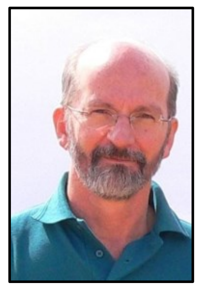
Investigating perception dynamics and uncertainty in temporal sensory data via independent components analysis (ICA)

John C. Castura
Douglas N. Rutledge
Allison K. Baker
Carolyn F. Ross

Compusense

AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Washington State University

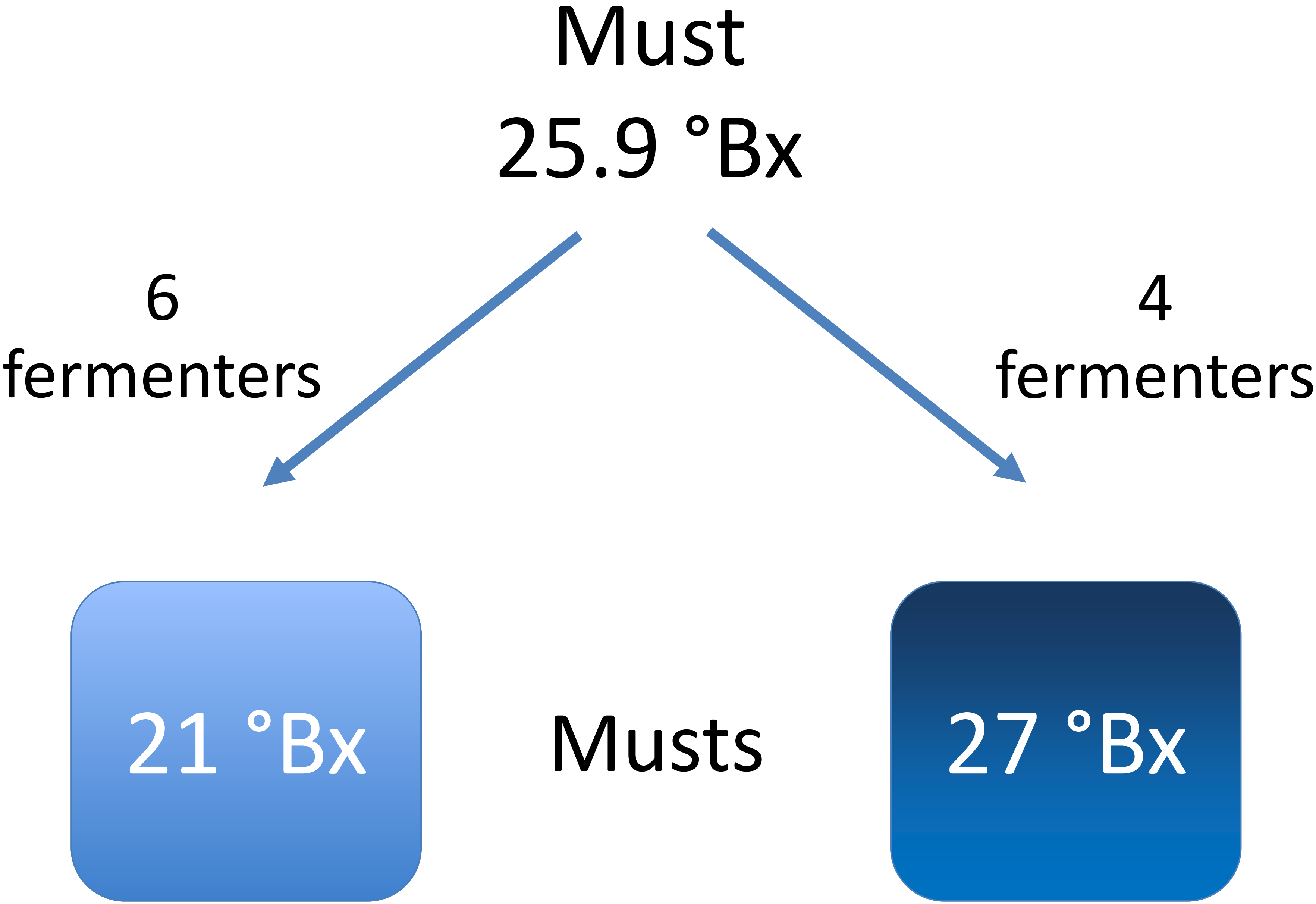
I am happy for you to photograph or tweet the slides from my talk

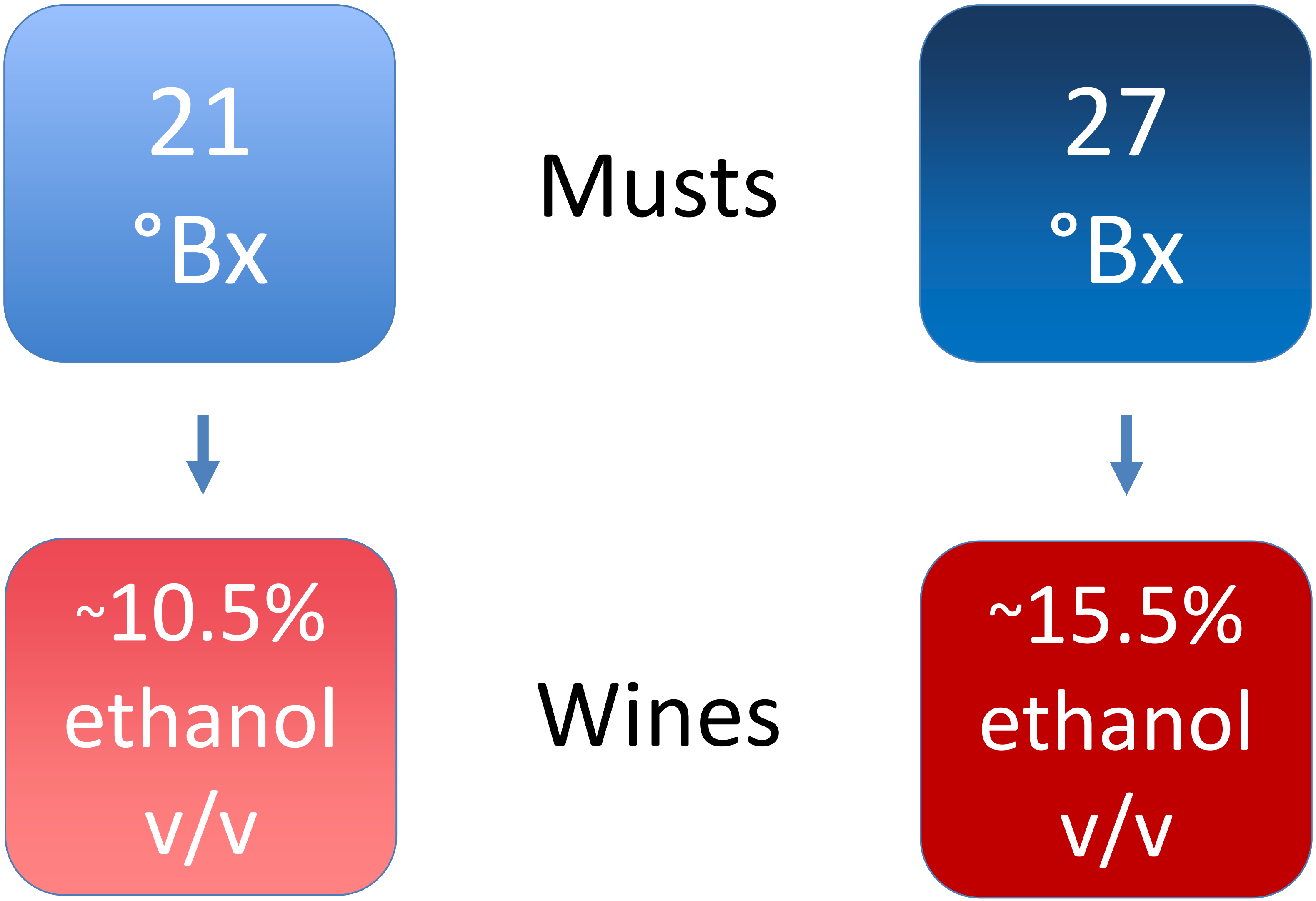
Tweet #Pangborn19

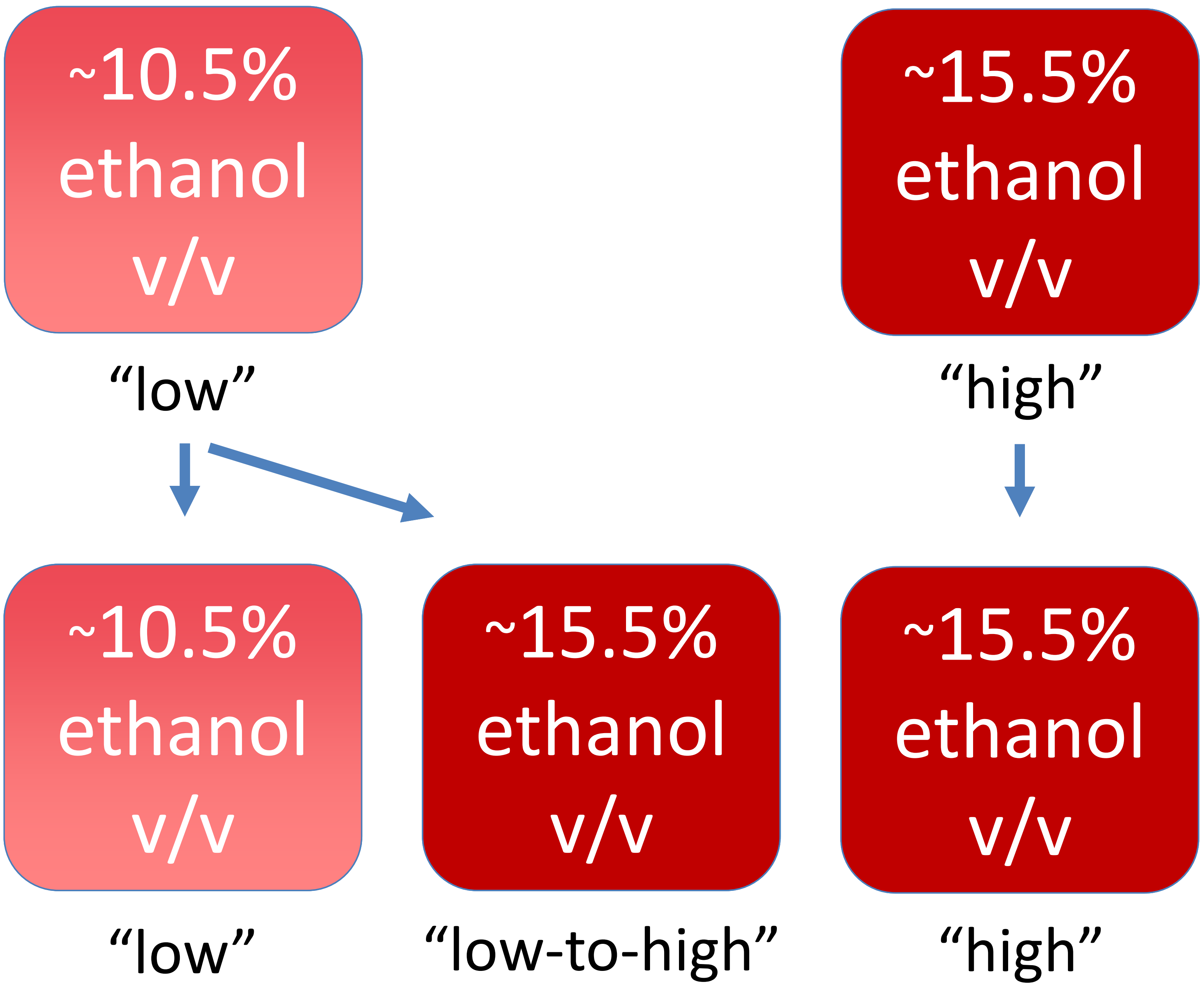
Organised by:


