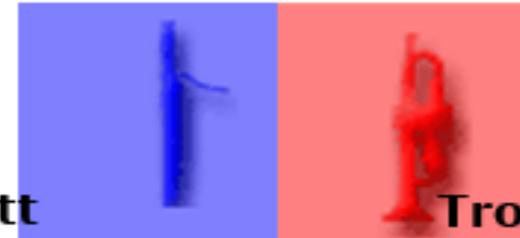
Oboe



Horn



Fagott



Trompete





If Instruments could talk ...

Blending and Similarity

The role of formants in timbre blending while playing simultaneously in unison

Timbral blending caused by overlapping of main formant areas:

Musical instrument timbres with **overlapping** main formant areas playing in unison are **not easy or not at all separable** from the total sound mixture.

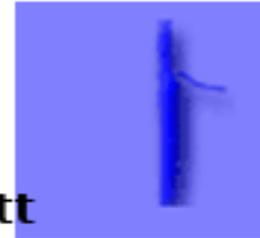
Oboe



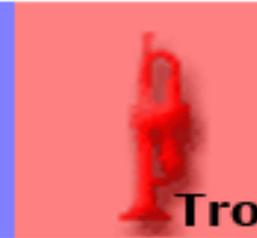
Horn



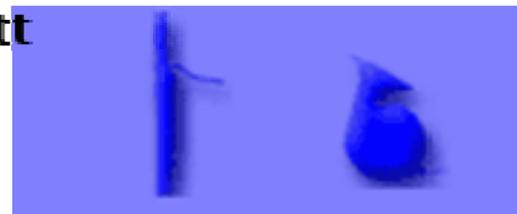
Fagott



Trompete



Fagott



Horn

Oboe



Trompete



If Instruments could talk ...

Blending and Similarity

Recommendations in orchestration treatises

Bassoon and French Horn blend well.

Roeser 1764, 22-23; Francoeur 1772, 55-56; Albrechtsberger 1790, 180; Vandenbroeck 1793, 9; Schubart 1806, 327; Marx 1851, 84, 145f. 148, 347; Gleich 1853, 21; Lobe 1878, 30, 31; Kling 1882, 33; Schubert 1885, 45; Prout 1888, 46-47, 106; Jadassohn 1889, 242, 254, 346, 348; Widor 1904, 46; Rimski-Korssakow 1912, 24, 57, 83, 88, 90; Riemann 1919, 48, 75; Körner, Rathke-Bernburger 1927, Tabelle; Heckel 1931, 23; Ribate 1943, 90 etc.

Oboe and Trumpet blend well.

Marx 1851, 208, 209, 347, 525; Rimski-Korssakow 1912, 35, 56, 88, 89, 92; Körner, Rathke-Bernburger 1927, Tabelle; Ribate 1943, 90; Koechlin 1955, Bd. 2, 193; Piston 1955, 427; Kunitz 1956, Bd. 3, 74; ders. 1958, Bd. 7, 555; ders. 1961, 20, 55; Kennan 1962, 169; Jacob 1962, 39-40, 65 etc.

Oboe and French Horn don't blend well.

Vandenbroeck 1793, 9; Marx 1851, 179f.; Prout 1888, 71, 106; Jadassohn 1889, 346; Volbach 1910, 43; Koechlin 1955, Bd. 2, 187, 235; Kunitz 1957, Bd. 6, 463, 472.

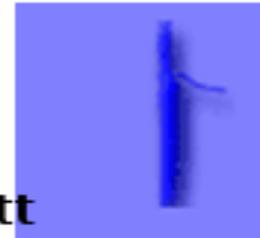
Bassoon and Trumpet don't blend well.

Schubart 1806 1969, 327; Koechlin 1955, Bd. 2, 196 U. 242; Kunitz 1958, Bd. 7, 557

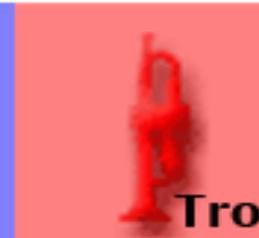
Oboe



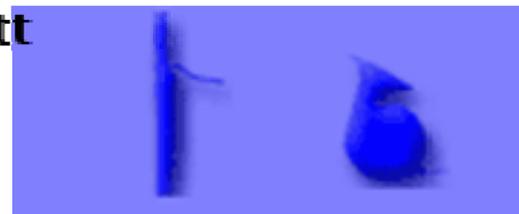
Horn



Fagott



Trompete



Fagott

Horn



Oboe

Trompete



If Instruments could talk ...

Take Home Message

In case of **simultaneously** playing musical instruments:
Formants are very helpful for predicting perceptual **timbre blending** or **separation**:

Matching main formants = timbres blend **homogenously**

Non-matching formants = timbres are perceived as **separated**

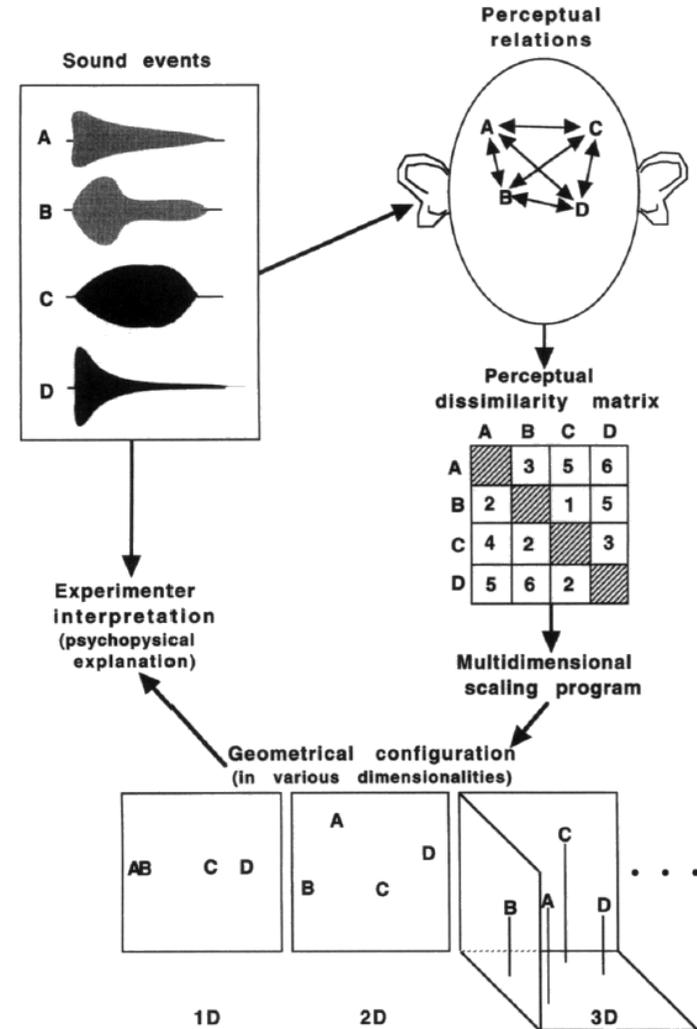
In case of **alternatingly** playing musical instruments:
Formants are very helpful for predicting **perceptual grouping** of melodies:

Matching main formants = only **one melody** gets perceived.

Non-matching formants = **two interwoven melodies** get perceived.



If Instruments could talk ... How to calculate Similarity?



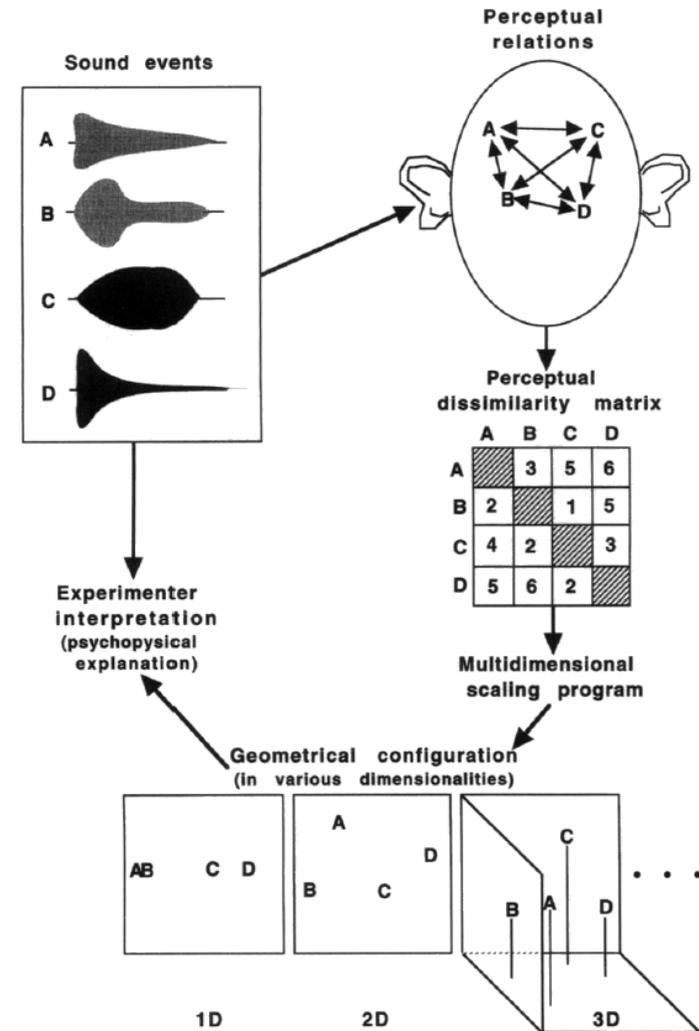
Calculation of a Timbre (Similarity) Space (McAdams 1999, p. 87)



If Instruments could talk ...

How to calculate Similarity?

1. Test subjects **compare timbres** (A-B comparison) and evaluate the perceived (dis)similarity on a **(dis)similarity scale**.



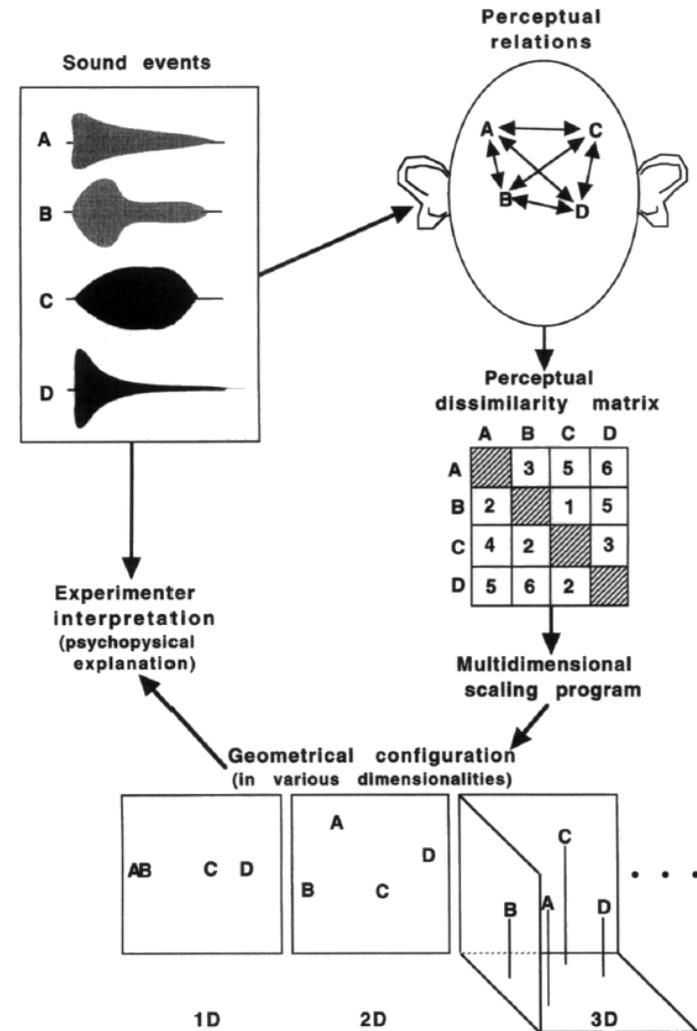
Calculation of a Timbre (Similarity) Space (McAdams 1999, p. 87)



If Instruments could talk ...

