Evolutionary Ecology Research, 2000, **2**: 375–383

Evolution of preference for consonances as a by-product

Masashi Kamo* and Yoh Iwasa

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-8581, Japan

ABSTRACT

Recent theoretical studies of evolution of visual signals in animals have revealed that biased preferences for symmetric patterns or simple coloration can evolve in the absence of positive fitness effects. In this paper, we study the evolution of biased preference for auditory signals. In music theory, intervals between a pair of auditory signals are classified into consonances and dissonances. Consonances are more comfortable to listen to than dissonances, and often have a frequency ratio close to a ratio of small integers. By examining the preferences shown by a three-layered network as a simplified model of an auditory system, we assess why we find consonances comfortable and dissonances uncomfortable. When the network was trained to accept monotones accompanied by harmonic tones and to reject random signals (noises), it developed a preference for consonances rather than dissonances. This suggests that the preference for consonances may have evolved as a by-product of training for a simple task, such as distinguishing mother's voices from noises, rather than as a result of being taught one-by-one. When the network was trained to favour a consonance and to reject a dissonance, it did not generalize the preference to other consonances or dissonances.

Keywords: auditory signals, consonance and dissonance, evolution of biased preference, generalization, harmonics, neural network.

INTRODUCTION

The female preference for male exaggerated ornaments may have evolved through direct positive fitness effects, or through indirect genetic effects, either through sexy son processes or through handicap principles (see papers cited in Andersson and Iwasa, 1996). Alternatively, the female preference may have evolved as a by-product of evolutionary processes independent of the male trait itself, and male traits may have simply exploited the existing sensory bias of the females (sensory exploitation hypothesis; Kirkpatrick and Ryan, 1991).

When an animal is trained to respond to a certain stimulus, it may respond to other stimuli that are similar but not exactly the same as the original training signal. This phenomenon is called 'generalization', and is important in understanding the results of learning experiments in animal psychology (Gutman and Kalish, 1956). Generalization is also important in understanding learning by neural network models.

* Author to whom all correspondence should be addressed. e-mail: kamo@bio-math10.biology.kyushu-u.ac.jp
Consult the copyright statement on the inside front cover for non-commercial copying policies.

Using a simple neural network model of a female recognition system, Enquist and Arak (1993, 1994, 1998) demonstrated that the preference for a certain set of visual signals can evolve as a by-product of training. The network trained to distinguish some input patterns from others often developed a propensity to favour certain patterns that had not been used in the training procedures (Enquist and Arak, 1993). This result was considered to explain supernormal stimuli in animal behaviour and subsequent exaggerated forms of sexually selected male ornaments, or of flower petals (Arak and Enquist, 1993). Kamo *et al.* (1998) studied a slight modification of the training procedures using the same network model as Enquist and Arak, but they observed a very different outcome. After training, the network evolved to show no sign of supernormal stimuli – the pattern used for training achieved the highest probability of being accepted by the trained network in all the cases examined. This example urges us to choose carefully the training procedures and network structures when we adopt neural network modelling in the study of sensory systems.

A similar model was developed to discuss an inherent preference for symmetric input patterns or simple colour (Enquist and Arak, 1994; Johnstone, 1994). Enquist and Arak (1994) argued that colourful or symmetric patterns, which are widely seen in animals and flowers, have evolved because of the sensory bias arising through the co-evolution between neural recognition systems and biological signals. Johnstone (1994) also showed that the preference for symmetry occurs as a by-product of selection in the context of mate recognition. However, all of these studies focused on the biased preference for visual signals.

Here, we study the biased preference for auditory signals. In the theory of harmonics, some combinations of tones are called consonances, whereas others are called dissonances (Table 1). Consonances are considered as intervals more comfortable to listen to than

Table 1. Harmonistic definition of intervals*

		Interval name	Number of semitones	Frequency ratio	Ideal ratio
Consonances	Perfect	perfect 1st	0	1.0	1 : 1
		perfect 8th	12	2.0	2 : 1
		perfect 5th	7	1.498	3 : 2
		perfect 4th	5	1.335	4 : 3
	Imperfect	major 3rd	4	1.260	5 : 4
		minor 3rd	3	1.189	6 : 5
		major 6th	9	1.682	5 : 3
		minor 6th	8	1.587	8 : 5
Dissonances		major 2nd	2	1.122	9 : 8
		minor 2nd	1	1.059	16 : 15
		augmented 4th	6	1.414	45 : 32
		diminished 5th	6	1.414	64 : 45
		major 7th	11	1.888	15 : 8
		minor 7th	10	1.782	16 : 9