ünchner
Rutschn


Wine


Must
25.9 °Bx
6
fermenters
4
fermenters
21 °Bx
Musts
27 °Bx


21 °Bx
Musts
27 °Bx
~10.5% ethanol v/v
Wines
~15.5% ethanol v/v


~10.5% ethanol v/v
"low"
~15.5% ethanol v/v
"high"
~10.5% ethanol v/v
"low"
~15.5% ethanol v/v
"low-to-high"
~15.5% ethanol v/v
"high"

3 wine treatments

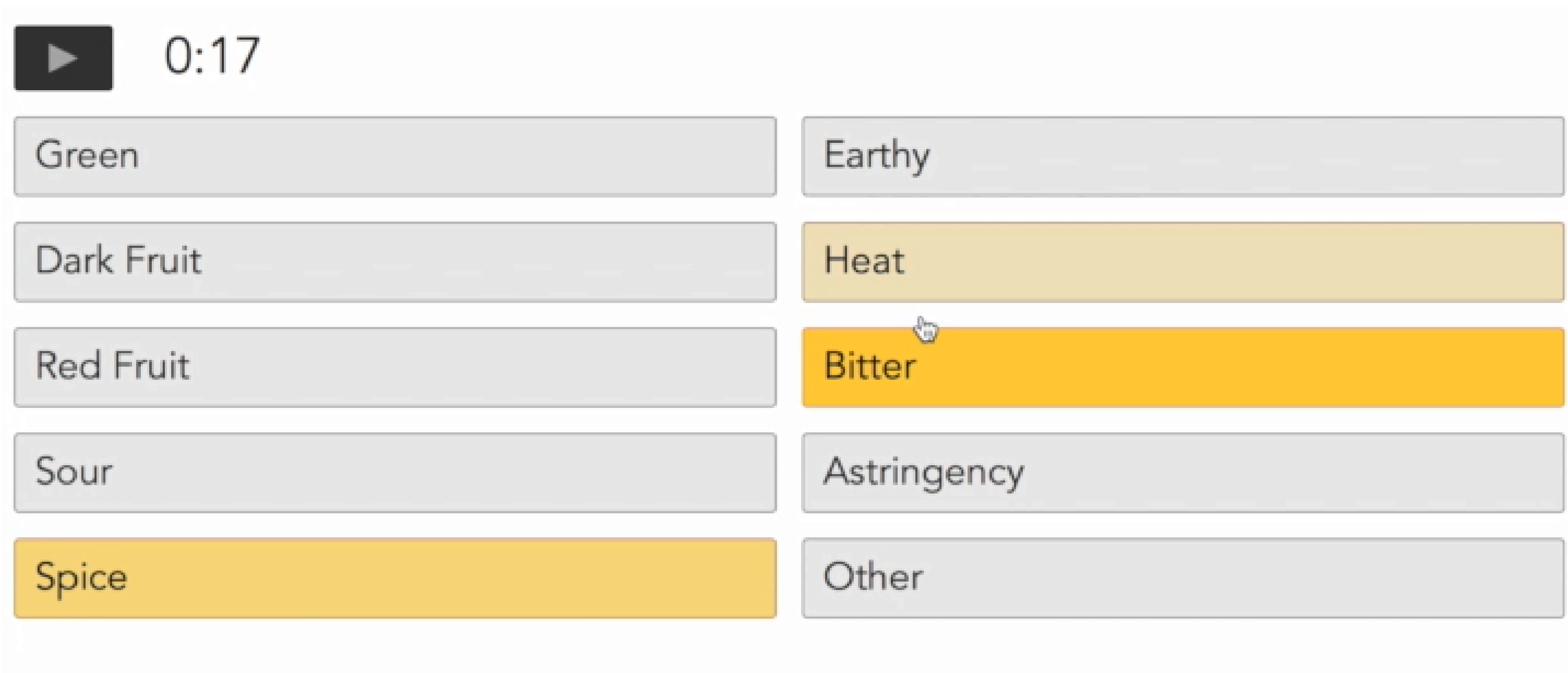
Low	Adjusted	High
~10.5% ethanol v/v	~15.5% ethanol v/v	~15.5% ethanol v/v
“low”	“low-to-high”	“high”

3 wine treatments

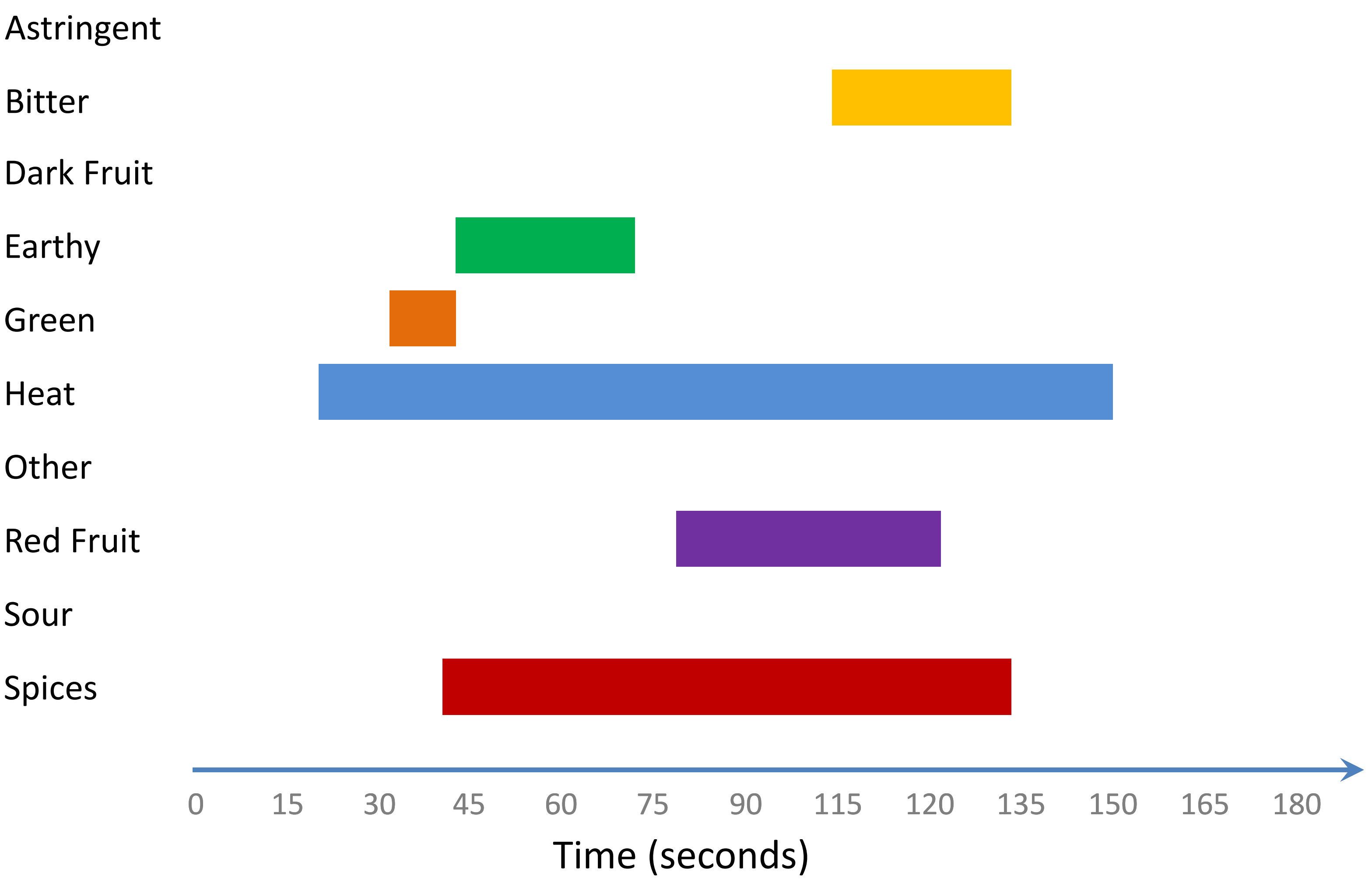
Low	Adjusted	High
~10.5% ethanol v/v	~15.5% ethanol v/v	~15.5% ethanol v/v
"low"	"low-to-high"	"high"

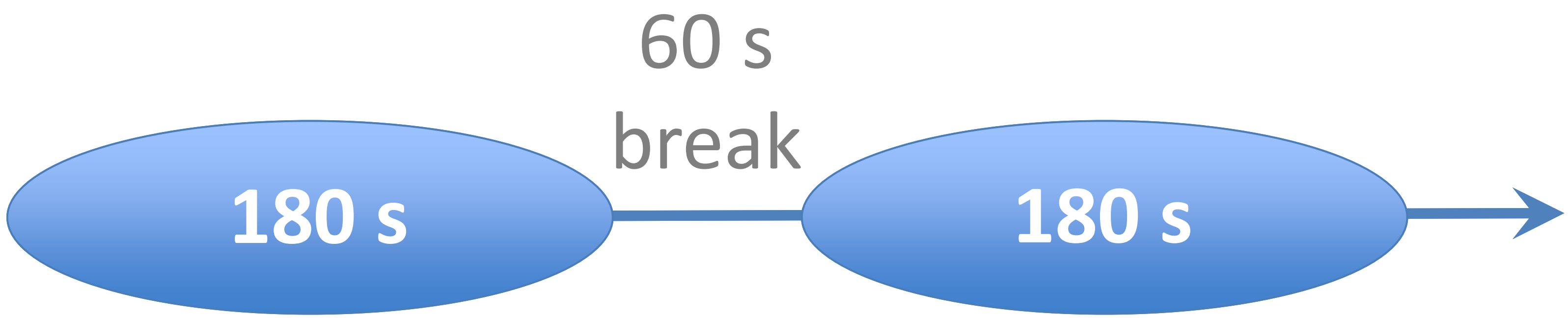
Temporal
Check All That Apply (TCATA)

Temporal Check-All-That-Apply (TCATA)



TCATA raw data





Sip 1
evaluation

Sip 2
evaluation

n = 13
(x4 replicates)

WineSips

High ethanol

Sip 1	H1
Sip 2	H2

Low ethanol

Sip 1	L1
Sip 2	L2

Adjusted (Low-to-High) ethanol

Sip 1	A1
Sip 2	A2

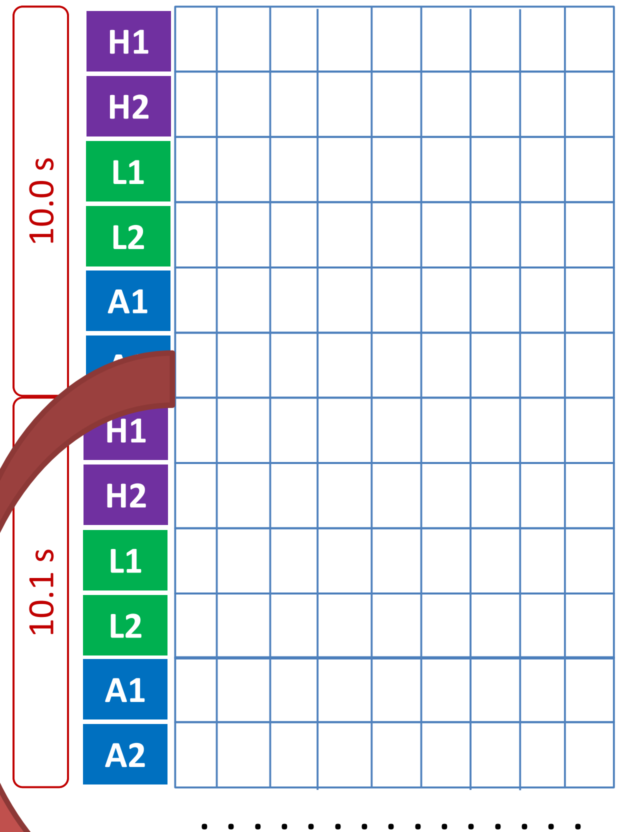
Exploratory data analysis (EDA) enables hypothesis generation and provides insights about experimental data.