How to calculate Similarity?

1. Test subjects compare timbres (A-B comparison) and evaluate the perceived (dis)similarity on a (dis)similarity scale.
2. The perceived (dis)similarities of all timbres are listed as **numbers** in a **perceived dissimilarity matrix**.



Calculation of a Timbre (Similarity) Space (McAdams 1999, p. 87)



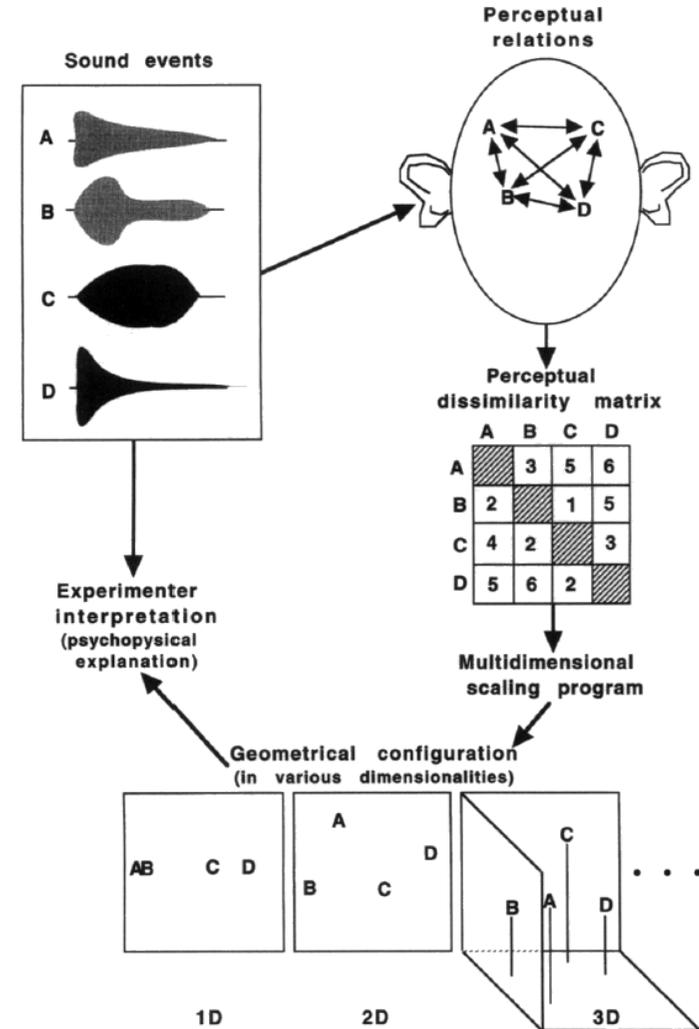
If Instruments could talk ...

How to calculate Similarity?

1. Test subjects compare timbres (A-B comparison) and evaluate the perceived (dis)similarity on a (dis)similarity scale.
2. The perceived (dis)similarities of all timbres are listed as numbers in a perceived dissimilarity matrix.

3. With the help of **Multidimensional Scaling** the number of perceptual dimensions are calculated.

The **closer** the entities on these dimensions, the **more similar** the timbre perception.



Calculation of a Timbre (Similarity) Space (McAdams 1999, p. 87)



If Instruments could talk ...

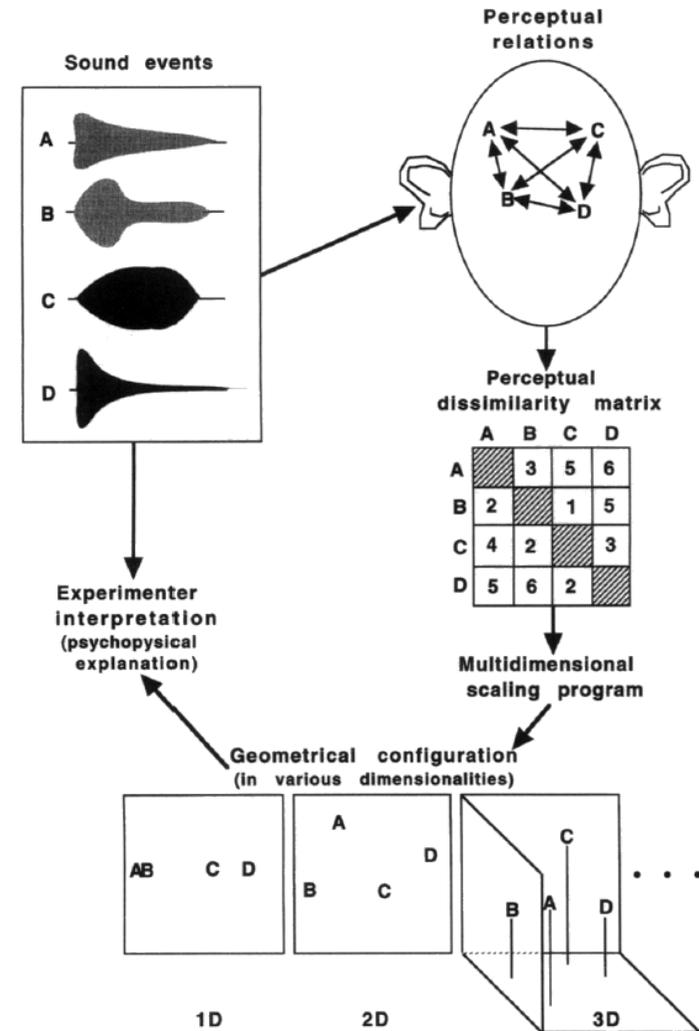
How to calculate Similarity?

1. Test subjects compare timbres (A-B comparison) and evaluate the perceived (dis)similarity on a (dis)similarity scale.
2. The perceived (dis)similarities of all timbres are listed as numbers in a perceived dissimilarity matrix.

3. With the help of **Multidimensional Scaling** the number of perceptual dimensions are calculated.

The **closer** the entities on these dimensions, the **more similar** the timbre perception.

4. The **dimensions** of the space get tested of **correlations** with timbre features.



Calculation of a Timbre (Similarity) Space (McAdams 1999, p. 87)

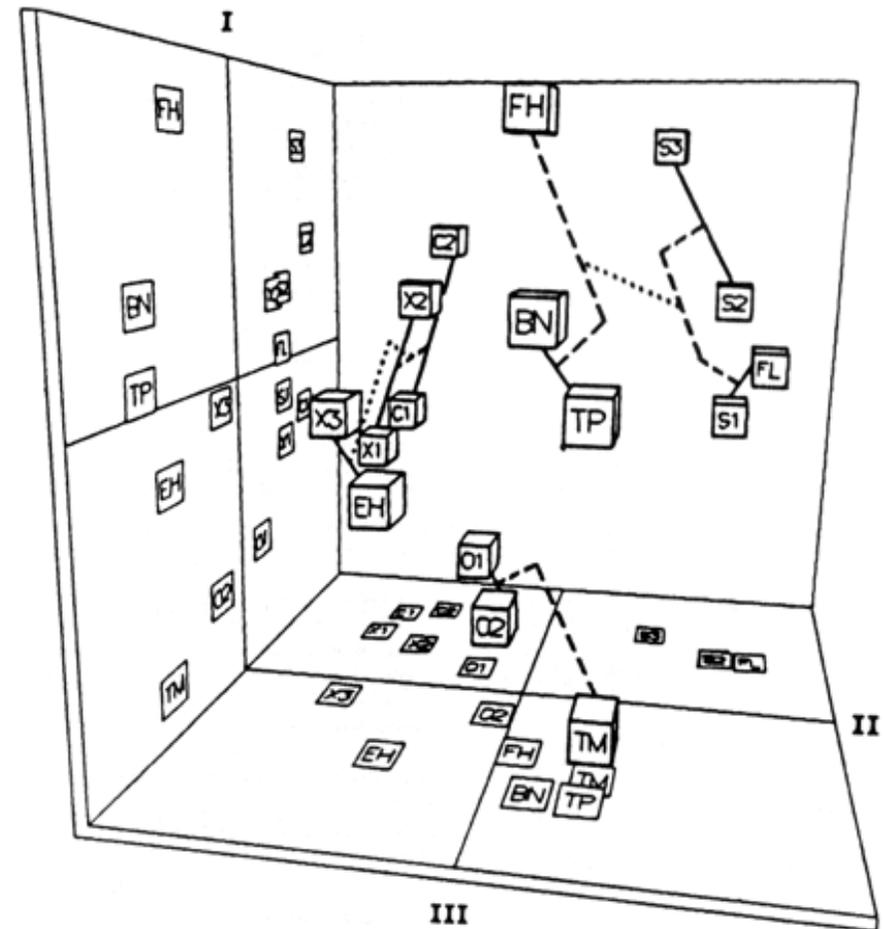


If Instruments could talk ...

How to calculate Similarity?

The first known Timbre Space has been built up by John Grey in 1975

- Dimension I: **spectral energy distribution**



Timbre (Similarity) Space
(Grey 1975, p. 62)

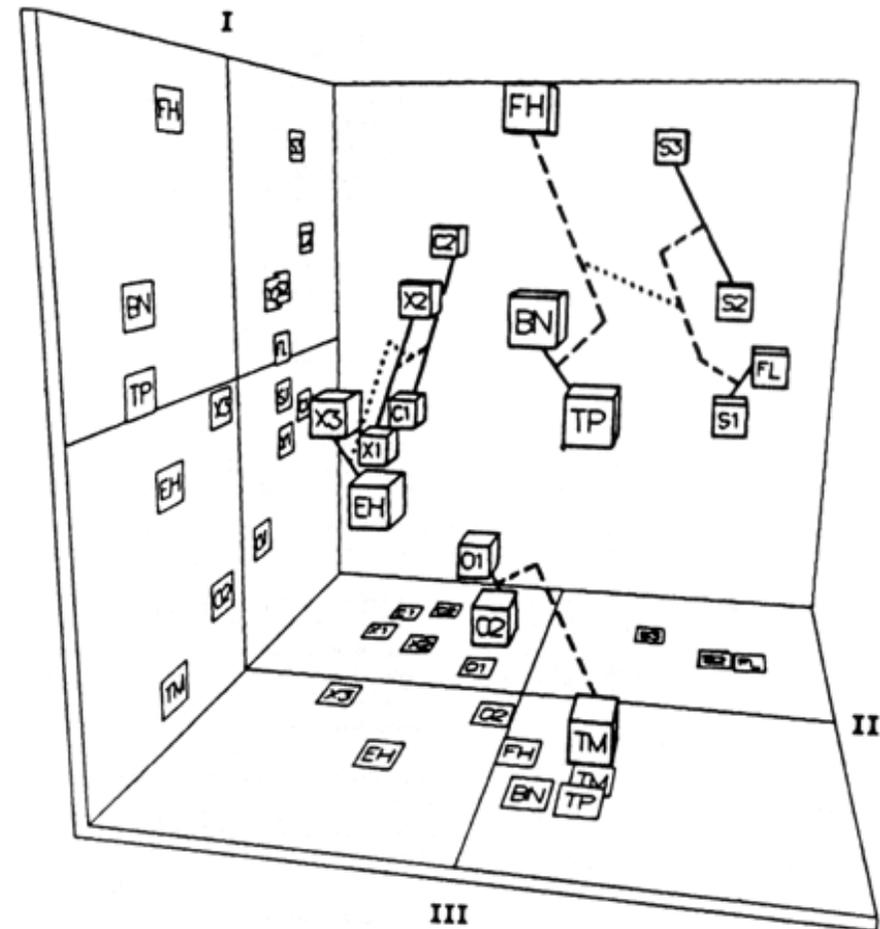


If Instruments could talk ...

How to calculate Similarity?

