Using a children's gameshow to study iterated learning and the emergence of combinatoriality

Jon W. Carr
Language Evolution and Computation Research Unit
University of Edinburgh

Duality of patterning

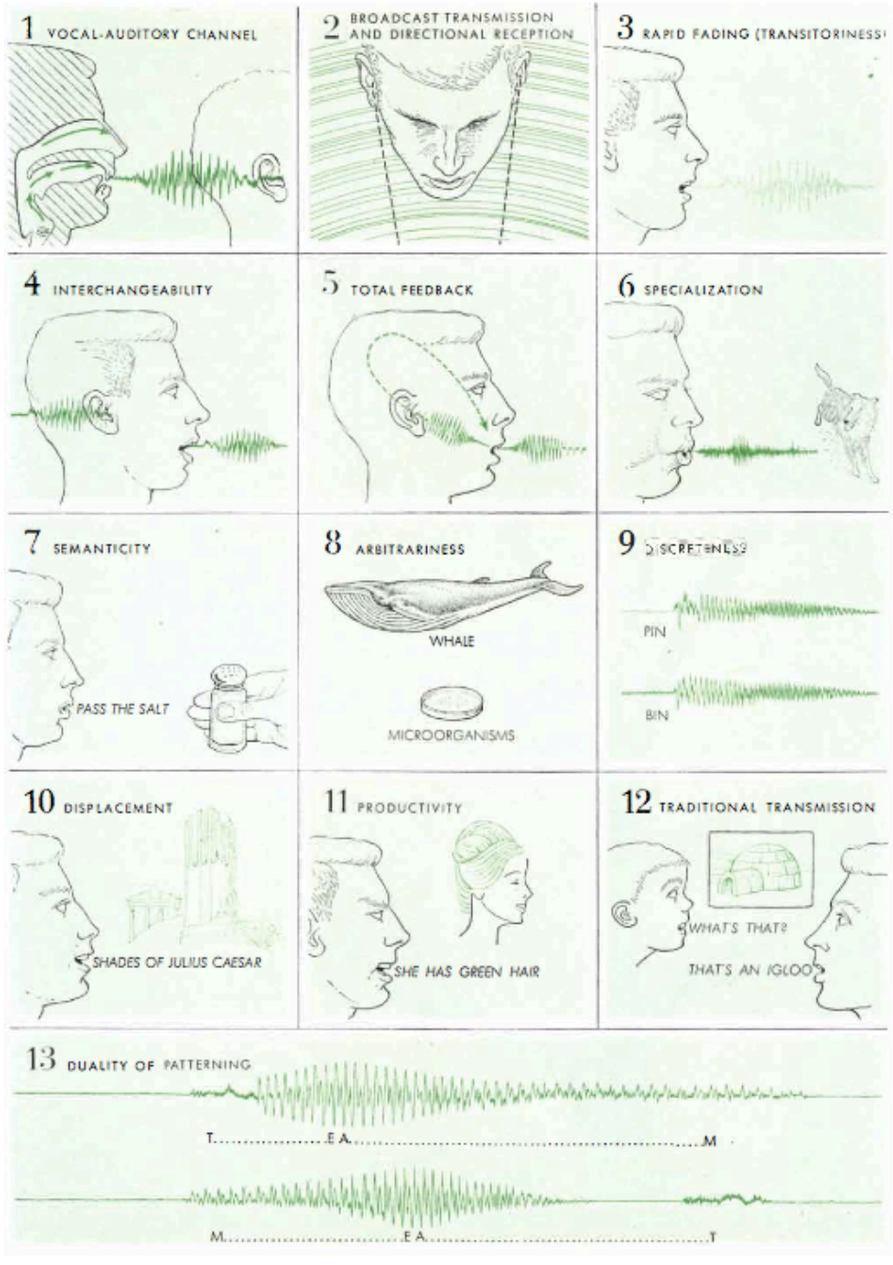
Hockett's classic article on the design features of language

The last feature on the list is **duality of patterning**, supposedly the feature that is specific to humans

