Wintec’s **Centre for Business, Information Technology and Enterprise** proudly present:

Research in Action

Emerging Technologies and Trends in IT

Programme

12pm

Bachelor student in industry poster session

Pepper creative VR demo available to try

12.30pm

CBITE IT Research

Cacophony Project - turning birdsong into data and using the best of breed IT technologies to dramatically improve our trapping ability.

SUPA: Strewn User-Preserved Authentication - a distributed authentication system for secure computing.

Computational modelling to track human emotion trajectories through time.

Software evolution using user reviews.

Process Mining - a special type of data mining to improve process performance.

Recommender Systems in E-Commerce.

1.20pm

Industry Research

AWARE group - Facial recognition for attendance.

Pepper Creative - VR and AR solutions to daily challenges.

1.50pm

Guss Wilkinson - Top Poster student award.

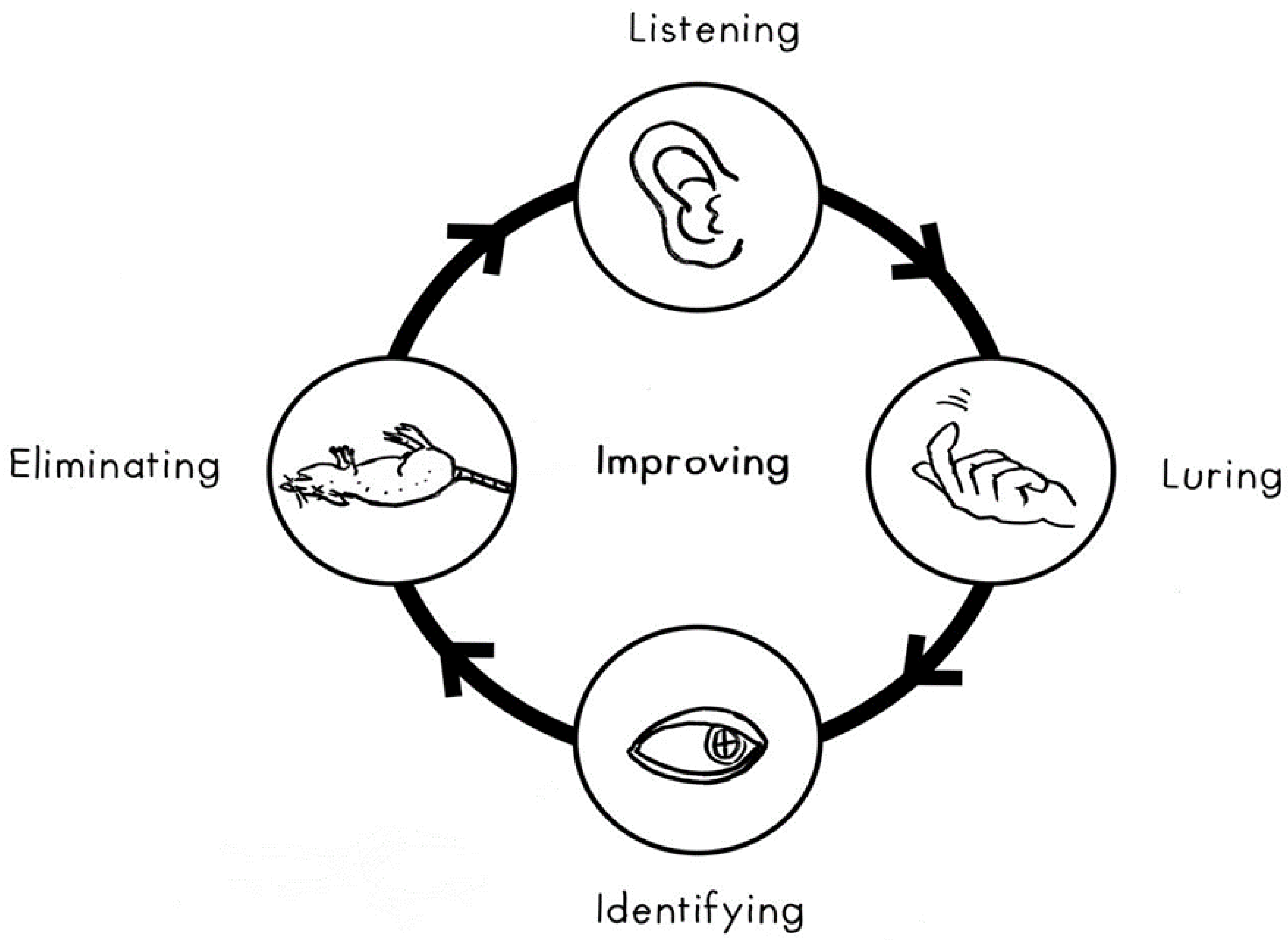
Research @ CBITE IT

- **Cacophony Project** - turning birdsong into data and using the best of breed IT technologies to dramatically improve our trapping ability.
- **SUPA: Strewn User-Preserved Authentication** - a distributed authentication system for secure computing.
- **Computational modelling to track human emotion trajectories through time.**
- **Software evolution using user reviews.**
- **Process Mining** - a special type of data mining to improve process performance.
- **Recommender Systems in E-Commerce.**

The Cacophony Project

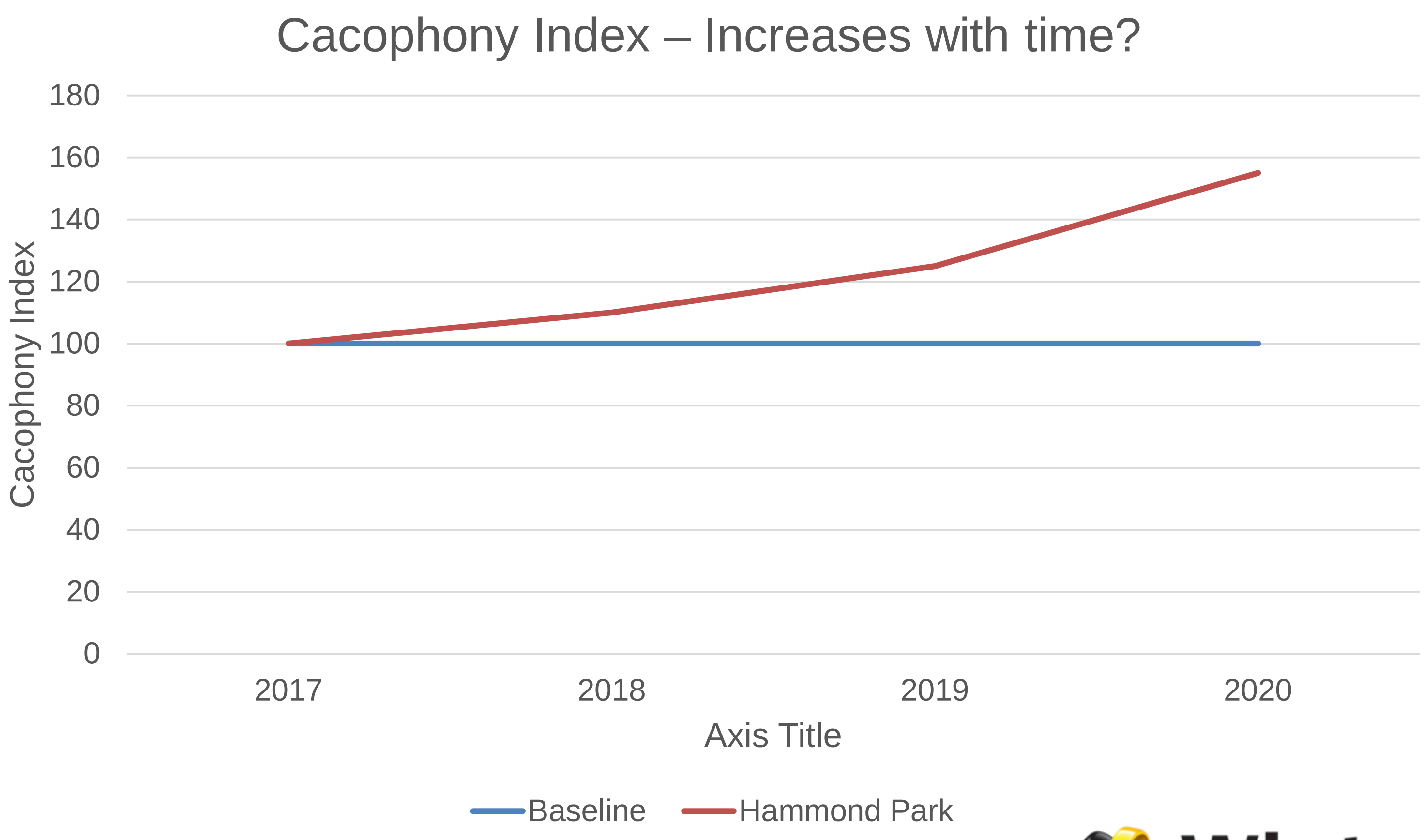
https://cacophony.org.nz/

"The Cacophony Project will turn birdsong into data and use the best of breed IT technologies to dramatically improve our trapping ability."


Listening
Luring
Identifying
Eliminating
Improving

Why listen?

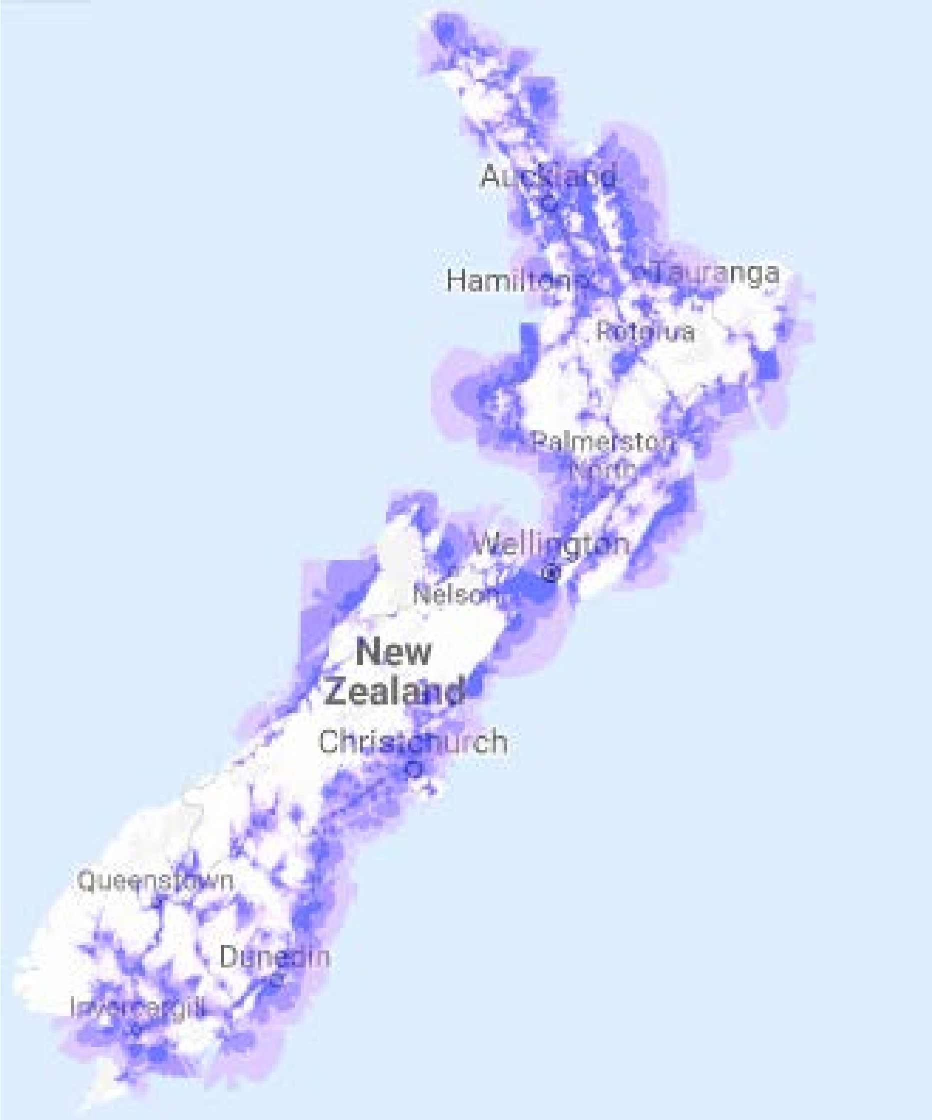
- Obtain a baseline ‘cacophony index’
  - Use any change from baseline to help determine if an intervention has had an effect?



The Cacophonometer


telecom

Spark mobile coverage (3G)



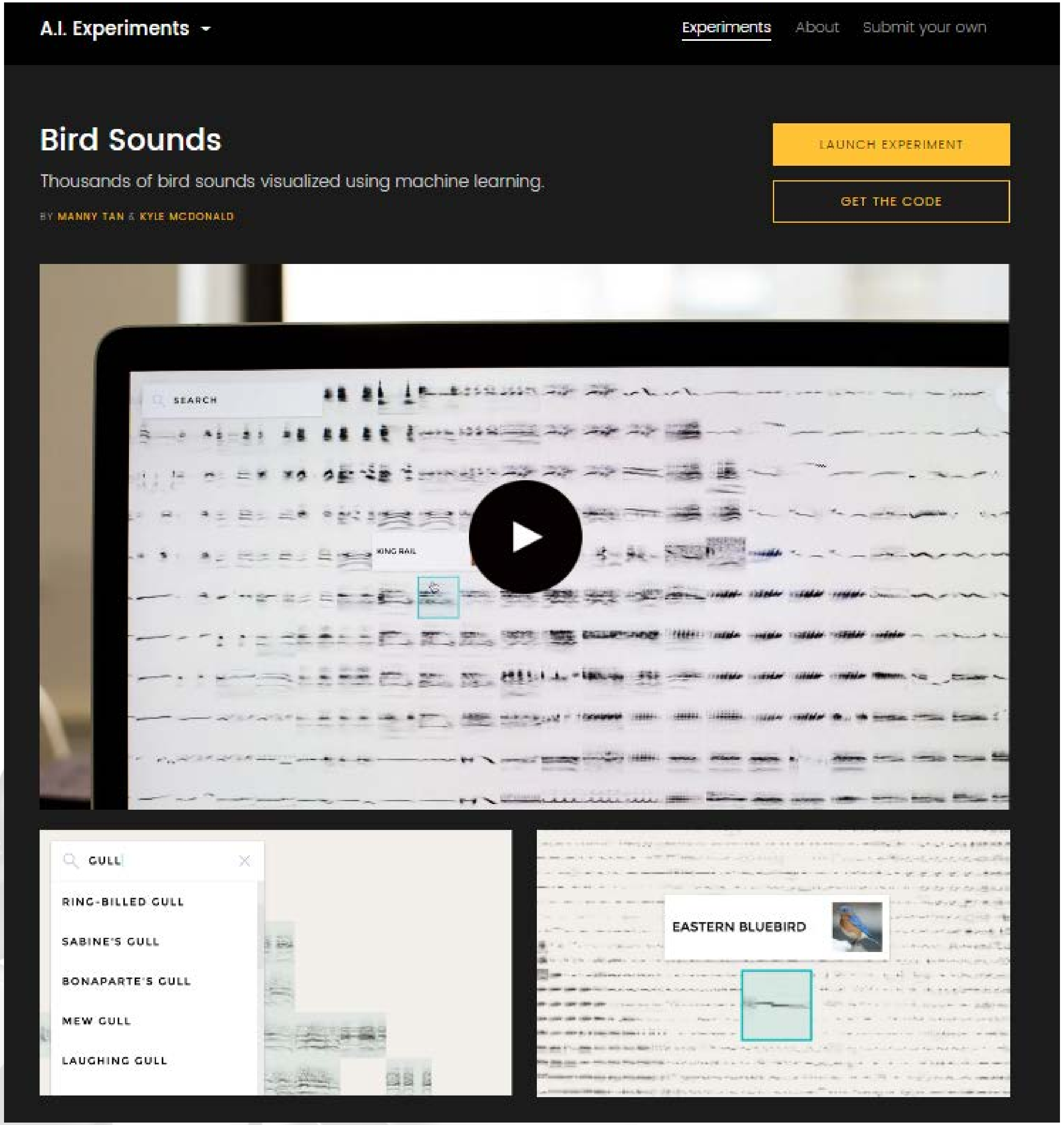
http://www.spark.co.nz/coverage

Results – Hammond Park

- Recording commenced in Hammond Park on 22 June
  - Recording for 1 minute approximately once per hour
  - Extra recordings around dawn and dusk