* Frequency ratios are based on equal temperament, in which the interval with n semitones has a frequency ratio of $2^{n/12}$. Those intervals with a frequency ratio close to a ratio of two simple integers are consonances and others are dissonances. Ideal ratios are based on 'just intonation' (Burns and Ward, 1982), and classification of consonance/dissonance is based on Shimofusa (1972).

dissonances. A typical consonance is an 'octave', exemplified by a pair of tones with a frequency ratio of exactly two. Another example is 'perfect fifth' (say 'C' and 'G') with a frequency ratio close to three halves. Many other examples of consonances have frequency ratios close to the ratios of two small integers (Table 1). In contrast, dissonances tend to have frequency ratios that are not close to a simple ratio. Why do we feel consonances more comfortable than dissonances? We investigate the question by analysing simple neural network models and by examining how the preference for certain pairs of tones over others appears in them.

Auditory signals in natural environments are unlikely to be pure tones – a basic tone (called a fundamental or a root) is almost always accompanied by its harmonic tones. For example, a sound of 220 Hz includes components of 440 Hz, 660 Hz, 880 Hz, etc. In the development of infants, important input auditory signals (e.g. the mother's voice) may be given in different pitches, but in each case the fundamental is accompanied by its harmonic tones. The infant might develop a preference for certain combinations of tones that are commonly included in the harmonic tones of its mother's voice. This may explain the preferences for consonances that have simple frequency ratios (Table 1), as these are more likely to be included in the harmonic tones of input signals. If so, the preference for consonances rather than dissonances might be a by-product of a simpler task, such as to discriminate important sounds (such as mother's voice) from noises. We call this the 'by-product hypothesis'.

Second, the training for a particular consonance and against a particular dissonance may be generalized to a preference for all consonances and against all dissonances. As explained later, consonances are characterized as sharing low harmonic components, unlike dissonances, and hence they may be more similar to each other than dissonances. We call this the 'harmonic-overlapping hypothesis'.

Alternatively, the preference for consonances over dissonances may be completely arbitrary, and their distinction may be determined by convention or by the historical accident that they were chosen by great composers in the past. If there is no basic reason to favour consonances, we must learn a preference for consonances over dissonances, one-by-one. We call this the 'taught one-by-one' hypothesis. In the following, we examine the by-product hypothesis and the harmonic-overlapping hypothesis by training neural network models. We treat the taught one-by-one hypothesis as the null hypothesis so that we favour it if both of the others are rejected.

NEURAL NETWORK MODEL

Auditory signals, or sounds, first arrive at the ear drum, go through the middle ear, and finally reach the cochlea. In the cochlea, sounds are transformed to electric signals by a group of hair cells on the basilar membrane. The hair cells are arranged in the order of their sensitive frequency. Cells on the apex of the basilar membrane react to low-frequency sounds, and cells on the base react to high-frequency sounds. This system is called 'cochlear tuning' (Nicholls *et al.*, 1992).

In this paper, we model the auditory system as a neural network that has three layers, as illustrated in Fig. 1. An input layer has 37 cells; cells on the left react to low frequencies and those on the right react to high frequencies. We set the ratio of the sensitive frequencies of adjacent input cells at $2^{1/12} = 1.059463$, which corresponds to one semitone in 'equal temperament'. An octave (the difference of double frequency) includes 12 semitones, and

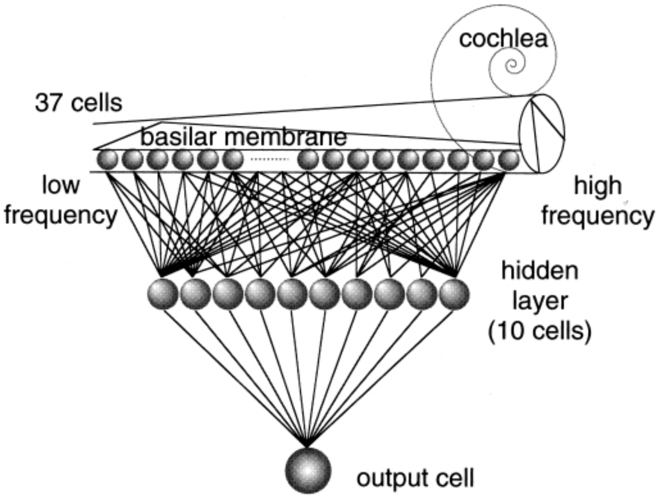

Fig. 1. The auditory system and the neural network used in this study. Each cell has a sensitive frequency, and cells are arranged in the input layer according to the frequency (cochlea tuning). The left-hand side of the input layer responds to the lower frequencies and the right-hand side to the higher frequencies. Music intervals between adjacent cells are a semitone, and the musical range of this network is three octaves.

the 37 input cells of the network can cover three full octaves. In the second layer, or hidden layer, there are 10 cells (or neurons), each of which is connected to all 37 cells in the input layer. Each connection has a weight, which may be modified by training. Finally, all 10 cells in the hidden layer are connected to the output layer with only one cell.