The first known Timbre Space has been built up by John Grey in 1975

- Dimension I: **spectral energy distribution**
- Dimension II: **onset-offset pattern** (especially attack transients and the synchronicity of upper harmonics)



Timbre (Similarity) Space
(Grey 1975, p. 62)

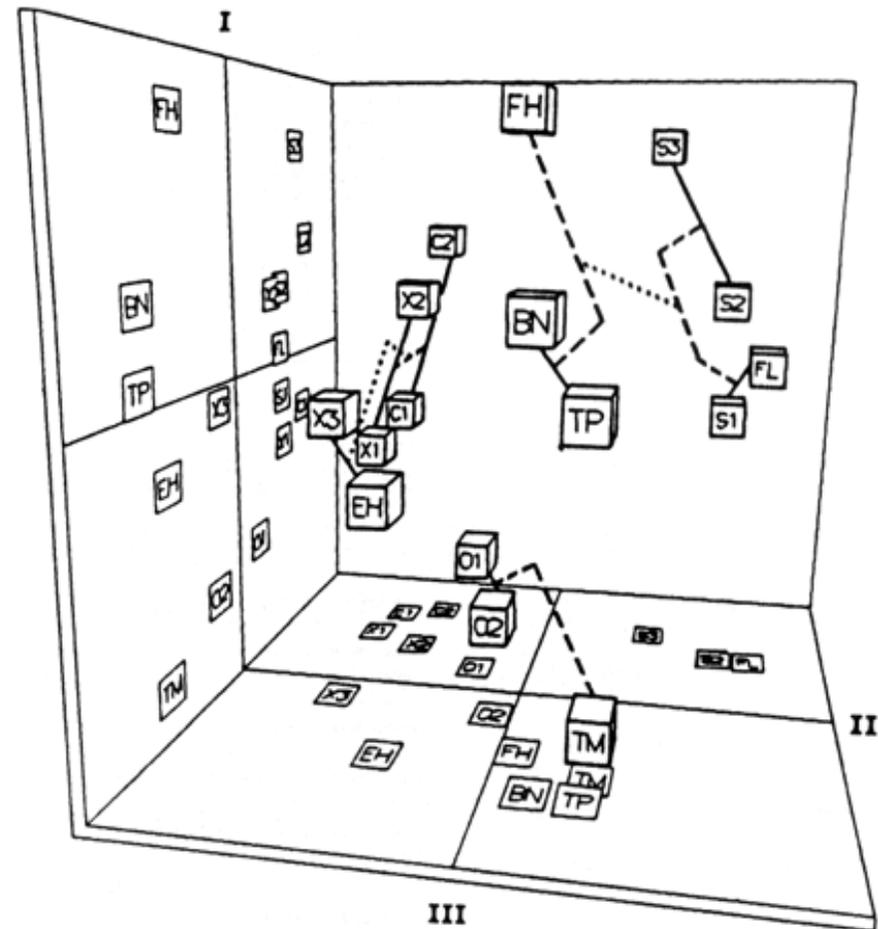


If Instruments could talk ...

How to calculate Similarity?

The first known Timbre Space has been built up by John Grey in 1975

- Dimension I: **spectral energy distribution**
- Dimension II: **onset-offset pattern** (especially attack transients and the synchronicity of upper harmonics)
- Dimension III: **temporal patterns** (fluctuations as well as the amount of inharmonicity in the attack part)



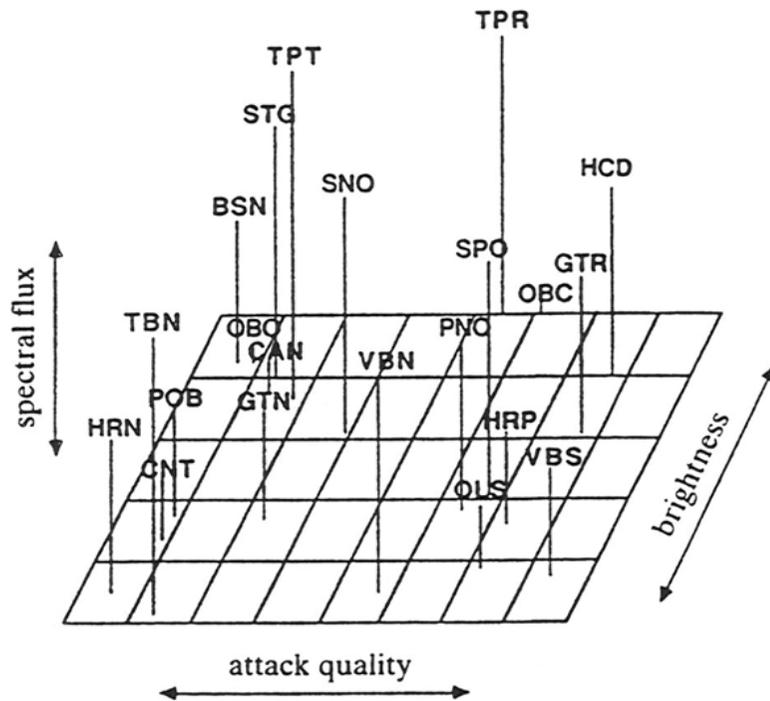
Timbre (Similarity) Space
(Grey 1975, p. 62)



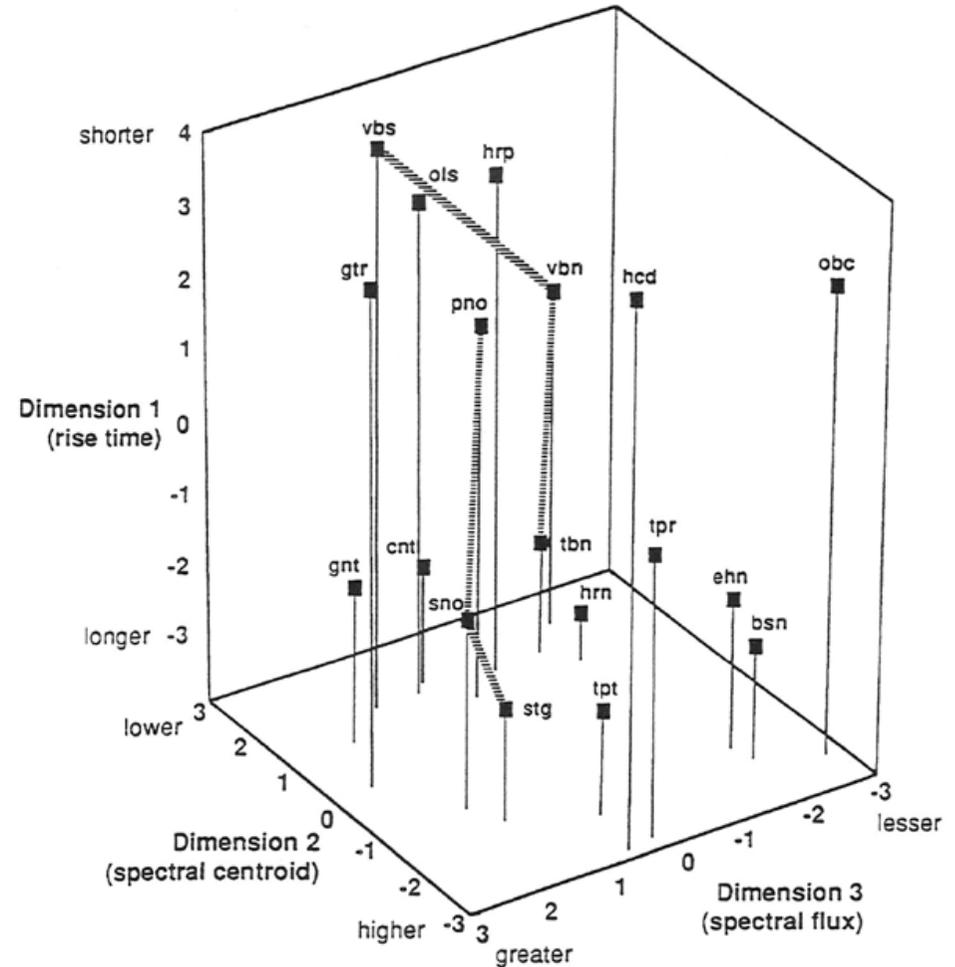
If Instruments could talk ...

How to calculate Similarity?

Further Timbre Spaces



Timbre Space based on synthetic FM sounds (Krumhansl 1989, p. 47)



Timbre Space based on synthetic FM sounds (McAdams et al. 1995, p. 185; McAdams 1999, p. 89)



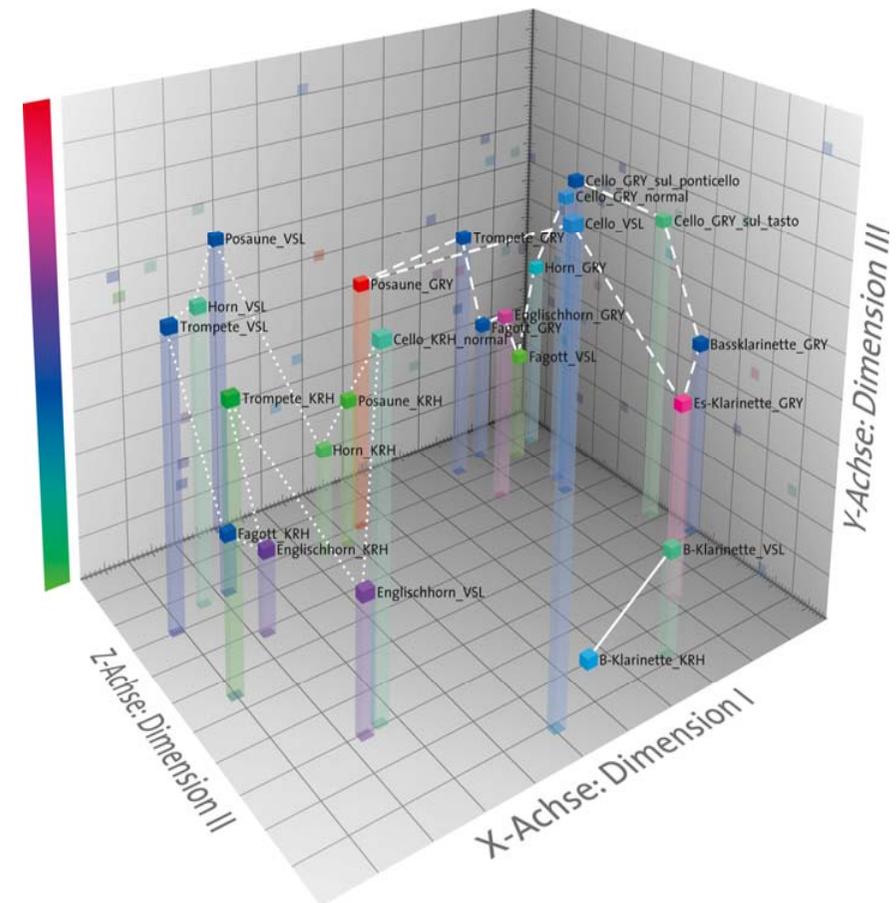
If Instruments could talk ...

How to calculate Similarity?

Meta Timbre Space

35 test subjects rated the (dis)similarity of **24** timbres (pitch: Eb4, 313 Hz):

- Grey (10 sounds, GRY)
- Krumhansl/McAdams (7 sounds, KRH)
- Vienna Symphonic Library (7 sounds, VSL)



[Meta Timbre Space based on sounds of Grey, Krumhansl/McAdams and Vienna Symphonic Library \(Siddiq, Reuter, Czedik-Eysenberg, Knauf 2015, p. 812\)](#)



If Instruments could talk ...