Exploratory data analysis (multivariate)

Citation proportions in multi-way array → Matrix

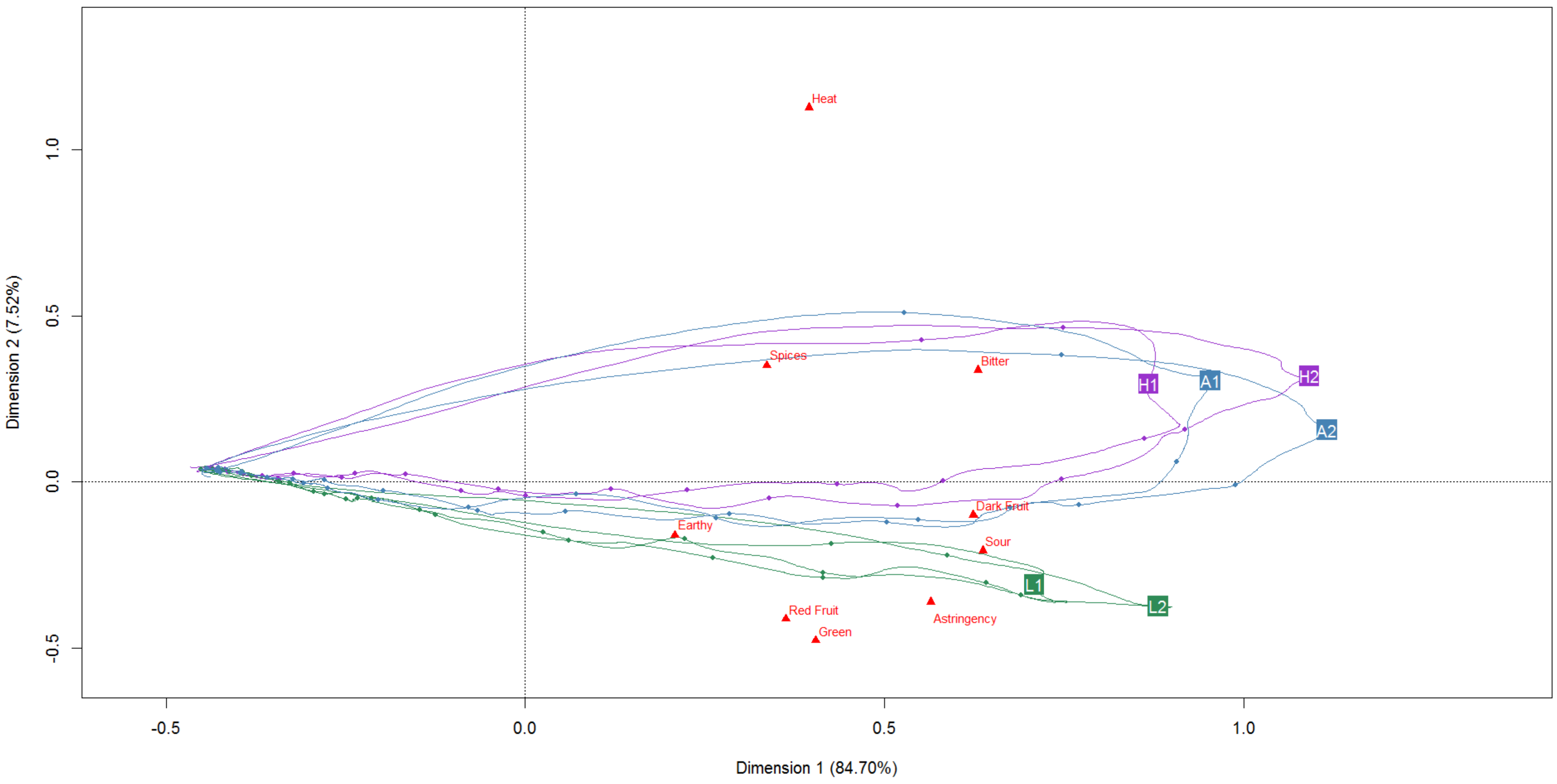
Columns:
Attributes

Rows: **WineSips**
x
Times

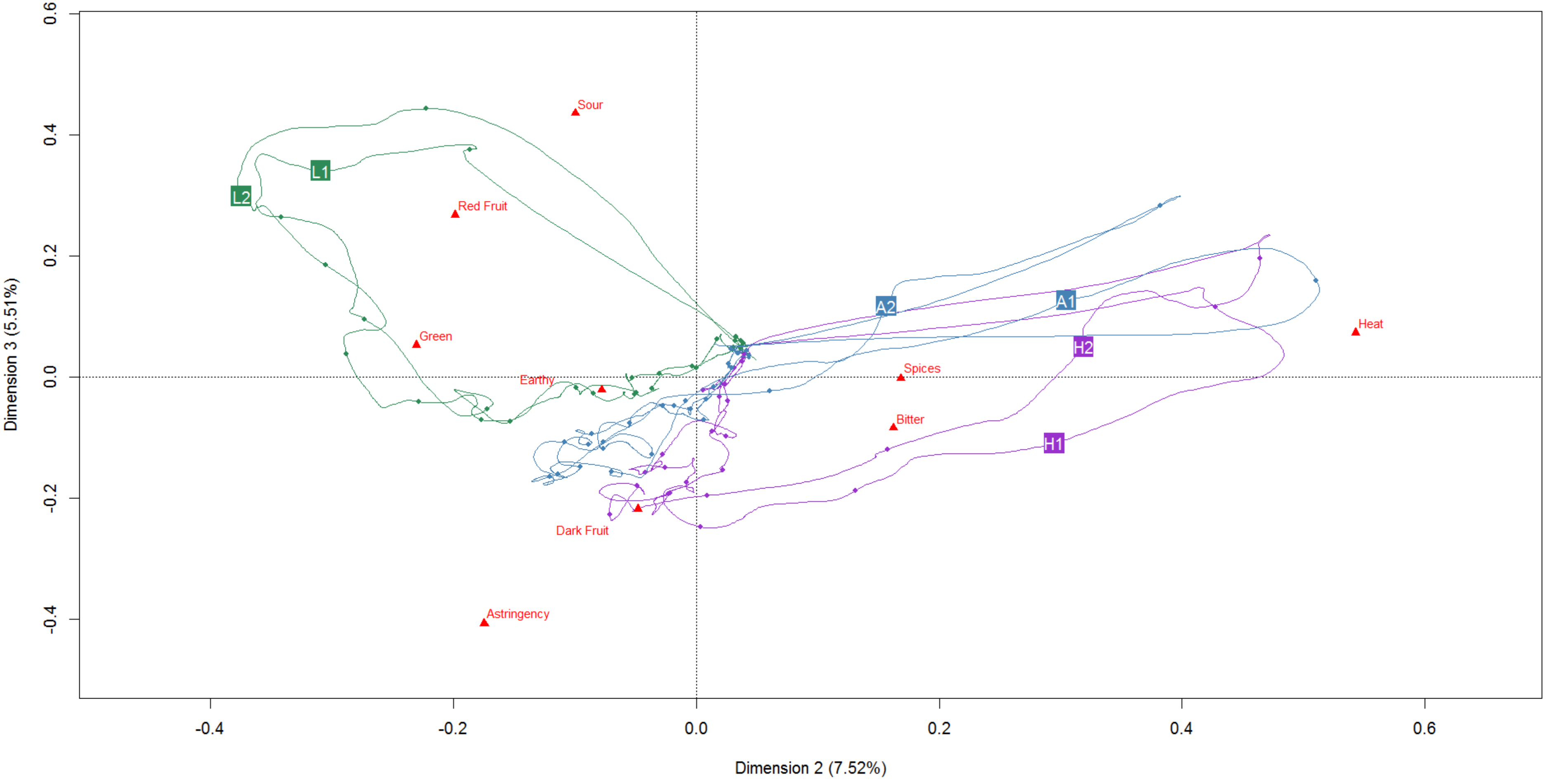


Principal component analysis

Plane of PC1 vs. PC2 for Syrah TCATA data



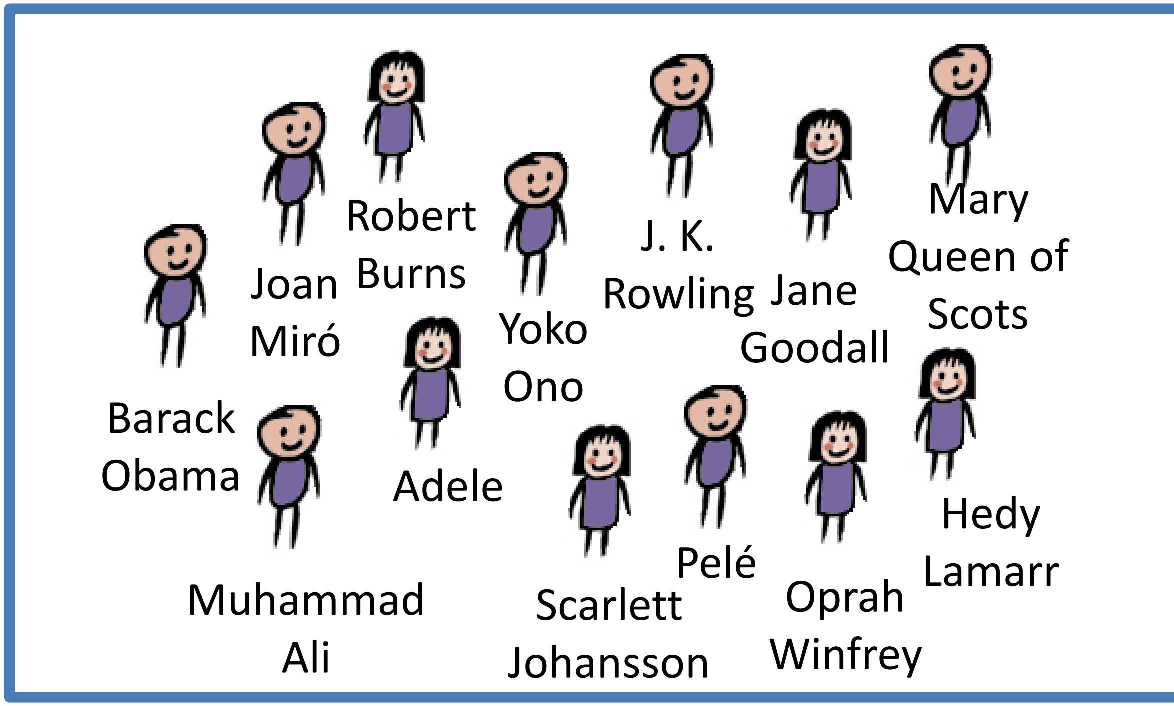
Plane of PC2 vs. PC3 for Syrah TCATA data



How well do these trajectories represent the evolution of flavours in the WineSips?