Morepork? 2-25am Thur 29 Jun 2017 v2.mp3

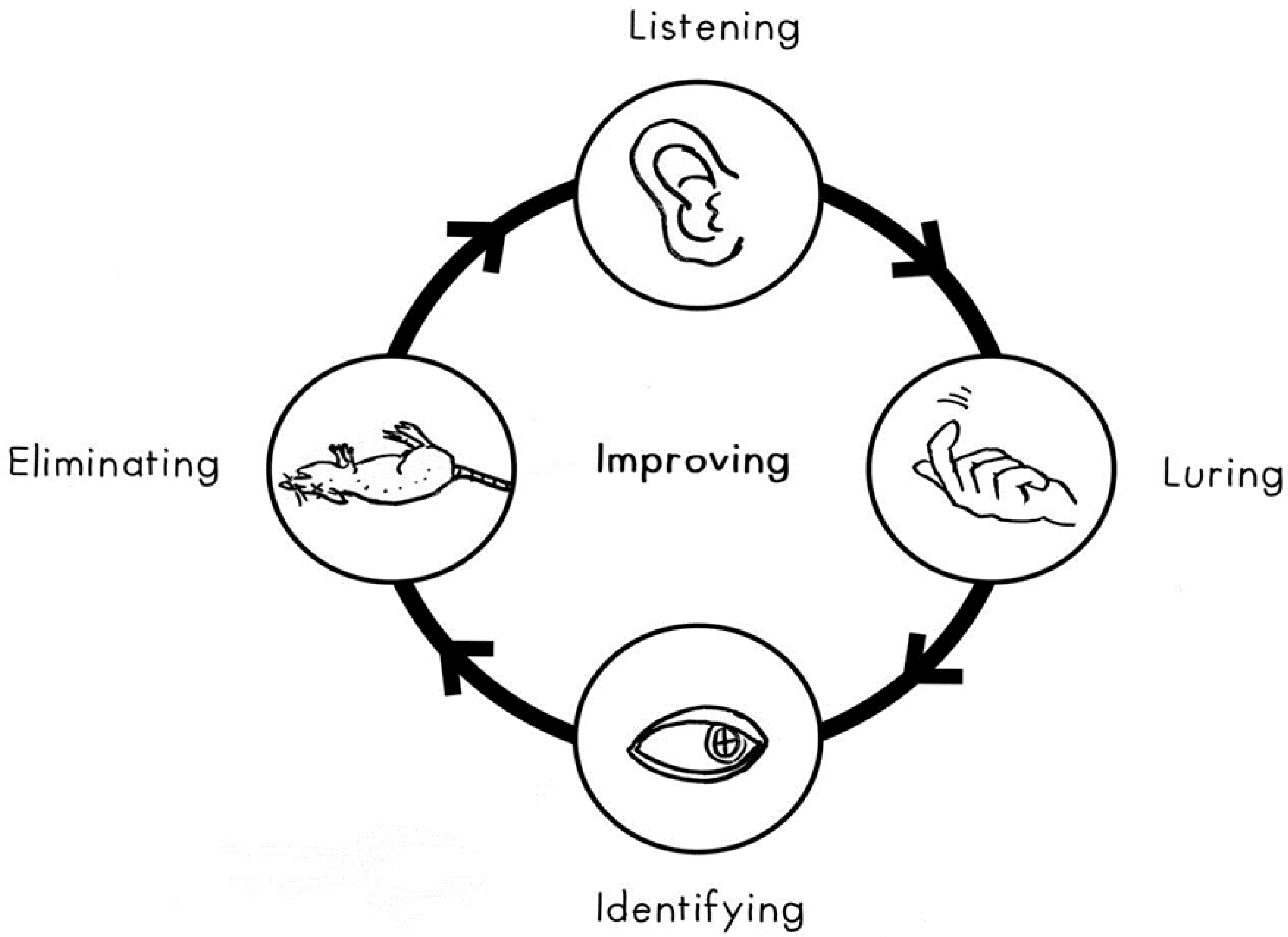
https://www.kidswaikato.co.nz/wp-content/uploads/2016/01/Hammond-Park-5-350x350.jpg



https://aiexperiments.withgoogle.com/bird-sounds

Bird sounds vary widely. This experiment uses machine learning to organize thousands of bird sounds. The computer wasn’t given tags or the birds’ names – only the audio. Using a technique called t-SNE, the computer created this map, where similar sounds are placed closer together.

Built by Kyle McDonald, Manny Tan, Yotam Mann, and friends at Google Creative Lab. Thanks to Cornell Lab of Ornithology for their support. The sounds are available in the Macaulay Library’s Essential Set for North America.



Listening
Luring
Identifying
Eliminating
Improving

Acknowledgements – The Team

- Menno Finlay-Smits - Project Manager/lead developer
- Cameron Ryan-Pears - Main engineer for project
- Tim Hunt (Wintec) - Cacophonometer lead developer
- Dave Lane - Open source design and Drupal CMS integration
- Grant Ryan - Project initiator/coordinator
- Jessica Lyons - Social media (Concentrate Ltd)
- Matthew Aitchison - Machine Learning
- Brent Martin - Machine learning (University of Canterbury)
- Elaine Murphy (DOC) - animal behaviour
- Roger McKenzie - Hardware technical advice
- Finn Maunsell - Cacophony Index - bird song analysis
- Gray Rathgen - Designer
- Kate Haley - Supporter
- Paul Campbell - Electronics design
- Tim Sjoberg (ZIP) - animal behaviour
- Mark Nikoria (Wintec) - data visualsation
- Michael Busby - Website design and development
- Max Johns - Content
- Matt Kavermann - Digital lures
- Alex James and Michael Plank (University of Canterbury) - Modelling and statistics
- Stephen Marsland (Massey University) - Bird song analysis
- Shaun Hendy - Science supporter

Friends of the Project

These kind folks and organisations have provided their support, promoted us to their networks, or have generally helped us along the way!

Main financial contributors
NEXT Foundation
Zero Invasive Predators
Jasmine
Spark Foundation
Other companies offering services for free for the project
Spark - providing free use of high speed mobile network, data hosting, mobile phones and high quality technical support!
Balanced Accounting - all accounting services
Concentrate Ltd - excellent social media services
Anderson Lloyd Lawyers - legal advice
Other supporters
PurePods Ltd
That Quirky Foundation
We're pretty sure that if you listen hard enough, you can hear the birds thanking our supporters.

Hamilton City Council
Riverlea Environment Society

Thank you!

Tim Hunt
tim.hunt@wintec.ac.nz

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

SUPA: Strewn User-preserved authentication

a distributed authentication system for secure computing

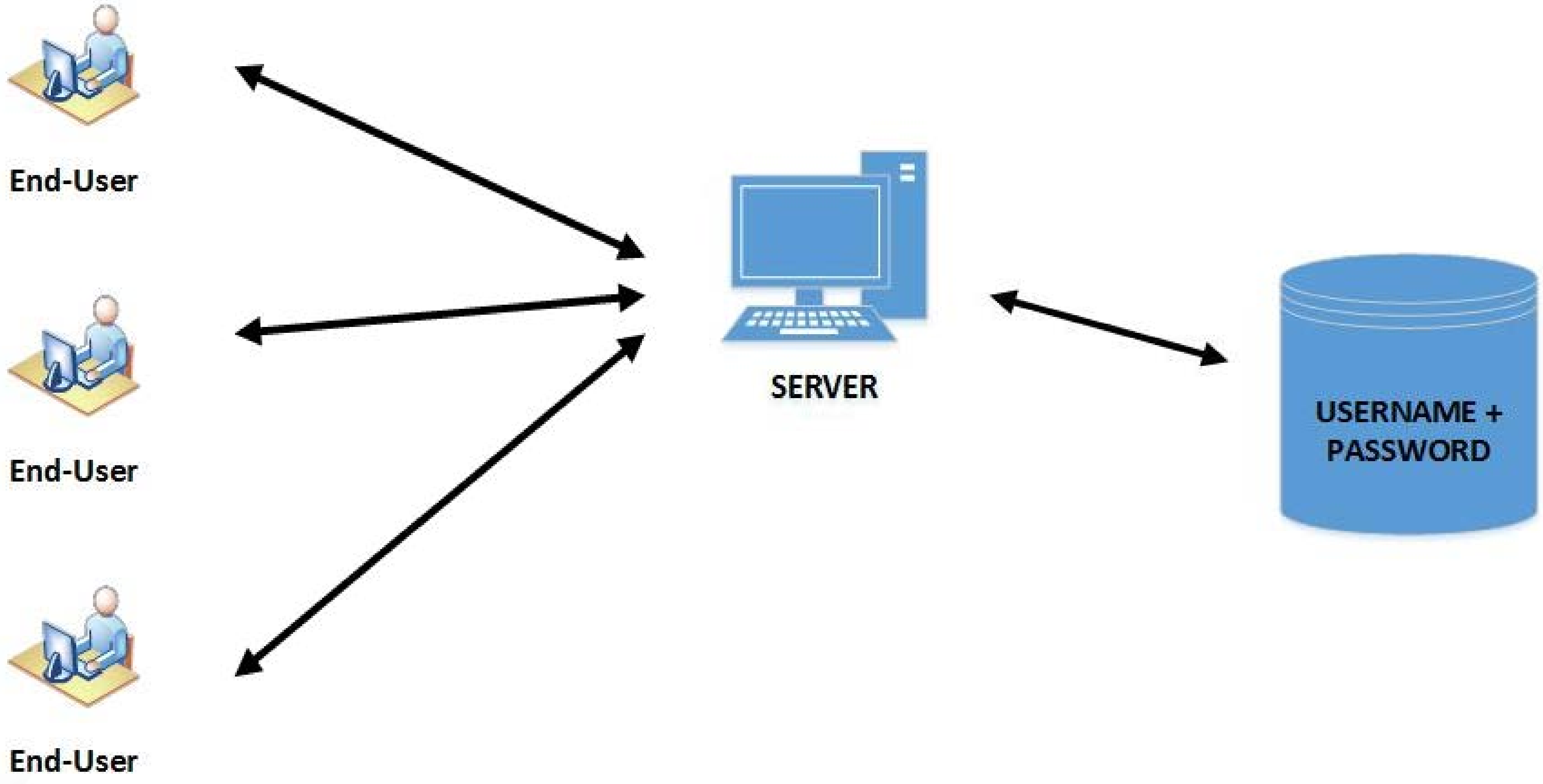
SUPA: What is it?

- An Authentication system
- Combines existing concepts to propose a new approach

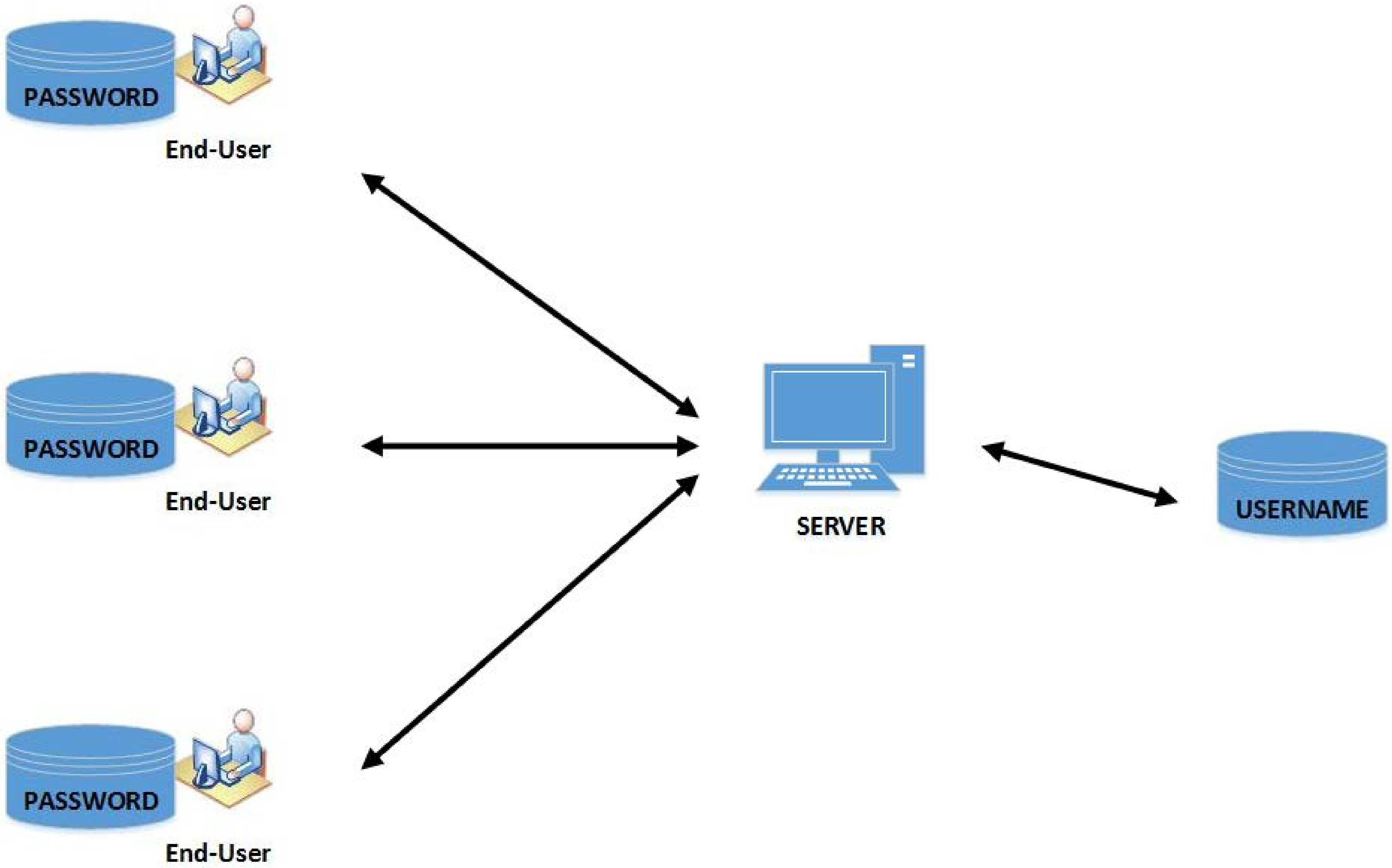
A research in progress

Traditional authentication approach

- Database with all users' secret credentials in the system.
- Alluring repository for attackers.



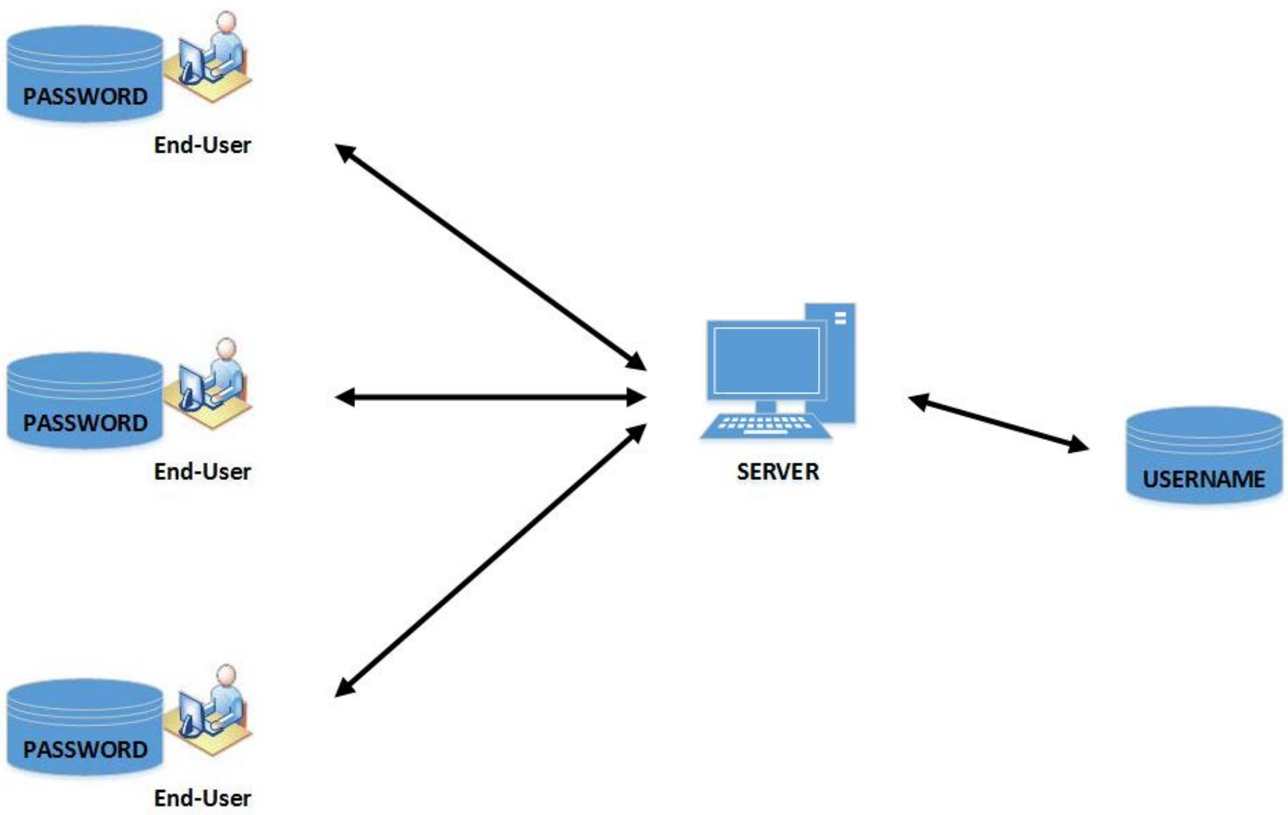
SUPA



PASSWORD
End-User
PASSWORD
End-User
PASSWORD
End-User
SERVER
USERNAME

SUPA

- Does not store users' passwords in the system.
- Passwords are stored at user-end.
- Thus, there is no database of 'secret credentials'.



But...

- Isn’t HASH irreversible?
- Why use SUPA then?