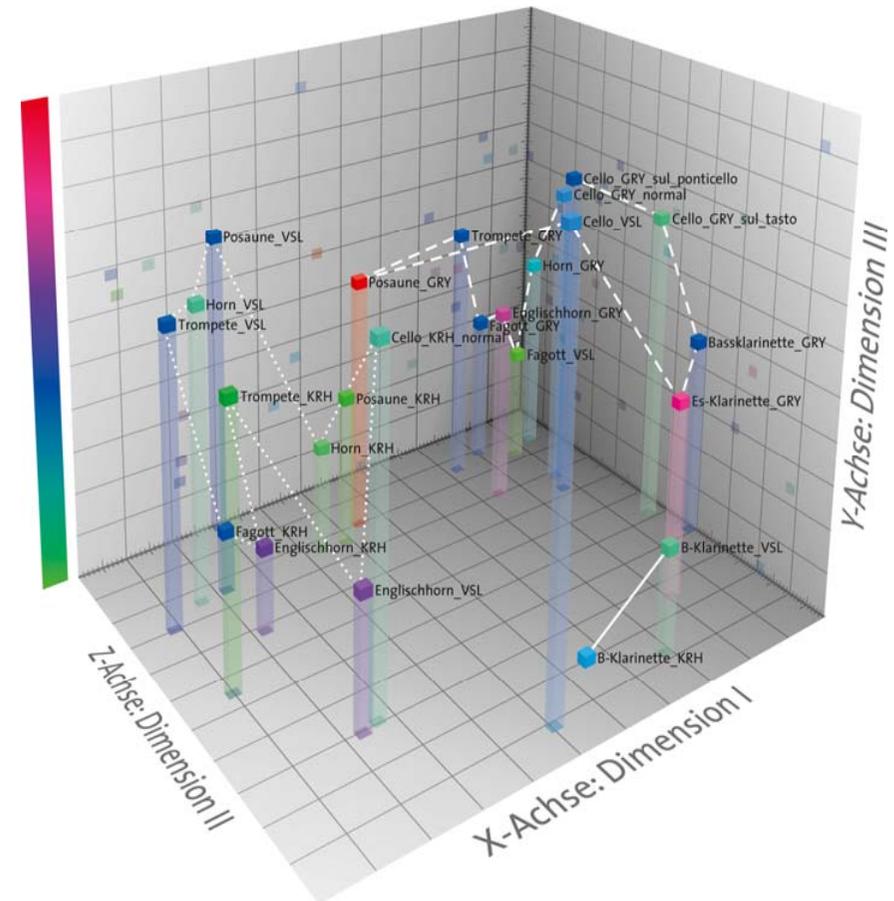
How to calculate Similarity?

Meta Timbre Space

35 test subjects rated the (dis)similarity of **24** timbres (pitch: Eb4, 313 Hz):

- Grey (10 sounds, GRY)
- Krumhansl/McAdams (7 sounds, KRH)
- Vienna Symphonic Library (7 sounds, VSL)

Via nonmetrical MDS (mdscale):
Meta Timbre Space with **4 Dimensions**
(stress= 0,0466)



[Meta Timbre Space based on sounds of Grey, Krumhansl/McAdams and Vienna Symphonic Library \(Siddiq, Reuter, Czedik-Eysenberg, Knauf 2015, p. 812\)](#)



If Instruments could talk ...

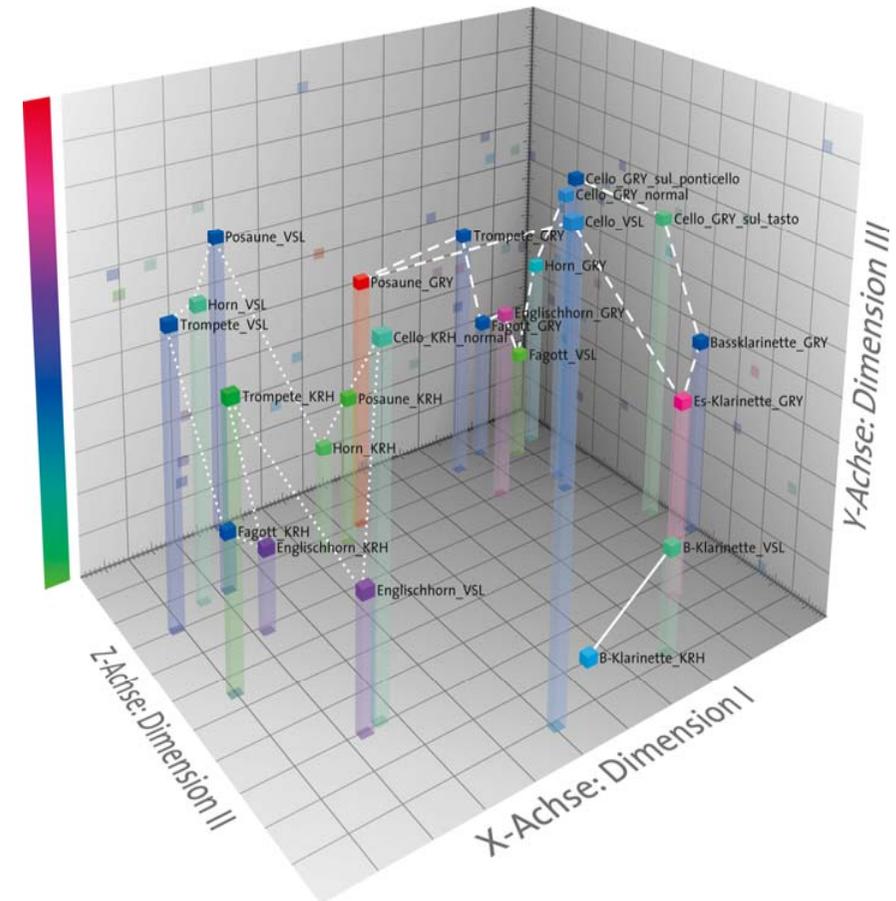
How to calculate Similarity?

Meta Timbre Space

In an agglomerative cluster analysis it turned out that timbres of the **same stimuli set** (GRY or KRH) show a larger similarity than timbres of the **same instrument**.

Timbre Spaces are more **stimuli-set-dependent** than instrument-dependent.

So Timbre Spaces are **hardly generalisable** or even **comparable**.



Meta Timbre Space based on sounds of Grey, Krumphansl/McAdams and Vienna Symphonic Library (Siddiq, Reuter, Czedik-Eysenberg, Knauf 2015, p. 812)



If Instruments could talk ...

Take Home Message

Timbre Spaces are very **descriptive**, **comprehensible** and **intuitive**,

but:

Timbre Spaces are **not** really **generalisable** or **comparable**.

Timbre Spaces are mostly based on **(re)synthesized timbres** in only **one single pitch**.

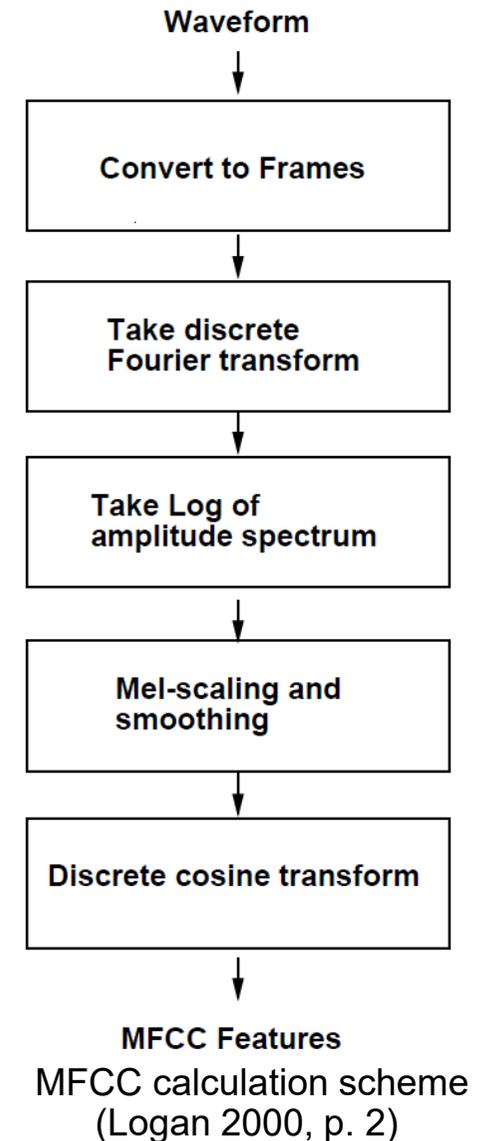
Dynamics, **articulations** etc. mostly have been neglected in Timbre Space studies.



If Instruments could talk ...

How to calculate Similarity?

With **MFCCs** (Mel Frequency Cepstrum Coefficients) Stephen Davis and Paul Memelstein developed a calculation method for **automatic speaker recognition** or **speech similarity evaluation** in 1980.



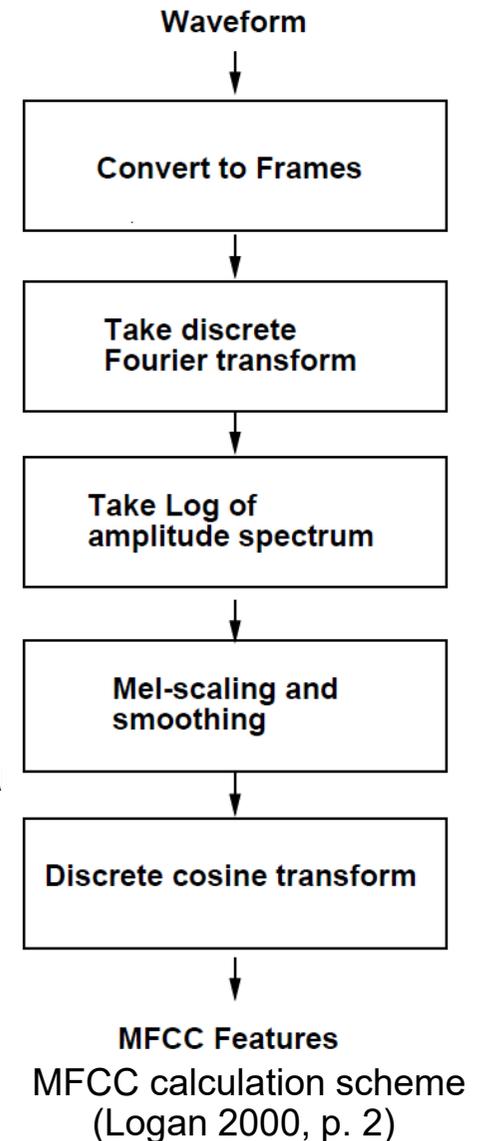


If Instruments could talk ...

How to calculate Similarity?

With **MFCCs** (Mel Frequency Cepstrum Coefficients) Stephen Davis and Paul Memelstein developed a calculation method for **automatic speaker recognition** or **speech similarity evaluation** in 1980.

In short, the method calculates for each 20 ms frame of the waveform a **mel-scale adapted Cepstrum** (spectrum of a spectrum) and compares the resulting envelope (as a 13-dimensional **vector**) with a set of standard envelopes (the coefficients).





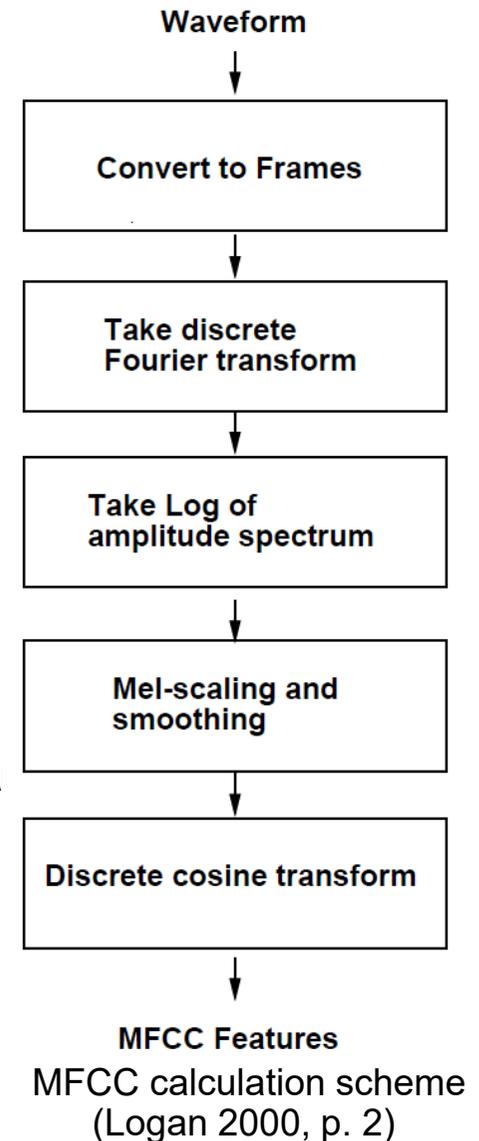
If Instruments could talk ...

How to calculate Similarity?

With **MFCCs** (Mel Frequency Cepstrum Coefficients) Stephen Davis and Paul Memelstein developed a calculation method for **automatic speaker recognition** or **speech similarity evaluation** in 1980.

