A Model of Perceptual Change by Domain Integration

Gert Westermann (**gert@csl.sony.fr**)
Sony Computer Science Laboratory
6 rue Amyot
75005 Paris, France

Abstract

A neural network model is presented that shows how the perception of stimuli can be changed due to their integration between different domains. The model develops representations based on the correlational structure between the stimuli of the domains. It shows how prototypicality, categorical perception, and interference effects between the domains can arise without the explicit labeling of categories. The model is extended to learn the sensori-motor integration between articulatory parameters and speech sounds and it is shown how it can, in accordance with the ideomotor principle, imitate sounds based on the developed mappings in a "babbling phase".

Introduction

The ability to categorize is one of the most fundamental cognitive processes. Nevertheless, uncovering the mechanisms that underlie this ability has challenged experimenters and modelers alike. The reason for this difficulty might be that categories can be formed in many different ways: in some cases, perhaps mainly in experimental situations, explicit information about the category of a stimulus is given. In other cases, no feedback might be available about the categorization choice, and in even others, no explicit categorization choice might be made at all.

At the same time, recent research has suggested that categorization itself can exert an influence on perception (Goldstone, 1995; Schyns *et al.*, 1998). While these effects have mainly been studied in a supervised paradigm, perceptual changes also occur prominently in categorization without supervision and without explicit labeling, for example, in being exposed to the phonemes of one's native language (Kuhl *et al.*, 1992).

Finally, there is clear evidence that in categorizing the world, we make use of all available information and integrate the information from different modalities, making categorization more robust and easier. For example, visual and auditory information are integrated in speech perception, leading to enhanced activity in the cortical areas responsible for both domains (e.g. Calvert *et al.*, 1999).

In this paper a neural network model is described that aims to integrate several aspects of categorization, namely, the combination of modalities and the perceptual changes that go hand in hand with categorization. The model suggests that some of the phenomena that are usually explained as the consequence of explicit categorization, e.g., prototype formation and categorical perception, can arise without such explicit categorization based on the correlational structure of the stimuli from different modalities, and that they can facilitate subsequent explicit categorization when it occurs.

The integration between modalities has previously been modeled by de Sa and Ballard (1998). Their neural network model consisted of one layer for each modality, and each layer made an explicit category decision. In a process of self-supervised learning both modalities learned to agree on their decision. While the model performed comparably to supervised models, it was necessary to determine the number of categories *a priori*, and due to absolute category boundaries in each modality, perceptual phenomena such as gradedness and varying sensitivities to differences between stimuli could not be modeled. The present model aims to give an account of how such perceptual phenomena that are usually linked with explicit categorization can occur without an explicit category decision and without labeling, and how such a model can be extended to also account for sensori-motor integration. The model is loosely inspired by neurobiological considerations.

The rest of the paper is organized as follows: first, the model is described in detail. Then, experiments with a simple data set are described that lead to perceptual change as the result of the integration between modalities. Finally, the application of the model to sensori-motor integration and the imitation of sounds is described.

The Domain-Integration Model

The model described here integrates the stimuli from two domains (modalities) into a unified percept. The architecture of the model is shown in fig. 1. Each domain is represented by a neural map, and Hebbian connections between the maps allow for the coordination between them. Usually, an input pair (one input per map) is presented to the maps simultaneously, and in the following the activation and weight update mechanisms are described.

Each neural map consists of a number n of units that are randomly positioned in the input space (in this paper, the input spaces for both domains are two-dimensional

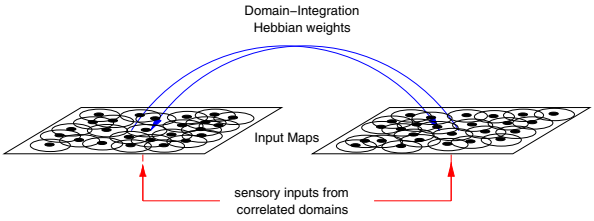

Figure 1: The architecture of the model.

to facilitate visualization). In the current model, the positions of these units remain fixed throughout learning. Each unit acts as a receptive field with a Gaussian activation function of a fixed width. Such receptive fields exist in many areas of the cortex. When an external input x is presented to the map, the Gaussian activation of the units is computed as

$$act_{i_{ext}} = e^{\frac{pos_i - x}{\sigma^2}} \tag{1}$$

where pos_i is the position of unit i, x is the input signal, and σ is the standard deviation (width) of the Gaussian. Each unit is connected with unidirectional Hebbian weights to all units on the map for the other domain. The Hebbian activation of a unit is the dot product of the weight value vector and the activation vector of the units on the other map:

$$act_{i_{hebb}} = \sum_k act_k w_{ik} \tag{2}$$

where the units on the other map are designated by the index k and w_{ik} is the weight from a unit k on the other map to the current unit i on this map.

The total activation of a unit is computed by summing the activations from the external stimulus and those from the Hebbian connections with the other map:

$$act_i = \gamma_e act_{i_{ext}} \quad \gamma_h act_{i_{hebb}} \tag{3}$$

where γ_e and γ_h are weighting parameters to control how much each partial activation contributes to the total activation of the unit.

