Comparison of Machine and Human Recognition of Isolated Instrument Tones

Ichiro Fujinaga
Schulich School of Music
McGill University

Overview

- Introduction
- Exemplar-based learning
    - k-NN classifier
    - Genetic algorithm
- Machine recognition experiments
- Comparison with human performance
- Conclusions

Introduction

"We tend to think of what we 'really' know as what we can talk about, and disparage knowledge that we can't verbalize." (Dowling 1989, 252)

- Western civilization's emphasis on logic, verbalization, and generalization as signs of intelligence
- Limitation of rule-based learning used in traditional Artificial Intelligence (AI) research
- The lazy learning model is proposed here as an alternative approach to modeling many aspects of music cognition

Traditional AI Research

"In AI generally, and in AI and Music in particular, the acquisition of non-verbal, implicit knowledge is difficult, and no proven methodology exists."
(Laske 1992, 259)

- Rule-based approach in traditional AI research
- Exemplar-based learning systems
  - Neural networks (greedy)
  - k-NN classifiers (lazy)
- Adaptive system based on a k-nearest neighbour (k-NN) classifier and a genetic algorithm

Exemplar-based learning

- The exemplar-based learning model is based on the idea that objects are categorized by their similarity to one or more stored examples
- There is much evidence from psychological studies to support exemplar-based categorization by humans
- This model differs both from rule-based or prototype-based (neural nets) models of concept formation in that it assumes no abstraction or generalizations of concepts
- This model can be implemented using k-nearest neighbour (k-NN) classifier and is further enhanced by application of a genetic algorithm

Applications of lazy learning model

- **Optical music recognition** (Fujinaga, Pennycook, and Alphonce 1989; MacMillan, Droettboom, and Fujinaga 2002)
- **Vehicle identification** (Lu, Hsu, and Maldague 1992)
- **Pronunciation** (Cost and Salzberg 1993)
- **Cloud identification** (Aha and Bankert 1994)
- **Respiratory sounds classification** (Sankur et al. 1994)
- **Wine analysis and classification** (Latorre et al. 1994)
- **Robot scene analysis** (Schaal et al. 2002)
- **Natural language translation** (Sato 1995)
- **Tomato classification** (Indriani et al. 2017)

Implementation of lazy learning

- The lazy learning model can be implemented by the k-nearest neighbour classifier (Cover and Hart 1967)
- A classification scheme to determine the class of a given sample by its feature vector
- The class represented by the majority of k-nearest neighbours (k-NN) is then assigned to the unclassified sample
- Besides its simplicity and intuitive appeal, the classifier can be easily modified, by continually adding new samples that it "encounters" into the database, to become an incremental learning system
- Criticisms: slow and high memory requirement

K-nearest neighbour classifier

"The nearest neighbor algorithm is one of the simplest learning methods known, and yet no other algorithm has been shown to outperform it consistently." (Cost and Salzberg 1993)

- The K-NN classifier is the simplest of all machine learning classifiers
- It is based on the principle that things that are similar, are close by

K-nearest neighbour classifier

"*Many sophisticated classification algorithms have been proposed... According to our experiments on the popular datasets, k-NN with properly tuned parameters performs on average best.*"
(Kordos, Blachnik & Strzempa 2010)

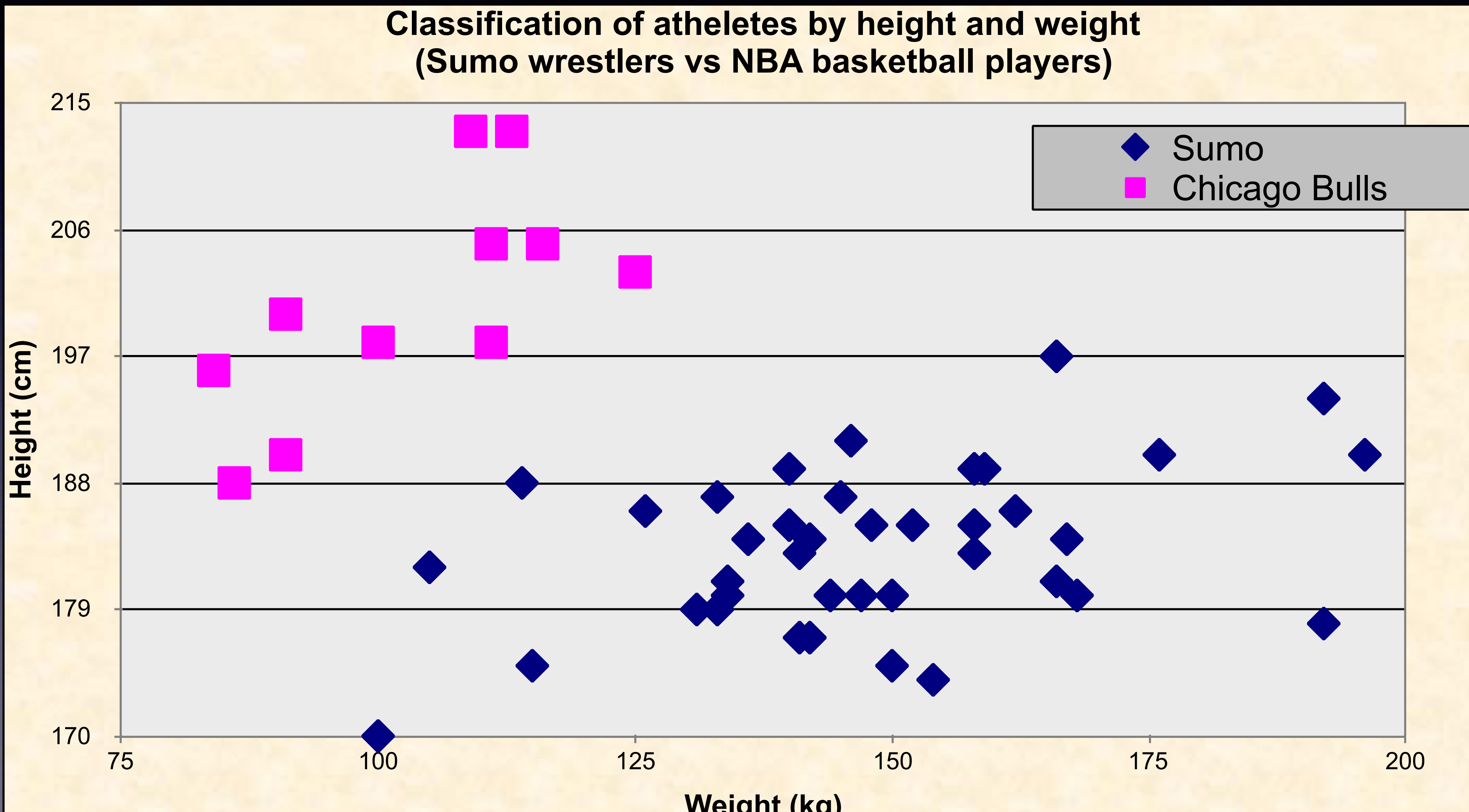
- Determine the class of a given sample by its feature vector:
    - Distances between feature vectors of an unclassified sample and previously classified samples are calculated
    - The class represented by the majority of k-nearest neighbours is then assigned to the unclassified sample