  - – Hash irreversibility
    - arguments exist
    - practically impractical (for the time being) but theoretically possible
    - relative to computing power

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

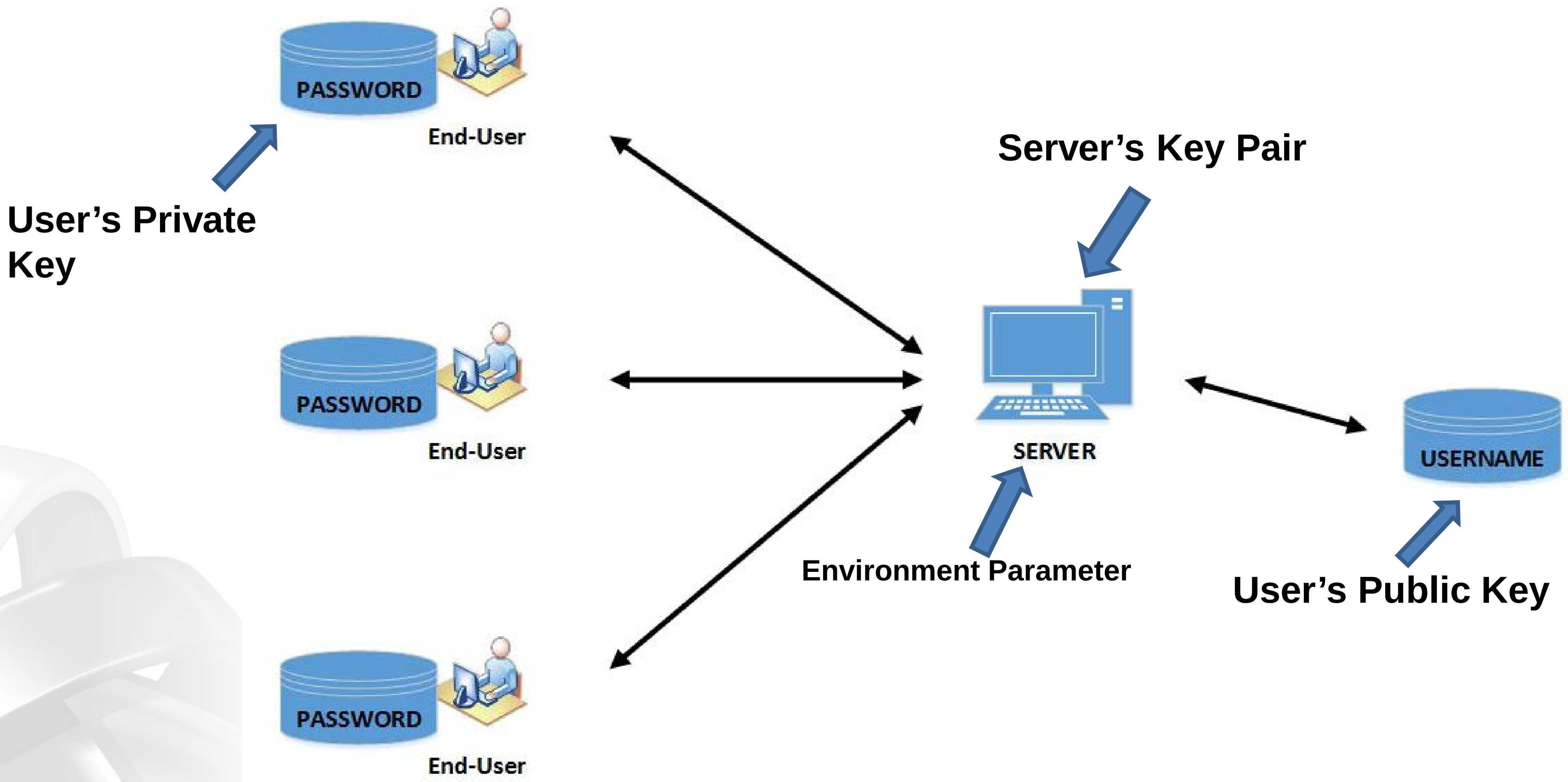
But...

- Isn’t HASH irreversible?
- Why use SUPA then? (cont..)
  - Hashing for the first time – at system-end or at user-end?
  - Key-logger?
  - Use key and salt to further strengthen the hash – but where to store salts and hashes?

I know the hash value, key and salt of your password – BINGO!!!

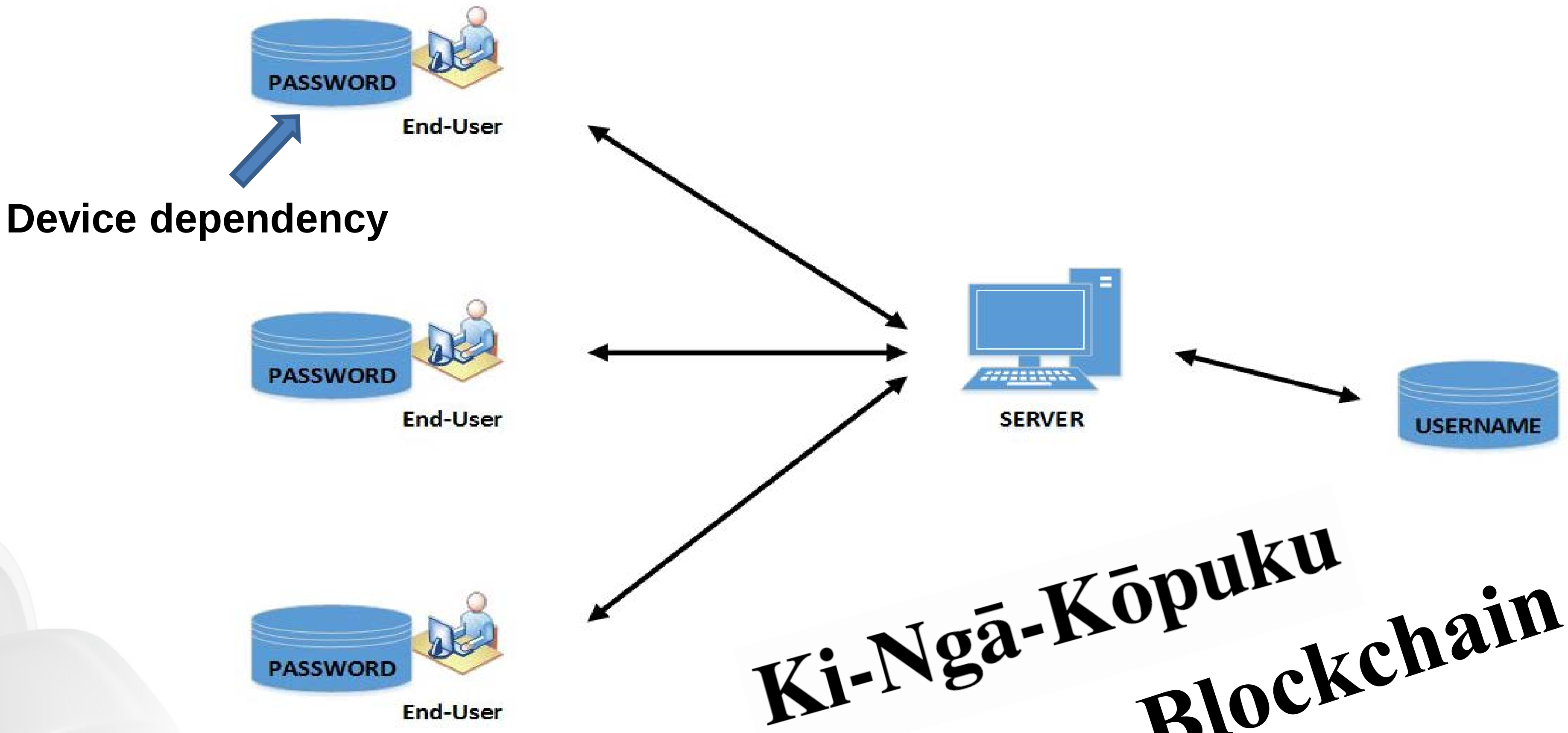
Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

SUPA



PASSWORD
End-User
User's Private Key
PASSWORD
End-User
PASSWORD
End-User
Server's Key Pair
SERVER
Environment Parameter
USERNAME
User's Public Key

SUPA - Limitations



Ki-Ngā-Kōpuku

Blockchain

Problem to solve:
- What if a user wants to log in from a new device.

Future developments

- Resolve device dependency
- Develop Working Prototype

Thank you!

Monjur Ahmed
monjur.ahmed@wintec.ac.nz

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Computational Modelling to Track Human Emotion Trajectories through Time

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

We can detect emotions…

Health monitoring

Driver drowsiness monitor

Lie detection

Automatic tutor

Controlling appliances by facial expressions

Entertainment/computer games

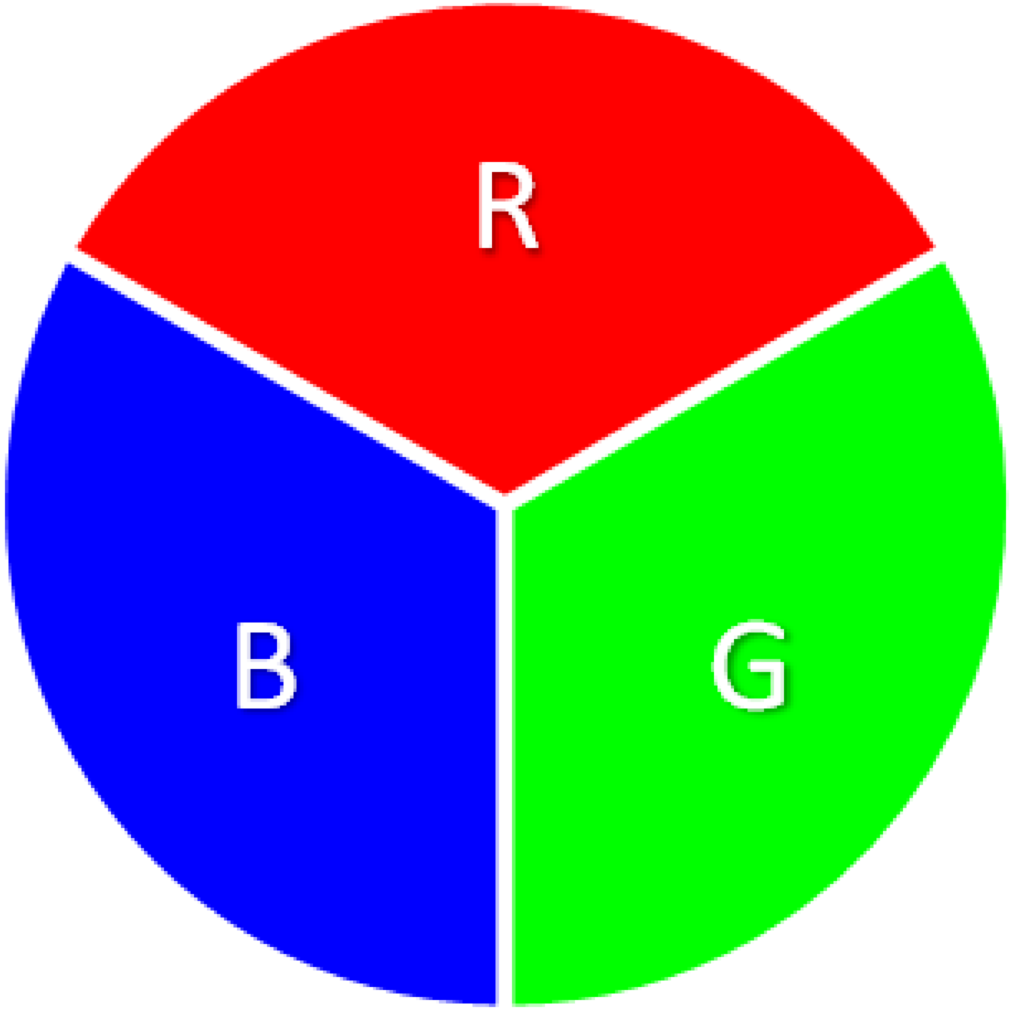
Infant's emotions analyzer

Aim of this study

- To develop a computational system that can
  - Recognise human emotions
  - Represent them in an appropriate space
  - Analyse the paths followed by them while a person experiences a set of emotional states through time.

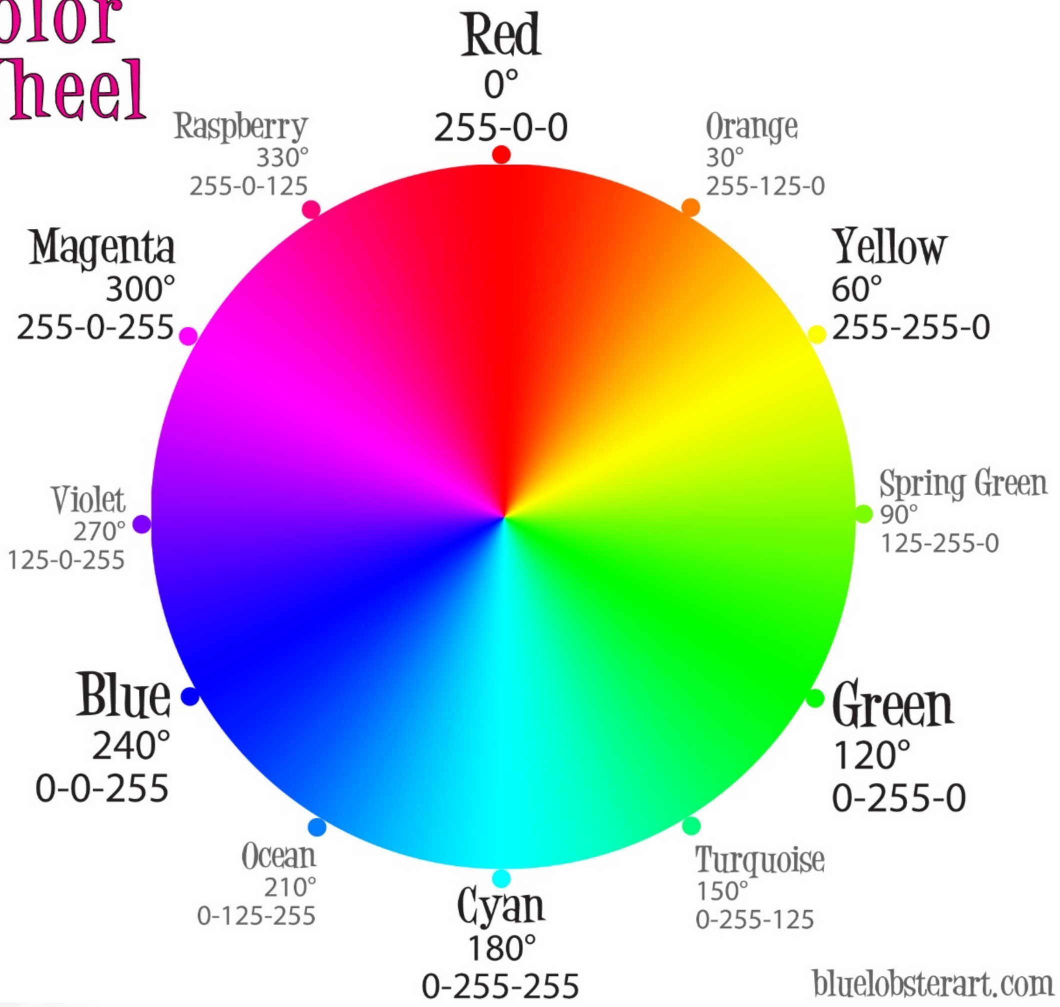
This is not so simple!

Emotions are complex!