Compositionality: speech is composed of **meaningful** recombinable units

Combinatoriality: words are composed of **meaningless** recombinable units

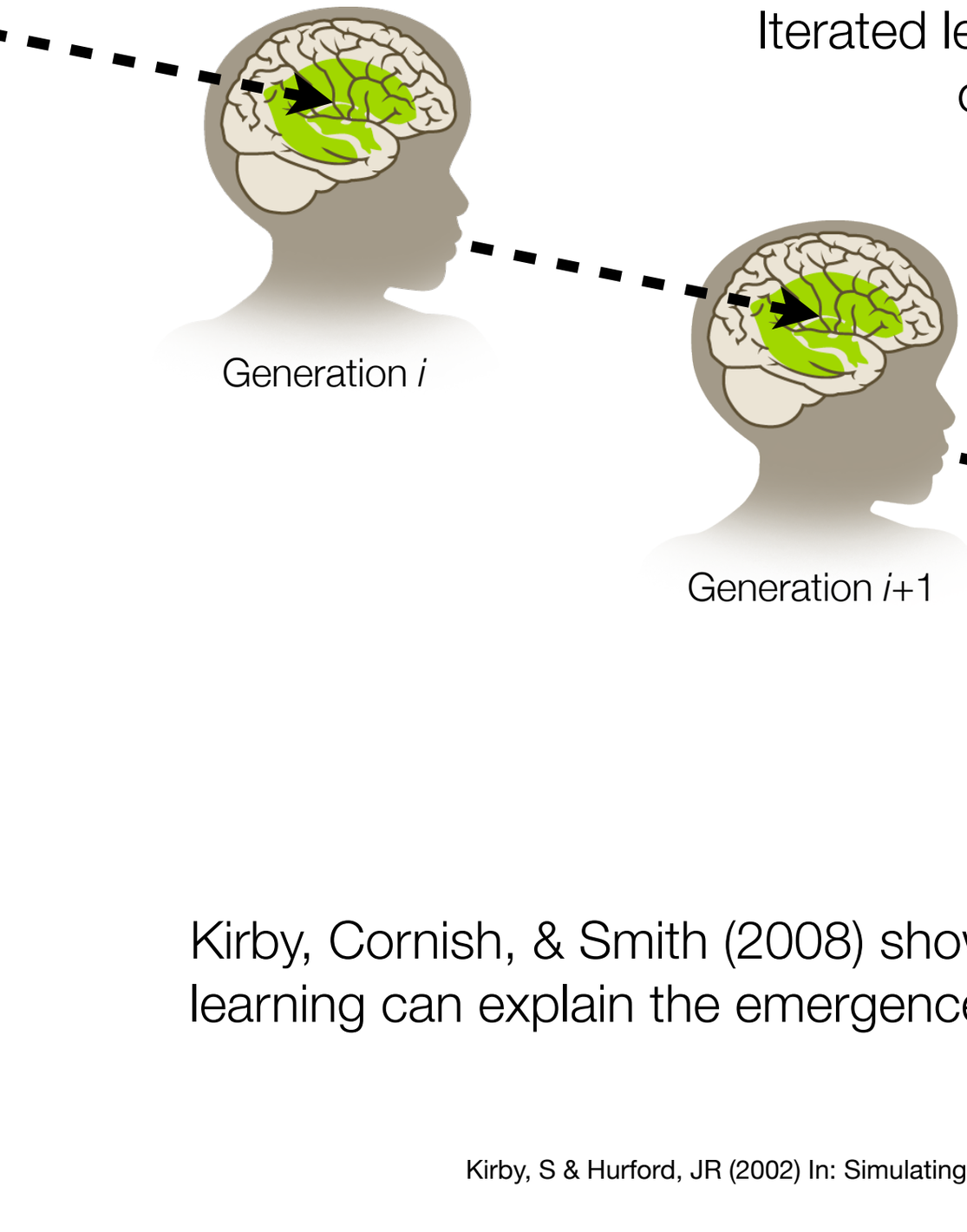
Both can be explained by **iterated learning**



Hockett, CF (1960) Sci Am, 203

Iterated learning

Iterated learning: languages adapt to the cognitive biases of their learners as they are culturally transmitted



Kirby, Cornish, & Smith (2008) showed that iterated learning can explain the emergence of compositionality

Kirby, S & Hurford, JR (2002) In: Simulating the evolution of language • Kirby, S, Cornish, H, & Smith, K (2008) Proc Natl Acad Sci, 105

Verhoef's slide whistle experiment

Participants had to learn an artificial whistled "language", and then reproduce it from memory.

These reproductions are used as training data for another participant.

After ten iterations the language began to exhibit combinatorial structure.

The "words" in the language begin to use a finite set of discrete recombinable units.

Together with Kirby et al. (2008), iterated learning can explain the emergence of both compositionality and combinatoriality.