An example of k-NN classifier
Basketball players and Sumo wrestlers

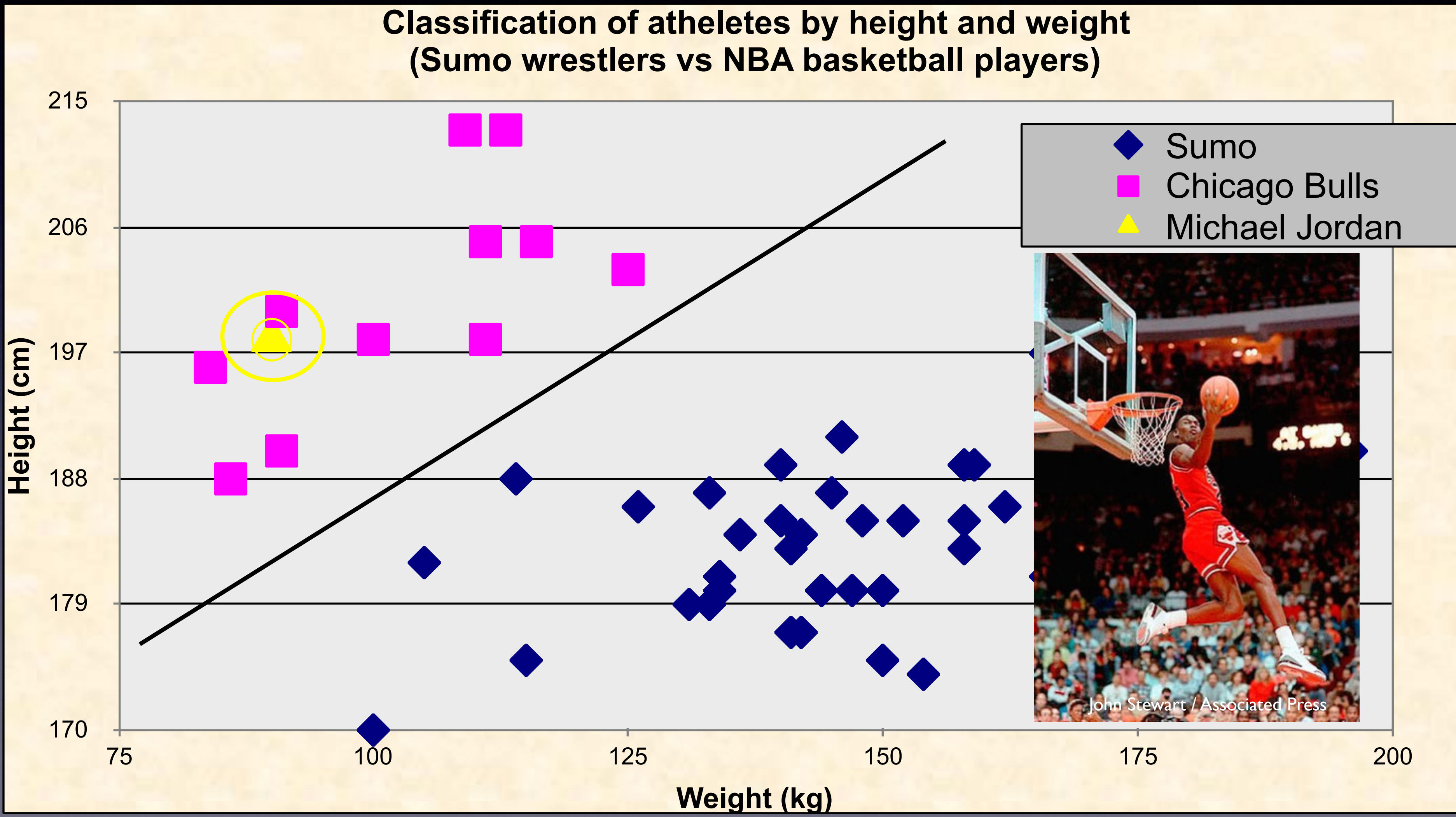
https://www.flickr.com/photos/29650319@N06/3172412470