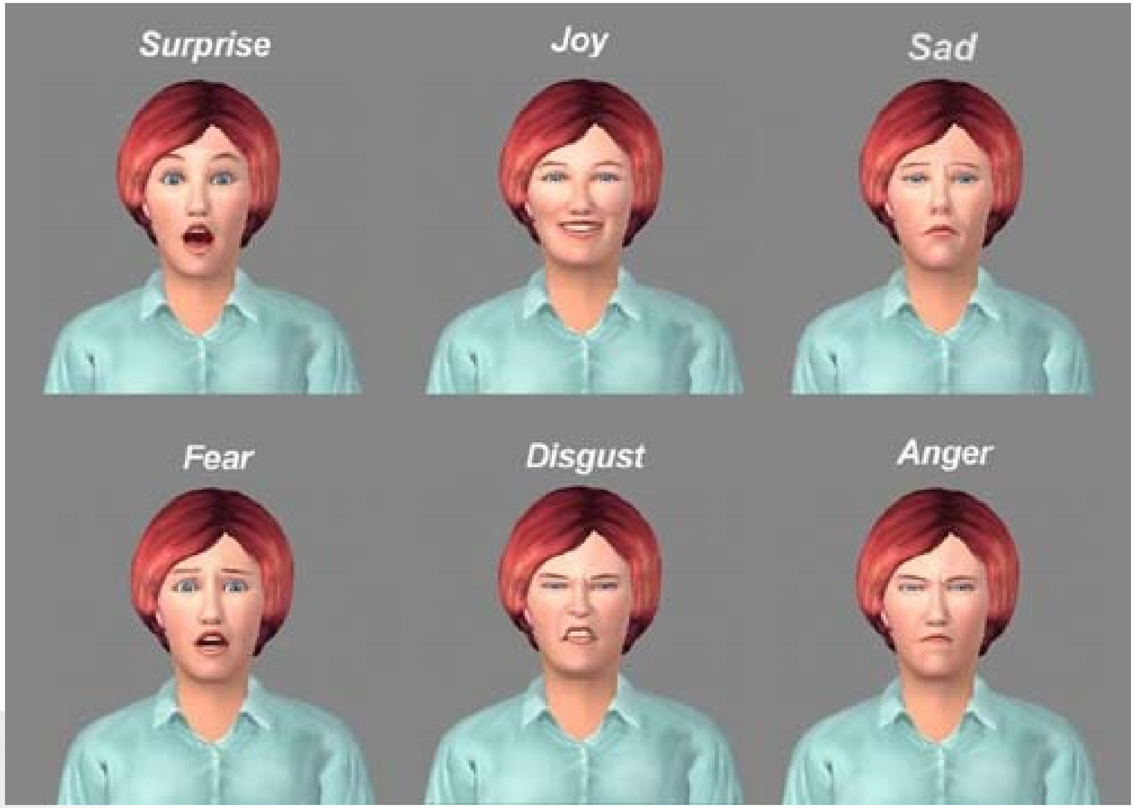


WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

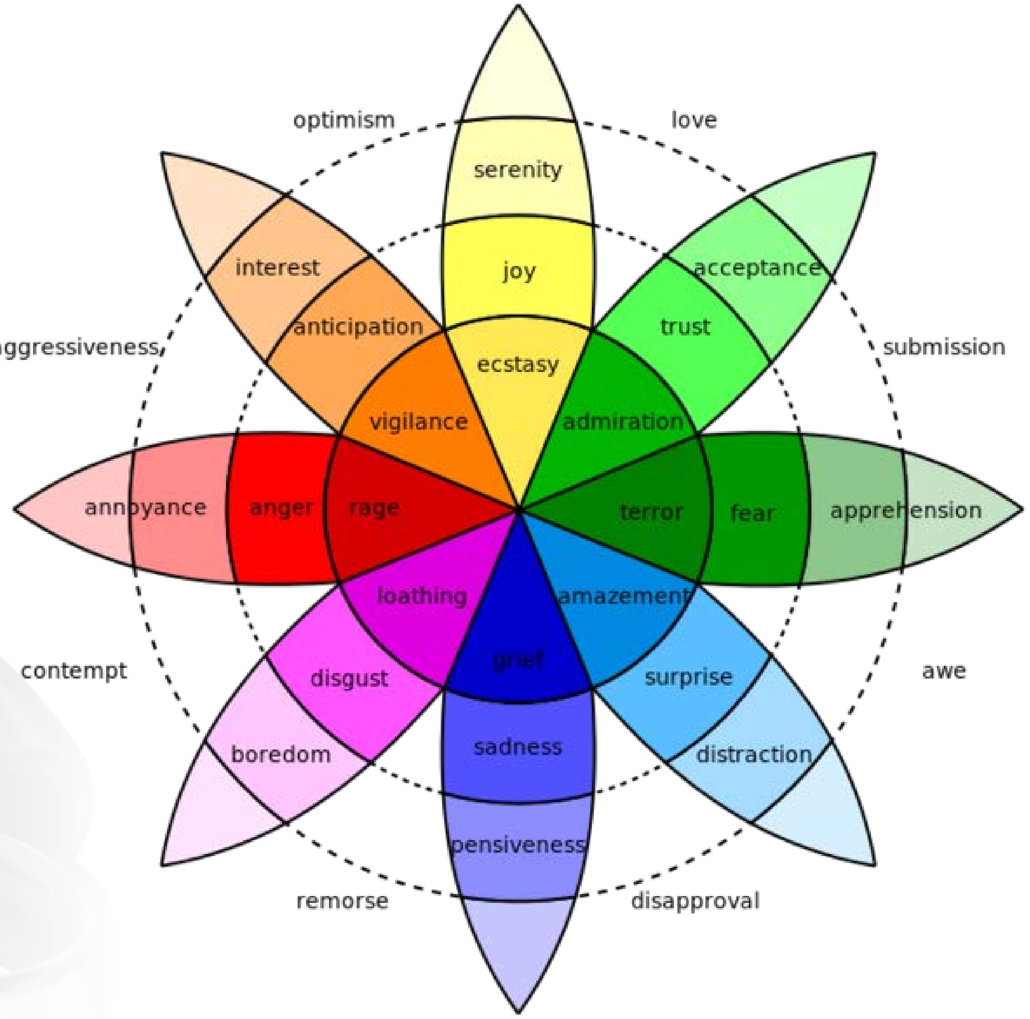
RGB Color Wheel



Six Basic Emotions



Complex Emotions

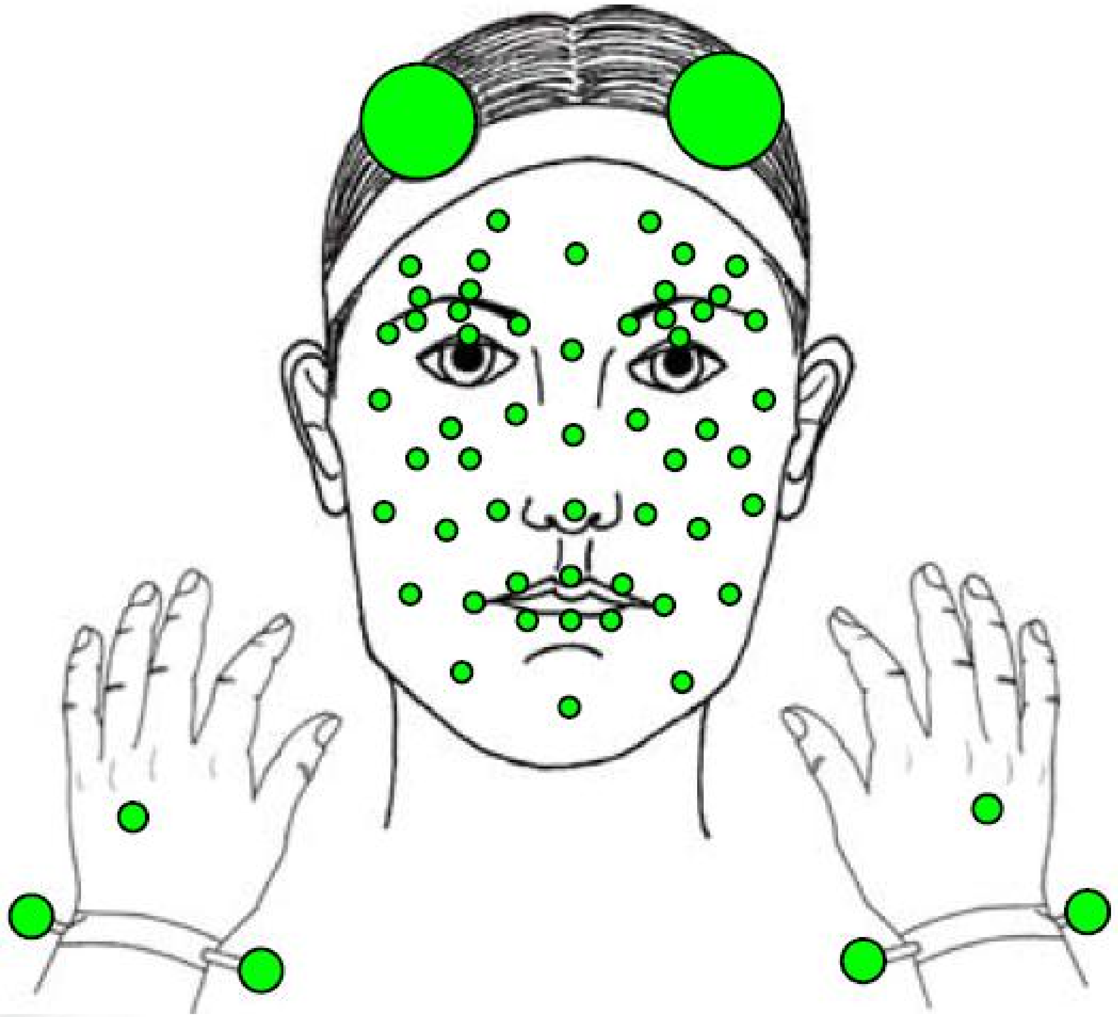


WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Research Questions

1. How can we map a set of facial points to basic as well as complex emotions?
2. Is it useful to represent facial changes over time in some emotion space?
3. What paths do emotions follow while moving from one state to another in an emotion space?

IEMOCAP Dataset



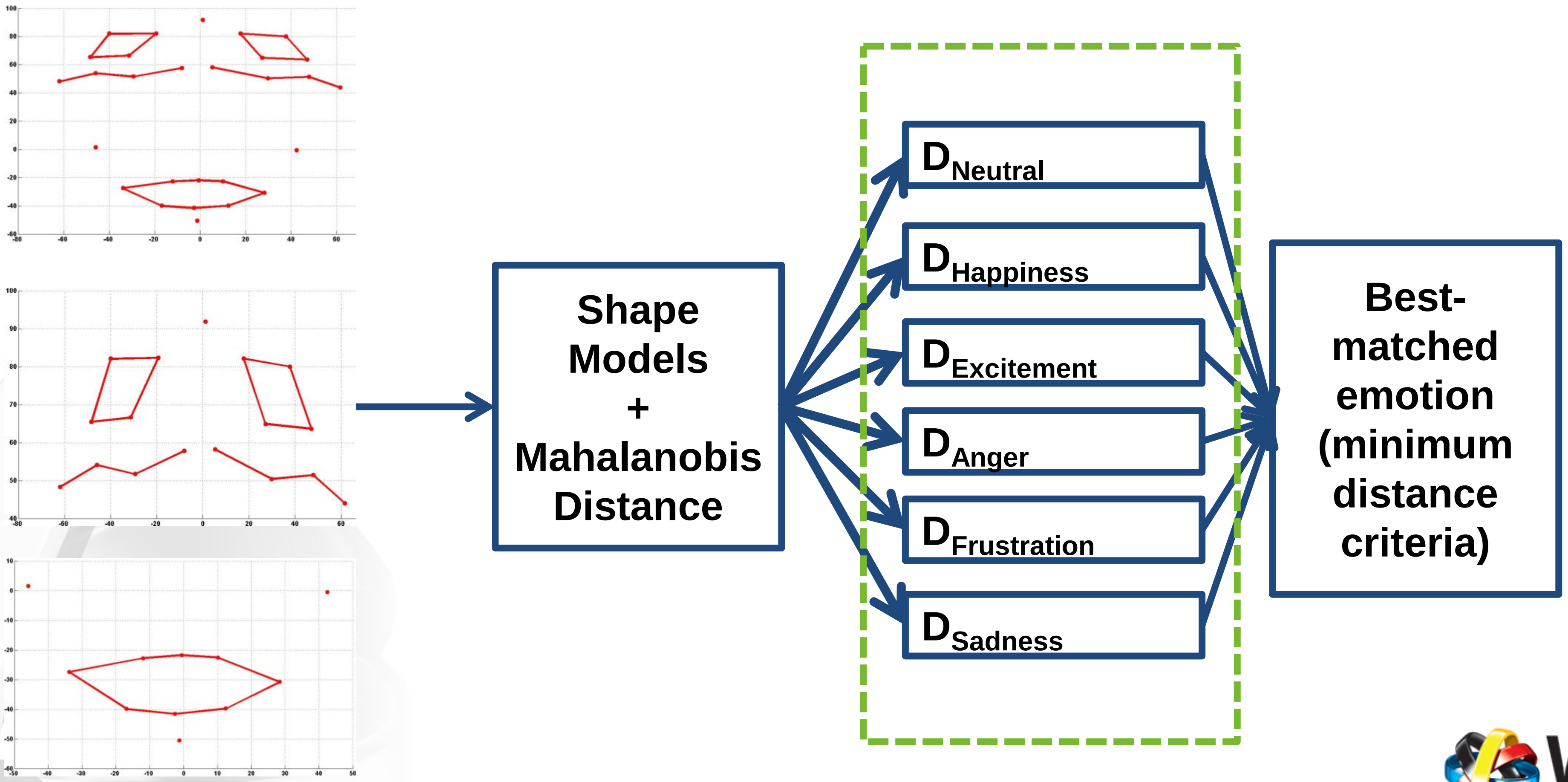
3D location of marker points recorded at 120fps

Our approach

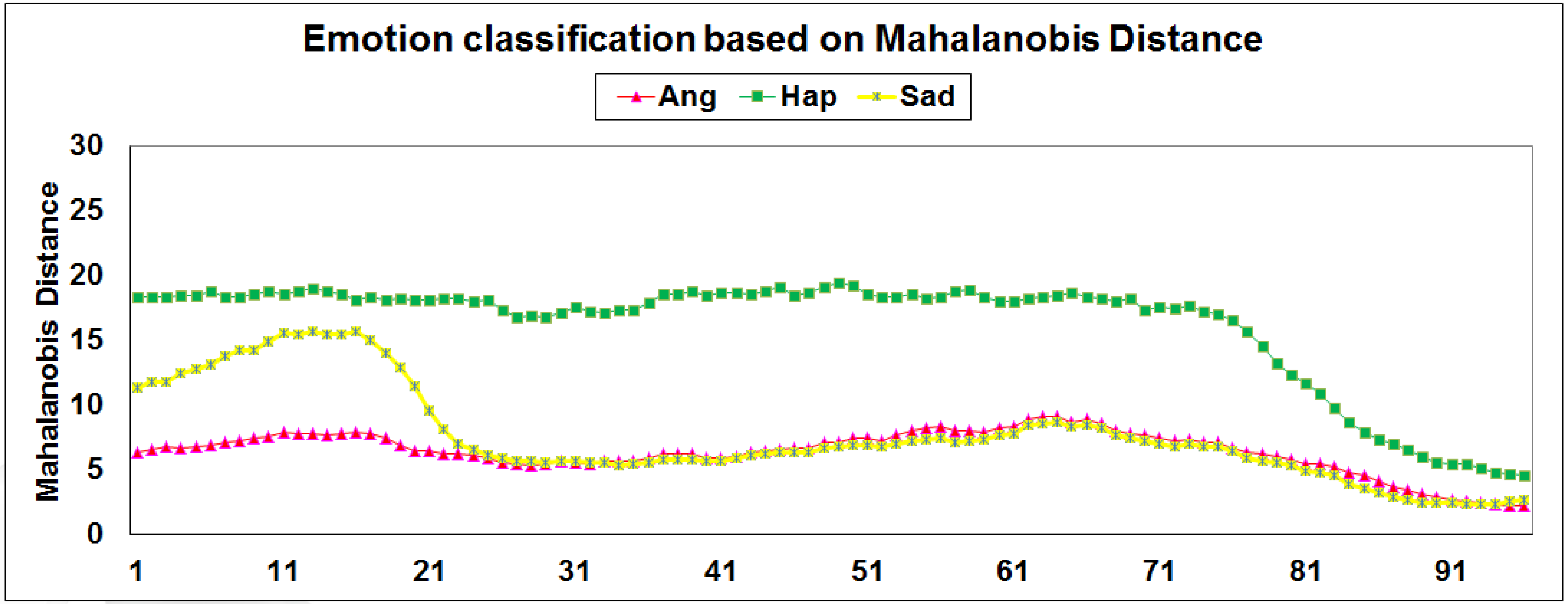
1. Basic Emotion Recognition
2. Modelling Complex Emotions
    1. Mapping in an appropriate space
3. Analysis of Emotion Dynamics

1. Basic Emotion Recognition

1) Basic Emotion Recognition



Blends of emotions



2. Modelling Complex Emotions

2) Modelling Complex Emotions

- Psycho-evolutionary theory
- Circumplex model of emotion representation
  - Activation-evaluation space
- Linear statistical techniques are inappropriate to model circular data.
  - Von-mises Circular Distribution