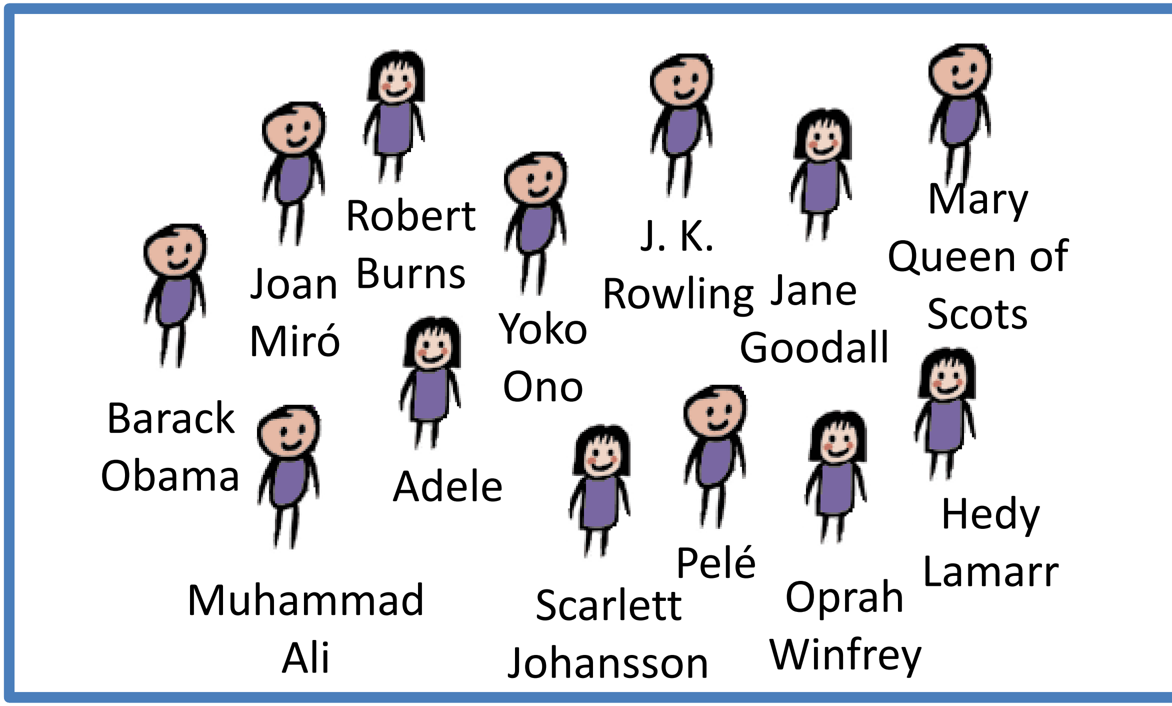
The Real Panel

(n=13)



The Real Panel

(n=13)

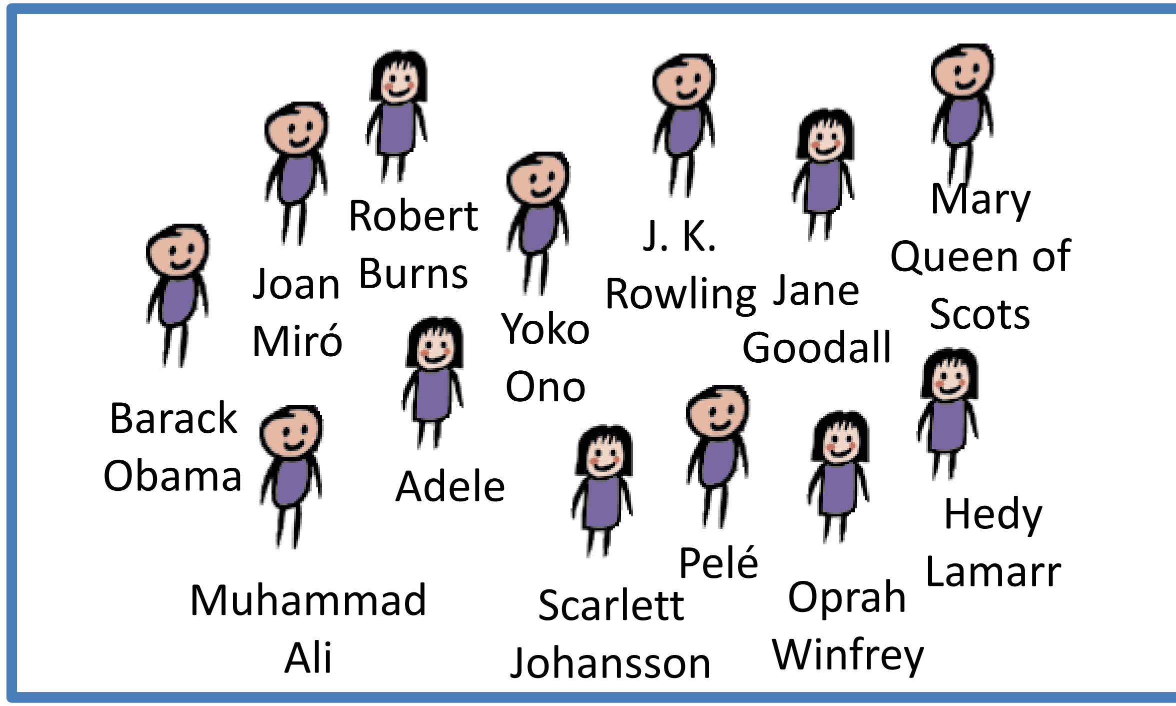


Virtual Panel 1

(n=13)

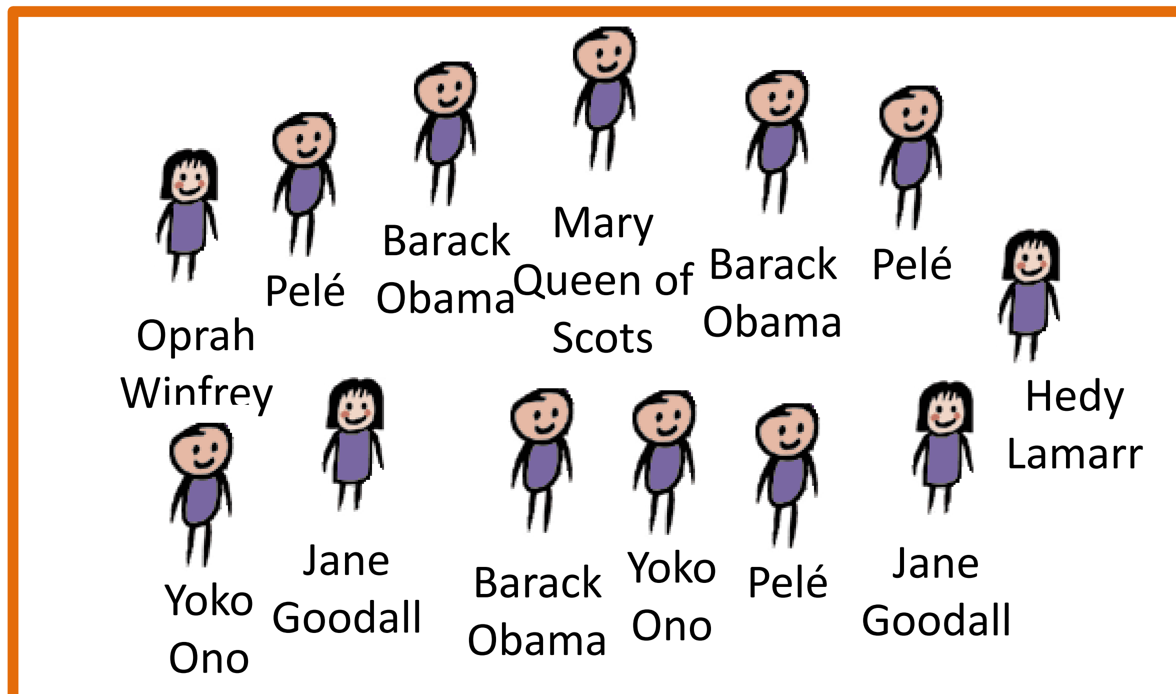
The Real Panel

(n=13)



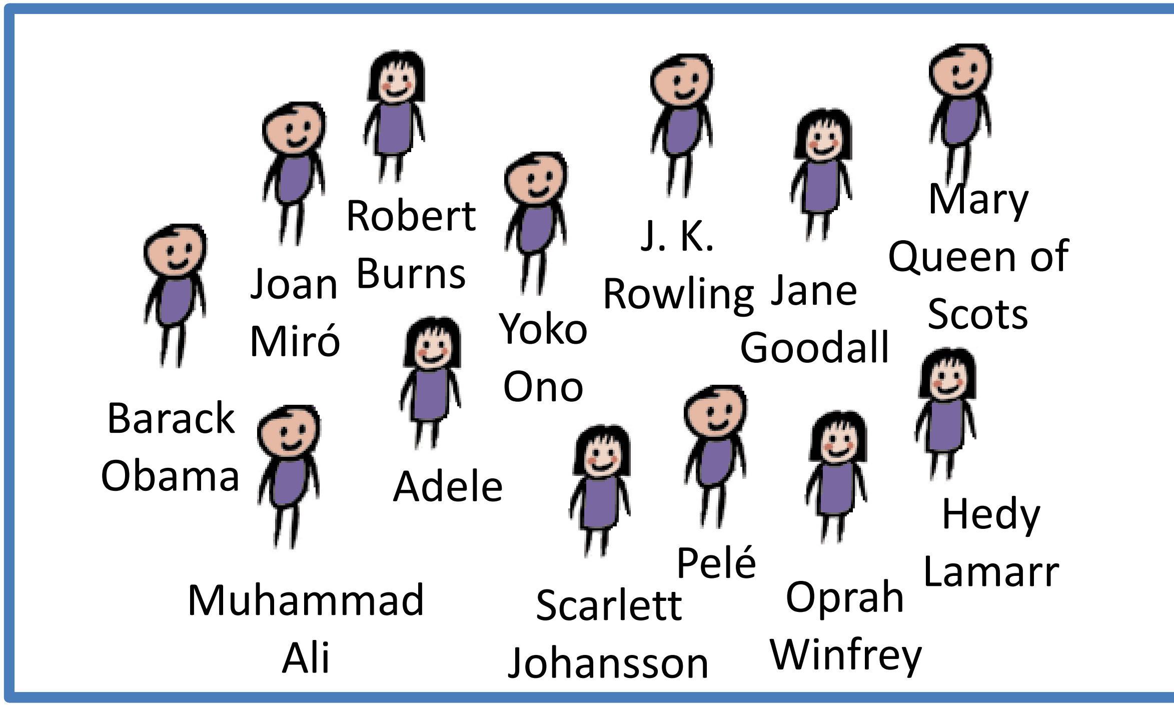
Virtual Panel 1

(n=13)