http://blogs.yahoo.co.jp/noa_kamiya/GALLERY/show_image.html?id=24844519&no=13

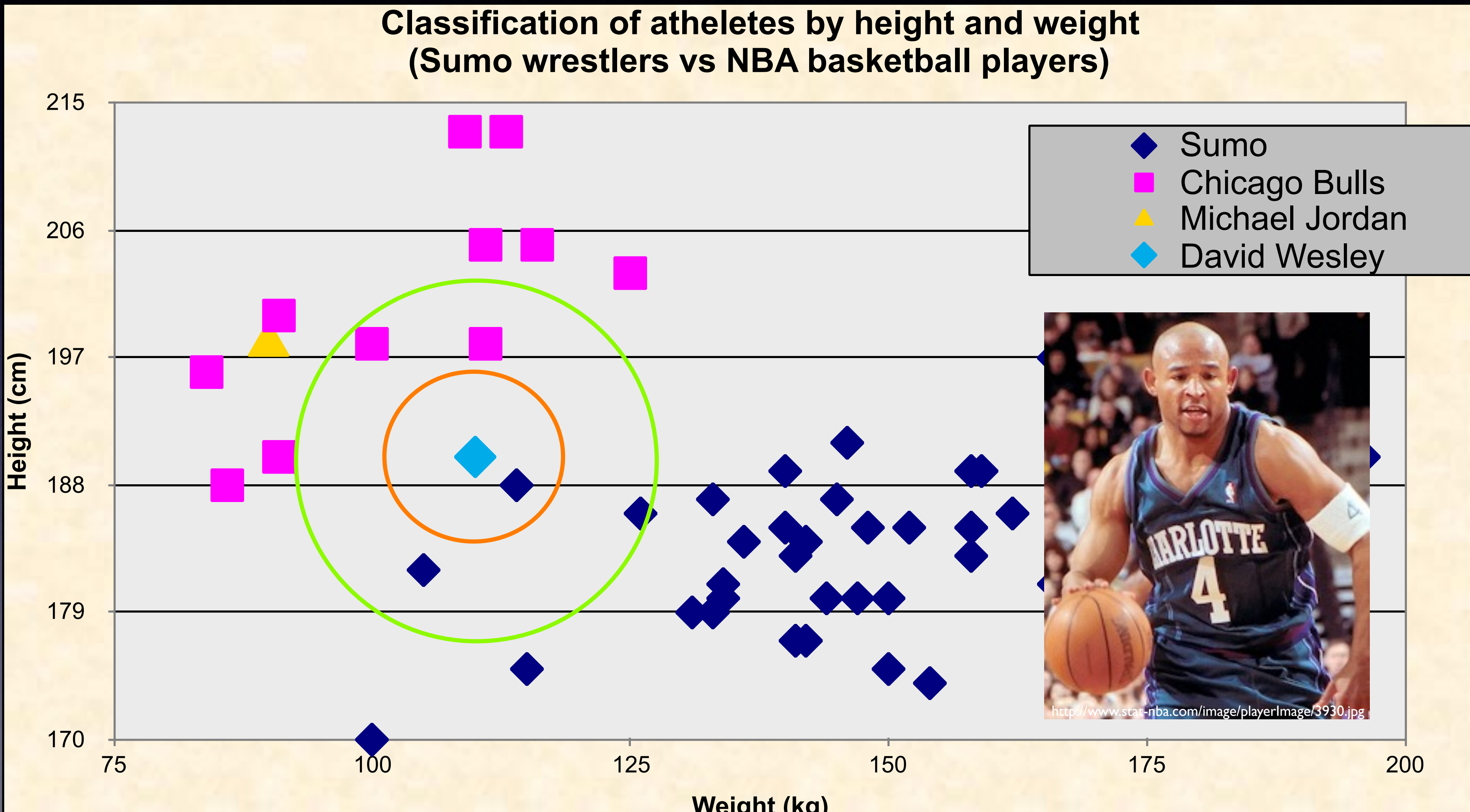
An example of k-NN classifier



Example of k-NN classifier
Classifying Michael Jordan

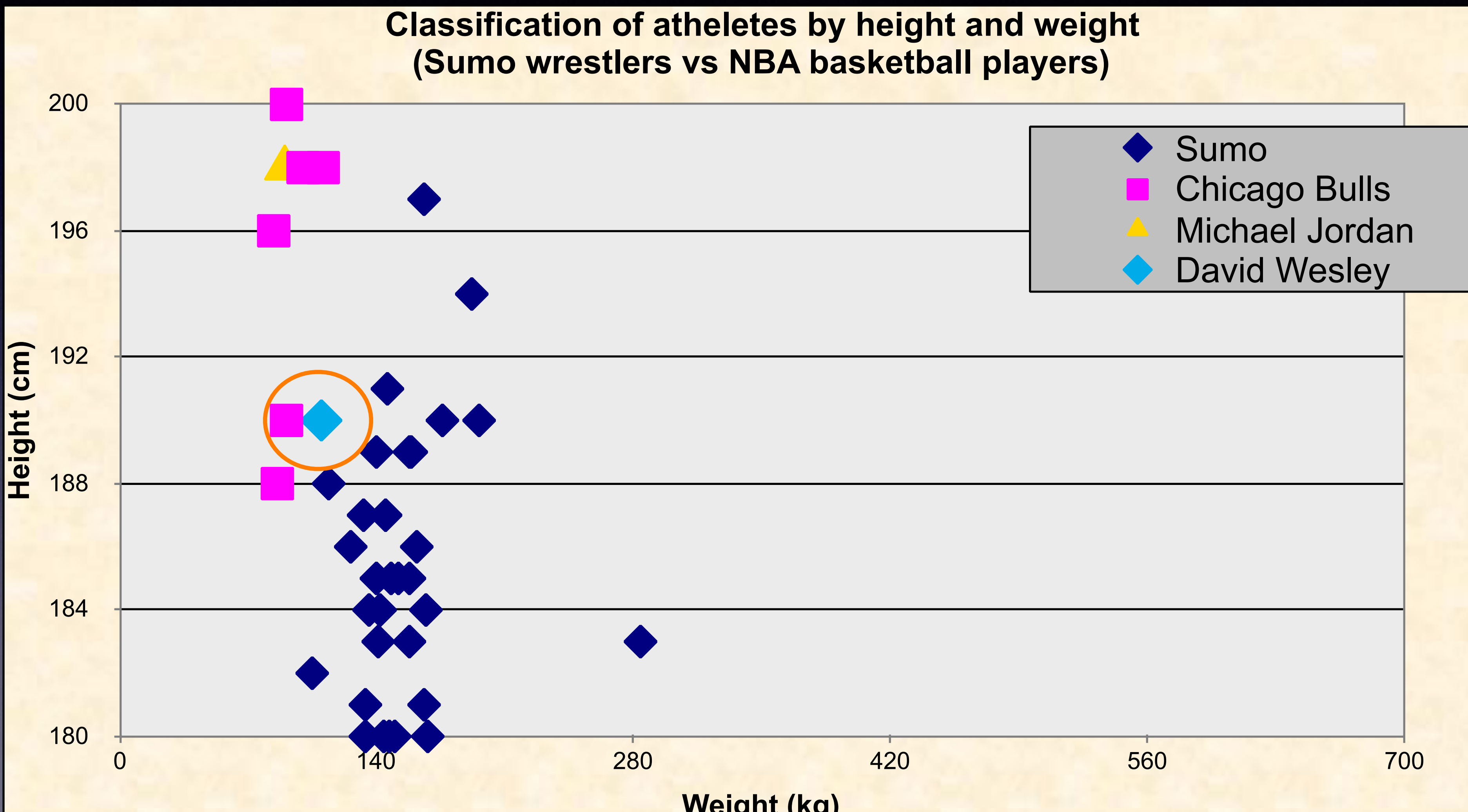


Example of k-NN classifier
Classifying David Wesley



Example of k-NN classifier

Reshaping the Feature Space



Distance measures

- The distance in a *N*-dimensional feature space between two vectors *X* and *Y* can be defined as:

$$\boxed{d = \sum_{i=0}^{N-1} \left| x_i - y_i \right|}$$

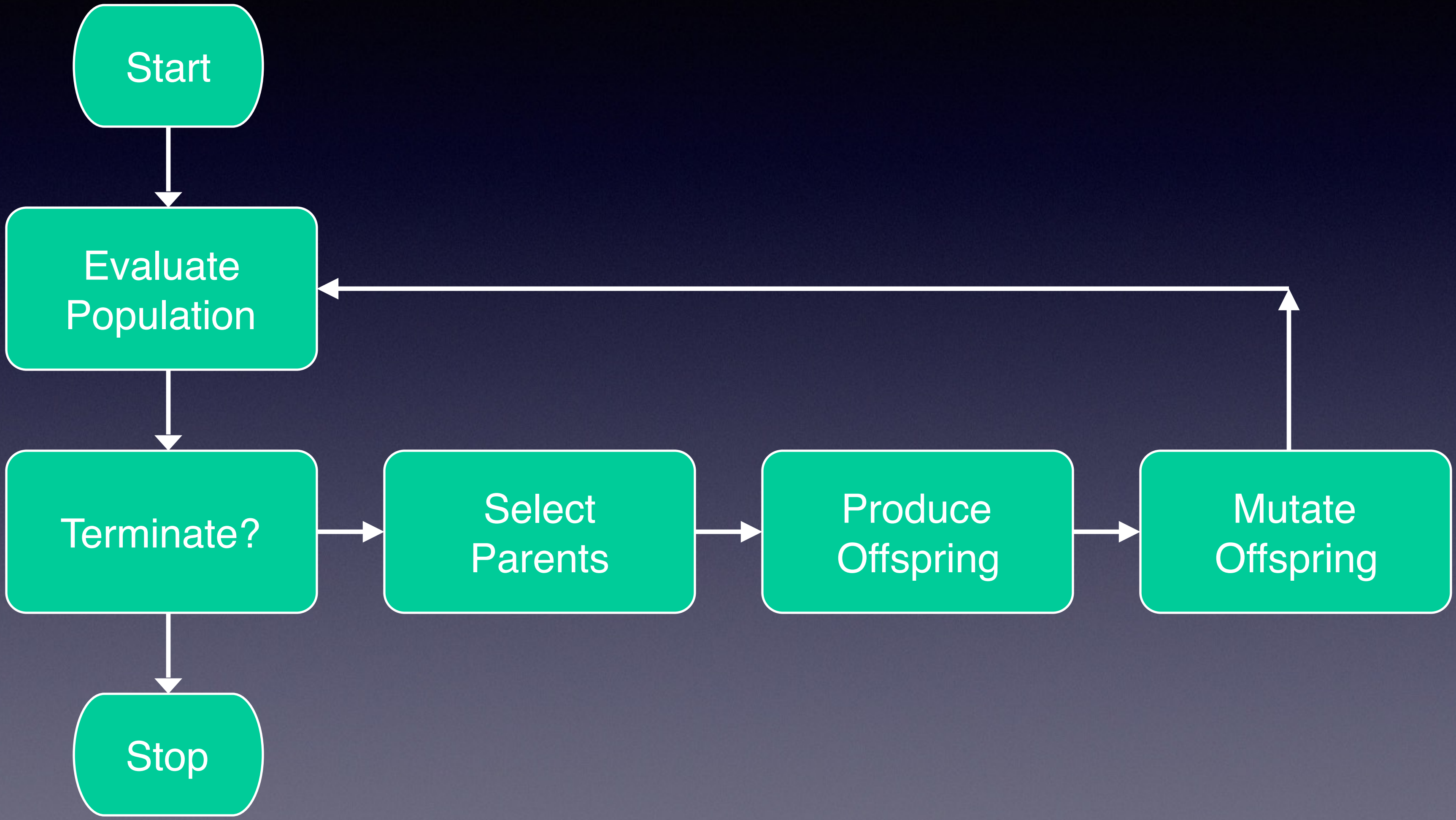
- A weighted distance can be defined as:

$$\boxed{d = \sum_{i=0}^{N-1} w_i \left| x_i - y_i \right|}$$

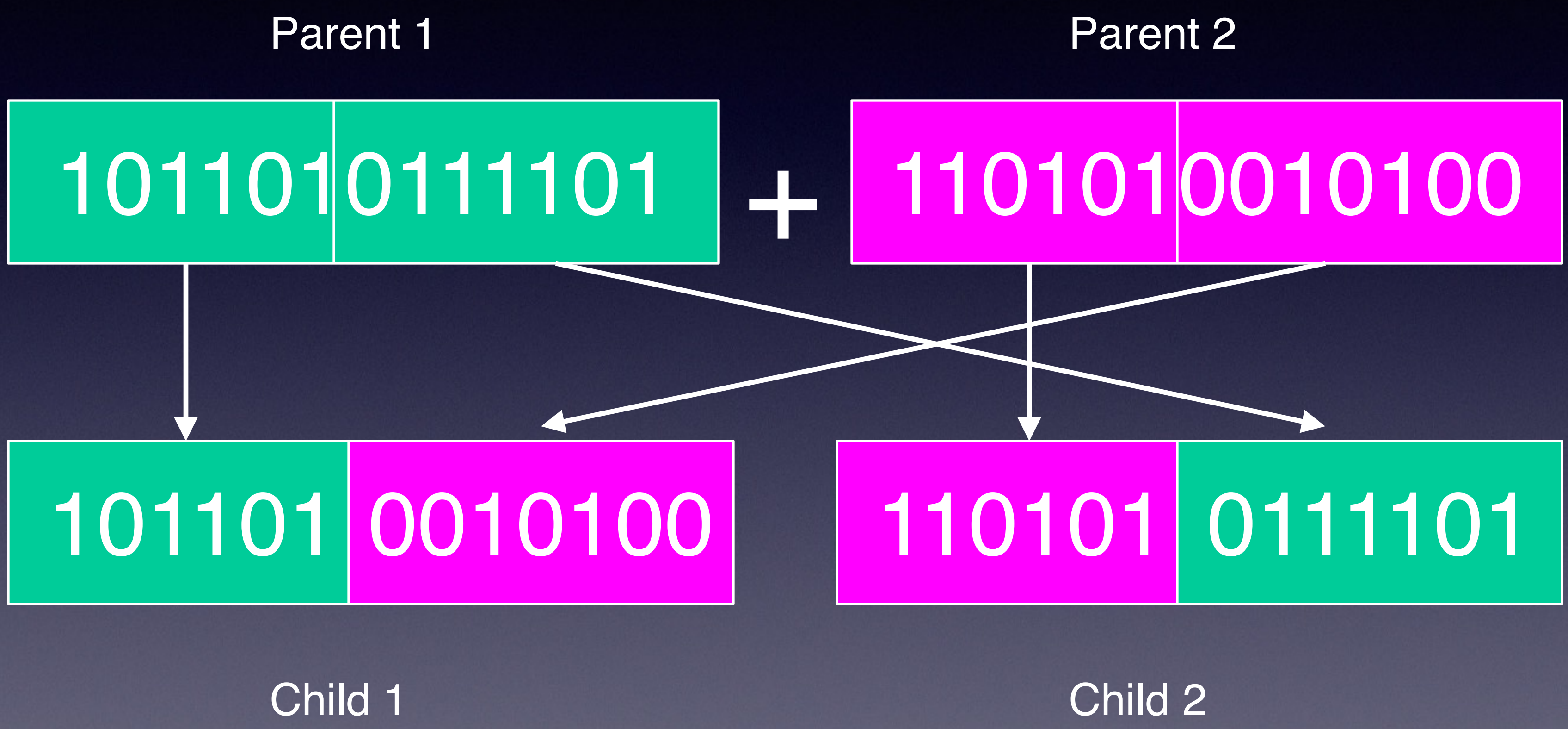
Genetic algorithms

- Optimization based on biological evolution
- Maintenance of population using selection, crossover, and mutation
- Chromosomes = weight vector
- Fitness function = recognition rate
- Leave-one-out cross validation