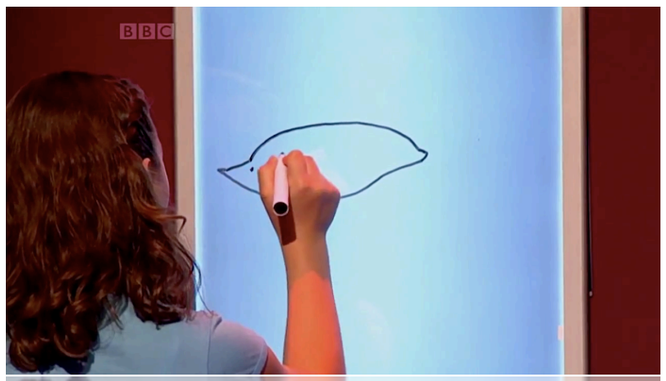
Verhoef, T (2012) Lang Cogn, 4

CBBC gameshow broadcast since November 2009

Three series, each with 52 episodes

Each episode pits two teams against each other in Chinese Whisper's based games

Teams are made up of six players, usually members of a family



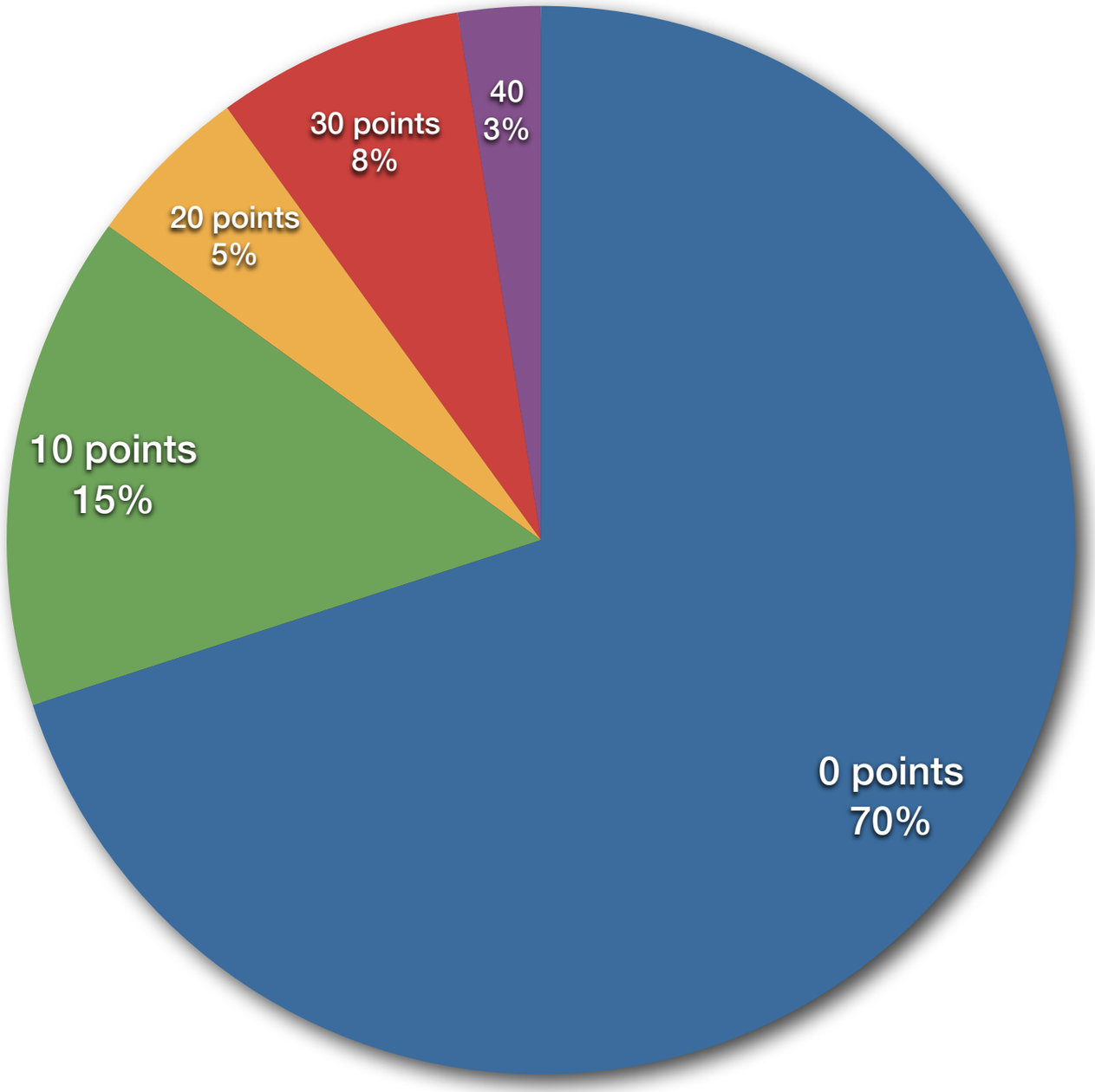
Quick on the draw

Mime time

The music round


BBC
the musical
10
20
30
40
50
1
2
3
4
5
6

Points scored



based on 40 teams

Benefits of the dataset

Cheap!

Large size – 312 chains, 1560 players

Pressure for faithful replication

More natural setup – participants are not locked away in some weird lab

Mathematical models

Computational models

Experimental models

Observational data

Similar setup to Verhoef (2012) – preexisting methods of analysis

Data is there – why not use it?

Limitations of the dataset

Initial input is already structured

Lack of experimental control

Data collection is constrained by the BBC's schedule

Noise – e.g. laughter from audience

Short chains of just 5 generations – may not be long enough to observe interesting phenomena

Prior experience of music – expectation of pop song

Reinterpretation based on prior experience

Players expect pop songs

Thus, emergent structure could be explained by players' memory of songs

Hypotheses

Hypothesis 1: As the songs are culturally transmitted they will tend to become easier to replicate. Learnability increases.

Hypothesis 2: As the songs are culturally transmitted they will tend to become more predictable by relying on a set of discrete recombinable units. Combinatoriality increases.