The initial values of the connection weights were generated by random numbers uniformly distributed between −1 and +1. The activity level of the jth cell in the hidden layer is:

$$x_j = f(u_j) \qquad j = 1, \ldots, 10 \tag{1a}$$

where

$$u_j = \sum_{i=1}^{37} w_{ji} x_i^{\text{input}} \tag{1b}$$

$f(u)$ is a sigmoidal function of u, and is given by $f(u) = 1/(1 + \exp[-u])$. w_{ji} is the weight on the connection between the ith cell in the input layer and the jth cell in the hidden layer. It is positive if a higher activity of the ith input cell enhances the activity of the jth cell of the hidden layer; but it is negative if the interaction is inhibitory. The output of the system is a sigmoidal function of the sum of signals coming from the hidden layer:

$$x_{\text{output}} = f(u_{\text{output}}) \tag{2a}$$

where

$$u_{\text{output}} = \sum_{j=1}^{10} w_{oj} x_j \tag{2b}$$

w_{oj} is the weight for the connection from the jth cell in the hidden layer and the output cell.

Starting from the initial network with randomly generated weights, we modified the weights of connections in the direction that would make the network show correct responses more often. We trained networks by back-propagation, which is a version of the gradient method for optimization (Bose and Liang, 1996). It provides a way to adjust the weights to minimize the sum of the square error between the output of the network and the desirable response (teach signal). High outputs of a network indicate that the input signal was accepted and low outputs indicate the inputs were rejected.

There is a difference between exact harmonics and the one expressed in equal temperament. For example, perfect fifth is the interval of two tones with frequency ratios of 3/2, which is included in the harmonics of a single tone as the second and the third harmonics. However, this is expressed in equal temperament as seven semitones (see Fig. 2), which is $2^{7/12} = 1.498$, which is close to but not exactly the same as 3/2. In this paper, we assume that generalization in the space of frequency works, and training with tones with a frequency ratio of 3/2 enhances the reaction of the auditory system to the intervals of ratio 1.498.

BY-PRODUCT OF TRAINING

Input signals in natural environments are composed of a basic tone ('fundamental') and its harmonic tones. Figure 2 illustrates all the differences between harmonic components of a monotone shorter or equal to an octave. The numerals indicate the number of semitones included in each interval. All these intervals correspond to consonances, but most intervals that do not appear in Fig. 2 are dissonances (see Table 1). The only exception was minor sixth (eight semitones), which is classified as a consonance (Shimofusa, 1972; see Table 1), but not included in Fig. 2. Since several low harmonics are usually more important than harmonics of still higher frequencies, we assumed that each basic tone is accompanied by six lower harmonic tones with equal amplitude.

The networks were trained to accept monotones accompanied by harmonic tones shown in Fig. 2 and to reject noises. Each accept signal was shifted for all cases in which two or more harmonic tones were included in the audible range of 37 input cells. There were 52 input signals in total.

As a simple model for random noises to reject, we generated uniform random numbers between 0 and 1 to all the cells in the input layer. In each step of weight modification, the network was presented with 50 accept signals and 500 reject signals.

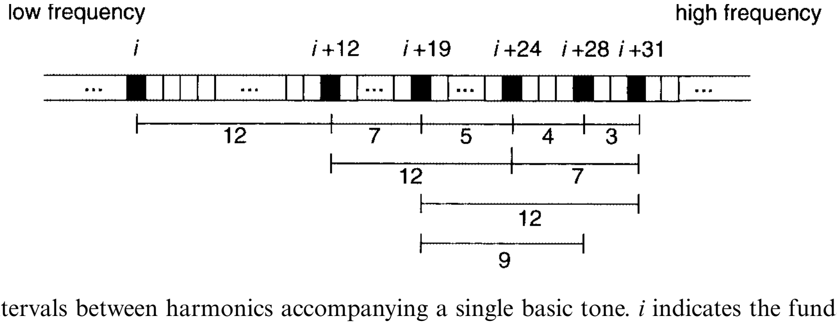

Fig. 2. Intervals between harmonics accompanying a single basic tone. *i* indicates the fundamental; the other black squares represent the harmonic tones. The numerals indicate the length of intervals, given as the number of semitones included. These intervals are all consonances.