Genetic Algorithm



Crossover in Genetic Algorithm



Applications of Genetic Algorithm in Music

- **Instrument design** (Horner *et al.* 1992, Horner et al. 1993, Takala *et al.* 1993, Vuori and Välimäki 1993, Poirson et al. 2007)
- **Compositional aid** (Horner and Goldberg 1991, Biles 1994, Johanson and Poli 1998, Wiggins 1998, Geem et al. 2001)
- **Expressive music performance** (Ramirez and Hazan 2005)
- **Granular synthesis regulation** (Fujinaga and Vantomme 1994)
- **Optimal placement of microphones** (Wang 1996)

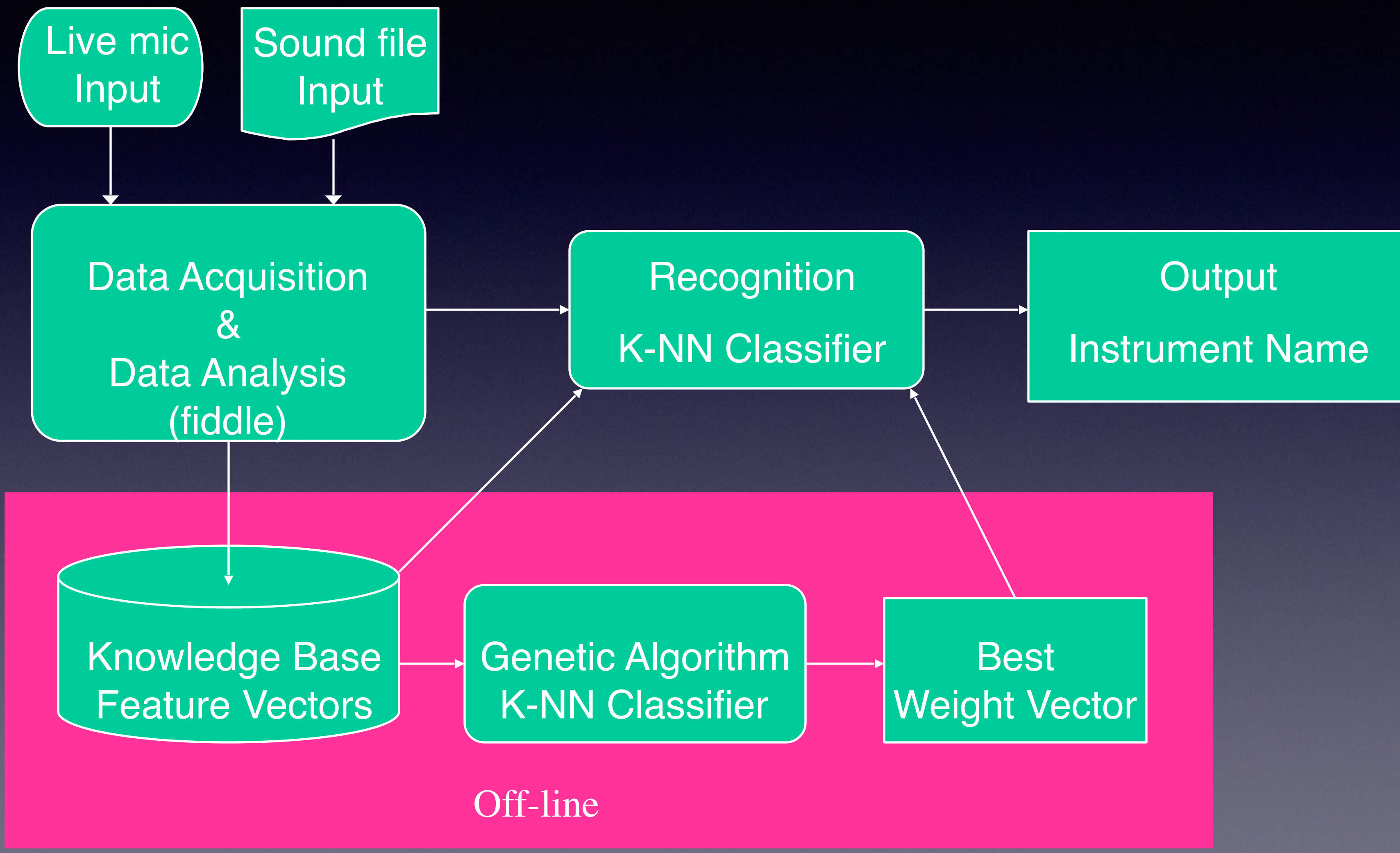
Realtime Timbre Recognition

- Original source: McGill Master Samples
- Up to over 1300 notes from 39 different timbres (23 orchestral instruments)
- Spectrum analysis of first 232ms of attack (9 overlapping windows)
- Each analysis window (46 ms) consists of a list of amplitudes and frequencies in the spectra