The Real Panel

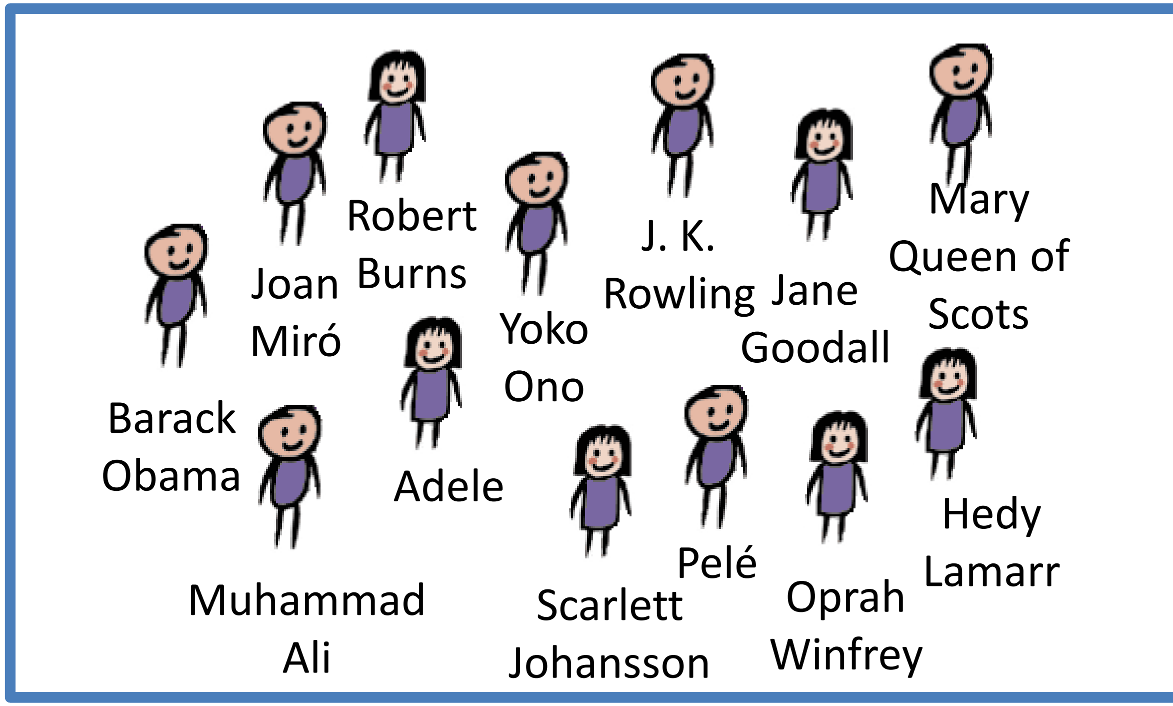
(n=13)



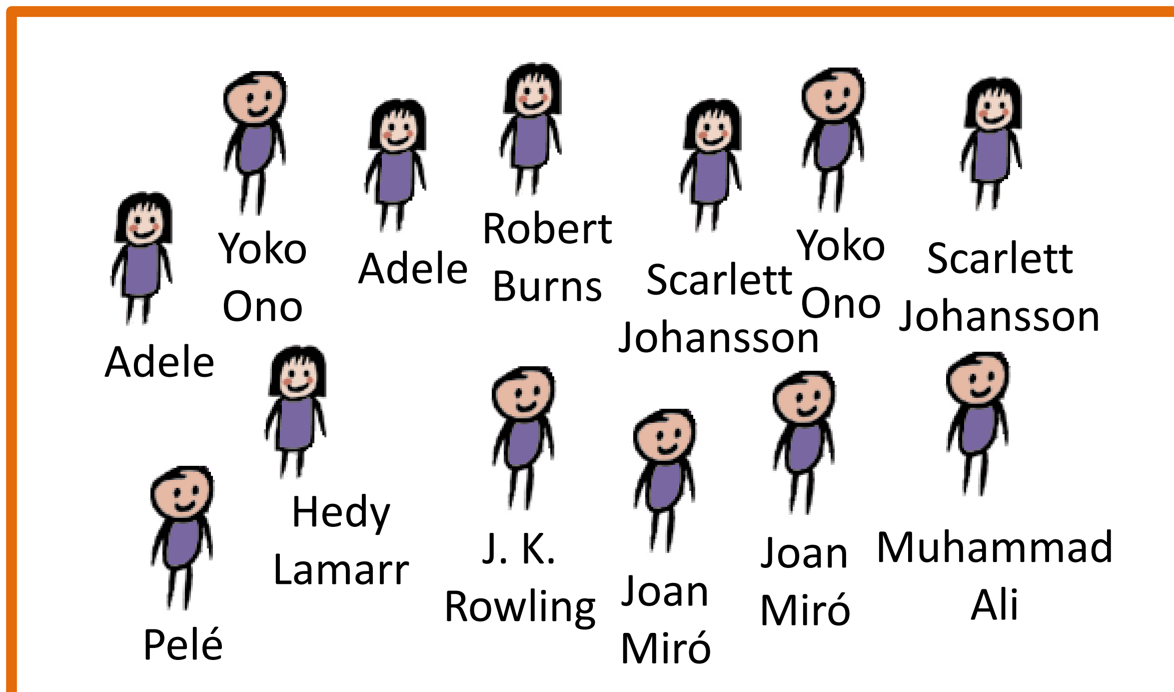
Virtual Panel 2

(n=13)

The Real Panel
(n=13)

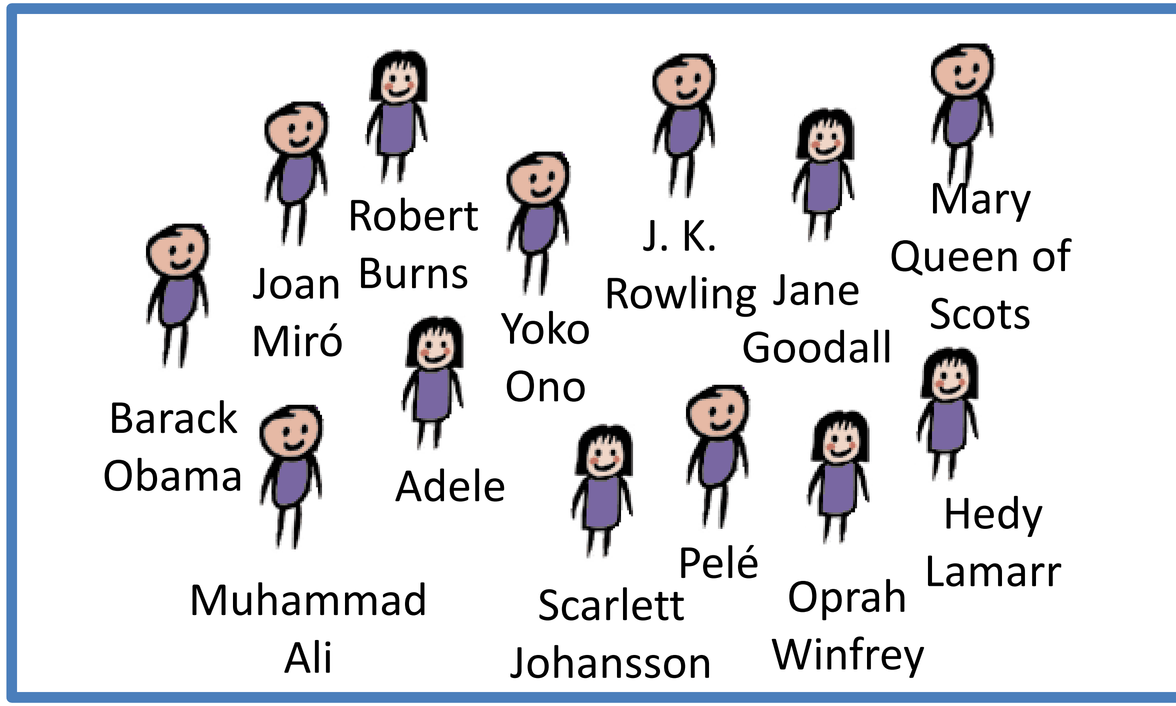


Virtual Panel 2
(n=13)



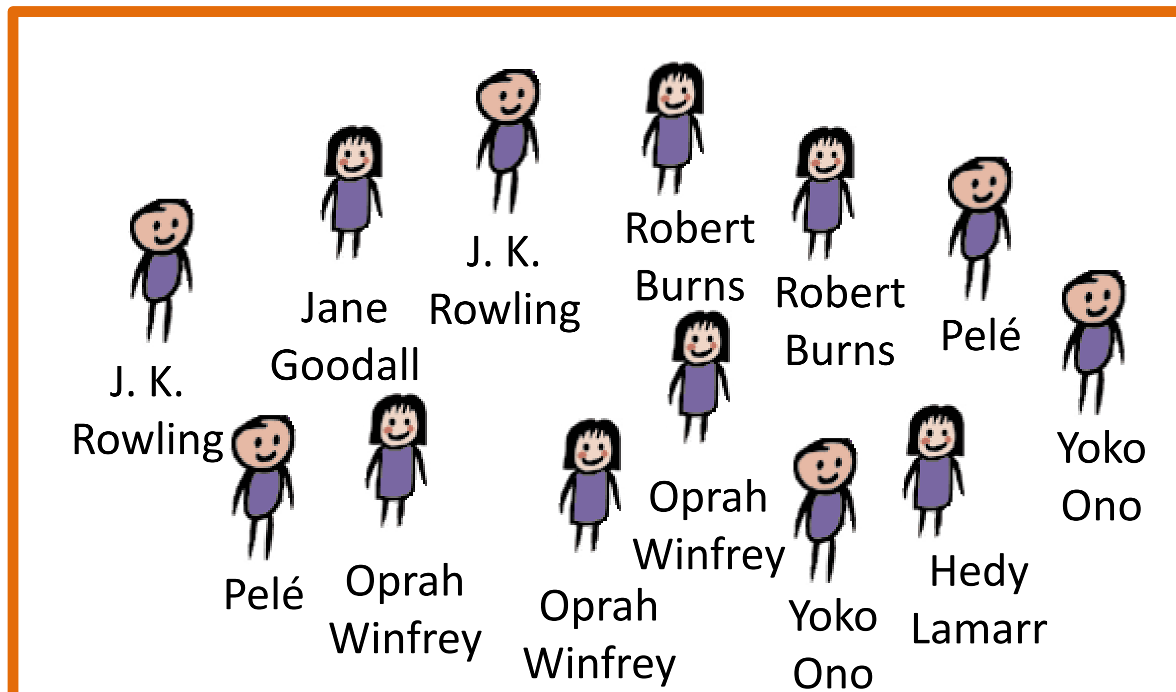
The Real Panel

(n=13)