In short, the method calculates for each 20 ms frame of the waveform a **mel-scale adapted Cepstrum** (spectrum of a spectrum) and compares the resulting envelope (as a 13-dimensional **vector**) with a set of standard envelopes (the coefficients).

This method **does not comply** with the mechanism of our audio perception, but the results are very convincing.

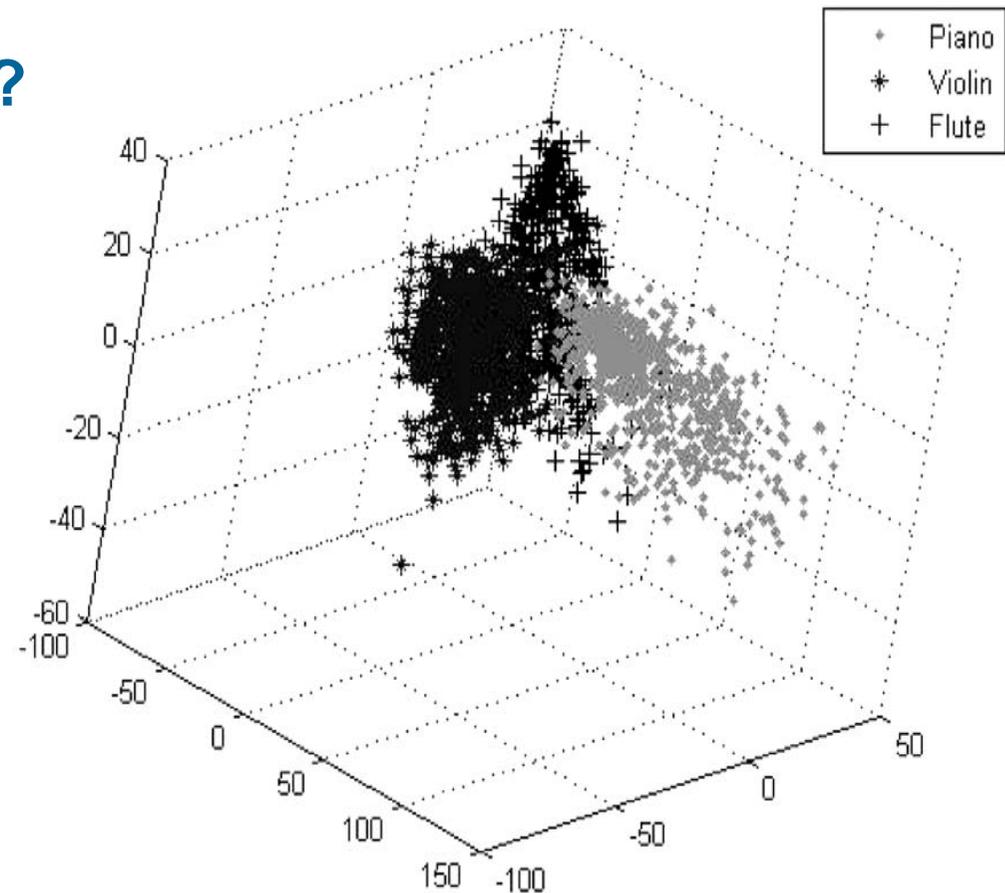




If Instruments could talk ...

How to calculate Similarity?

This method turned out to be also applicable for **automatic recognition** and **categorization of musical instruments** and music. Today MFCCs are the **standard method** for calculating audio similarity.



Computational timbre description/separation by MFCCs
(Loughran, Walker, O'Neill, O'Farrel 2008, p. 3)

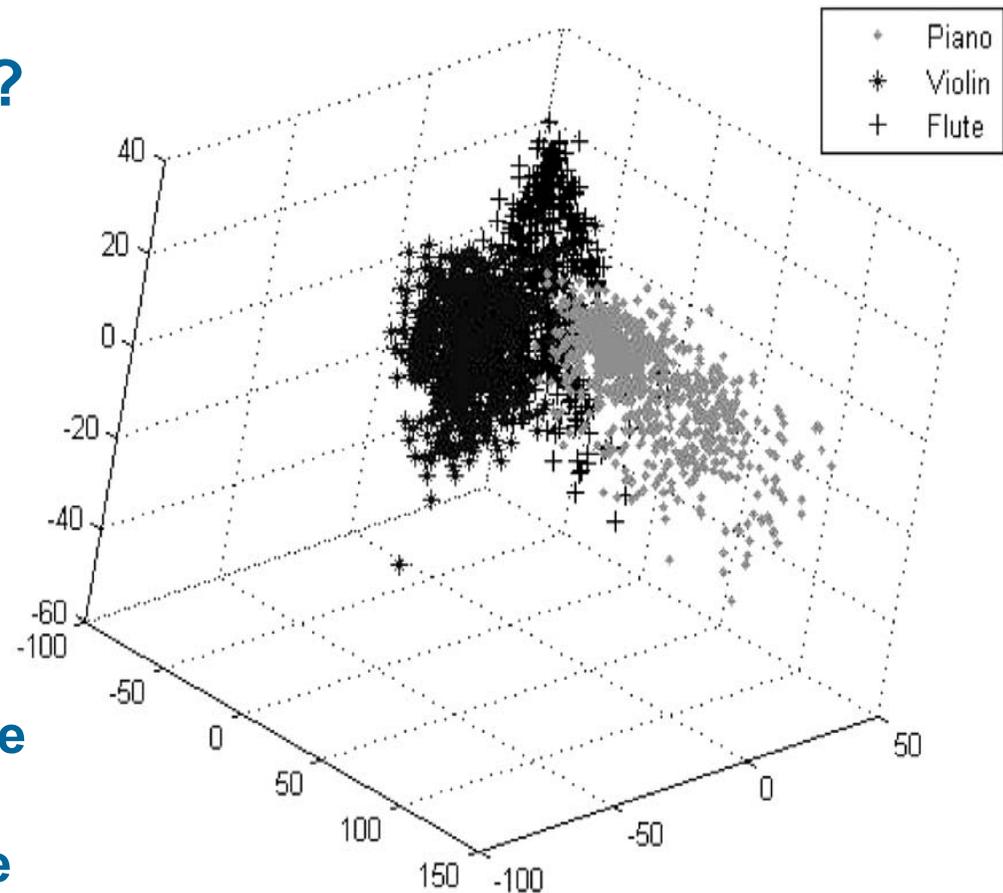


If Instruments could talk ...

How to calculate Similarity?

This method turned out to be also applicable for **automatic recognition** and **categorization of musical instruments** and music. Today MFCCs are the **standard method** for calculating audio similarity.

Question: Is there a more intuitive alternative for timbre similarity calculation on the basis of timbre features, which are more suitable with human audio perception?



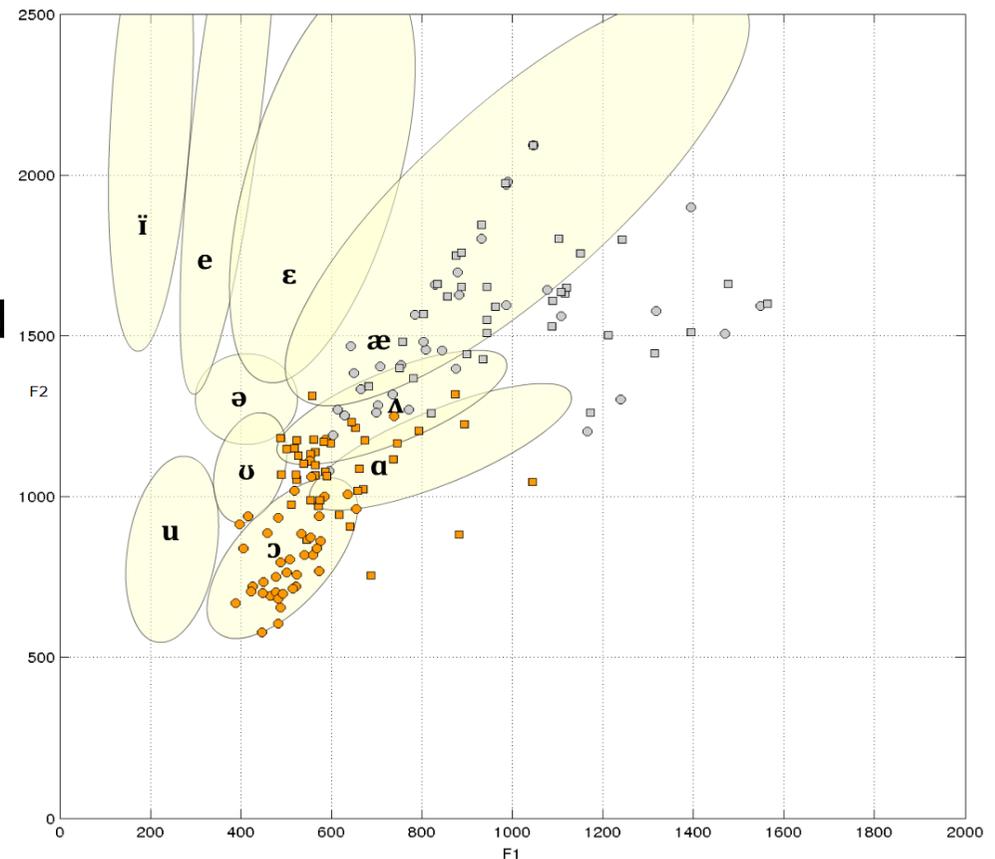
Computational timbre description/separation by MFCCs
(Loughran, Walker, O'Neill, O'Farrel 2008, p. 3)



If Instruments could talk ...

Formants vs. MFCCs

With the help of Praat the **first and second formant (F1 and F2)** of conventional western orchestral wind instruments have been measured in **all reachable pitches** and in **two different dynamics** (*ff* and *pp*).



Formant map with the sounds of bassoon (orange) and oboe (grey) in all achievable pitches in *ff* and *pp* (Reuter, Czedik-Eysenberg, Siddiq, & Oehler, 2017).

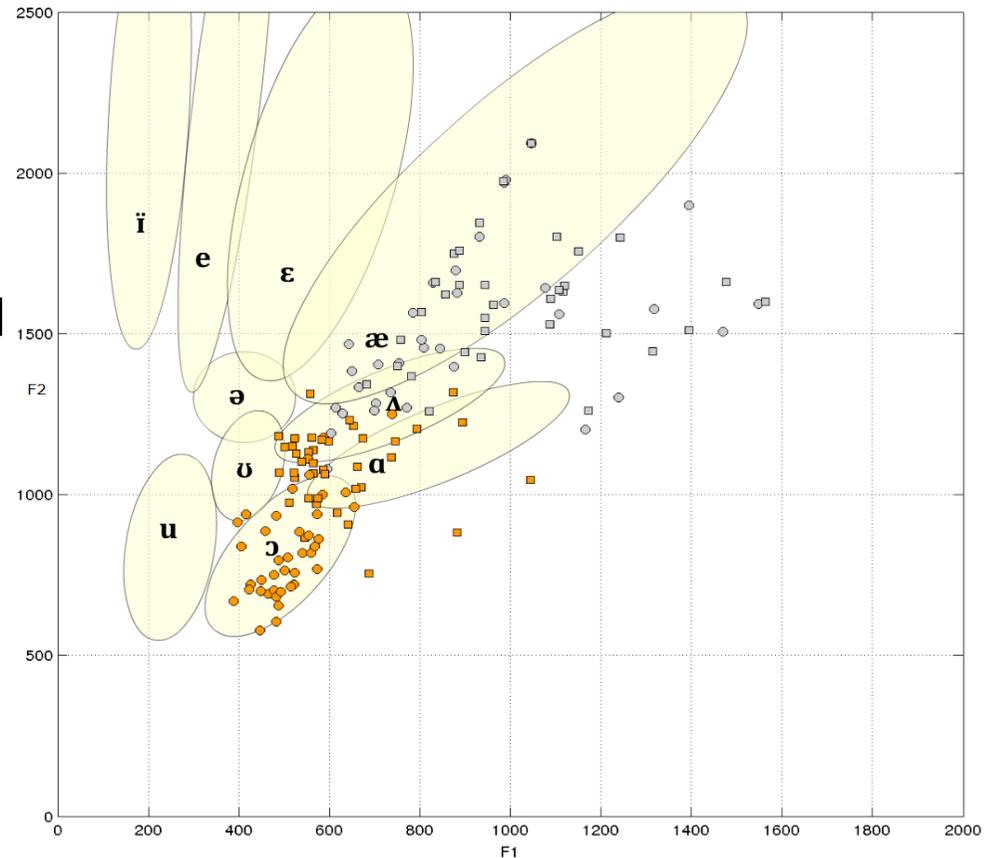


If Instruments could talk ...