Features

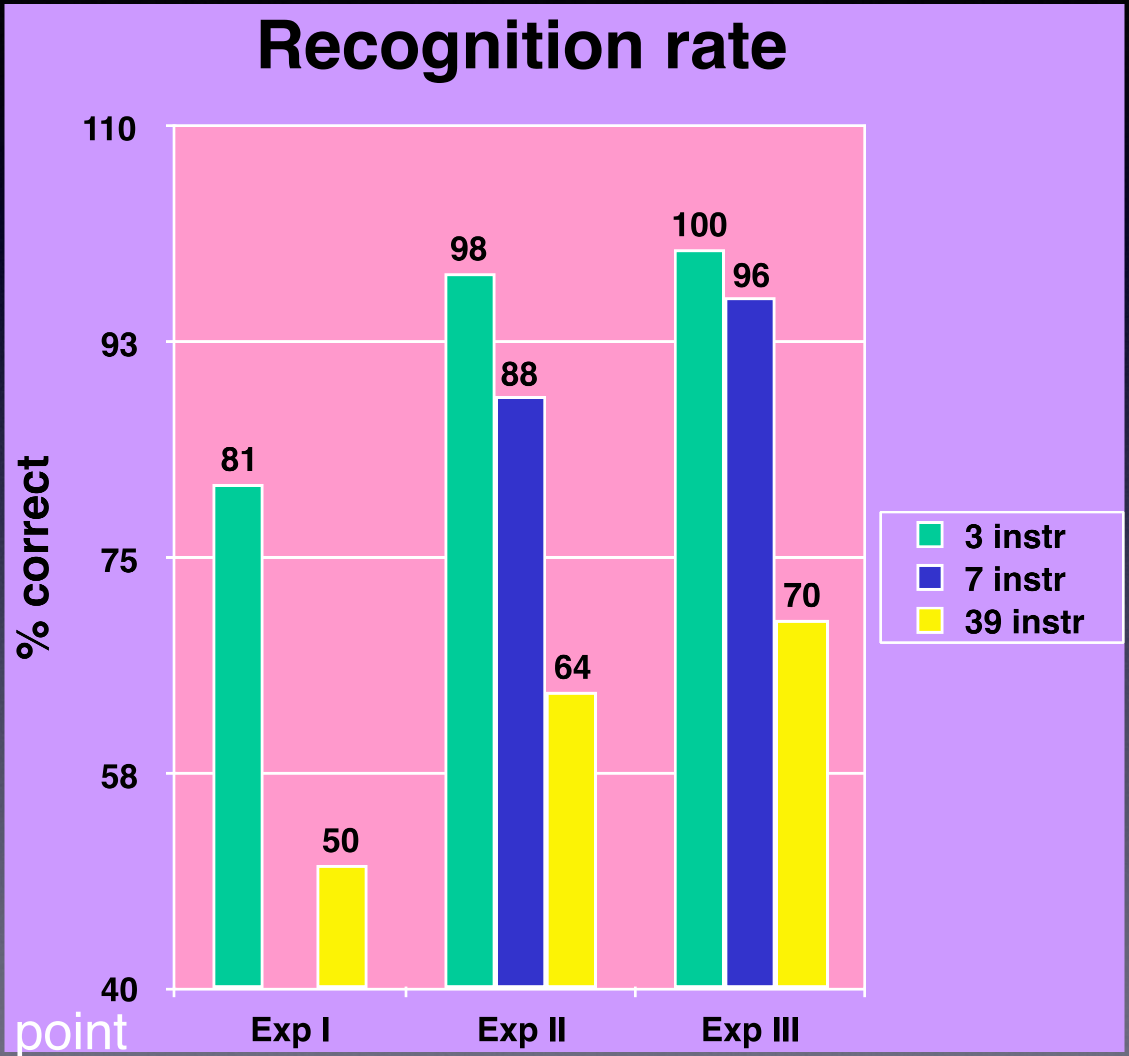
- Static features (per window)
  - pitch
  - mass or the integral of the curve (zeroth-order moment)
  - centroid (first-order moment)
  - variance (second-order central moment)
  - skewness (third-order central moment)
  - amplitudes of the harmonic partials
  - number of strong harmonic partials
  - spectral irregularity
  - tristimulus
- Dynamic features
  - means and velocities of static features over time

Overall Architecture for Timbre Recognition

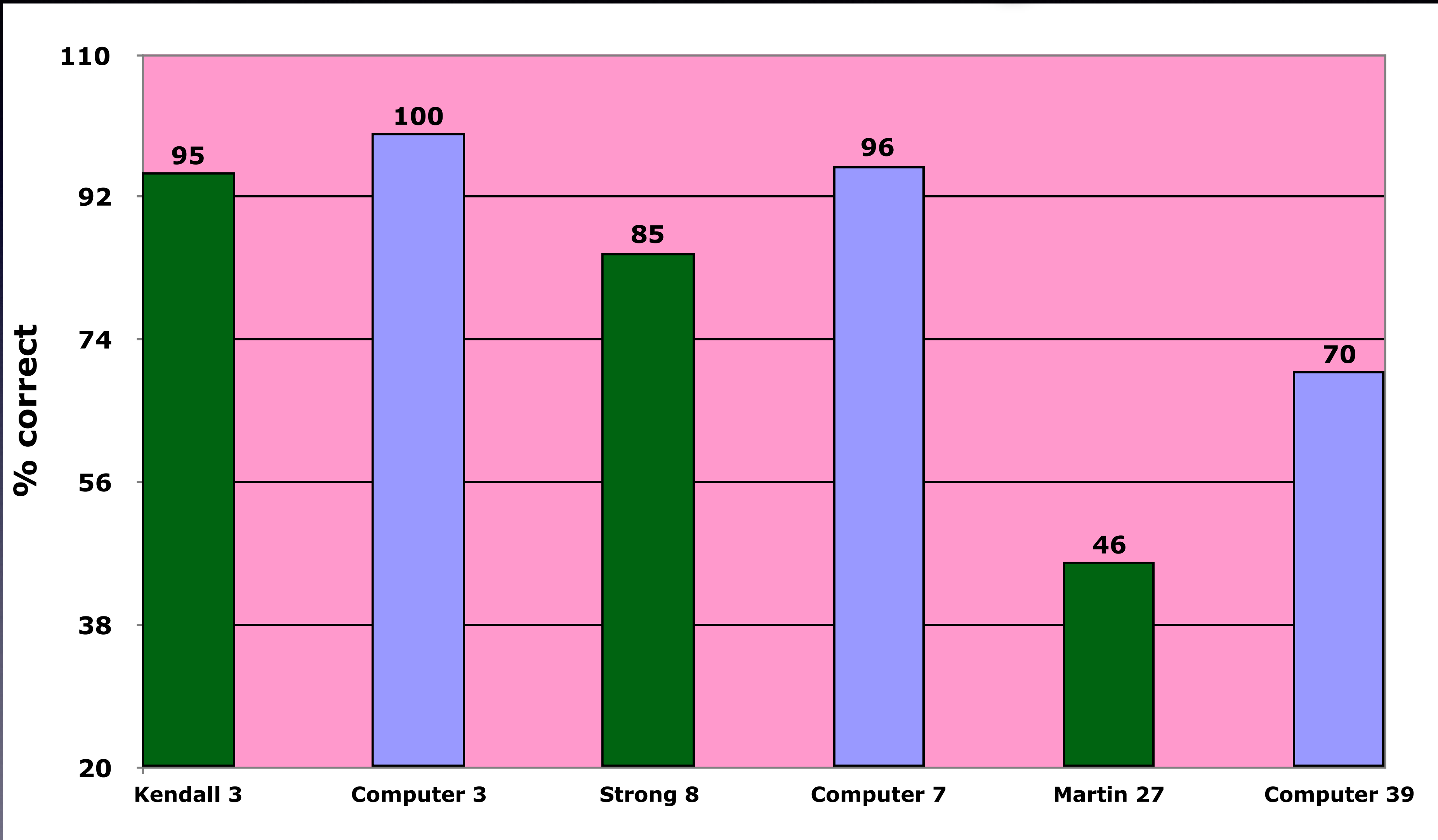


Results

- Experiment I
  - SHARC data
  - static features
- Experiment II
  - McGill samples
  - *Fiddle*
  - dynamic features
- Experiment III
  - more features
  - redefinition of attack point