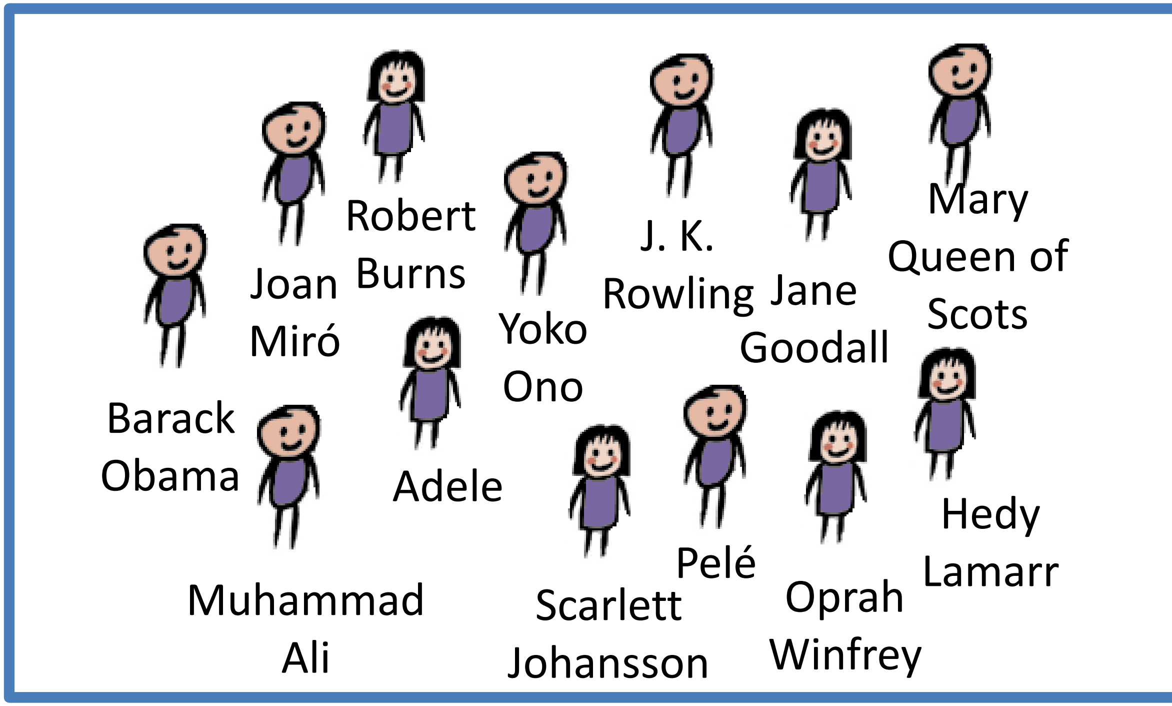


Virtual Panel 3

(n=13)



The Real Panel
(n=13)



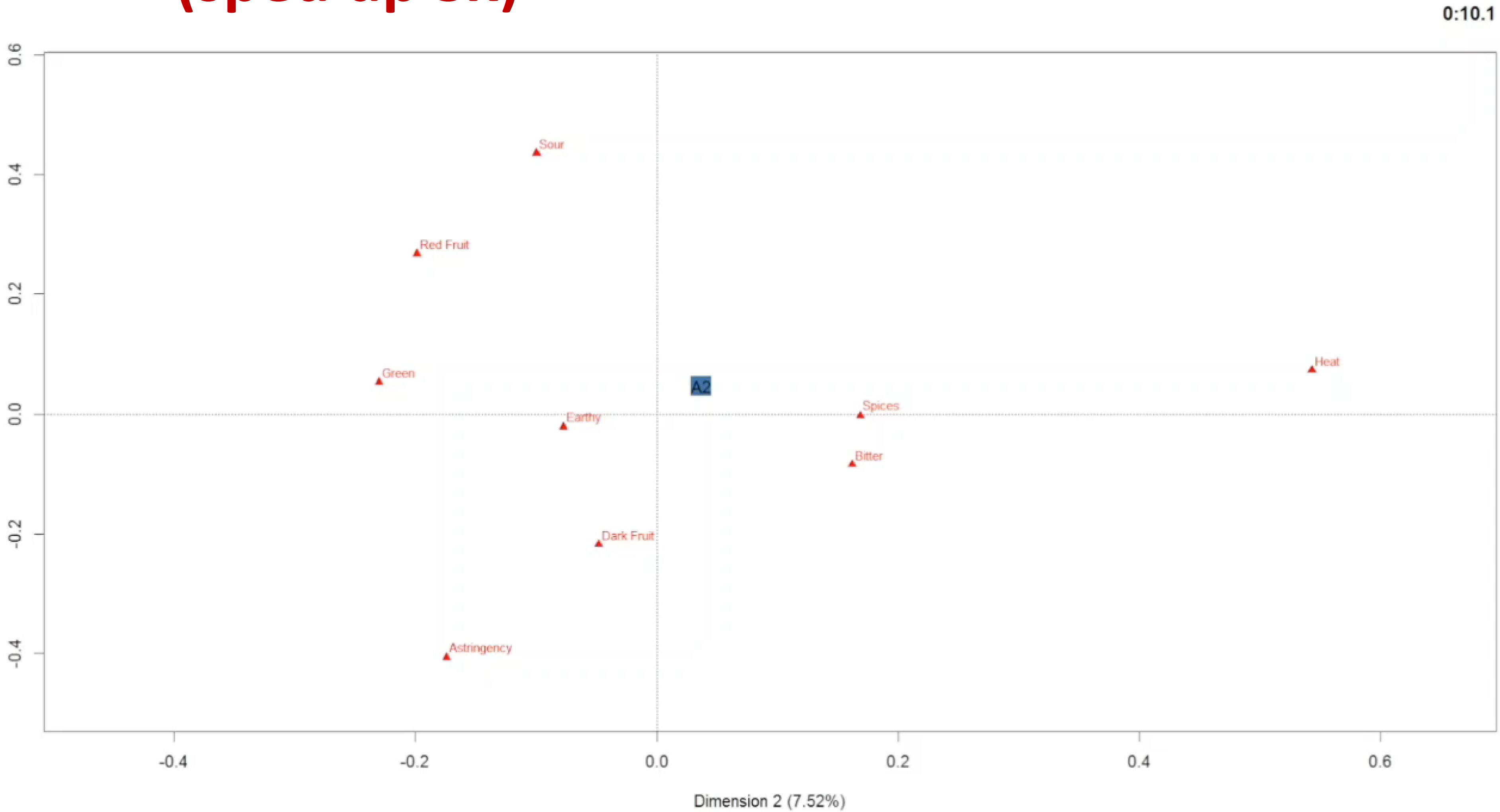
…and so on…

1 real panel (n=13)

+

499 virtual panels (n=13)

Plane of PC2 vs. PC3 for Syrah TCATA data
(sped up 5x)



Can PCA do anything useful with TCATA data?

The answer seems to be...

Yes!

Syrah data

Temporal Check-All-That-Apply Characterization of Syrah Wine

Allison K. Baker, John C. Castura, and Carolyn F. Ross

Abstract: Temporal Check-All-That-Apply (TCATA) is a new dynamic sensory method for which analysis techniques are still being developed and optimized. In this study, TCATA methodology was applied for the evaluation of wine finish by trained panelists ($n = 13$) on Syrah wines with different ethanol concentrations (10.5% v/v and 15.5% v/v). Raw data were time standardized to create a percentage of finish duration, subsequently segmented into thirds (beginning, middle, and end) to capture panel perception. Results indicated the finish of the high ethanol treatments lasted longer (approximately 12 s longer) than the low ethanol treatment ($P \leq 0.05$). Within each finish segment, Cochran's Q was conducted on each attribute and differences were detected amongst treatments ($P \leq 0.05$). Pairwise tests showed the high ethanol treatments were more described by astringency, heat/ethanol burn, bitterness, dark fruit, and spices, whereas the low ethanol treatment was more characterized by sourness, red fruit, and green flavors ($P \leq 0.05$). This study demonstrated techniques for dealing with the data generated by TCATA. Furthermore, this study further characterized the influence of ethanol on wine finish, and by extension wine quality, with implications to winemakers responsible for wine processing decisions involving alcohol management.

Keywords: ethanol, syrah, TCATA, wine finish

Practical Application: Effective temporal descriptive sensory methods are essential in order to accurately characterize attributes that change over time. This paper introduces the Temporal Check-All-That-Apply (TCATA) methodology for performing a dynamic characterization of wine finish, which has a complex evolution of sensations. The methods of standardization, time segmentation, and statistical analyses expand upon the current TCATA methodology and the understanding of wine finish.

Introduction

Sensory perception is a temporal rather than static experience. As such, it is important to have effective techniques for capturing nuances of sensory attributes as they develop and diminish over time. Common methods for capturing dynamic sensory evaluation data include time-intensity (TI; ASTM 2013) and temporal dominance of sensations (TDS; Pineau and others 2003). Each provides valuable information regarding changing sensory properties over time but each has challenges. While TI provides descriptive curves from which parameters can be extracted for analysis, only 1 attribute is evaluated at a time. This limitation can lead to a halo-dumping effect, particularly when evaluating a complex product (Clark and Lawless 1994). TDS allows for multi-attribute evaluation during which a panelist indicates 1 dominant attribute at any given time from a list (Pineau and others 2009, 2012). However, limiting each panelist at each moment to the selection of only the most dominant attribute also limits their ability to characterize nondominant (but nonetheless important) aspects of a product (Ares and others 2015). For example, in a complex product like red wine, astringency and alcohol burn tend to dominate perception, overshadowing smaller differences in flavors that may be present in the finish.

Recently, Temporal Check-All-That-Apply (TCATA) has been introduced (Castura and others 2016). It extends Check-All-That-Apply (CATA) methodology for temporal data collection. It has been used with trained and semi-trained panelists who evaluate a sample by continually checking and unchecking the attributes from a prepopulated list. Unlike in TDS, TCATA allows concurrent term selection. Specifically, instead of indicating 1 attribute that is dominant at any given time, panelists can indicate all attributes perceived at any given time. This method provides binary data with no direct indication of attribute intensities; however, previous studies have reported correlation between attribute intensity and CATA frequency in the evaluation of milk desserts and beer (Bruzzone and others 2012; Reinbach and others 2014).

We sought to determine whether TCATA provides an effective temporal sensory evaluation of a complex product and complex perception: wine finish. Wine finish incorporates several sensory modalities, consisting of the flavors, tastes, and mouthfeels that remain after a wine sample is swallowed or expectorated (Jackson 2002). Related words describing wine finish or a component thereof include aftertaste, persistence, and length (or long). Economic studies on premium wines suggest that a long and pleasant finish is indicative of a high quality wine (Lecocq and Visser 2006; Benfratello and others 2009).