The training continued until the average error rate became less than 5%. After training, intervals of two notes within an octave (a difference less than or equal to 12 semitones) without harmonic tones were given to the network. There were 366 test patterns, 211 of which were consonances and 155 of which were dissonances. If networks have no systematic preference for a group of inputs over the others, we expect that the average response of independently trained networks to the former to be higher than the average response to the latter with 50% chance. Hence, we can apply a binomial test to detect a systematic bias.

Figure 3 illustrates the average responses of the trained networks to the consonances and to the dissonances defined by music theory (Table 1). The former was larger than the latter for 46 networks among 50 replicates, which was statistically significant (binomial test, $z = -5.9397$, $P < 0.001$).

Null intervals (the relationship with the same tone) were excluded from the analysis in Fig. 3, although they are consonances (perfect first). If these were included, there was a much larger difference in the responses to the two groups of input intervals than in Fig. 3. This result supports the by-product hypothesis.

GENERALIZATION BETWEEN CONSONANCES BY HARMONIC OVERLAPPING

The second issue is whether the preference for a particular consonance is generalized to other consonances. If two fundamentals are consonances, the frequency ratio is a ratio of small integers, and hence their sets of harmonics tend to share common tones. Sharing common harmonic tones is less likely to occur if two fundamentals are dissonances. Using the property of having common harmonic tones, the network may be able to distinguish the two classes of intervals. If one consonant pair of tones is favoured in training, the network might develop a preference for other consonances that did not appear in training.

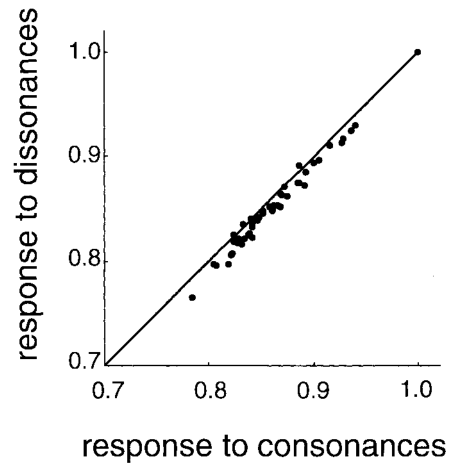

Fig. 3. The average responses of 50 trained networks trained independently. The average responses to consonances and those to dissonances, defined in Table 1. The average response to consonances was higher than that to dissonances (binomial test, $z = -5.9397$, $P < 0.001$). The line has a slope of 1, intercept of 0 and is not the regression.

To test the possibility of generalization, we used accept signals of intervals of seven semitones (perfect fifth) and rejected signals of six semitones (diminished fifth) (Table 1). We used all cases in which the fundamentals of both input signals were included in the first octave. Each network experienced 13 input patterns in each trial (six pairs of tones with perfect fifth and seven pairs of tones with diminished fifth in an octave).

Figure 4 illustrates the average response to consonances and the average response to dissonances that were not used in the training. The networks were trained with the tones with harmonics, but were tested by the tones without harmonics. The average response to consonances was not higher than the average response to dissonances (55 times among 100 replicates, statistically not significant; binomial test, $z = 1.0$, $P = 0.097$). This implies that the trained preference was not generalized to other consonances.

DISCUSSION

We examined how the preference for some pairs of tones might arise. A special feature of the auditory system, compared with the visual, olfactory and other sensory systems, is that input signals are almost always accompanied by their harmonics. Characteristic biases commonly occurring in auditory signals might be explained as a coincidental outcome of this feature.