Data collection

Play episode on BBC iPlayer

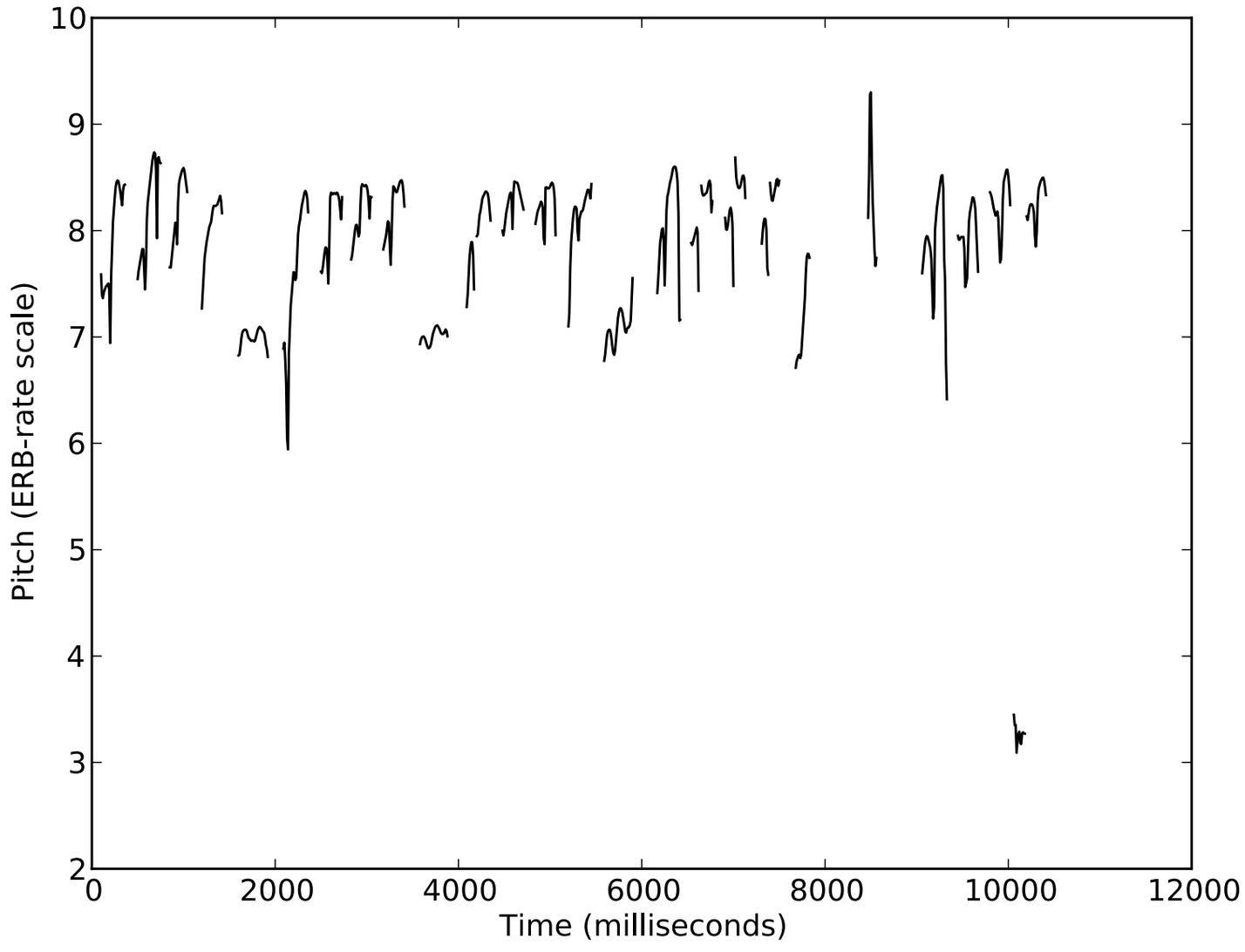
Capture audio using Audio Hijack Pro

Isolate songs and remove noise using Audacity

Convert the songs into pitch tracks using Praat

bbc.co.uk/iplayer/ • rogueamoeba.com/audiohijackpro/ • audacity.sourceforge.net • www.fon.hum.uva.nl/praat/

Data collection



bbc.co.uk/iplayer/ • rogueamoeba.com/audiohijackpro/ • audacity.sourceforge.net • www.fon.hum.uva.nl/praat/

Measuring learnability

Compute the derivative dynamic time warping (**DDTW**) distance between consecutive players' songs

This quantifies the transmission error between two players' songs

Computed for each set of consecutive players

Transmission error is expected to fall over time as learnability increases

Sakoe, H & Chiba, S (1978) IEEE T Acoust Speech, 26 • Keogh, EJ & Pazzani, MJ (2001) 1st SIAM Internat Conf Data Mining

Measuring combinatoriality – clustering

Segment pitch track. Segments indicated by:

– period of noise bounded by silence

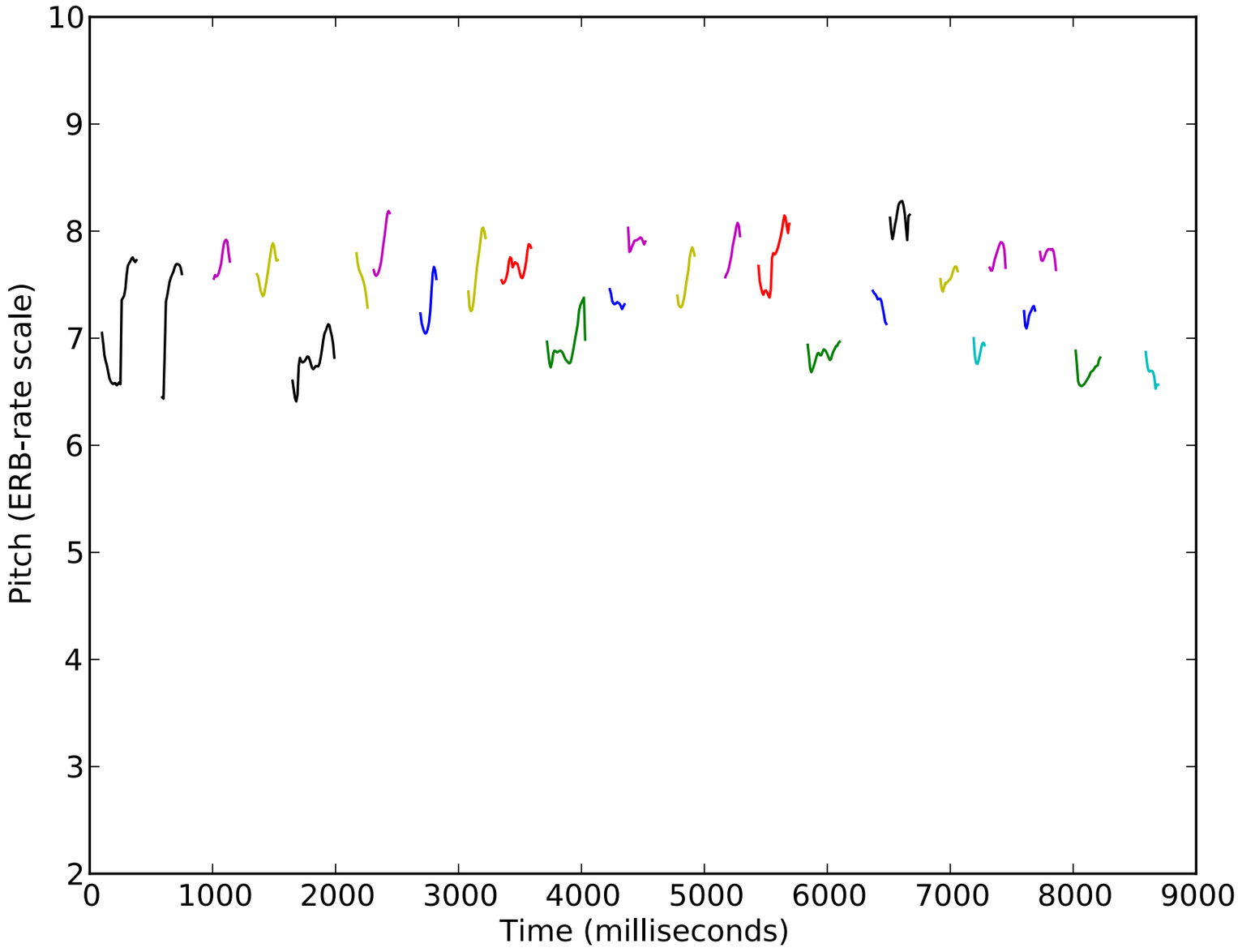
– a sudden dramatic change in pitch

Cluster segments based on their similarity (using **DTW** as distance metric)

Average linkage agglomerative hierarchical clustering

Clustering forms a set of **building blocks**, each with at least one member

Measuring combinatoriality – clustering



Measuring combinatoriality – entropy

Songs that are more combinatorial should be more **compressible**

The compressibility of a song can be estimated with the information theoretic measure of **Shannon entropy**

The entropy of a song is calculated as:

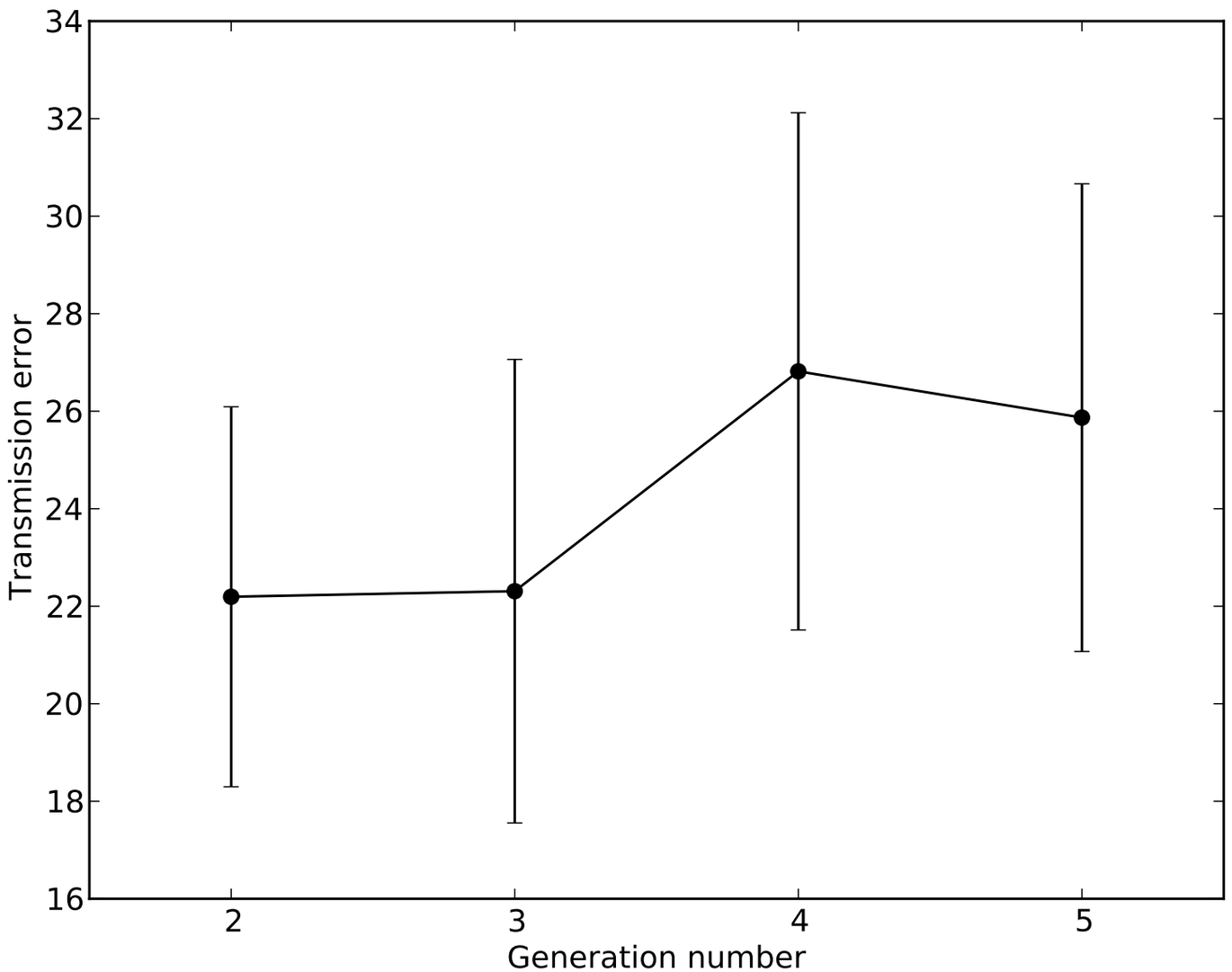
$$H = -\sum_{b \in B} P(b) \cdot \log_2 P(b)$$