Secondarily, this study was a designed experiment with wines systematically manipulated to allow for conclusions regarding a causal relationship between ethanol and wine finish. Numerous previous studies have described the influence of wine ethanol concentration on aroma and flavor. Increases in ethanol concentration (ranging from 0% v/v to 18% v/v) is known to decrease aroma intensity and volatile recovery (Le Berre and others

MS 20160069 Submitted 1/13/2016, Accepted 4/8/2016. Authors Baker and Ross are with School of Food Science, Washington State Univ., PO Box 646376, Pullman, Wash. 99164, U.S.A. Author Castura is with Compusense Inc., 255 Speedvale Ave. W., Guelph, Ontario, N1H 1C5, Canada. Direct inquiries to author Ross (E-mail: cfross@wsu.edu).

doi: 10.1111/1750-3841.13328

Food Quality and Preference 54 (2016) 90–100

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

ELSEVIER

Using contrails and animated sequences to visualize uncertainty in dynamic sensory profiles obtained from temporal check-all-that-apply (TCATA) data

CrossMark

John C. Castura [a,*], Allison K. Baker [b], Carolyn F. Ross [b]

[a] Compusense Inc., 255 Speedvale Ave. W., Guelph, Ontario N1H 1C5, Canada
[b] School of Food Science, Washington State University, Pullman, WA 99164-6376, USA

A R T I C L E I N F O

Article history:
Received 7 March 2016
Received in revised form 24 June 2016
Accepted 27 June 2016
Available online 29 June 2016

Keywords:
Temporal check-all-that-apply
TCATA
Contrails
Animated sequences
Bootstrap scores
Data concentration ellipses
Virtual panels

A B S T R A C T

Approaches for analyzing temporal check-all-that-apply (TCATA) data are further developed and illustrated using data arising from a Syrah wine finish evaluation. Raw and smoothed trajectories are obtained using principal component analysis. Virtual panels are obtained from a partial bootstrap, and the attribute citation proportions are then projected into the solution space to form contrails. Trajectories are overlaid on the contrails, allowing smoothing to be evaluated. Separation between two contrails provides evidence that the trajectories differ. At individual time slices, data concentration ellipses are overlaid on bootstrap scores. Separation of ellipses provides evidence of differences among treatments. Difference trajectories and difference ellipses can also be plotted; if the difference ellipse excludes the origin it indicates a difference between the treatments. Animated sequences summarize changes in product characterization over time in a manner that facilitates review. A glossary of terms introduced in the paper is provided in an appendix.

1. Introduction

Temporal check-all-that-apply[1] (TCATA; Castura, Antúnez, Giménez, & Ares, 2016) is a temporal sensory method, and its data are used to describe the **temporal evolution** of sensory attributes in the products under evaluation. Assessors, who could be either trained or untrained, are tasked with checking and unchecking attributes from a list during the evaluation period, such that the attributes that are selected at any given time describe the product at that time. TCATA methodology has been used to investigate temporal sensory evolution in a range of product categories: food, such as yogurt (Castura et al., 2016), salami, cheese, French bread, and marinated mussels (Ares et al., 2015); beverages, such as chocolate-flavoured milk (Oliveira et al., 2015) and red wine (Baker, Castura, & Ross, 2016); and non-food, such as cosmetic products (Boinbaser, Parente, Castura, & Ares, 2015).

Familiar heuristic approaches for visualizing data from temporal dominance of sensations (TDS; Pineau et al., 2009) studies have been leveraged to show **TCATA curves** as smoothed **attribute citation proportions** over time. For each TCATA attribute, the **citation rate** of a product of interest can be contrasted with the average citation rate of the other products (Castura et al., 2016, Figs. 3 and 4). The data visualization described above can be considered to be a generalization of **difference curves**, which contrasts attribute citation rates for one product against another product over time, thus emphasizing statistically significant differences (Castura et al., 2016, Fig. 5).

Univariate TCATA curves can be onerous to review when there are many products and many attributes. For this reason it can be useful to consider TCATA data from a multivariate perspective, which provides both data reduction and interpretation advantages (Johnson & Wichern, 2007, chap. 8). Castura et al. (2016, Figs. 6 and 7) submit a contingency table of TCATA citation frequencies (rows: Product * Time; columns: Attributes) to correspondence analysis (CA) and join adjacent **time slices** to create a separate curve per product. Each curve can then be smoothed so as to show trends without overfitting the data. Each curve is called a **trajectory**, following terminology for multivariate changes in TDS dominance rates in Lenfant, Loret, Pineau, Hartmann, and Martin (2009). A sense of **temporal progression** is given by placing markers along the trajectories at specified time intervals (e.g. every 5 s).

A check-all-that-apply (CATA) contingency table can also be analyzed via principal component analysis (PCA) on the covariance matrix (Meyners, Castura, & Carr, 2013). It is possible to conduct

* Corresponding author.
E-mail address: jcastura@compusense.com (J.C. Castura).

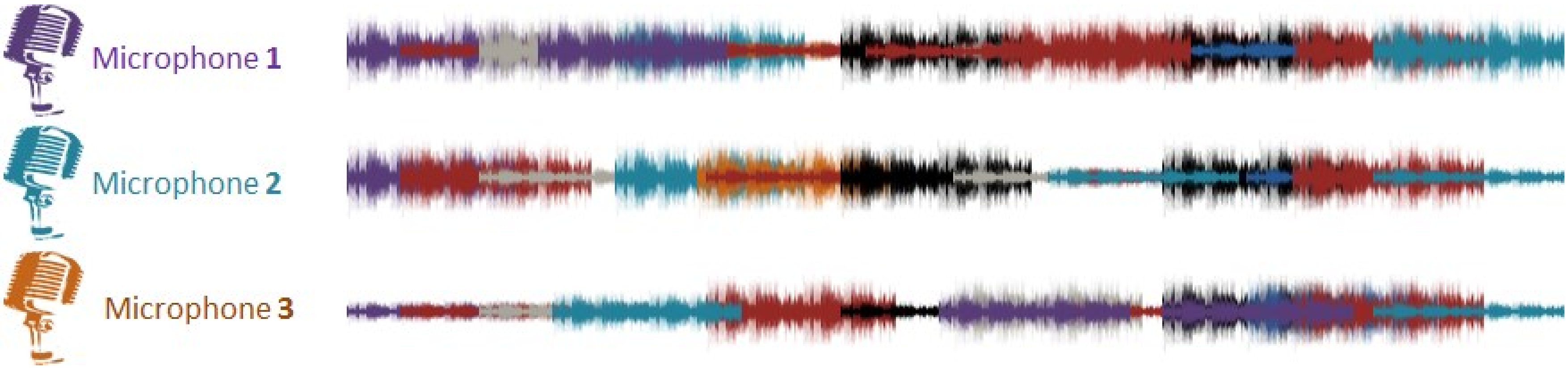
[1] The first use of each glossary term (defined in Appendix A) is shown in bold.

http://dx.doi.org/10.1016/j.foodqual.2016.06.011
0950-3293/© 2016 The Authors. Published by Elsevier Ltd.

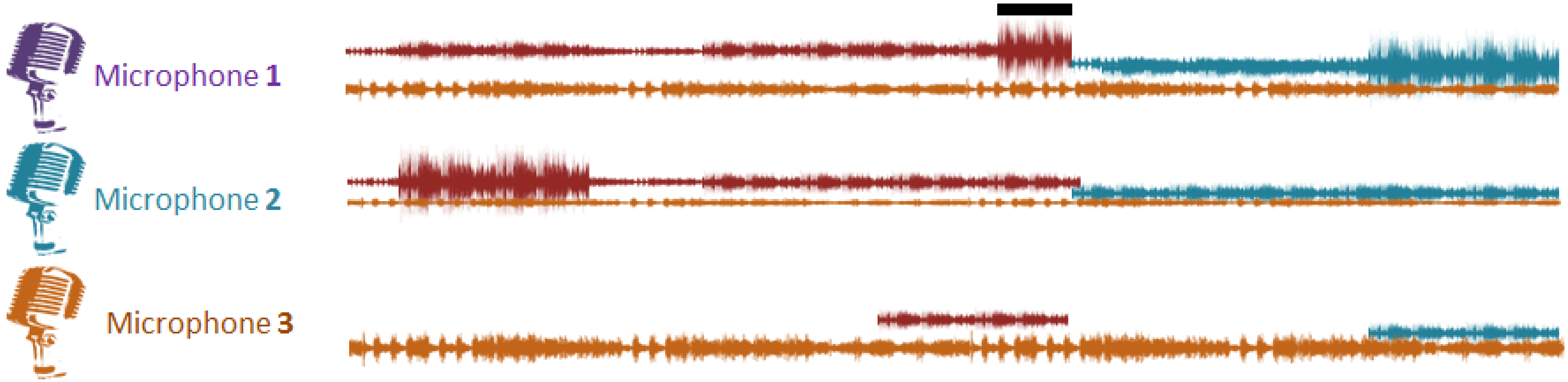
Independent Components Analysis (ICA) with the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm

What is ICA?


Microphone
1
Aaaa-oooooo
Microphone
2
Microphone
3



Blind Source Separation



Blind Source Separation

I’m Jim! Last year I traversed the Panama Canal!

That is so interesting, Jim. I think I see someone I know…

Differences between PCA and ICA

PCA

- Recoding of X to Y
- PCs are each linear combination of attributes
- PC1 is the direction of greatest dispersion of the samples
- Variance explained orders PCs by ‘importance’

ICA

- Signal separation
- No order of importance for ICs
- Need to estimate number of sources (via KMO measures)
- Assumes mixing process is linear

$$\mathbf{X} = \mathbf{A} \times \mathbf{S}$$

X **are data observed**

Cocktail party:

mixed sounds recorded by the array of microphones

$$X = A \times S$$

S are the source signals

Cocktail party:

sound sources (uncontaminated)

$$X = A \times S$$

A is a mixing matrix

Cocktail party:

weighting coefficients that define mixing

$$\mathbf{X} = \mathbf{A} \times \mathbf{S}$$

$\mathbf{X}$ (data) are more Gaussian than $\mathbf{S}$ (sources)

$\mathbf{S}$ (sources) are independent of one another

Strategy: solve $\mathbf{S} = \mathbf{W} \times \mathbf{X}$,

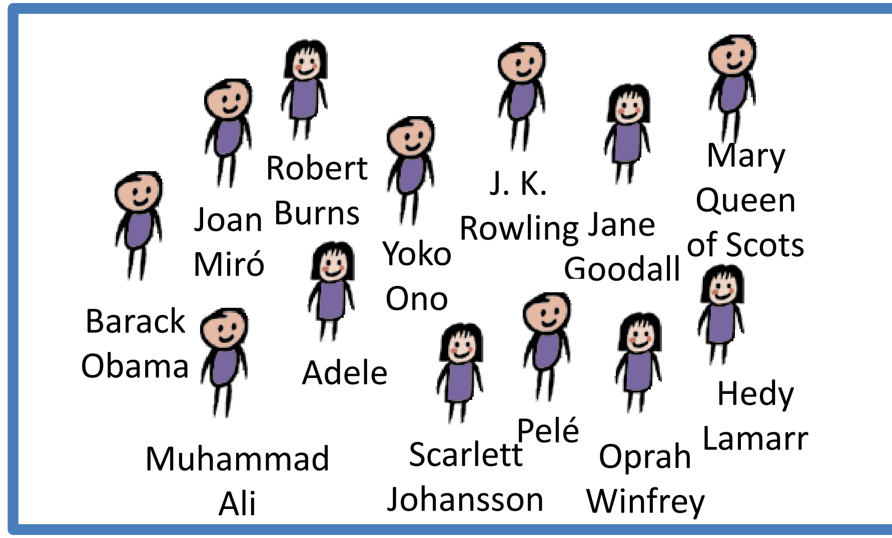
where $\mathbf{W}$ is the pseudoinverse of $\mathbf{A}$,

and the objective is to maximize non-Gaussianity.