Formants vs. MFCCs

With the help of Praat the **first and second formant (F1 and F2)** of conventional western orchestral wind instruments have been measured in **all reachable pitches** and in **two different dynamics** (*ff* and *pp*).

In a field between the two dimensions **F1** and **F2** each **pitch** and **dynamic** is symbolized by a **point** positioned by the middle frequency of the first and second formant each.



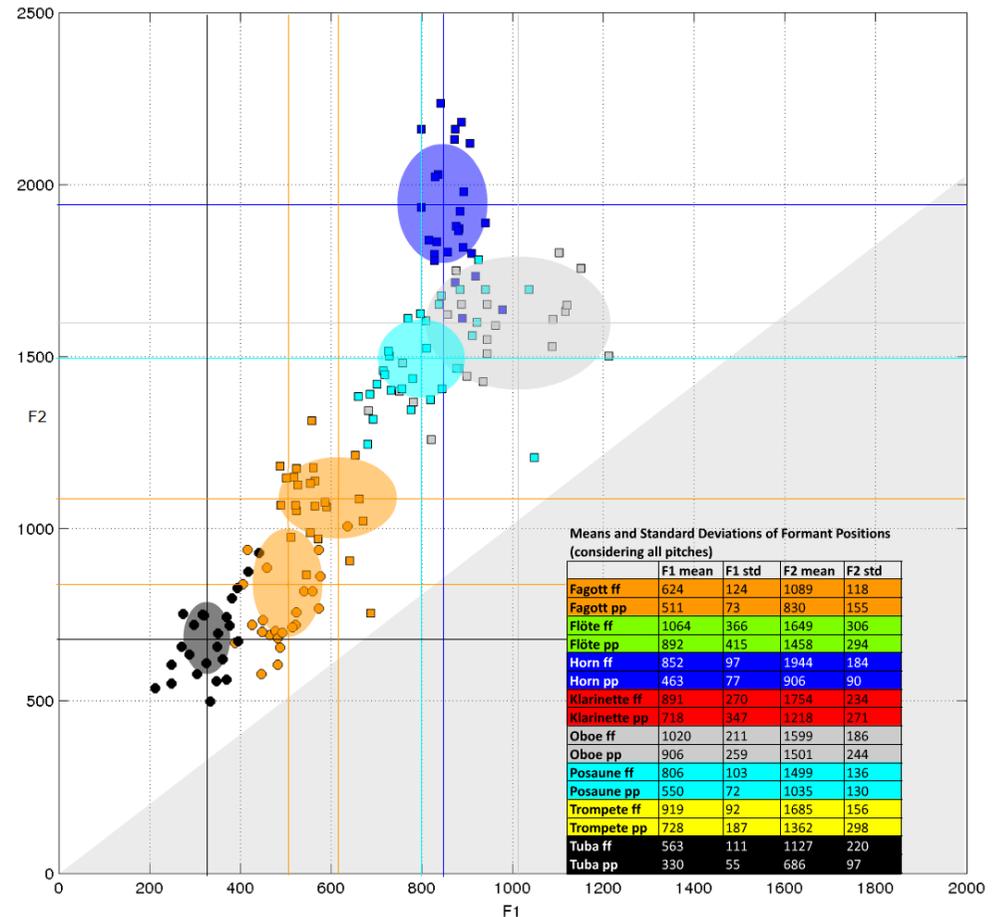
Formant map with the sounds of bassoon (orange) and oboe (grey) in all achievable pitches in *ff* and *pp* (Reuter, Czedik-Eysenberg, Siddiq, & Oehler, 2017).



If Instruments could talk ...

Formants vs. MFCCs

With the help of the **first two formant areas (F1 and F2)** timbres of wind instruments with concise formant structures can be visually and auditively **discriminated** and **matched** with corresponding **vowel timbres.**



Formants, their mean and standard deviation of oboe, trombone, bassoon and tuba (Reuter, Siddiq, Czedik-Eysenberg, Oehler 2016)

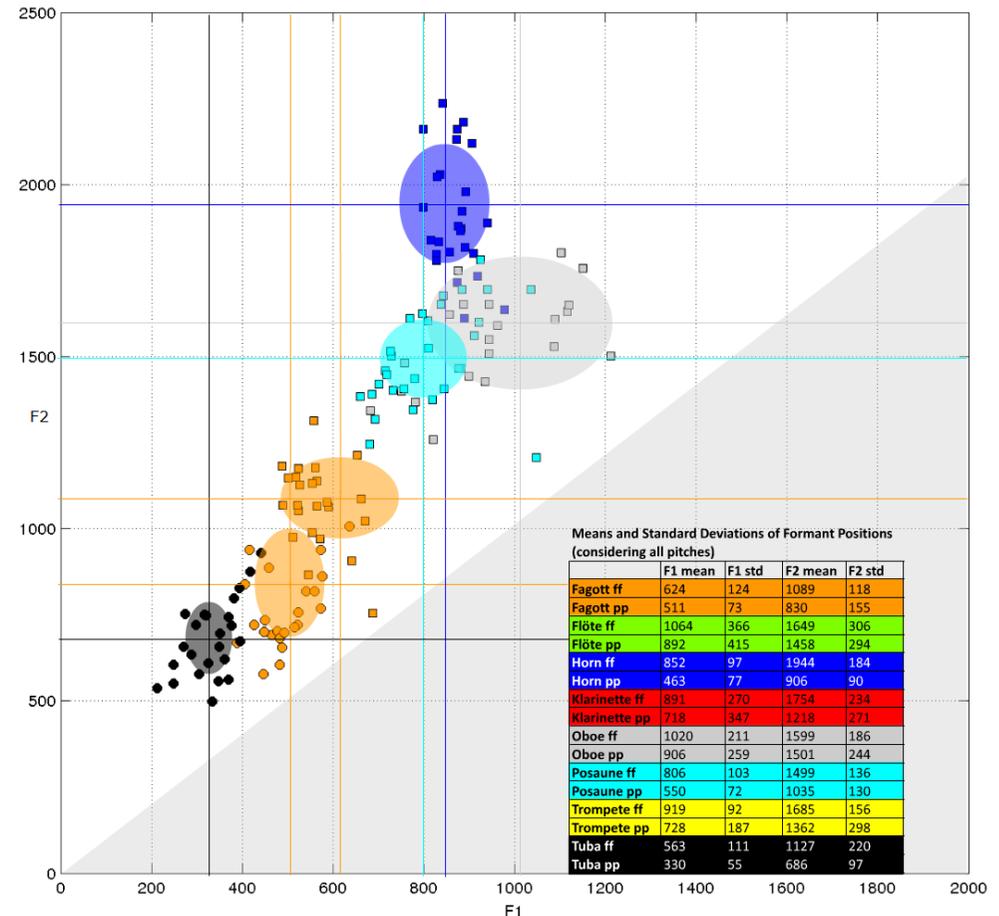


If Instruments could talk ...

Formants vs. MFCCs

With the help of the **first two formant areas (F1 and F2)** timbres of wind instruments with concise formant structures can be visually and auditively **discriminated** and **matched** with corresponding **vowel timbres**

Question: It looks intuitively, but does it really work?



Formants, their mean and standard deviation of oboe, trombone, bassoon and tuba (Reuter, Siddiq, Czedik-Eysenberg, Oehler 2016)



If Instruments could talk ...

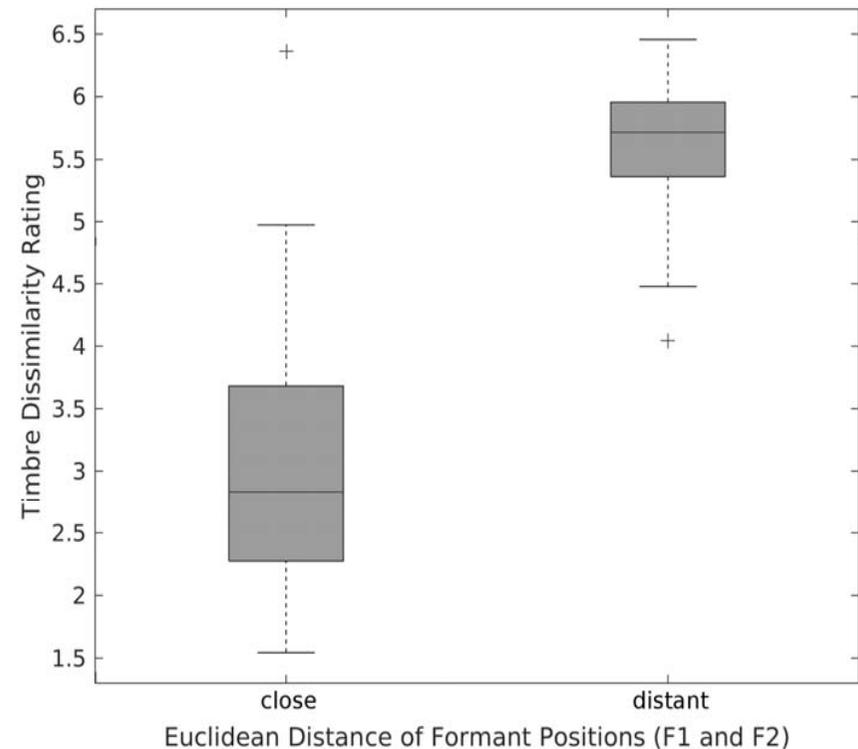
Formants vs. MFCCs

22 participants listened to **40** loudness-adjusted timbre combinations:

20 * **very close** formant regions and

20 * **very distant** formant regions

Rating them on a **(dis)similarity scale (1-8)**; (8 = maximum dissimilarity).



The distance of the formant positions (X-axis: close vs. distant) correlates strongly with ratings of perceived timbre similarity (Y-axis: 1 = very similar; 7 = very dissimilar). (r = 0.759, t-test with p < 0.001, 95% CI [-3.1381, -1.8960])



If Instruments could talk ...

Formants vs. MFCCs

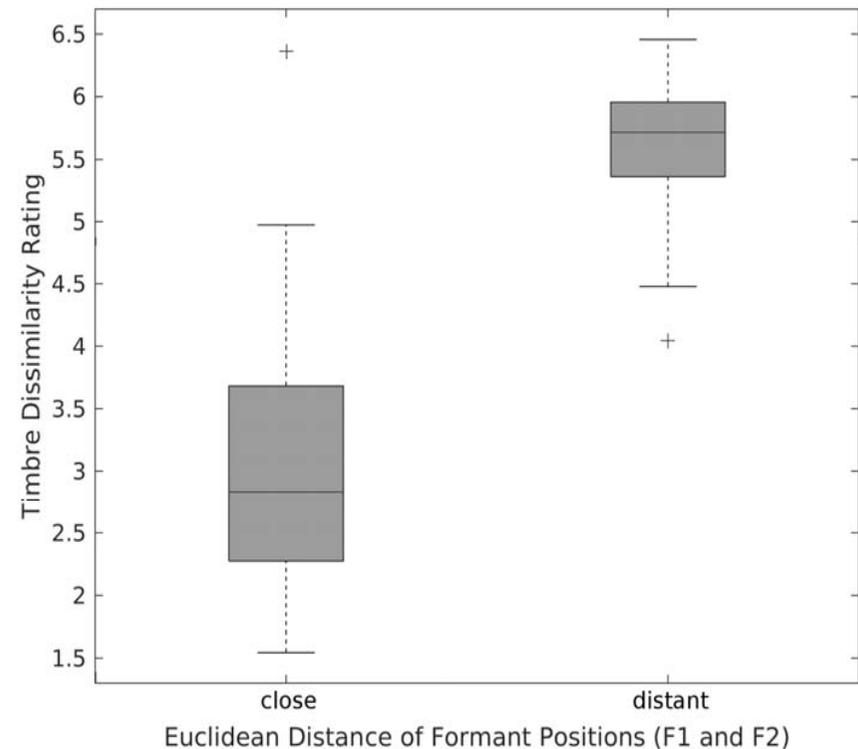
22 participants listened to **40** loudness-adjusted timbre combinations:

20 * **very close** formant regions and

20 * **very distant** formant regions

Rating them on a **(dis)similarity scale (1-8)**; (8 = maximum dissimilarity).