$$P(b) = \frac{n_b}{N}$$

Entropy is expected to fall over time as structure increases

Shannon, CE (1948) Bell Syst Tech J, 27

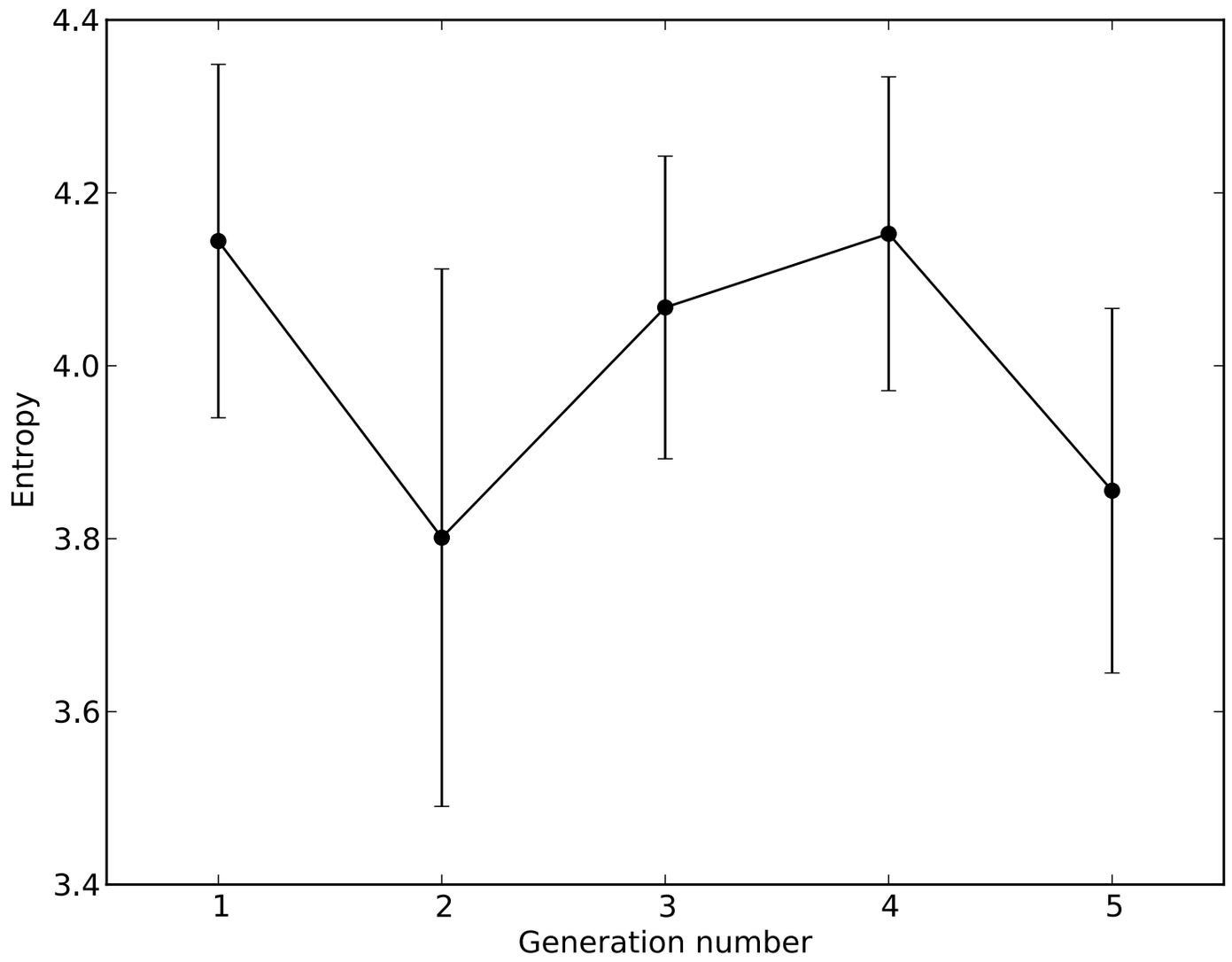
Results – learnability



Page's trend test $L = 937$, $m = 39$, $n = 4$, p = n.s.

