Modeling the Acquisition of Statistical Regularities in Tone Sequences

Amaury Hazan, Piotr Holonowicz, Ines Salselas,
Perfecto Herrera, Hendrik Purwins

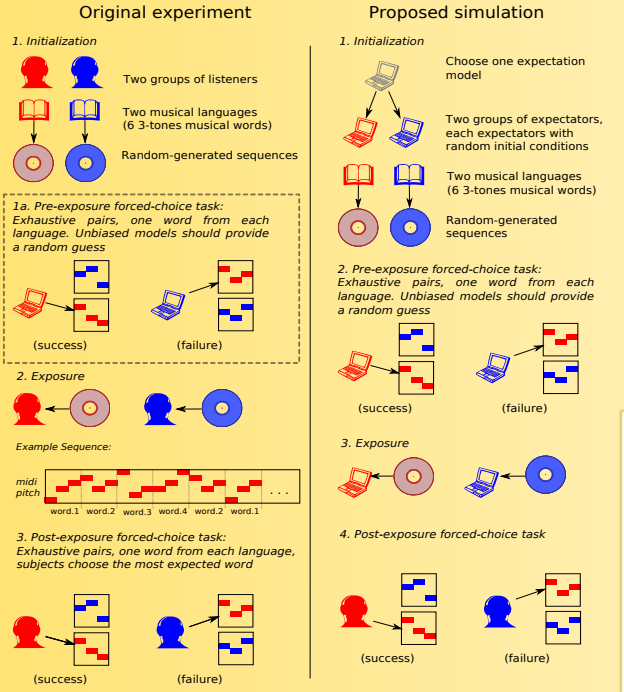
Music Technology Group
Universitat Pompeu Fabra, Barcelona

Alicja Knast, Simon Durrant

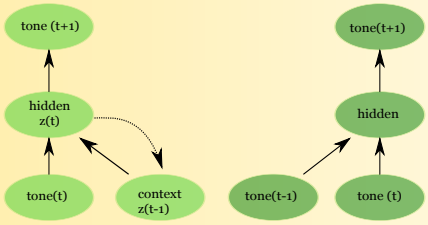
Interdisciplinary Centre of Computer Music Research
University of Plymouth

Sequence learning is an important process involved in many cognitive tasks, and is probably one of the most important processes governing music processing. In this work we build and evaluate computational models addressed to solve a tone-sequence learning task in a framework which simulates forced-choice tasks experiments. The specific approach we have selected is that of Artificial Neural Networks in an on-line setting, which means the network weights are always updated when new events are presented.
Here, we aim at simulating the findings obtained by Saffran, Johnson, Aslin and Newport (1999). We propose a validation loop that follows the experimental setup that was used with human subjects, in order to characterize the networks' accuracy to learn the statistical regularities of tone sequences. Tone-sequence encodings based on pitch class, pitch class intervals and melodic contour are considered and compared. The experimental setup is extended by introducing a pre-exposure forced-choice task, which makes it possible to detect an initial bias in the model population prior to exposure. Two distinct models (i.e. Simple Recurrent Network or a Feedforward Network with a time window of one event) lead to similar results. We obtain the most consistent learning behavior using an encoding based on Pitch Classes, which is not a relative representation. More importantly, our simulations and additional behavioral experiments highlight the impact of tone sequence encoding in both initial model bias and post-exposure discrimination accuracy. Furthermore, we suggest that melodic encoding and representation should be further investigated when inspecting and modeling behavioral experiments involving musical sequences.

Experiment



Overview of the experimental sequence.
At the left side is the original experiment performed by Saffran et all. (1999). The dotted rectangle illustrates an extra stage that has been added during replication of the original experiment by Knast . At the right side is the simulation.

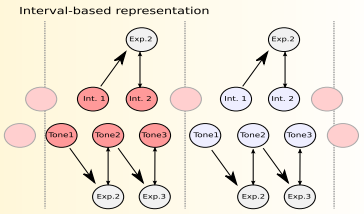


Simple Reccurent Network (SRN) Feed-forward Neural Network (FNN)

Artificial Neural Networks for statistical sequence learning:

1. The FNN has a feedback, the outputs are coupled with the inputs. Thus to perform next-event prediction task we need so many inputs as the number of past events we want the network to take into account.