The Real Panel

(n=13)

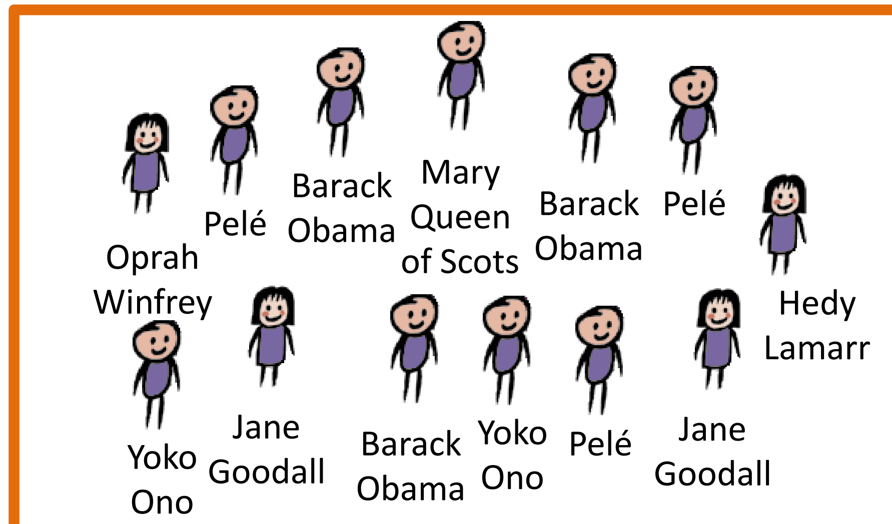
$$X=A\times S$$



Virtual Panel 1

(n=13)

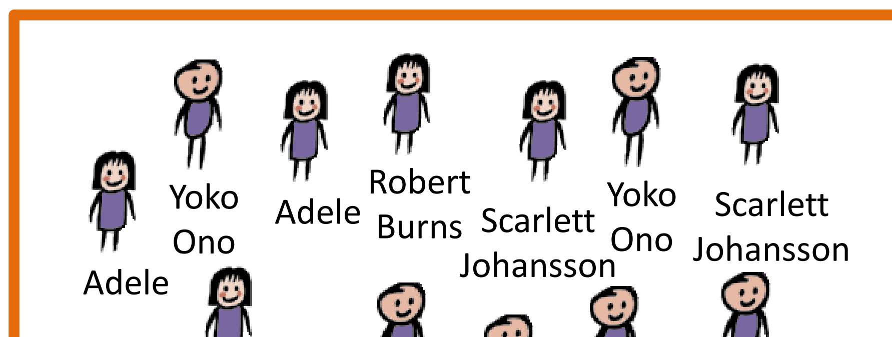
$$X\times(S^{T}\times(S^{T}\times S)^{-1})=A$$



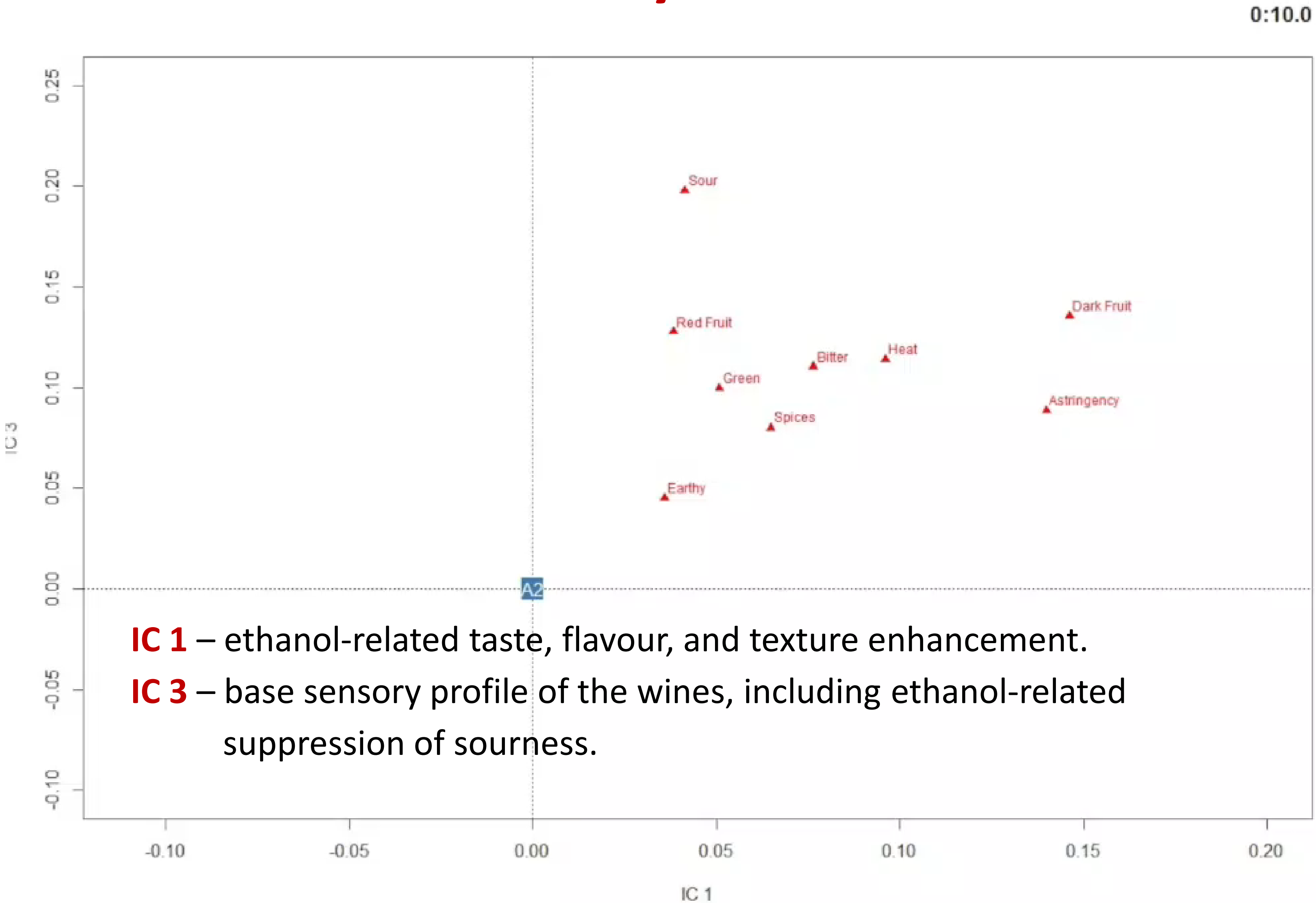
Virtual Panel 2

(n=13)

$$X\times(S^{T}\times(S^{T}\times S)^{-1})=A$$



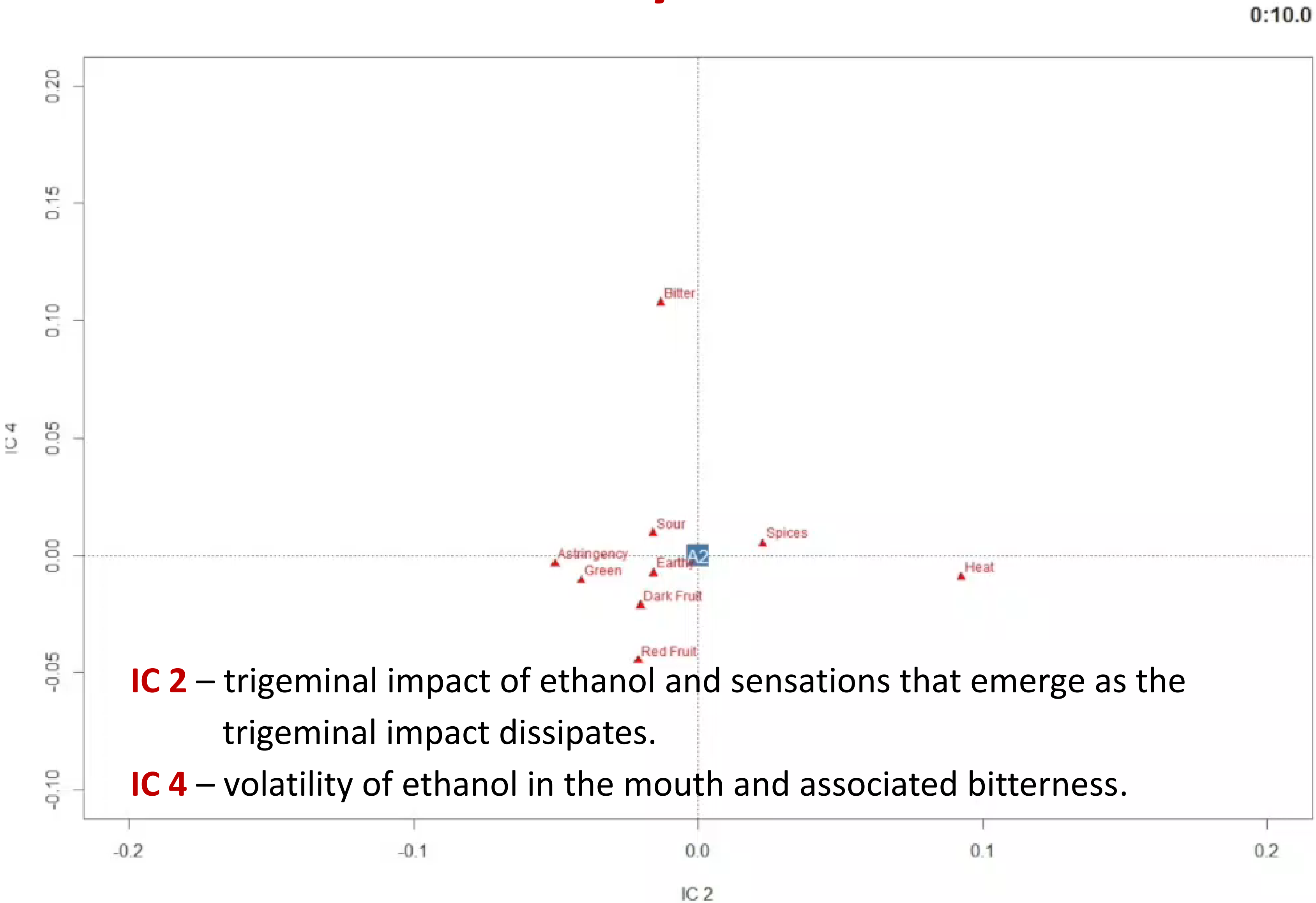
Plane of IC1 vs. IC3 for Syrah TCATA data



IC 1 – ethanol-related taste, flavour, and texture enhancement.

IC 3 – base sensory profile of the wines, including ethanol-related suppression of sourness.

Plane of IC2 vs. IC4 for Syrah TCATA data



IC 2 – trigeminal impact of ethanol and sensations that emerge as the trigeminal impact dissipates.

IC 4 – volatility of ethanol in the mouth and associated bitterness.

Can ICA do anything useful with TCATA data?

The answer seems to be…

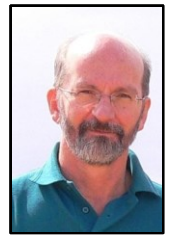
Yes!

Thank you for your attention!

John C. Castura

Compusense

Douglas N. Rutledge



AgroParisTech
INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT
PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES

Allison K. Baker

Washington State University

Carolyn F. Ross