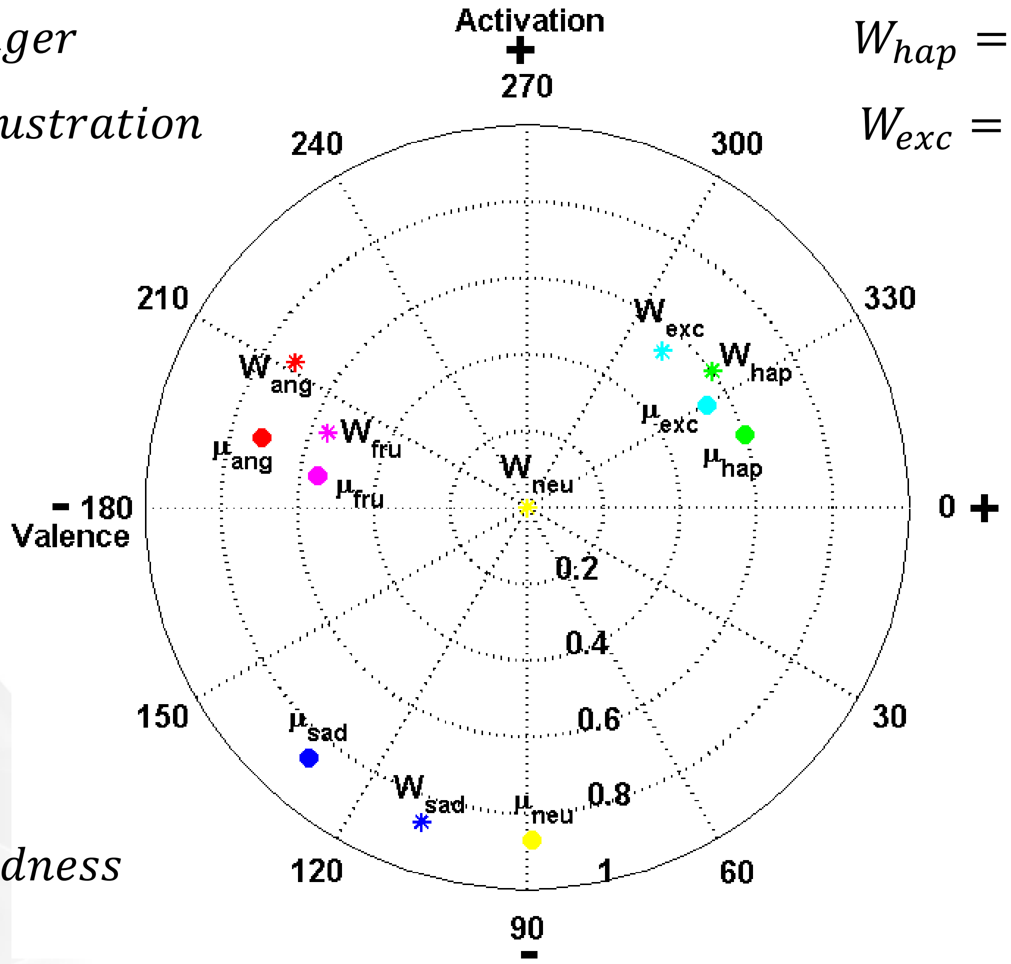
Mapping Basic Emotions into the AES

$W_{ang} = Whissel\ Anger$

$W_{fru} = Whissel\ Frustration$

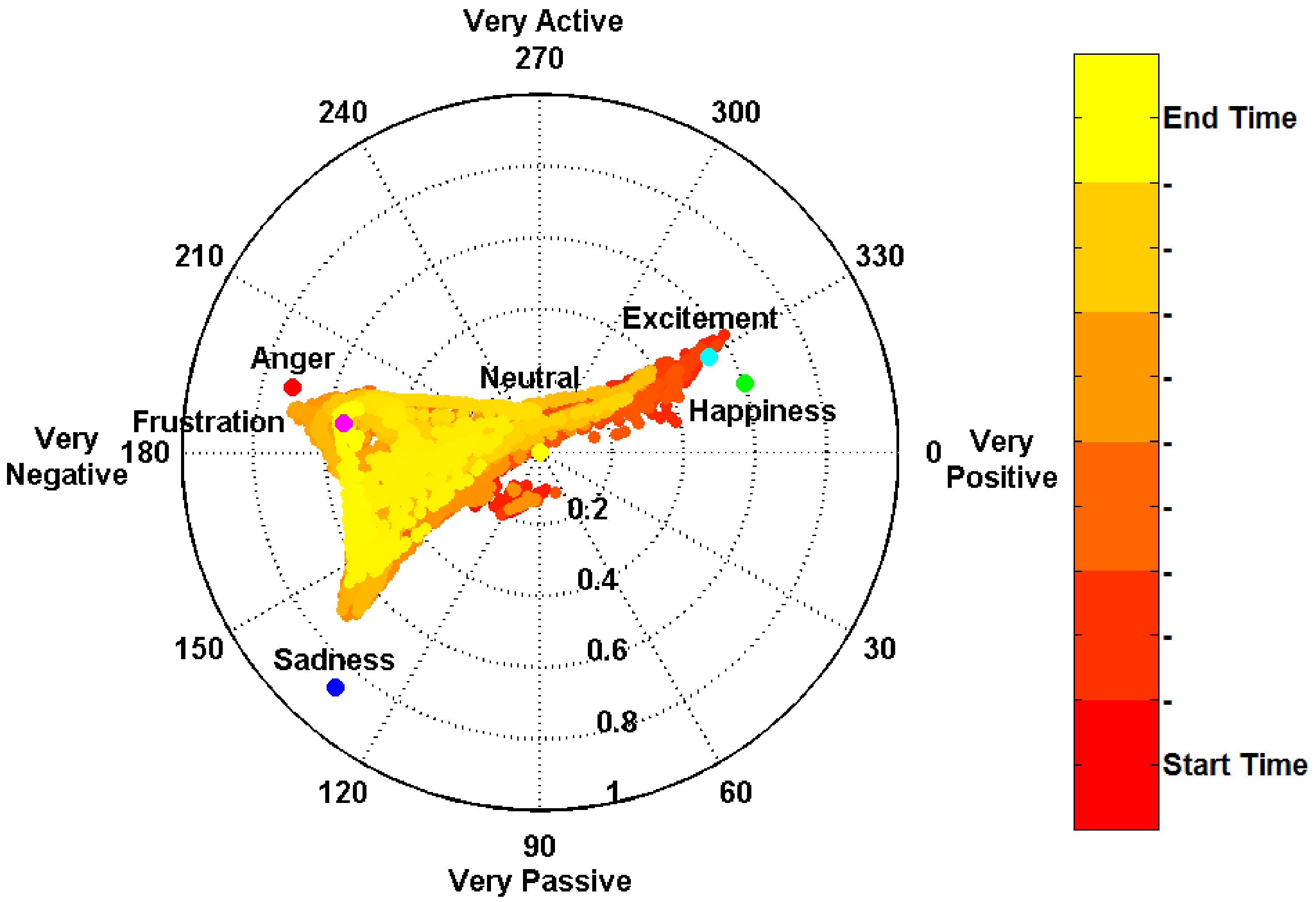
$W_{hap} = Whissell\ Happiness$

$W_{exc} = Whissell\ Excitement$



$W_{sad} = Whissell\ Sadness$

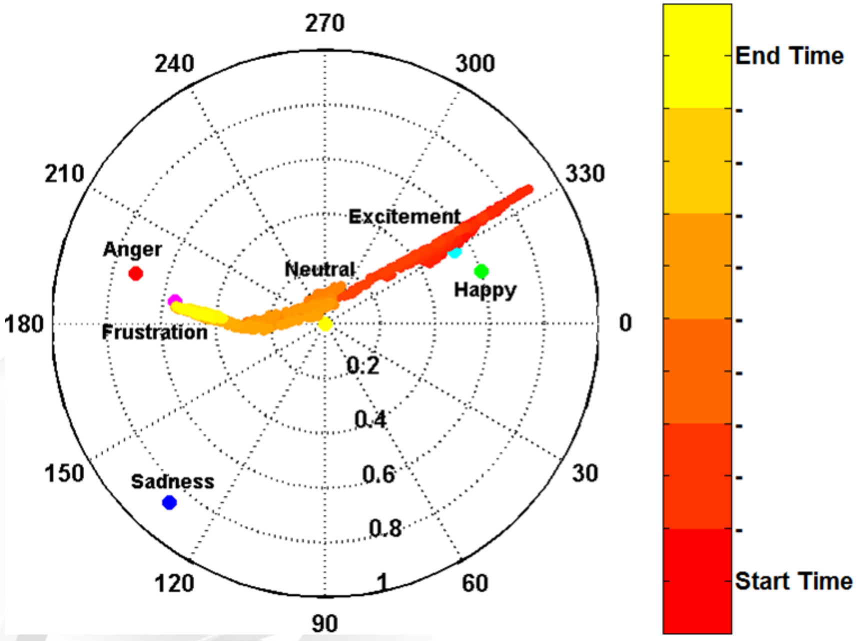
Mapping Complex Emotions into the AES

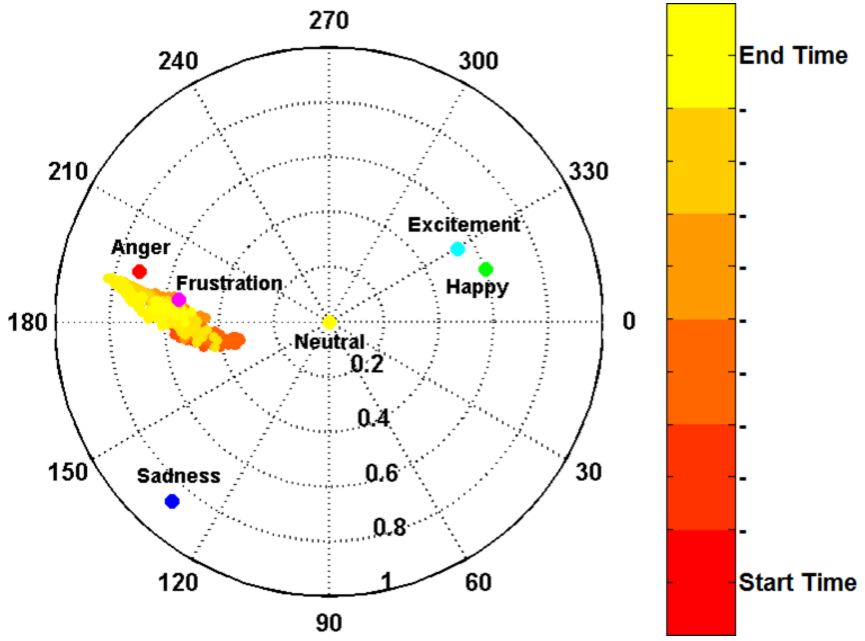


3. Analysis of Emotion Dynamics

3) Analysis of Emotion Dynamics

1) The paths between emotions form ‘smooth’ trajectories in the space.





2) Negatively correlated emotions **3) Positively correlated emotions**

WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Tracking Emotions through Time

The facial expressions of female actor is being tracked

Thank you!

Ayesha Hakim
ayesha.hakim@wintec.ac.nz

Software Evolution Using User Reviews

WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Motivation

- Challenge of Market-Driven Requirements Engineering (MDRE).
- Online Application Development Platforms (OADPs).
- The Case of Android Operating System(OS).
- Application of text mining and Natural Language Processing techniques to analyse user reviews.

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Users' Feedback in OADPs

- App reviews have served as an important source of generating requirements for future versions of apps. (Carreno , L. V. G. and Winbladh, K. "Analysis of user comments: An approach for software requirements evolution," Proc. - Int. Conf. Softw. Eng., pp. 582–591, 2013.)

- Features have been extracted from apps' descriptions (release notes) to assess the relationship between apps names and the features they provide as evident in their descriptions. (Finkelstein, A., Harman, M., Jia, Y., Martin, W., Sarro, F., & Zhang, Y. (2014). App Store Analysis: Mining App Stores for Relationships between Customer, Business and Technical Characteristics. UCL Department of Computer Science, London, Tech. Rep. RN/13/21, 2014)

Study Focus

- Reconciliation of developers responses to requests expressed in users’ reviews in subsequent releases of apps.

Research Questions

- RQ1. Do Android developers respond to end-users' requests for application improvements?

- RQ2. How are end-users' requests addressed over time?

Research Setting (Repository)

- Issues identified by the Android community are submitted to the Android OS issue tracker hosted by Google.
- We extracted a snapshot of the issue tracker, comprising 21,547 issues logged between January 2008 and March 2014.
- Issues were labelled as defect (15,750 issues), enhancement (5,354 issues), others (5 issues), and null (438 issues).

Enhancement Requests

Version (Release)	Last release date	Number of days between releases	Total issues logged	Mean issues per day
Early versions (1.0, 1.1)	09/02/2009	451	173*	0.4
Cupcake (1.5)	30/04/2009	80	64	0.8
Donut (1.6)	15/09/2009	138	141	1.0
Éclair (2.0, 2.01, 2.1)	12/01/2010	119	327	2.8
Froyo (2.2)	20/05/2010	128	349	2.7
Gingerbread (2.3, 2.37)	09/02/2011	265	875	3.3
Honeycomb (3.0, 3.1, 3.2)	15/07/2011	156	372	2.4
Ice Cream Sandwich (4.0, 4.03)	16/12/2011	154	350	2.3
Jelly Bean (4.1, 4.2, 4.3)	24/07/2013	586	1,922	3.3
KitKat (4.4)	31/10/2013	99	781	7.9
		2,176	5,354	2.7
* Total number of issues logged between the first beta release on 16/11/2007 and Android version 1.1 released on 09/02/2009				

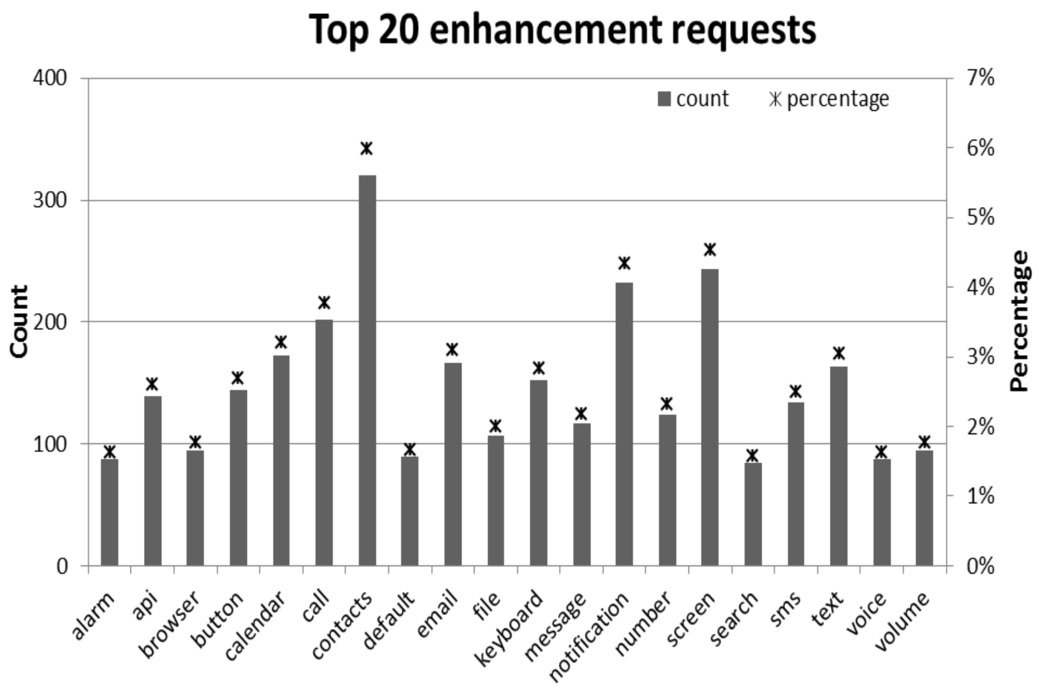
Features Mentioned Most

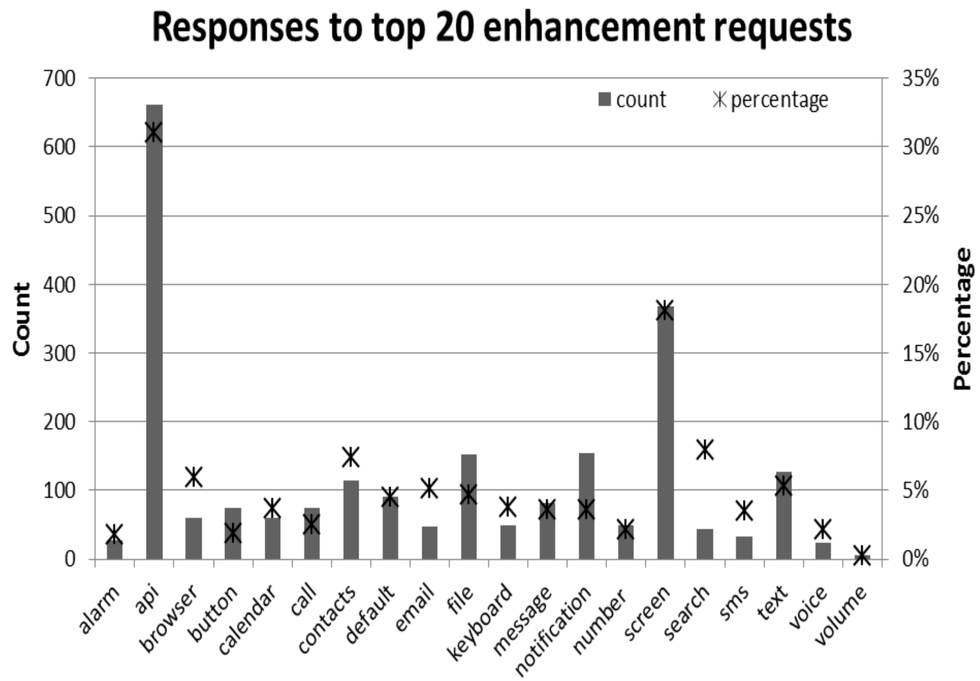
- Corpus linguistic part-of-speech (POS) tagging: we created a program that incorporated the Stanford API to enable us to extract noun phrases from both the enhancement requests and release notes. (Toutanova et al. 2003. "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," in: *ACLHLT*. Edmonton, Canada: Association for Computational Linguistics, 173-180.)

- Computational linguistic n-gram analysis: we extended the POS tagging by computing the total of each noun as unigrams (e.g., if “SMS” appeared at least once in 20 enhancement requests our program would output SMS = 20). (Manning, C. D., and Schtze, H. 1991. *Foundations of Statistical Natural Language Processing*. London: MIT Press.)

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

RQ1. Do Android developers respond to end-users' requests for application improvements?