Generalization is an important property of all neural systems, including auditory systems. This is also important in forming a preference for consonances. For example, the input signals include harmonics three times higher in frequency than the basic tone, and is exactly 1.5 times that of the first octave, thus causing a preference for the perfect fifth. However, the perfect fifth, expressed in terms of equal temperament or seven semitones, is

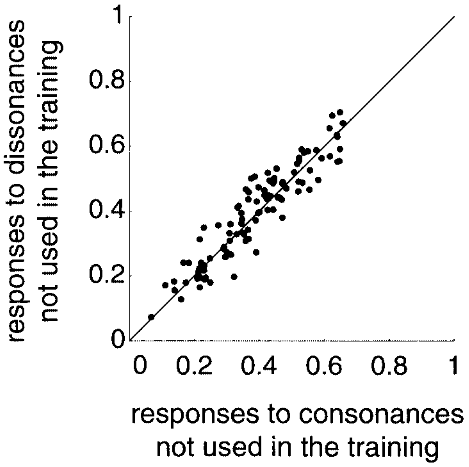

Fig. 4. The average responses of 100 independently trained networks testing the harmonic-overlapping hypothesis. The networks were trained to favour perfect fifth (seven semitone intervals) and to reject diminished fifth (six semitones). Test signals were intervals without harmonics. Of 100 networks, only 55 showed a preference for consonances not used in the training, which was not statistically significant (binomial test, $z = 1.0$, $P = 0.097$). The line has a slope of 1, intercept of 0 and is not the regression.

$2^{7/12} = 1.4983$, which is close to but not exactly 1.5. The latter is just intonation. However, thanks to the generalization ability of the auditory system, the network can develop a preference for seven semitones from the input signal. The difference between just intonation and the closest interval of equal temperament may be larger than this, but it is still small (Table 1).

In our trials, networks trained to accept monotones with harmonics and reject random noises developed a preference for consonances spontaneously. Hence, preference for consonances can be acquired without being taught one-by-one. It is simply a by-product of training procedures to distinguish monotones from random noises, caused by the fact that most auditory signals are accompanied by harmonic tones with a frequency equal to multiples of the basic frequency. We conclude tentatively that the preference for consonances arises as a by-product of training for monotones against random noises.

On the other hand, we found that the networks trained to accept a perfect fifth and to reject a diminished fifth did not develop a preference for consonances that were not used in the training (Fig. 4). This implies that training to favour a particular consonance and to reject a particular dissonance was not generalized to other consonances or dissonances.

In summary, the preference for consonances is likely to have arisen inadvertently. This suggests that preferences between auditory signals may not be completely conventional or arbitrary. If so, the same process should work for animals that have a similar auditory range, and they too might find consonances more comfortable to listen to than dissonances.

ACKNOWLEDGEMENTS

This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science, Sports and Culture, Japan and also by CREST, Japan Science and Technology Corporation. We thank the following for their very helpful comments: D. Falush, H. Hakoyama, Y. Morimoto, S. Okamoto and N. Yamamoto.

REFERENCES

Andersson, M. and Iwasa, Y. 1996. Sexual selection. *Trends Ecol. Evol.*, **11**: 53–58.

Arak, A. and Enquist, M. 1993. Hidden preferences and the evolution of signals. *Phil. Trans. R. Soc. Lond. B*, **340**: 207–213.

Bose, N.K. and Liang, P. 1996. *Neural Network Fundamentals with Graphs, Algorithms, and Applications*. New York: McGraw-Hill.

Burns, E.M. and Ward, W.D. 1982. Interval, scale and tuning. In *The Psychology of Music* (D. Deutch, ed.), pp. 301–334. Tokyo: Nishimura Shoten Co. (Japanese translation).

Enquist, M. and Arak, A. 1993. Selection of exaggerated male traits by female aesthetic senses. *Nature*, **361**: 446–448.

Enquist, M. and Arak, A. 1994. Symmetry, beauty and evolution. *Nature*, **372**: 169–172.

Enquist, M. and Arak, A. 1998. Neural representation and the evolution of signal form. In *Cognitive Ecology* (R. Dukas, ed.), pp. 21–87. Chicago, IL: The University of Chicago Press.

Gutman, N. and Kalish, H.I. 1956. Discriminability and stimulus generalization. *J. Exp. Psychol.*, **51**: 79–88.

Johnstone, R.A. 1994. Female preference for symmetrical males as a by-product of selection for mate recognition. *Nature*, **372**: 172–175.

Kamo, M., Kubo, T. and Iwasa, Y. 1998. Neural network for female mate preference, trained by a genetic algorithm. *Phil. Trans. R. Soc. Lond. B*, **353**: 399–406.

Kirkpatrick, M. and Ryan, M.J. 1991. The evolution of mating preference and the paradox of the lek. *Nature*, **350**: 33–38.

Nicholls, J.G., Martin, A.R. and Wallace, B.G. 1992. *From Neuron to Brain*, 3rd edn. Sunderland, MA: Sinauer Associates.

Shimofusa, K. 1972. *Harmonics* (*Waseigaku*). Tokyo: Ongakunotomosha Publishing (in Japanese).