2. The SRN uses a context layer that is a copy of the hidden layer from step *t-1*. In each step the input is a new tone *t* plus the neurons from context layer. So it maintains a sort of state that allows it to perform the tasks such as sequence prediction.

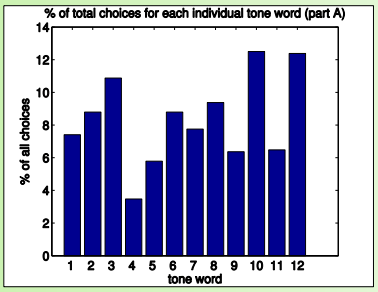


Forced-choice task simulation.
1. In both cases, of experiments on humans and simulations, we ask the subject to select from two tone words the one that is more expected.

2. In the simulation, since the subject is a neural network that produces a prediction of a single tone or interval, the more expected tone-word is chosen after exposure of all its parts (the 3 tones or 2 intervals, dependant on chosen encoding).

3. The choice is made by computing sums of cosine distances between expected tones (intervals) and the tones (intervals) coming from "listened" words and taking the closer one.

Results and discussion



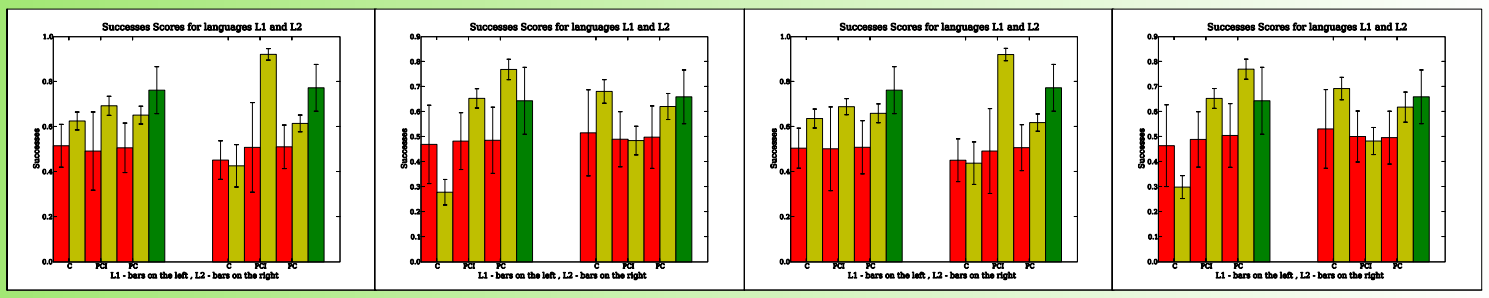
Pre-exposure distribution of selected tone words among Language L1 (tone words 1-6) and Language L2 listeners (tone words 7-12).

1. We repeated experimental loop 100 times in order to ensure the statistical significance of the results. The best parameters for neural networks were: 2 hidden units and learning rate of 0.01, 1 training epoch (to simulate the real conditions - exposure for each stimuli only once).

2. Both models performed equally, the SRN could not take advantage of the longer context memory. That might confirm that to solve the task only a computation of the transition probabilities between consecutive events, is necessary.

3. Pre-exposure distribution is non-uniform,the tone words are not chosen by human subjects with equal probability. It seems that there might be additional information carried by the tone words that has an impact on process of learning.

4. There is a bias towards one language that is dependant on encoding. With the Pitch Classes the bias is minimal, however interval based encodings are more strongly affected, particularly the Contour based encoding.



SRN model, Experiment 1 | SRN model, Experiment 2 | FNN model, Experiment 1 | FNN model, Experiment 2

Forced-choice accuracy obtained with SRN (left frame) and FNN (right frame) predictors for distinct tone sequence encodings, compared with the subjects' response in (Saffran et al. 1999). The simulation of both Experiment 1 (left: words versus non-words) and Experiment 2 (right: words versus part-words). For each experiments and model, the results are shown for Language L1 on the left and for Language L2 on the right. For each encoding, the pre-exposure (red bars) and post-exposure (light-brown bars) mean score is plotted, along with its standard deviation over the 100 runs. Contour encoding is denoted C, Pitch Class Interval encoding is denoted PCI, and pitch class encoding is denoted PC. For each language, the dark green bar shows the ground truth post-exposure accuracy obtained by (Saffran et al. 1999), denoted GT.