Contacts, screen, ***notification***, ***call*** **and** ***calendar*** **were the top five features requested for enhancement, whereas** ***api, screen, notification, file,*** **and** ***text*** **were most prominent in Google's responses.**

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

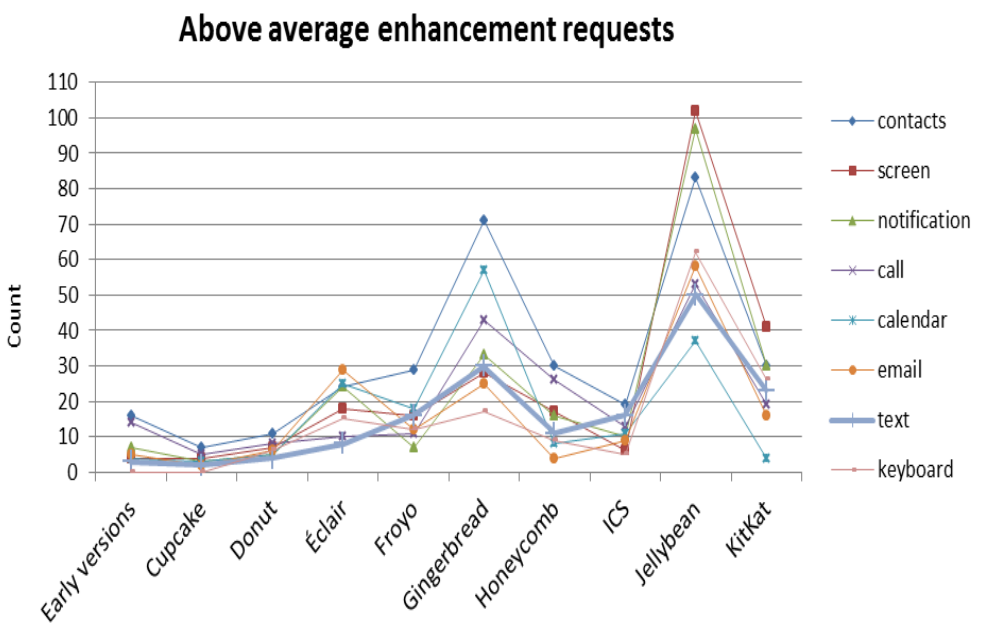
Triangulation of Responses to Requests

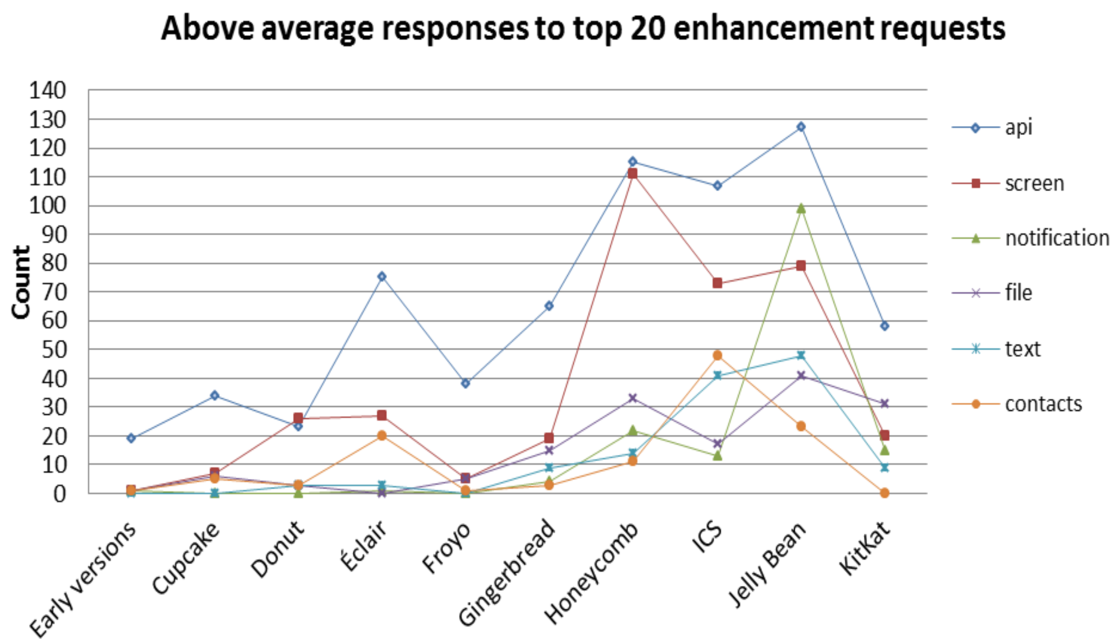
- We assess the potential association between the total numbers of requests and responses for each feature in the top 20 requested features.

- We found that there was a statistically significant and strong correlation between the number of users' requests and the number of Google's responses to these requests (r=0.54, p-value= 0.01).

- This result suggests that the Android development team generally responded to the top enhancement requests made by the Android community.

RQ2. Developers Responses to Requests Over Time





Screen and api issues were given the greatest amount of attention over the Android OS releases.

Triangulation of Requests and Reponses– Android Versions

- Correlations between enhancement requests and responses across all Android OS major versions.

Version	*r*	p-value	Version	*r*	p-value
Early	0.23	0.32	Gingerbread	0.00	0.99
Cupcake	0.18	0.45	Honeycomb	0.40	0.08
Donut	-0.08	0.73	**ICS**	**0.49***	**0.03***
Éclair	0.21	0.39	**Jellybean**	**0.57***	**0.01***
Froyo	0.22	0.34	KitKat	0.27	0.26

Statistically significant outcomes are marked with a star (*) and strong and medium correlations are shown in Bold typeface.

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Triangulation of Requests and Reponses – Top 20 Features

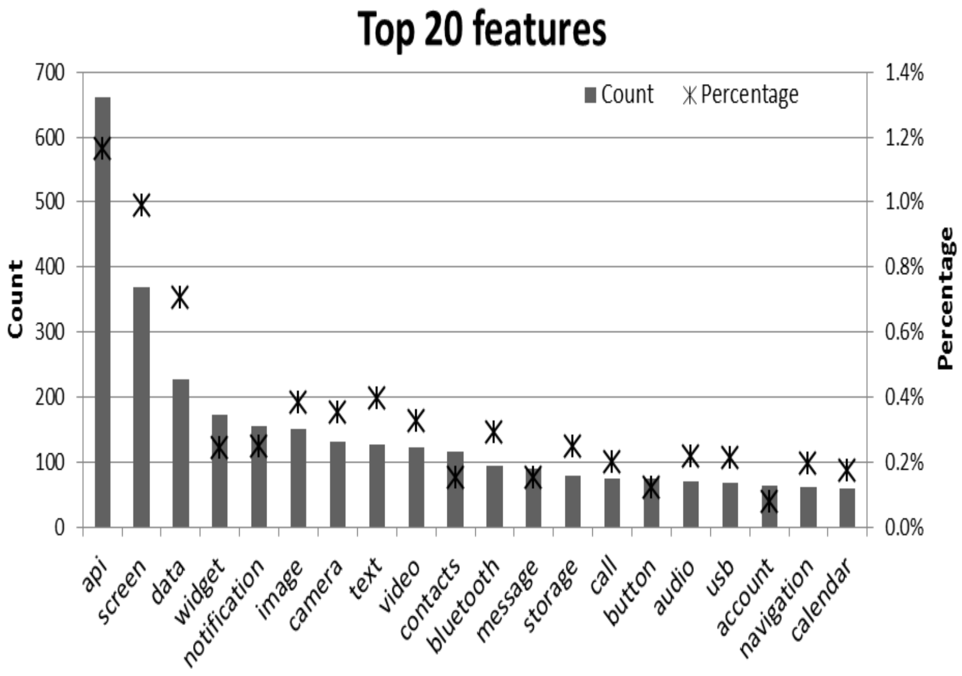
Feature	*r*	p-value	Feature	*r*	p-value
alarm	0.23	0.53	keyboard	0.34	0.34
api	0.30	0.39	**message**	**0.66**	**0.04***
browser	0.14	0.69	notification	0.55	0.10
button	0.53	0.11	**number**	**0.77**	**0.01***
calendar	0.17	0.64	screen	0.46	0.19
call	**0.69**	**0.03***	search	-0.55	0.10
contacts	0.07	0.85	sms	0.22	0.53
default	0.48	0.16	**text**	**0.71**	**0.02***
email	0.09	0.81	voice	0.16	0.67
file	0.42	0.23	**volume**	**0.79**	**0.01***

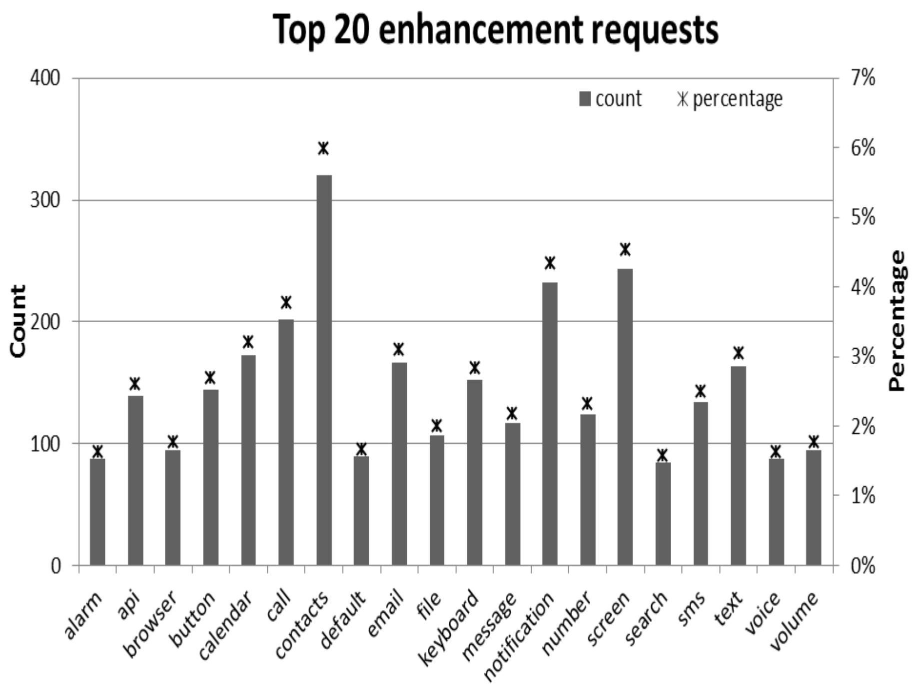
Results of Triangulation

- End-user facing features were given the most attention - overall, enhancement requests for ***call, message, number*****,** ***text*** **and** ***volume*** were given particular attention over Android OS releases.

- Latter releases of Android benefitted the most from user reviews.

Summary





Overlap of 45% as 9 of the enhancement requests are found in the top 20 features found in the release notes.

Acknowledgments

- Licorish, S. A., Tahir, A., **Bosu, M. F**. & MacDonell, S.G. (2015). On Satisfying the Android OS Community: Users' Feedback Still Central to Developers' Portfolio. In Proceedings of the 24th Australasian Software Engineering Conference (ASWEC 2015)

- We thank Google for making their issue logs publicly available.

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

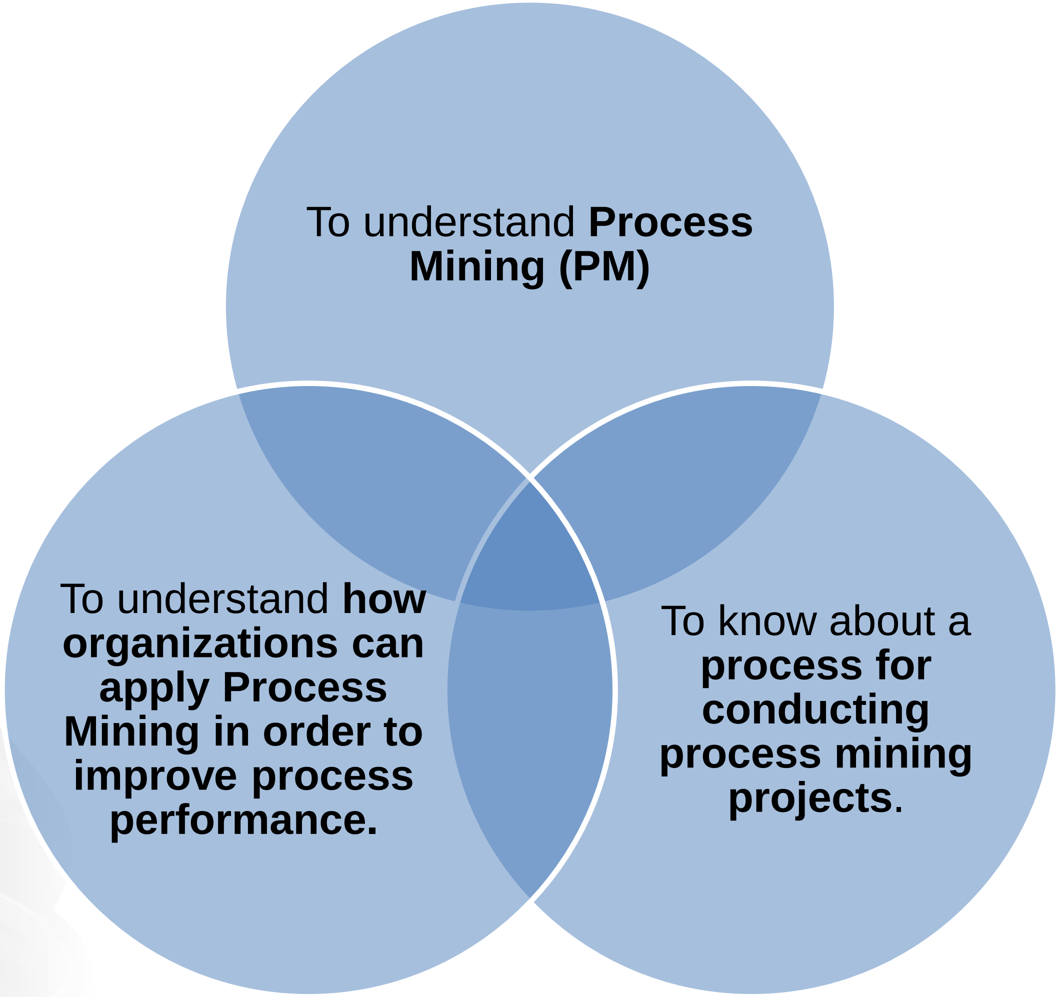
Thank you!

Michael Bosu
michael.bosu@wintec.ac.nz

Process Mining

a special type of data mining to improve process performance

Presentation Outcomes



Context

Increased use of information systems to support process execution

+

Detailed information about the implementation of processes being recorded

+

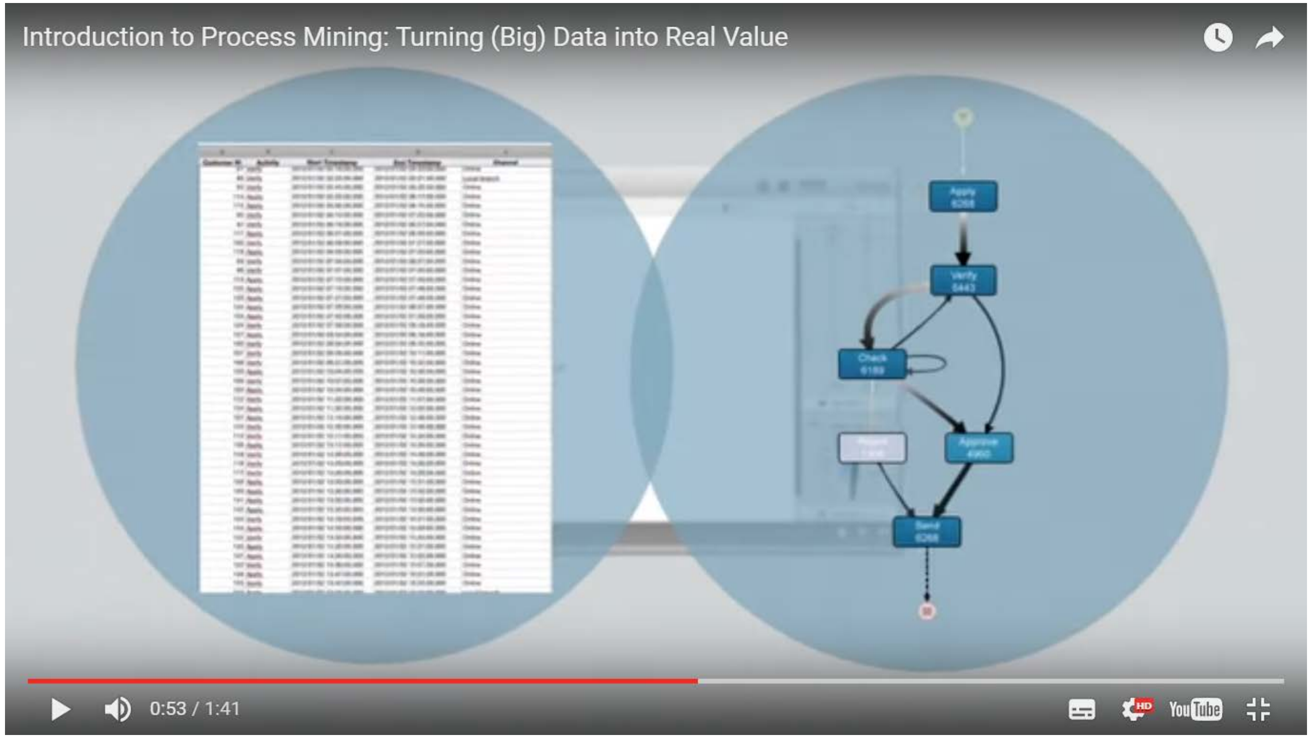
Extract data from existing information systems mediums such as event logs, transaction logs, etc

=

Process Mining: Knowledge regarding the organization’s actual business processes

Process mining

- Process Mining is a relatively young discipline that sits between **data mining and process modeling & analysis**.
- The idea is to **discover, monitor and improve real processes** (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's information systems (VAN DER AALST, 2011).