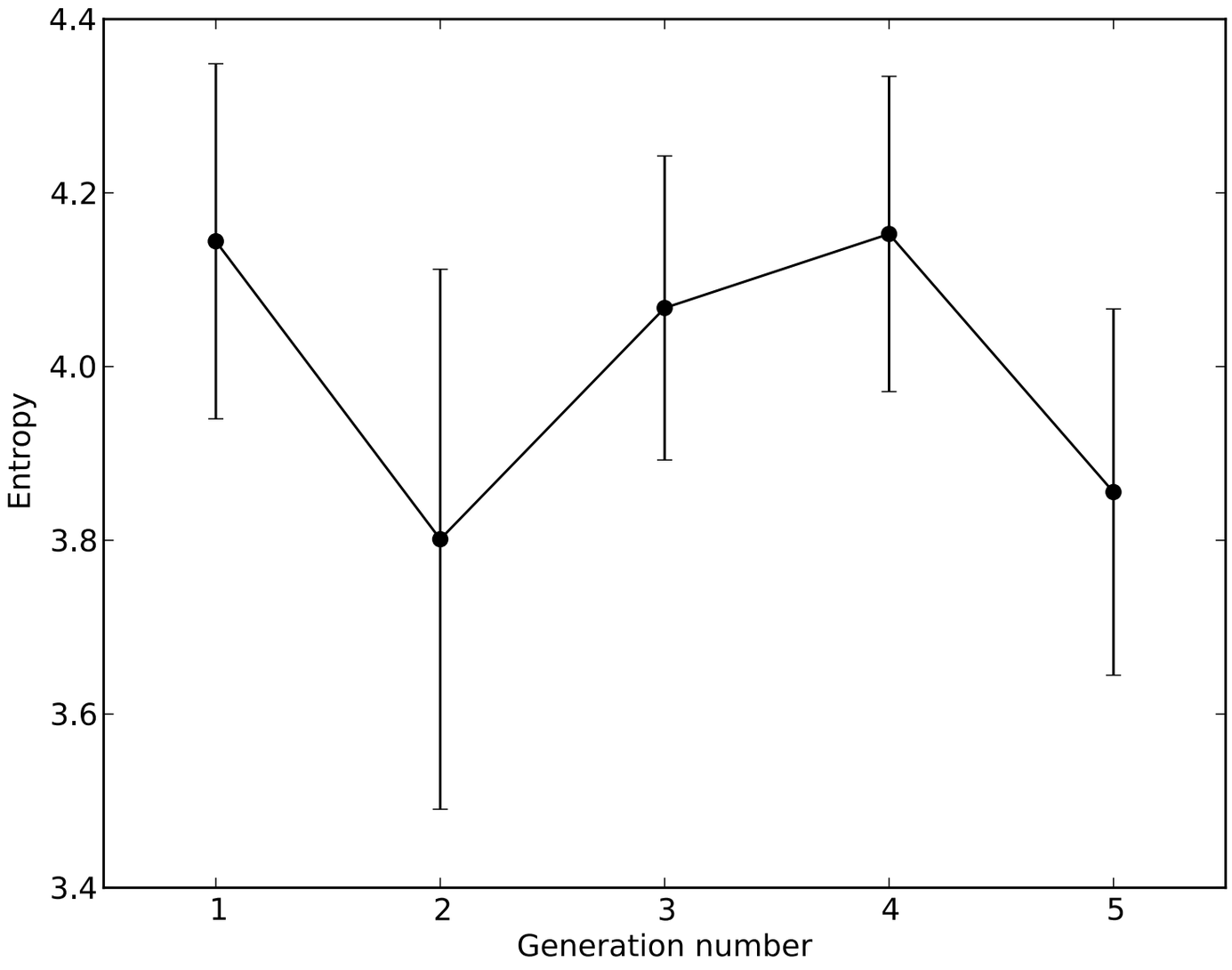
Page, E (1963) J Am Stat Assoc, 58

Results – combinatoriality

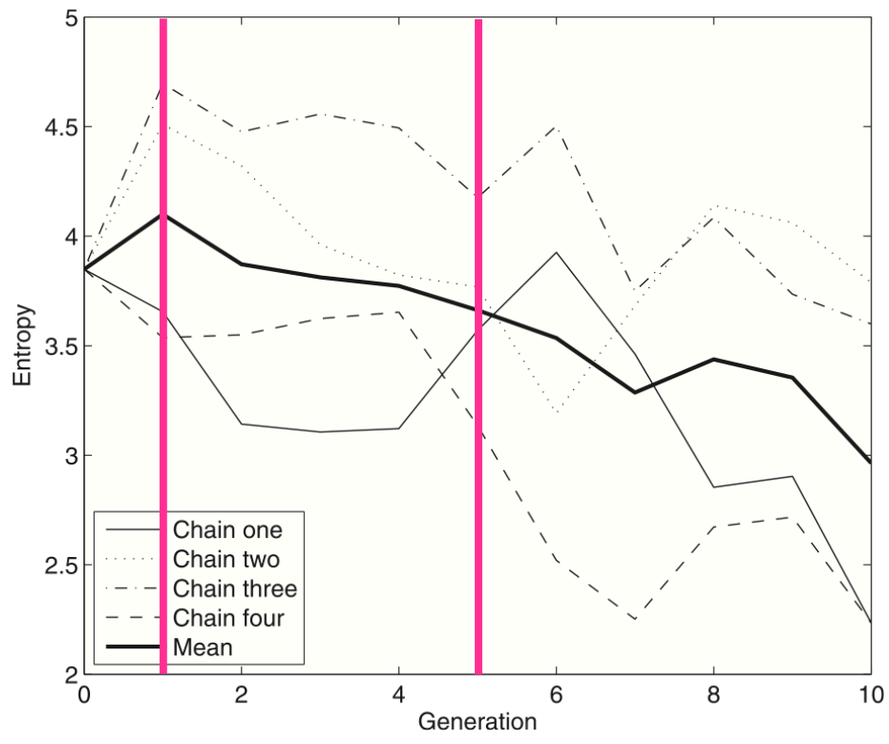


Page's trend test $L = 1758$, $m = 38$, $n = 5$, $p = 0.0597$ (n.s.)