Results: The **distance** of the formant positions **correlates strongly** with ratings of **perceived timbre similarity** ($r = 0,759$, $p < 0,001$).



The distance of the formant positions (X-axis: close vs. distant) correlates strongly with ratings of perceived timbre similarity (Y-axis: 1 = very similar; 7 = very dissimilar). ($r = 0.759$, t-test with $p < 0.001$, 95% CI [-3.1381, -1.8960])



If Instruments could talk ...

Formants vs. MFCCs

A **comparison** between formants and MFCCs show ...

Timbre feature	r	p
Formant 1	0.7514	< 0.0001
Formant 2	0.7477	< 0.0001
Formant 3	0.4227	< 0.0001
MFCC 1	0.6384	< 0.0001
MFCC 2	0.5959	< 0.0001
MFCC 3	0.3513	0.0262
MFCC4	0.0261	0.8731
MFCC5	0.0781	0.6320
MFCC6	0.1129	0.4879
MFCC7	-0.2638	0.1000
MFCC8	0.2317	0.1503
MFCC9	0.0771	0.6364
MFCC10	0.0463	0.7766
MFCC11	-0.0444	0.7858
MFCC12	-0.0093	0.9548
MFCC13	-0.1722	0.2881

Correlation of individual formant positions and MFCCs with the perceived timbre similarity.



If Instruments could talk ...

Formants vs. MFCCs

A **comparison** between formants and MFCCs show

- a **strong correlation** between formants distances and perceived timbre similarity.

Timbre feature	r	p
Formant 1	0.7514	< 0.0001
Formant 2	0.7477	< 0.0001
Formant 3	0.4227	< 0.0001
MFCC 1	0.6384	< 0.0001
MFCC 2	0.5959	< 0.0001
MFCC 3	0.3513	0.0262
MFCC4	0.0261	0.8731
MFCC5	0.0781	0.6320
MFCC6	0.1129	0.4879
MFCC7	-0.2638	0.1000
MFCC8	0.2317	0.1503
MFCC9	0.0771	0.6364
MFCC10	0.0463	0.7766
MFCC11	-0.0444	0.7858
MFCC12	-0.0093	0.9548
MFCC13	-0.1722	0.2881

Correlation of individual formant positions and MFCCs with the perceived timbre similarity.



If Instruments could talk ...

Formants vs. MFCCs

A **comparison** between formants and MFCCs show

- a **strong correlation** between formants distances and perceived timbre similarity.
- a **weaker correlation** of **MFCCs 1-3** with the listeners' similarity scores.

Timbre feature	r	p
Formant 1	0.7514	< 0.0001
Formant 2	0.7477	< 0.0001
Formant 3	0.4227	< 0.0001
MFCC 1	0.6384	< 0.0001
MFCC 2	0.5959	< 0.0001
MFCC 3	0.3513	0.0262
MFCC4	0.0261	0.8731
MFCC5	0.0781	0.6320
MFCC6	0.1129	0.4879
MFCC7	-0.2638	0.1000
MFCC8	0.2317	0.1503
MFCC9	0.0771	0.6364
MFCC10	0.0463	0.7766
MFCC11	-0.0444	0.7858
MFCC12	-0.0093	0.9548
MFCC13	-0.1722	0.2881

Correlation of individual formant positions and MFCCs with the perceived timbre similarity.



If Instruments could talk ...

Formants vs. MFCCs

A **comparison** between formants and MFCCs show

- a **strong correlation** between formants distances and perceived timbre similarity.
- a **weaker correlation** of **MFCCs 1-3** with the listeners' similarity scores.
- **no significant correlations** of the **MFCCs 4-13** to the listeners' judgements.

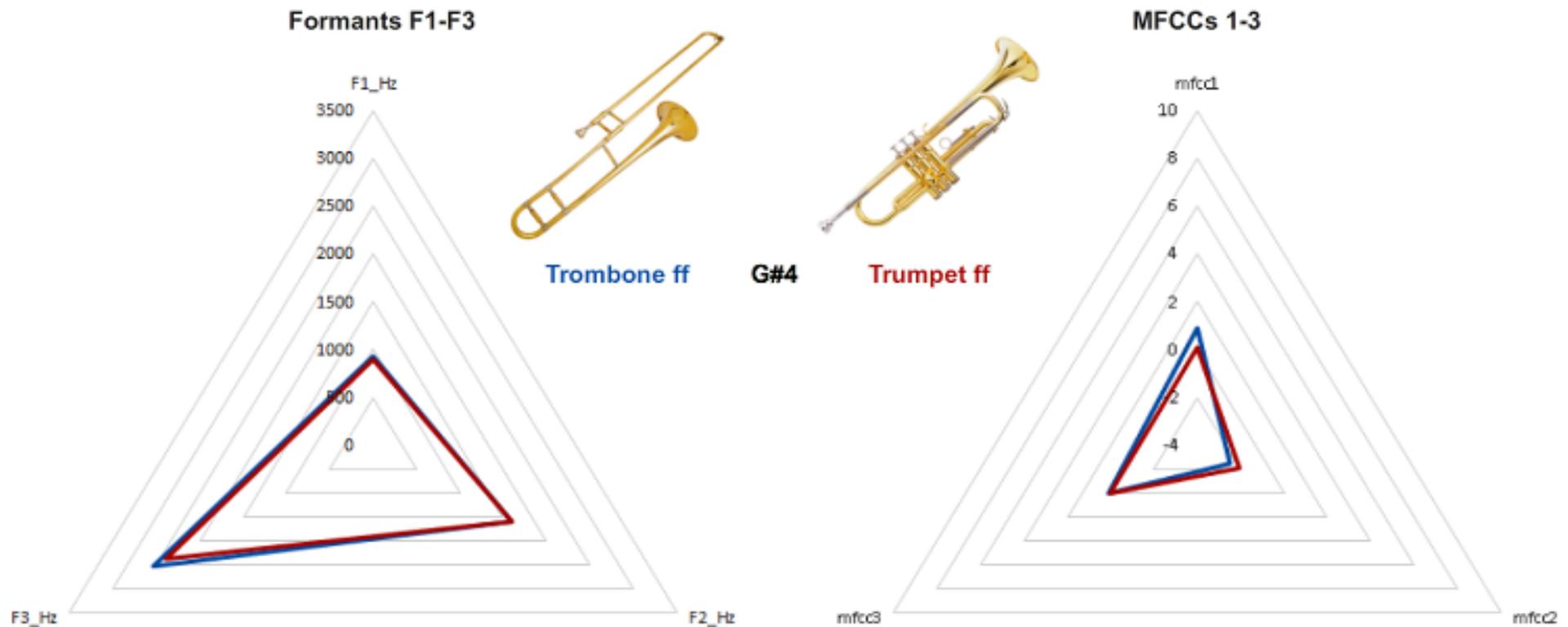
Timbre feature	r	p
Formant 1	0.7514	< 0.0001
Formant 2	0.7477	< 0.0001
Formant 3	0.4227	< 0.0001
MFCC 1	0.6384	< 0.0001
MFCC 2	0.5959	< 0.0001
MFCC 3	0.3513	0.0262
MFCC4	0.0261	0.8731
MFCC5	0.0781	0.6320
MFCC6	0.1129	0.4879
MFCC7	-0.2638	0.1000
MFCC8	0.2317	0.1503
MFCC9	0.0771	0.6364
MFCC10	0.0463	0.7766
MFCC11	-0.0444	0.7858
MFCC12	-0.0093	0.9548
MFCC13	-0.1722	0.2881

Correlation of individual formant positions and MFCCs with the perceived timbre similarity.



If Instruments could talk ...

Formants vs. MFCCs

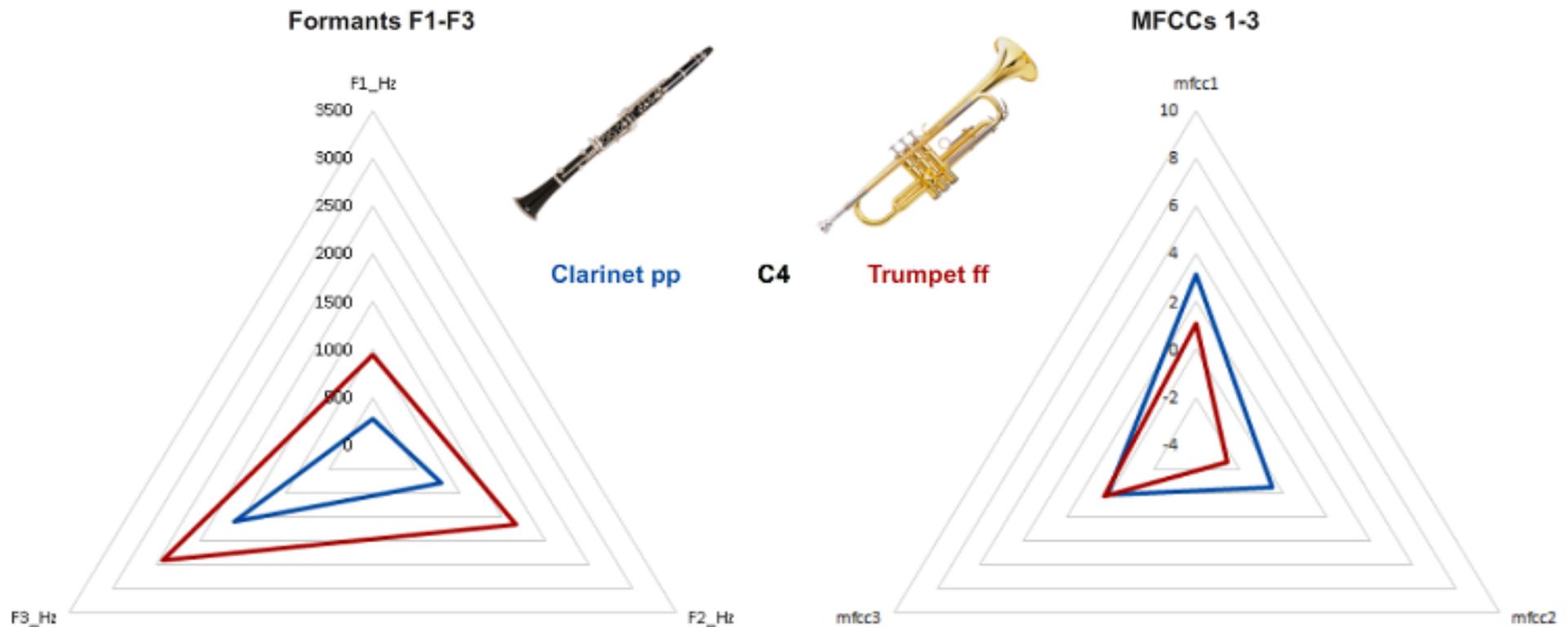


Spider's web representation with the respective axes F1, F2, F3 or MFCC1, MFCC2, MFCC3 of the pair with the strongest perceived timbre similarity (trombone *ff* and trumpet *ff* on G#4) (Reuter, Czedik-Eysenberg, Siddiq, Oehler 2018, p. 369)



If Instruments could talk ...