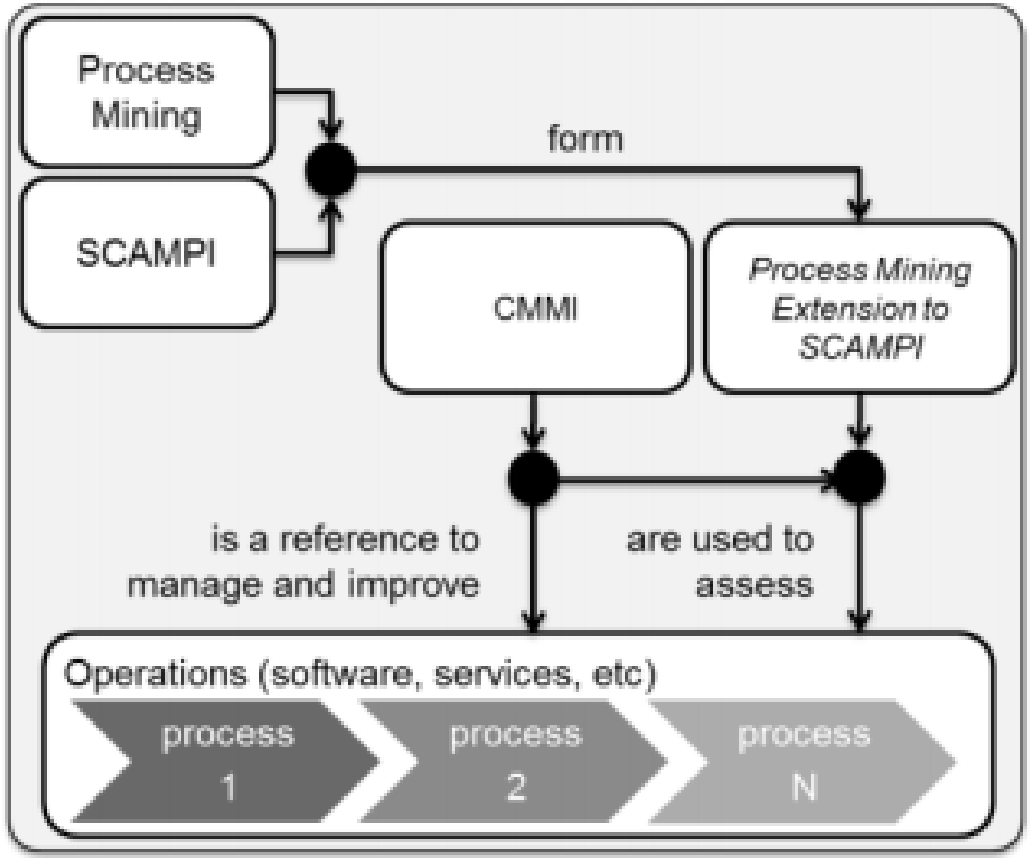


source: www.processmining.org

Background: “Process Mining Extension to SCAMPI”

- An extension to SCAMPI-Standard CMMI Appraisal Method for Process Improvement.
- Developed as an outcome of a PhD degree (Technical Report*)

*available at http://www.biblioteca.pucpr.br/pergamum/biblioteca/img.php?arquivo=/000059/000059bf.pdf

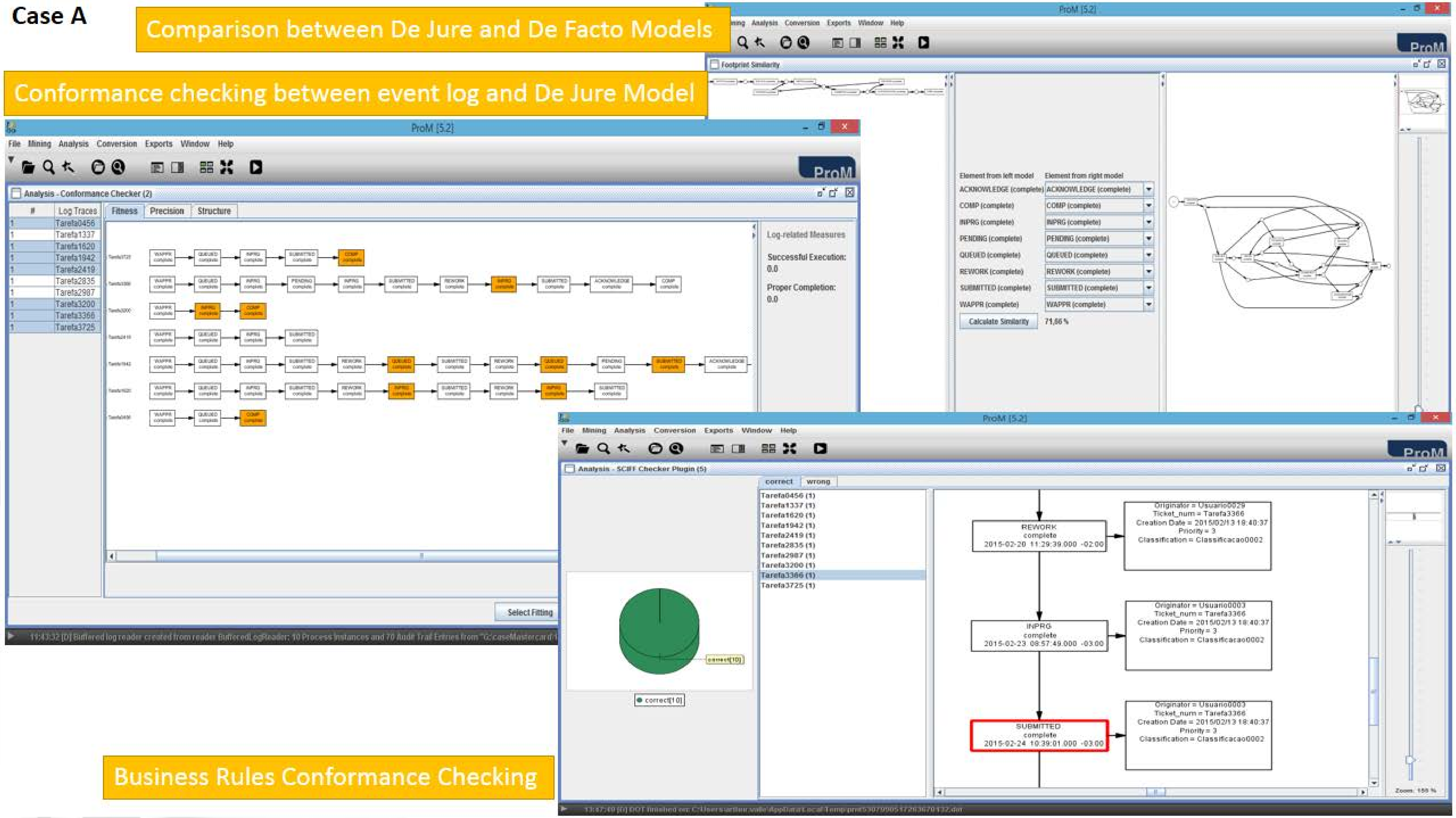


Background: “Process Mining Extension to SCAMPI”

Case A

Comparison between De Jure and De Facto Models

Conformance checking between event log and De Jure Model



Business Rules Conformance Checking

Background: Applied Data Science Research Group:

It aims to ***research and apply data science related technologies****, including but not limited to data mining/text mining /process mining and analytics,* ***in order to provide valuable solutions to industry, academia and community*** *in Waikato region.*

As a differentiation, the group does not develop new tools and algorithms but **identifies relevant functionalities from existing tools and algorithms and integrates them in simple-to-use processes that companies can apply to obtain value from their data**.

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Our current DataScience related research

Objective: To design, implement and apply a process to perform process mining projects

Method: Nine methodological steps, inspired by the V-model:

- Step1: research problem.
- Step2: research objectives.
- Step3: literature review.
- Step4: requirements identification and process implementation.
- Step5: testing the process.
- Step6 and 7: verification and validation of the process.
- Step8: interpretation and discussion of results.
- Step9: conclusion (i.e checking research problem).

Expected results: A process to conduct process mining projects is designed, implemented, tested and applied

Process to conduct Process Mining projects

1. Scoping and Planning:

- 1.1 Identify business processes and associated information systems, and gather basic knowledge
- 1.2 Determine goals and research questions
- 1.3 Determine the required team, data, tools and techniques.

2. Data Understanding:

- 2.1 Locate and explore required data in the system's log
- 2.2 Evaluate the data in the system's logs and select appropriate dataset

3. Data Processing:

- 3.1 Extract the set of required event data
- 3.2 Prepare the extracted dataset, by cleaning, constructing, merging, mapping, formatting and transforming the data
- 3.3 Familiarize and filter log

4. Process Mining and Analysis:

- 4.1 Apply process mining techniques to answer (research) questions

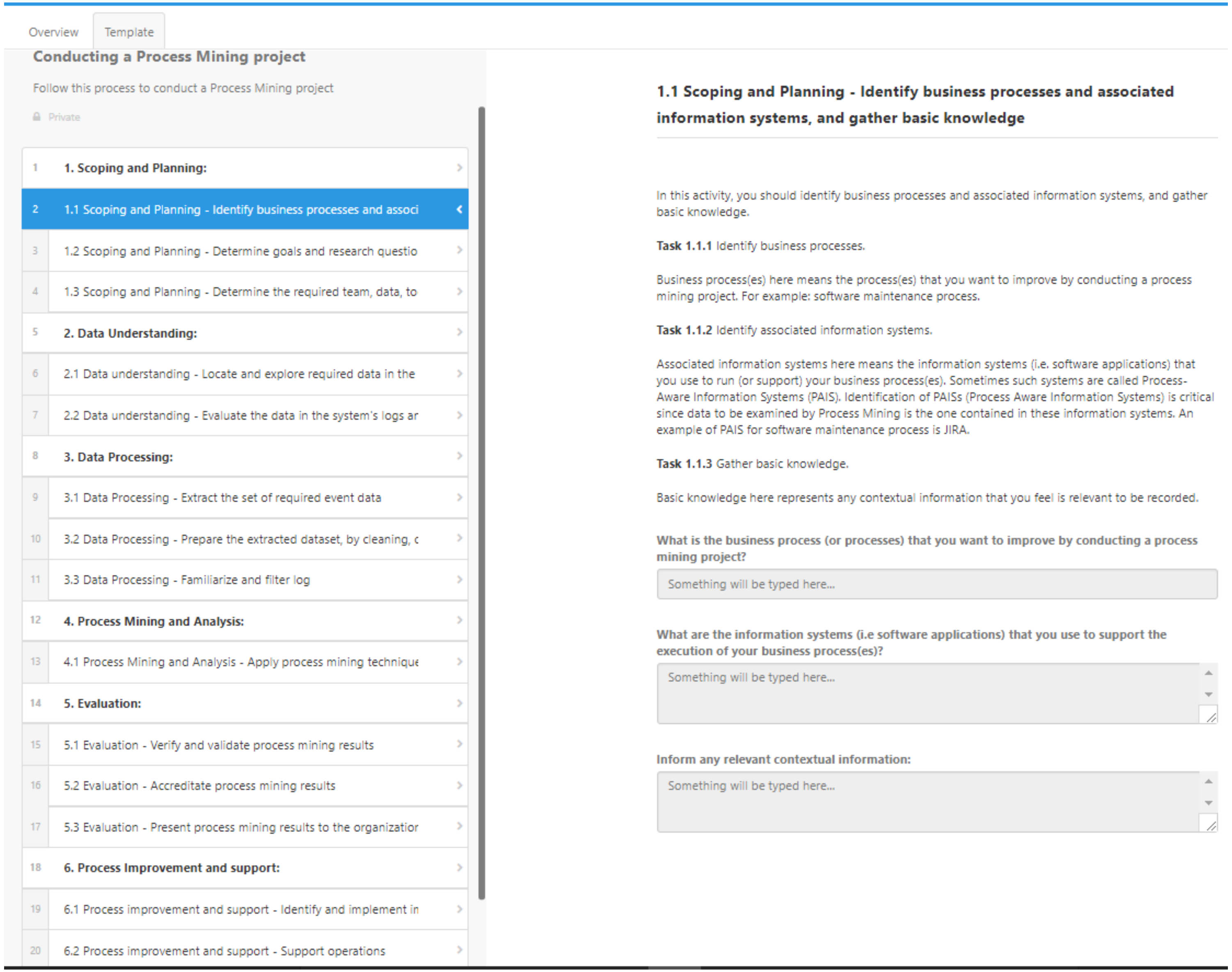
5. Evaluation:

- 5.1 Verify and validate process mining results
- 5.2 Accreditate process mining results
- 5.3 Present process mining results to the organization

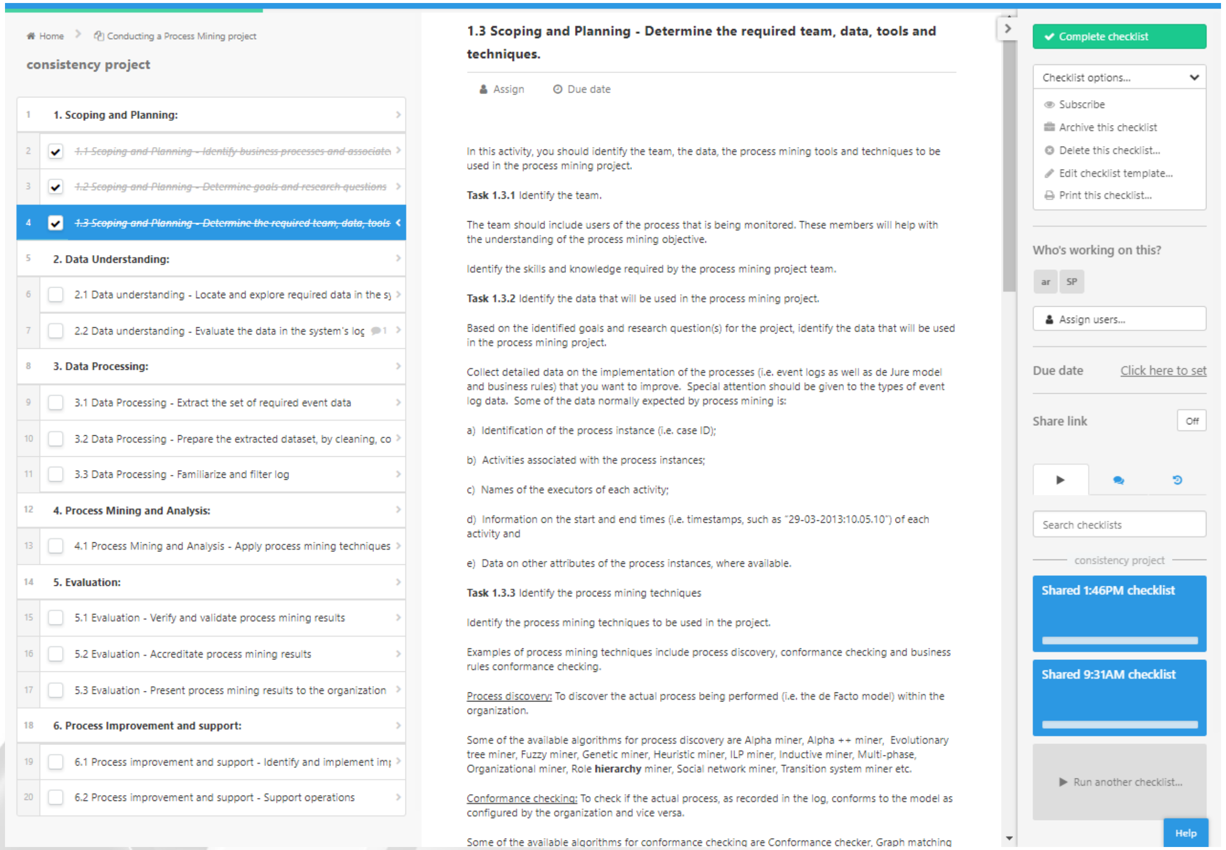
6. Process Improvement and support:

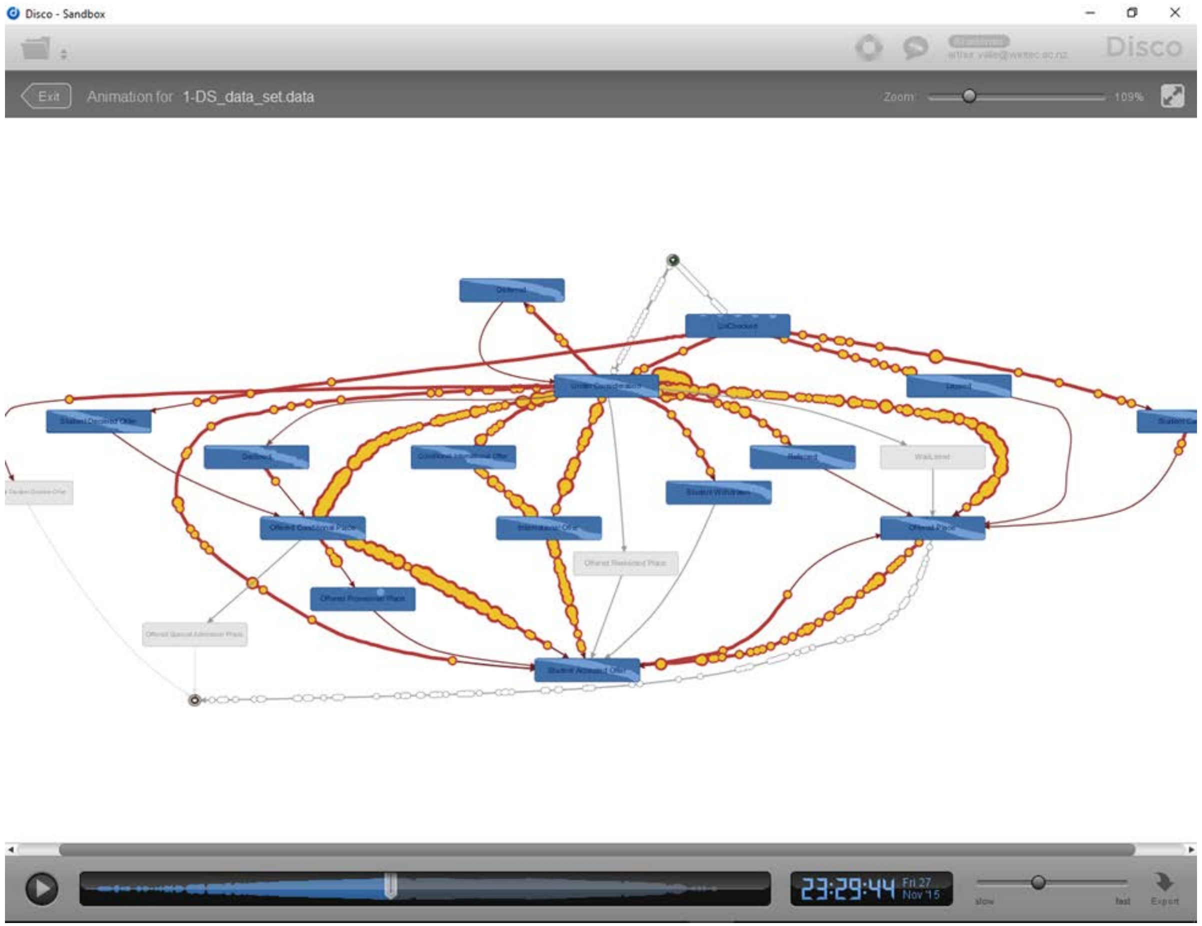
- 6.1 Identify and implement improvements
- 6.2 Support operations

Implementation of the process



Current status: Step5- ”testing” the process





Future work

- To complete the remaining steps (6 to 9) of the research.
- To conduct case studies in organizations (**let us know if you have interest**).
- To involve current and prospect students.
- To provide valuable solutions (i.e consultancy, training, etc) to industry and community.