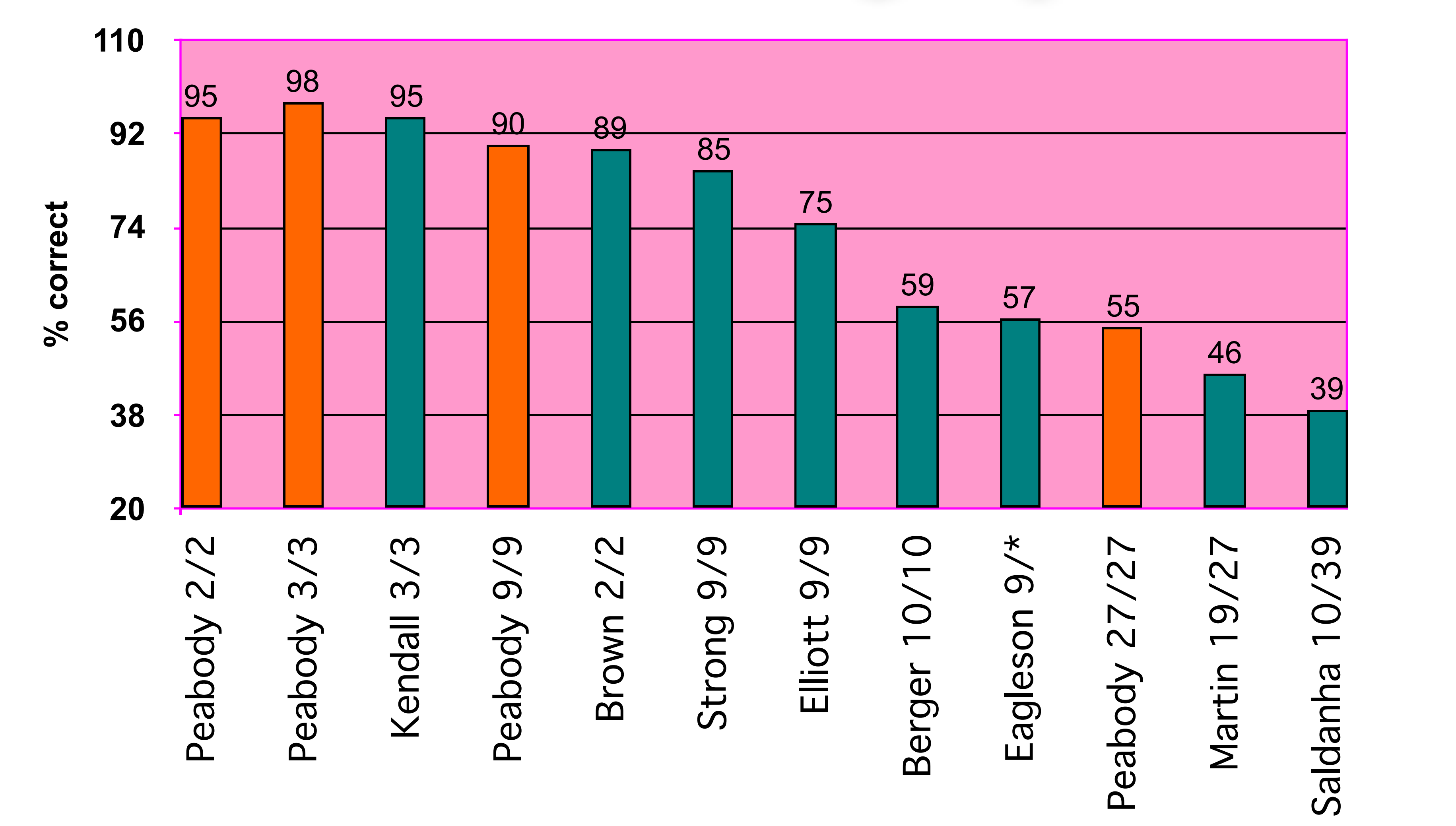
Human vs Computer



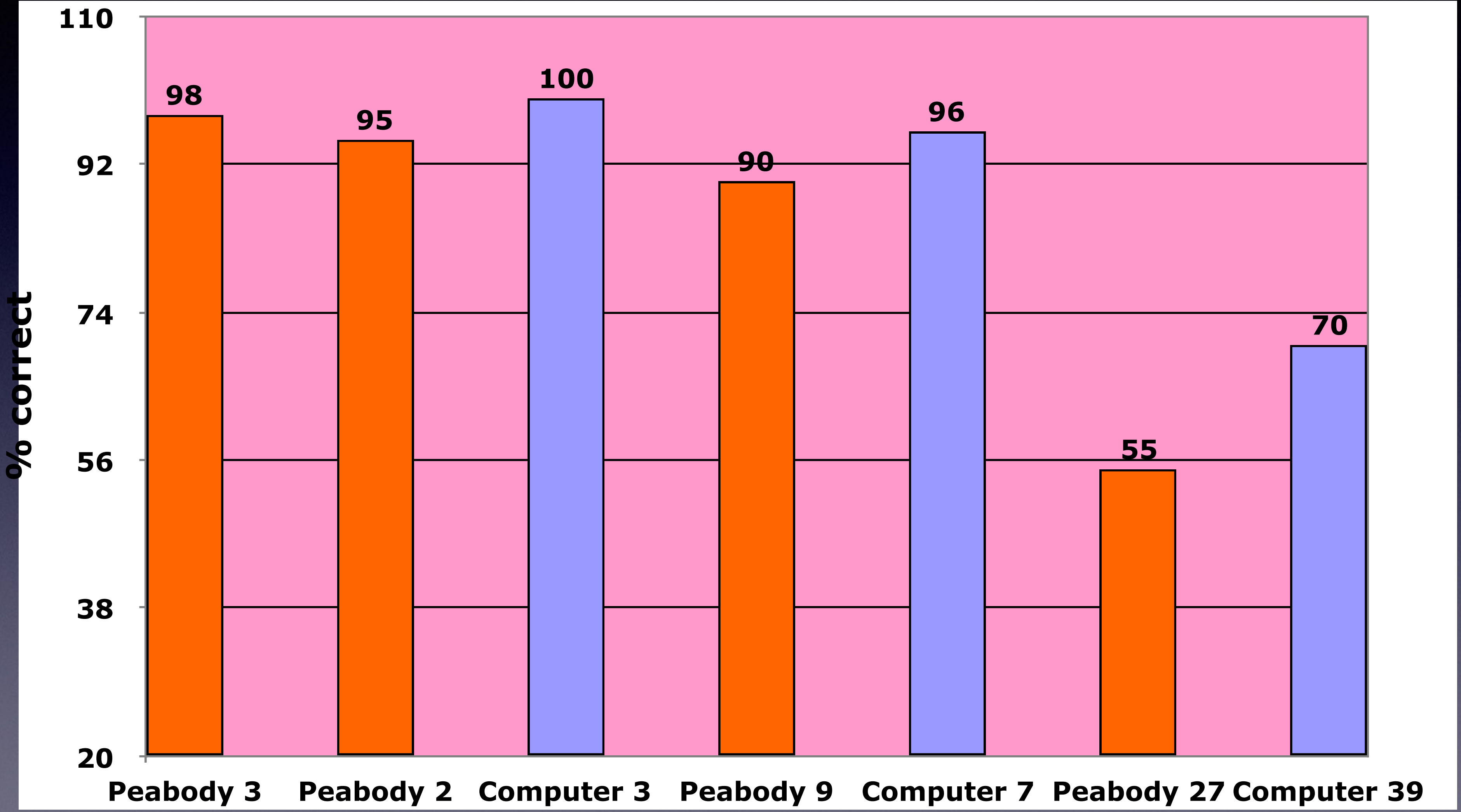
Peabody experiment

- 88 subjects (undergrad, composition students and faculty)
- Source: McGill Master Samples
- 2-instruments (oboe, saxophones)
- 3-instruments (clarinet, trumpet, violin)
- 9-instruments (flute, oboe, clarinet, bassoon, saxophone, trombone, trumpet, violin, cello)
- 27-instruments:
  - violin, viola, cello, bass
  - piccolo, flute, alto flute, bass flute
  - oboe, english horn, bassoon, contrabassoon
  - Eb clarinet, Bb clarinet, bass clarinet, contrabass clarinet
  - saxes: soprano, alto, tenor, baritone, bass
  - trumpet, french horn, tuba
  - trombones: alto, tenor, bass