Formants vs. MFCCs



Spider's web representation with the respective axes F1, F2, F3 or MFCC1, MFCC2, MFCC3 of the pair with the least perceived timbre similarity (clarinet *pp* and trumpet *ff* on C4) (Reuter, Czedik-Eysenberg, Siddiq, Oehler 2018, p. 369)

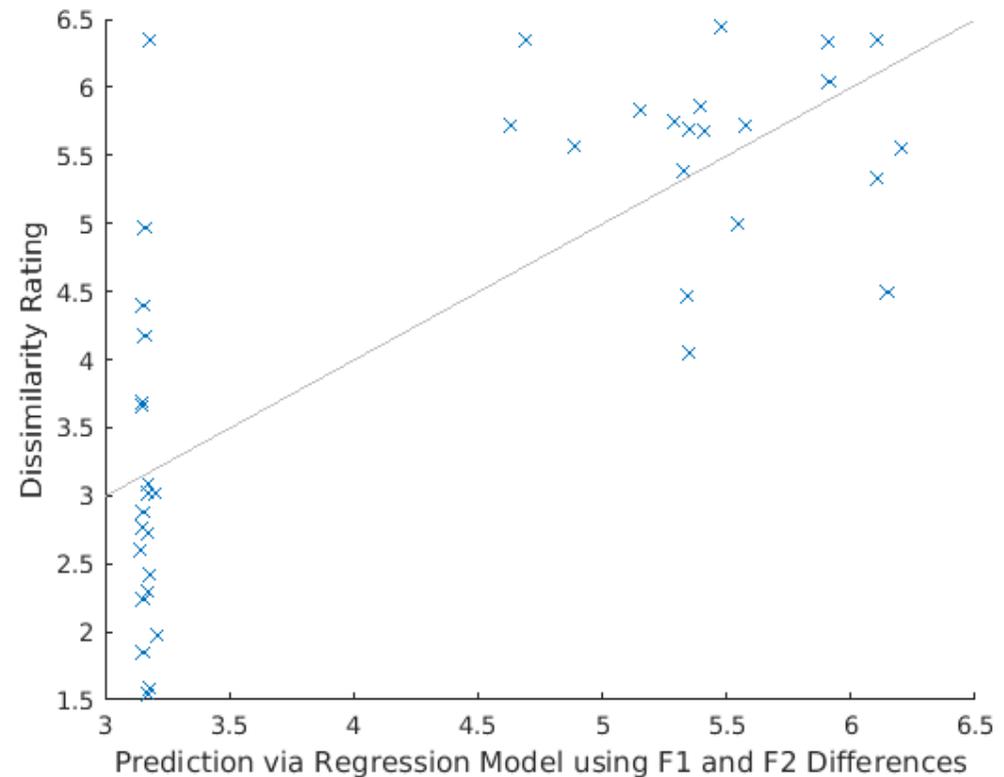


If Instruments could talk ...

Formants vs. MFCCs

Predicting timbre similarity perception (based on formants)

Regression models trained via machine learning (5-fold cross-validation)



Prediction model based on F1 and F2
($R^2 = 0.53$, RMSE = 1.08, MSE = 1.16, MAE = 0.86)
(Reuter, Czedik-Eysenberg, Siddiq, Oehler 2018, p. 369)



If Instruments could talk ...

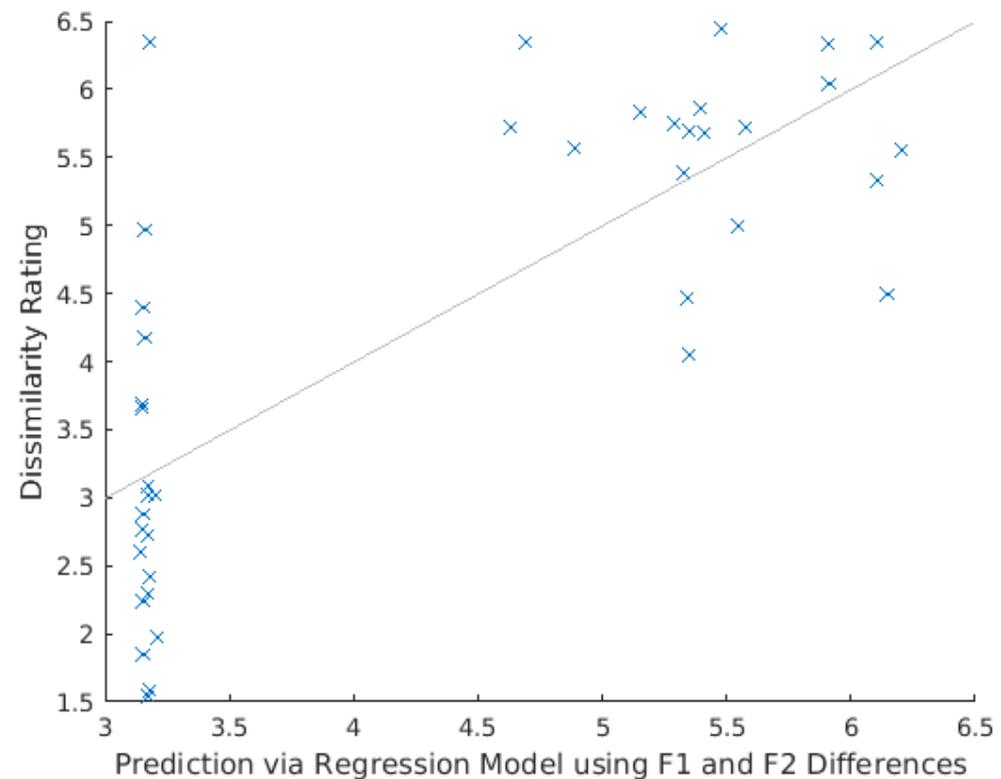
Formants vs. MFCCs

Predicting timbre similarity perception (based on formants)

Regression models trained via machine learning
(5-fold cross-validation)

Best result for timbre similarity prediction based on **formants**:

Prediction model based on **F1** and **F2** ($R^2 = 0.53$)



Prediction model based on F1 and F2
($R^2 = 0.53$, RMSE = 1.08, MSE = 1.16, MAE = 0.86)
(Reuter, Czedik-Eysenberg, Siddiq, Oehler 2018, p. 369)

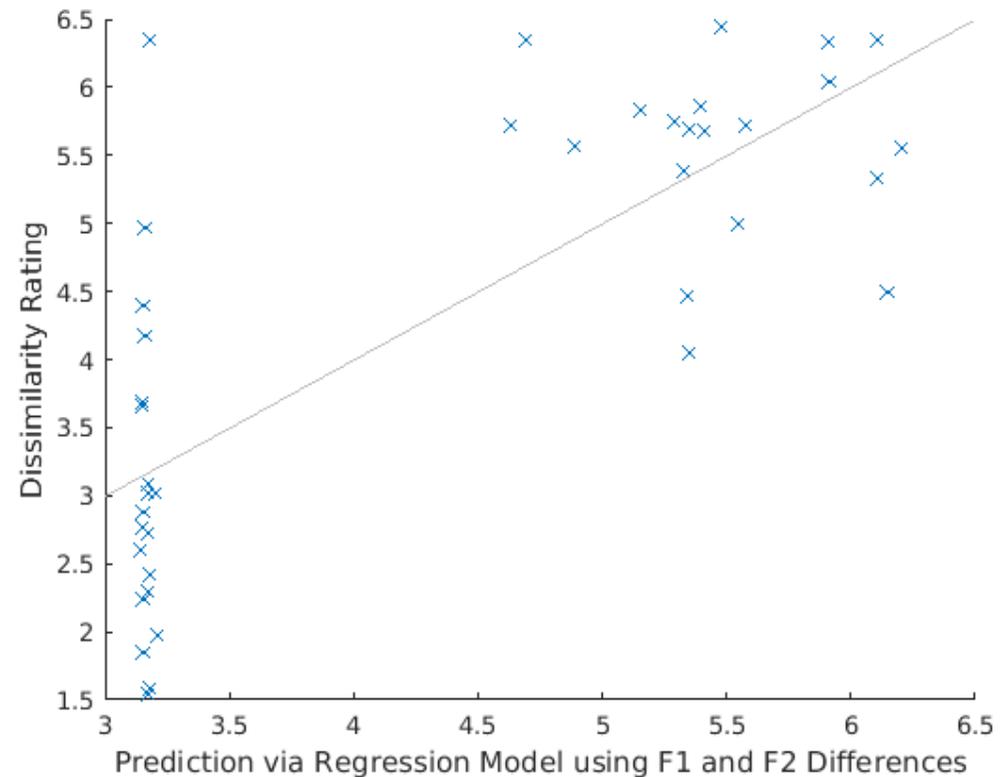


If Instruments could talk ...

Formants vs. MFCCs

Predicting timbre similarity (based on MFCCs)

Regression models trained via
machine learning
(5-fold cross-validation).



Prediction model based on MFCC1 and MFCC2
($R^2 = 0.56$, RMSE = 1.05, MSE = 1.10, MAE = 0.81)
(Reuter, Czedik-Eysenberg, Siddiq, Oehler 2018, p. 370)



If Instruments could talk ...

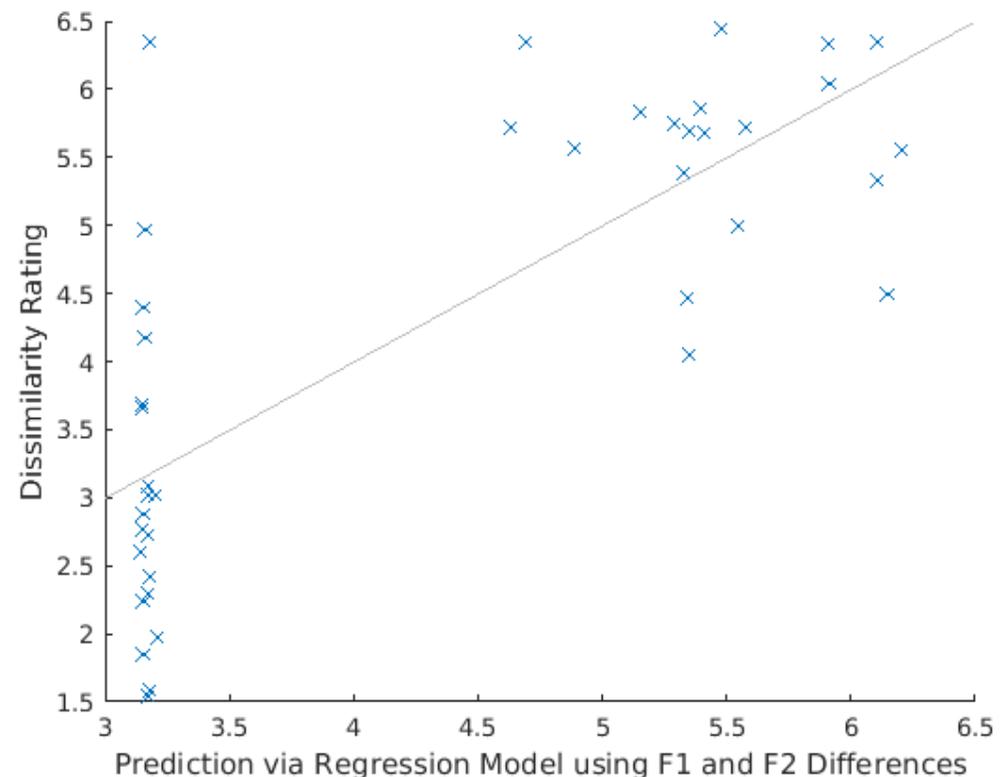
Formants vs. MFCCs

Predicting timbre similarity (based on MFCCs)

Regression models trained via
machine learning
(5-fold cross-validation)

Best result for timbre similarity
prediction based on **MFCCs**:

Prediction model based on
MFCC1 and **MFCC2** ($R^2 = 0.56$)



Prediction model based on MFCC1 and MFCC2
($R^2 = 0.56$, RMSE = 1.05, MSE = 1.10, MAE = 0.81)
(Reuter, Czedik-Eysenberg, Siddiq, Oehler 2018, p. 370)



If Instruments could talk ...

Take Home Message - Conclusion

Formants enable us to **recognize and categorize** musical instrument timbres like vowel sounds.

Formants allow **predictions** about **timbre blending** and **auditory grouping**.

Formant distances enable us to make **timbre similarity calculations** (similar to MFCCs).

Further **advantages** of formants:

- Formants need **only two values** for timbre description
- Formants **compactly** and **intuitively** describe a **distinctive audible spectral content**
- Formants provide a **solid foundation** (> 90 years of research history)

謝謝