Thank you!

Arthur Valle
arthur.valle@wintec.ac.nz

Recommender Systems in E-Commerce

The Need for Recommendation

- *"We are leaving the age of information and entering the age of recommendation."*

Chris Anderson

- *"We have 6.2 million customers; we should have 6.2 million stores. There should be the optimum store for each and every customer."*

Jeff Bezos, founder and CEO of Amazon.com in an interview for Business Week during March 1999.

Recommender Systems

- Systems for recommending items (e.g. books, movies, CD's, web pages, newsgroup messages) to users based on their preferences and similarities with other users
- Many on-line services provide recommendations (e.g. Amazon, MovieLens, Youtube, Facebook)

Recommender Systems

- Systems for recommending items (e.g. books, movies, CD's, web pages, newsgroup messages) to users based on their preferences and similarities with other users
- Many on-line services provide recommendations (e.g. Amazon, MovieLens, Youtube, Facebook)

Wintec
WAIKATO INSTITUTE OF TECHNOLOGY
Te Kuratini o Waikato

Recommender Systems

- Systems for recommending items (e.g. books, movies, CD's, web pages, newsgroup messages) to users based on their preferences and similarities with other users
- Many on-line services provide recommendations (e.g. Amazon, MovieLens, Youtube, Facebook)
- Recommender systems have shown great success to substantially increase sales at on-line stores

- Amazon.com generates X percent of their sales through the recommendation lists (30 < X < 70)

- Netflix (DVD rental and movie streaming) generates X percent of their sales through the recommendation lists (30 < X < 70)

Recommendation Approaches

- **Collaborative Filtering (CF)**

Recommendation is based on previously rated data

- **Content-based**

Recommendation is based on the content of items

- **Knowledge-based**

Recommendation is based on the user’s requirements

- **Demographic**

Recommendation is based on user’s demographic information

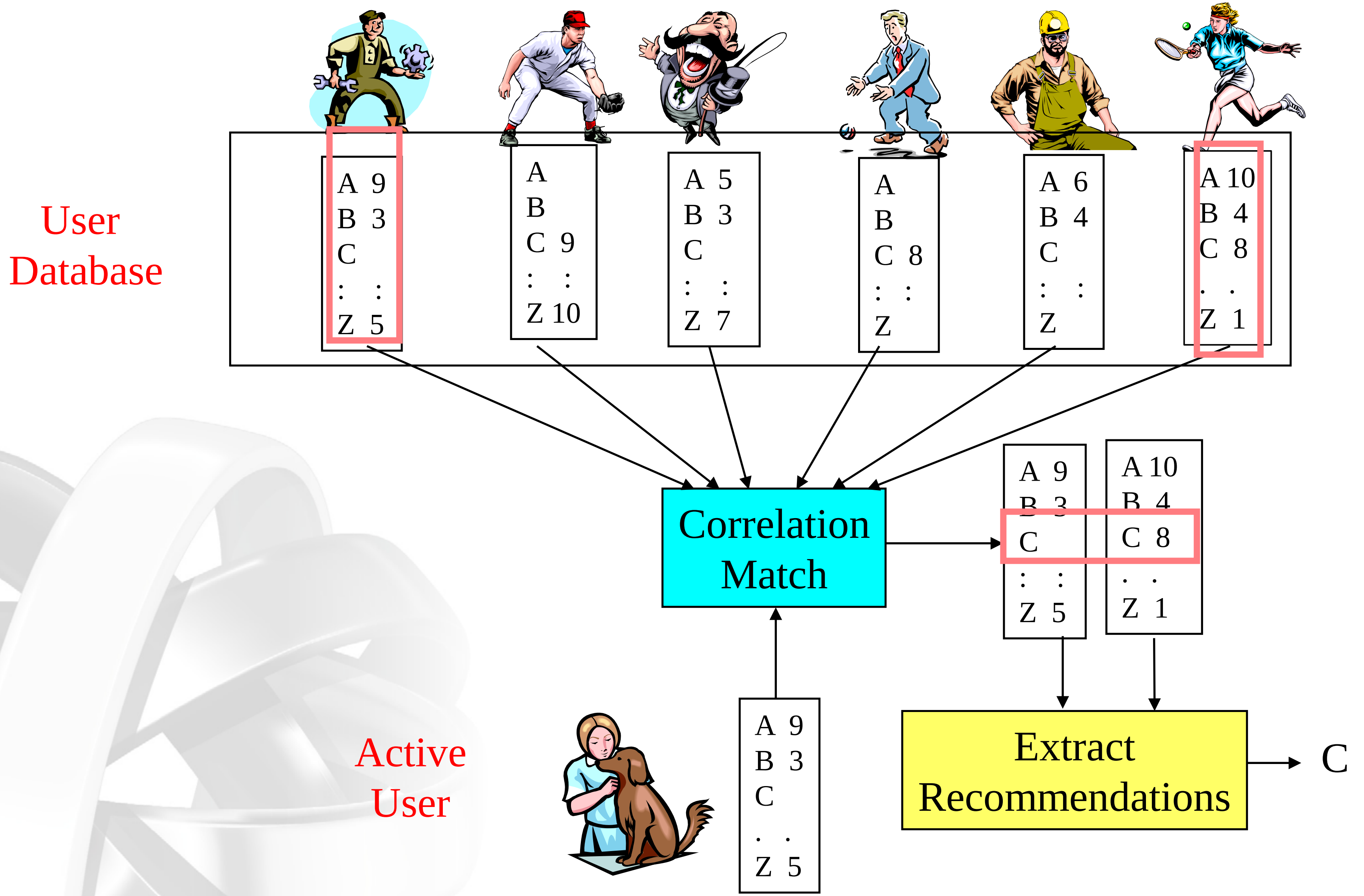
- **Hybrid approaches**

A combination of previous approaches

Collaborative Filtering

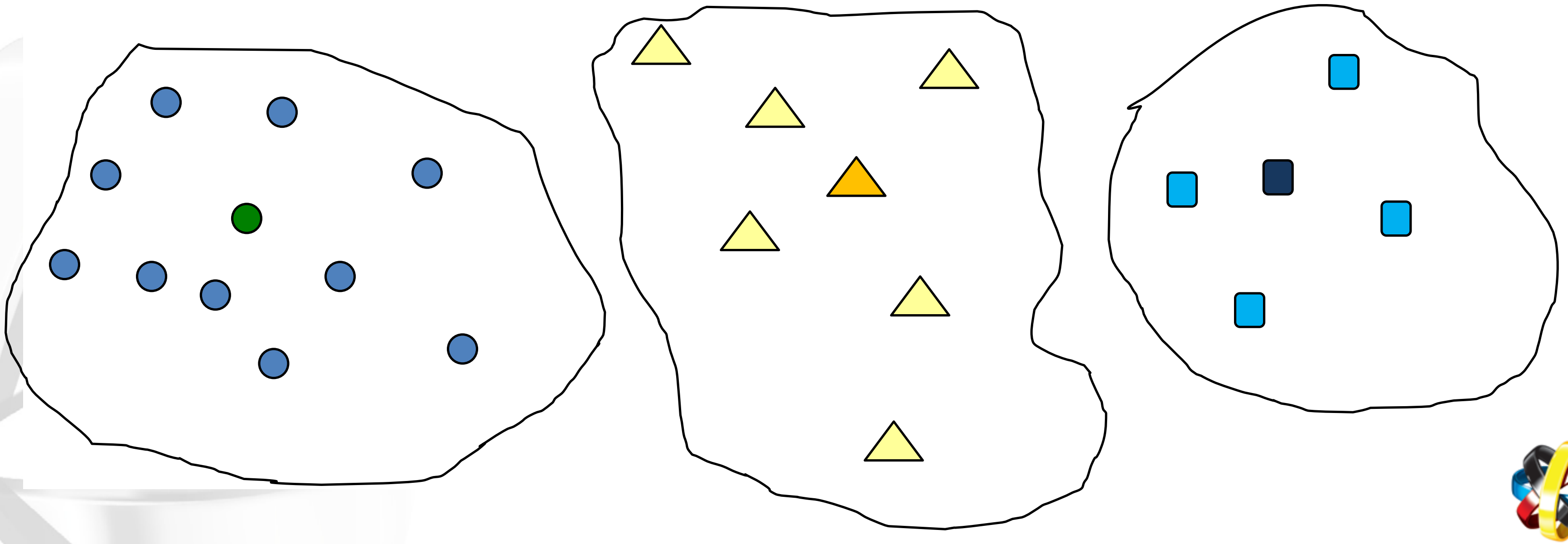
- Maintain a database of many users' ratings of a variety of items.
- For a given user, find other similar users whose ratings strongly correlate with the current user.
- Recommend items rated highly by these similar users, but not rated by the current user.
- Almost all existing commercial recommenders use this approach (e.g. Amazon).

Collaborative Filtering



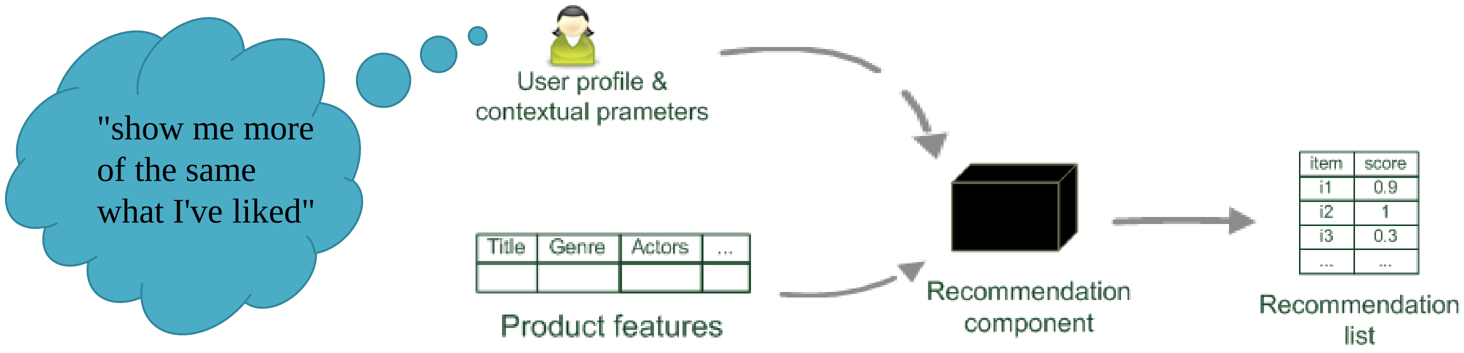
Clustering Users

- Users are clustered based on their similarities
- Choosing suitable cluster-heads has a major impact on the performance



Content-Based Filtering

- Recommendations are based on the content of items rather than on other users' opinions.
- Use machine learning algorithms to induce a profile of the users preferences from examples based on the features describing the content.

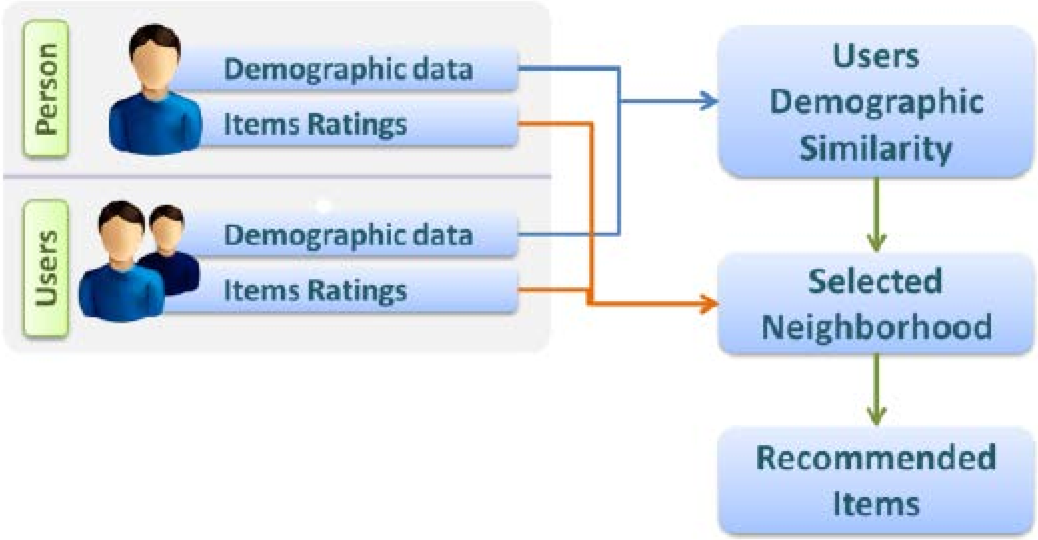


Knowledge-based recommender systems

- Constraint-based
  - based on explicitly defined set of recommendation rules
  - fulfill recommendation rules
- Case-based
  - model past experiences, storing both the problem description and the solution applied in that context
  - retrieve items that are similar to specified requirements
- Both approaches are similar in their recommendation process
  - users specify the requirements
  - systems try to identify solutions
  - if no solution can be found, users change requirements

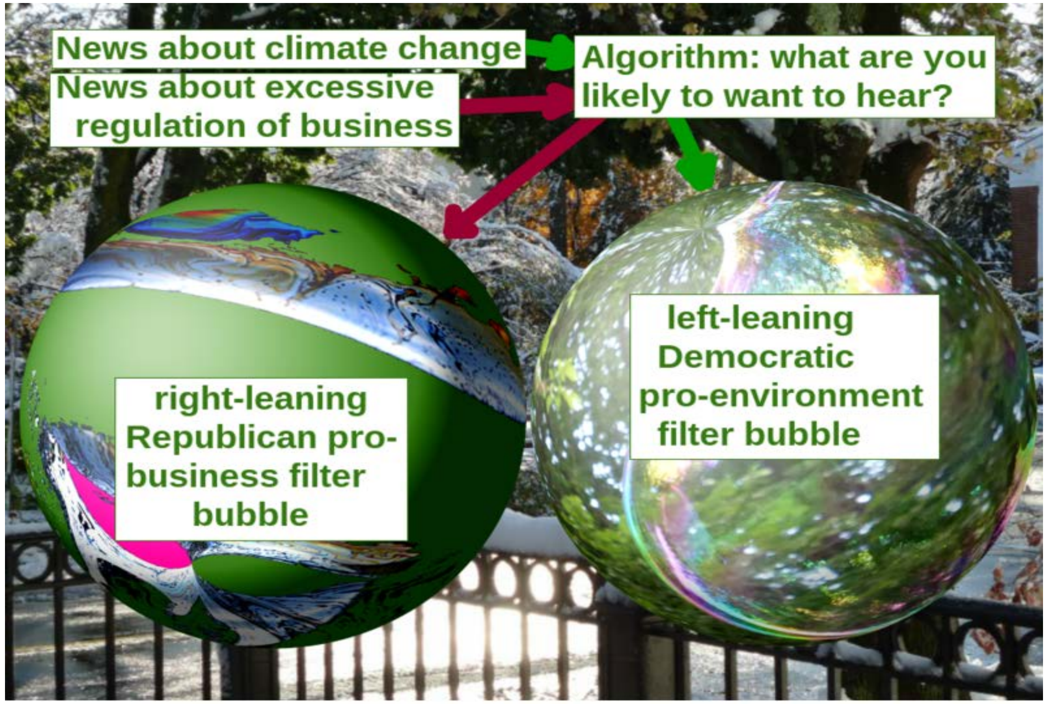
Demographic Recommender Systems

- They use demographic information of users to find similar users. Then, a list of items that have good feedback from similar users are recommended to the target user.



Filter Bubble

- Personal ecosystem of information that's been catered by these algorithms
- Recommender systems close us off to new ideas, subjects, and important information



Recommender Systems in e-commerce

Value for the customer

- Find things that are interesting
- Narrow down the set of choices
- Help to explore the space of options
- Discover new things
- Entertainment

Value for the provider

- Additional and probably unique personalized service for the customer
- Increase trust and customer loyalty
- Increase sales, click trough rates, conversion etc.
- Opportunities for promotion
- Obtain more knowledge about customers

Challenges

- Having huge amounts of data, millions of customers and millions of items
- Recommendations must be offered in real-time
- Cold start: New customers are initially characterized on the basis of limited information
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithms must respond immediately to new information

Challenges

- Having huge amounts of data, millions of customers and millions of items

- The total number of people who use YouTube – 1,300,000,000

- 300 hours of video are uploaded to YouTube every minute

- Almost 5 billion videos are watched on Youtube every single day

- YouTube gets over 30 million visitors per day

- Recommendations must be offered in real-time
- Cold start: New customers are initially characterized on the basis of limited information
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithms must respond immediately to new information

Challenges

- Having huge amounts of data, millions of customers and millions of items
- Recommendations must be offered in real-time
- Cold start: New customers are initially characterized on the basis of limited information
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithms must respond immediately to new information

Thank you!

Reza Rafeh
reza.rafeh@wintec.ac.nz

Industry Research

- Facial recognition for attendance - Kane O’Donnell, **AWARE group**
- VR and AR solutions to daily challenges - Lance Bauerfind, **Pepper Creative**

- **Guss Wilkinson** - Top Poster student award

Research in Action

Emerging Technologies and Trends in IT