Results – combinatoriality



Verhoef (2012)



$L = 1427$, $m = 4$, $n = 10$, $p < 0.001$

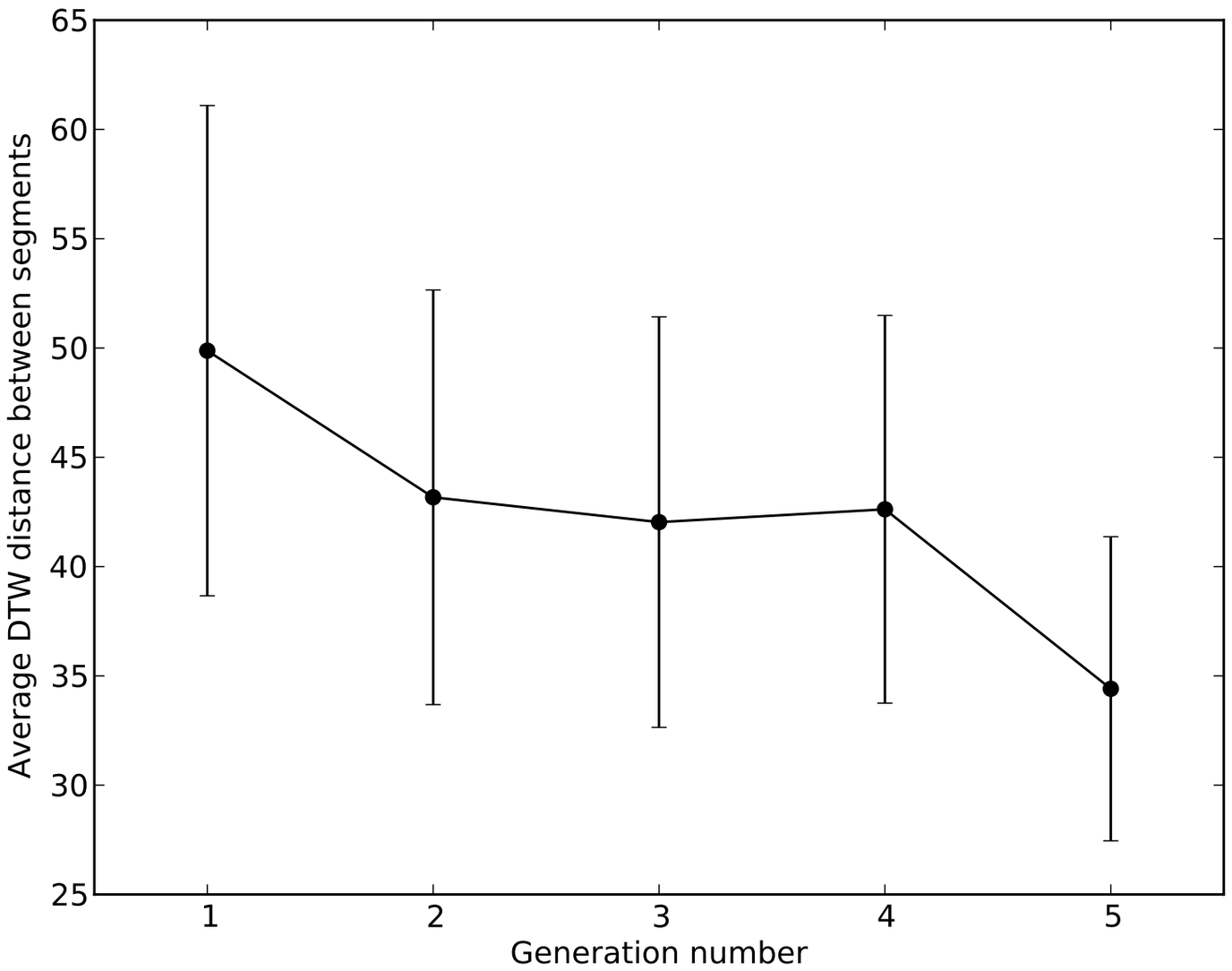
Page's trend test $L = 1758$, $m = 38$, $n = 5$, $p = 0.0597$ (n.s.)

Reasons for the lack of interesting results

In the case of learnability, there may be a **ceiling effect** – the songs become maximumly learnable very quickly.

In the case of combinatoriality, there may not be enough generations to see any interesting effects.

Combinatoriality – alternative metric



Page's trend test $L = 1777, m = 38, n = 5, p = 0.015$

Discussion and future directions

The results are currently inconclusive

May require a lot more data before the overall trend comes into focus

Still need to tweak the algorithms – especially the clustering

This dataset shouldn’t stand alone – should be used to support the conclusions of randomized controlled experiments

Maybe it’s worth looking for other kinds of dataset that are of an iterated nature

Thanks!

Questions or comments?

References

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 88–96.

Keogh, E. J., & Pazzani, M. J. (2001). Derivative dynamic time warping. In V. Kumar & R. Grossman (Eds.), *Proceedings of the 1st SIAM international conference on data mining*.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–147). London, UK: Springer Verlag.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the USA*, *105*, 10681–10686.

Page, E. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, *58*, 216–230.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*, 43–49.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*, 357–380.