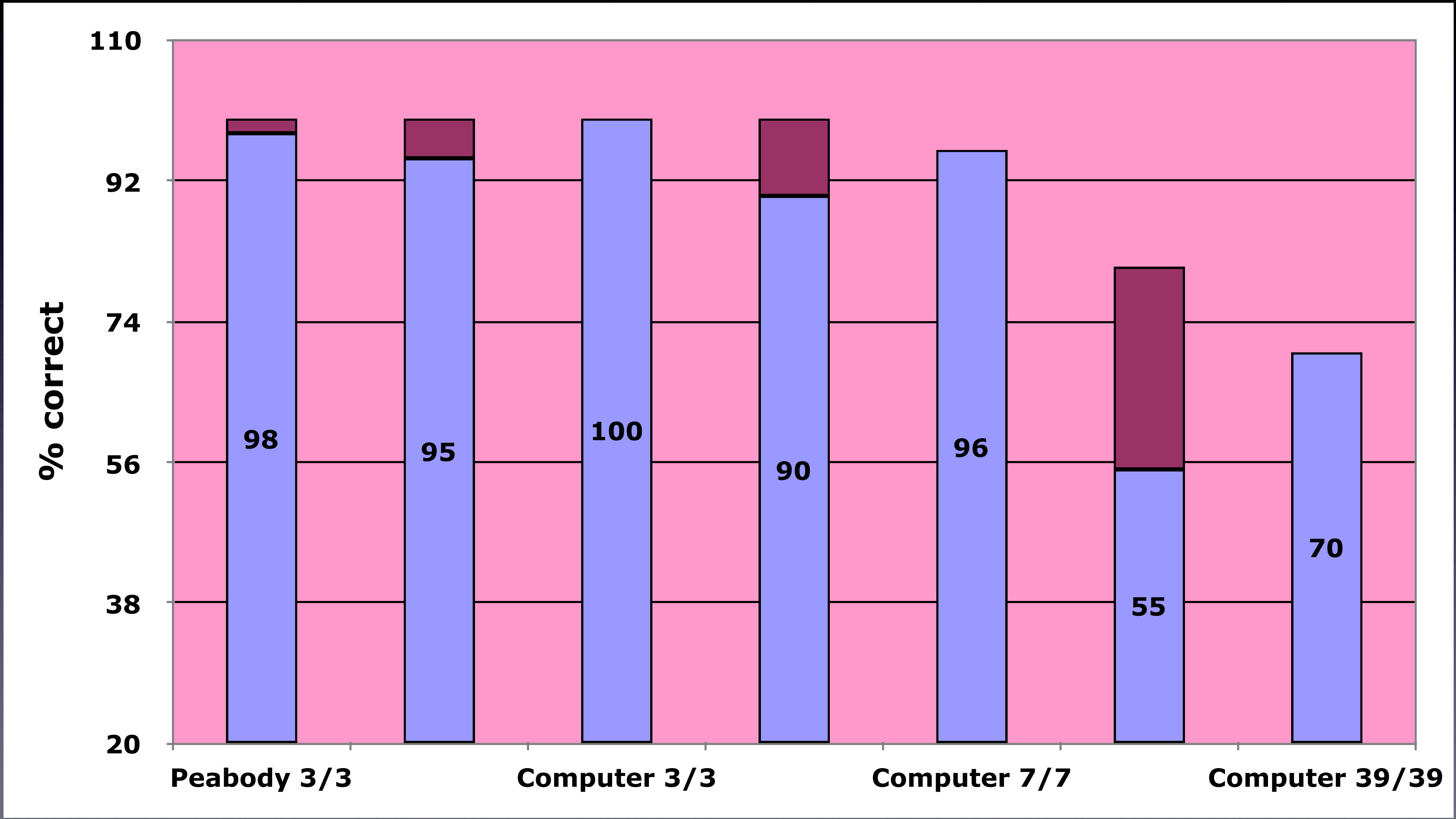
Peabody vs other human groups



Peabody subjects vs Computer



The best Peabody subjects vs Computer



Future Research for Timbre Recognition

- Performer identification
- Speaker identification
- Tone-quality analysis
- Multi-instrument recognition
- Expert recognition of timbre

Conclusions

- Realtime adaptive timbre recognition by k-NN classifier enhanced with genetic algorithm
- A successful implementation of the exemplar-based learning system in a time-critical environment
- Recent human experiments poses new challenges for machine recognition of isolated tones

Recognition rate for different lengths of analysis window