The activation update after presentation of a pattern is synchronous for all units, and the activation values are scaled to a maximum of 1.0.

One input to a map will typically activate several units, and the response r_i to an input x, that is, how the neural map "perceives" that input, is computed by a population code: the response is the vector sum of all units i, weighted by their activation values:

$$\mathbf{r}_x = \frac{\sum_i act_i \mathbf{pos}_i}{\sum_i act_i} \tag{4}$$

Such population codes have been found to play a role for example in the encoding of motor commands in the monkey cortex (Georgopoulos *et al.*, 1988) where the direction of arm reaching is predicted accurately by adding the direction vectors of direction sensitive neurons, weighted by their firing rate. In computational models, population codes have been successfully used to show the emergence of a perceptual magnet effect for phonemes (Guenther and Gjaja, 1996).

The Hebbian connections between the maps are updated with the covariance learning rule (Sejnowski, 1977):

$$\Delta w_{ik} = \alpha(act_i - \bar{act}_i)(act_k - \bar{act}_k) \tag{5}$$

where $\bar{act}_i$ and $\bar{act}_k$ are the average activation values of units i and k over a certain time interval. This rule strengthens the connections between units when their activation values are positively correlated, weakens them when the activations are negatively correlated, and does not change the weights when the activations are decorrelated.

This correlation-based weight update has the consequence that units that respond to stimuli that consistently co-vary across the domains develop higher activations due to the growing Hebbian weights: co-varying inputs in the two domains result in the same units on both maps to have co-varying activation values, and thus to develop strong Hebbian connections. This results in such units not only receiving external, but also strong Hebbian activation and becoming more active than other units that do not reliably co-vary with units from the other domain map. Given that the population code response is weighted by unit activations, this means that such units "pull" the response towards them and induce a perceptual change (fig. 2). Therefore, an input-pair with normal (previously observed) correlational structure will become more prototypical so that other, nearby inputs will be displaced towards it.

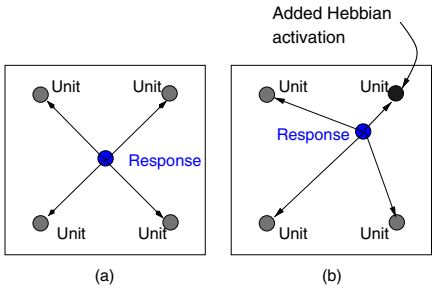

Figure 2: The response to an input is influenced by external Hebbian activation. (a): Without Hebbian activation, the response lies in the middle between four equally activated units. (b): When one unit is activated more due to Hebbian activation, the response is displaced towards that unit.

In the following section, experiments with this model are described that investigate the nature of the induced perceptual changes based on the integration between the two input domains.

Experiments

The domain-integration model was tested with a simple data set (fig. 3) to investigate the nature of the developed perceptual changes and the role of correlations between data from the two domains in this process.

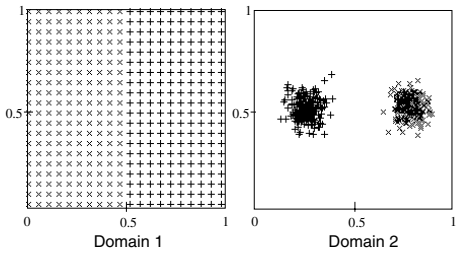

Figure 3: The data used to evaluate the model. The correlational structure between data items splits each domain into two classes, denoted by x and +, respectively.

Domain 1 consists of 400 evenly distributed two-dimensional data in the range from 0 to 1. Domain 2 consists of two clusters of 200 data each with Gaussian distributions around the centers $(0.25, 0.5)$ and $(0.75, 0.5)$. In training, the "left half" of the data in domain 1 (i.e., between 0.0 and 0.5) co-occurred with data from the "right" cluster of domain 2, and the "right half" in domain 1 (0.5 to 1.0) with data from the "left" cluster in domain 2.

Although this data set is artificial, it could be interpreted as, for example, a continuous variation of width and height of an object (domain 1) and associated sounds at certain frequencies and volumes (domain 2) in a modality-integration experiment.

The neural maps for each domain consisted of 200 randomly placed units. All data pairs were presented to the model a single time in randomized order. The Hebbian connections between the maps had initial values of 0 and were updated after presentation of each data pair. The parameter settings were $\alpha = 0.01$, and for each map, $\sigma = 0.05$, $\gamma_e = 1.0$, and $\gamma_h = 0.02$.

Development of Prototypes

Fig. 4 shows the initial and final responses to the data set. Each data input creates a response on its neural map (eq. (4)). Fig. 4A shows the initial response of the neural maps to the data from each domain. With all Hebbian connections being zero, the response is only determined by the actual input signal to the map and gives a rather faithful representation of the original data set in fig. 3. Due to the random location of units the original data is slightly distorted in the response.

During the training of the model, the Hebbian connections between units responding to co-varying data in both domains are strengthened and those responding to non-co-varying data are weakened or remain unchanged (eq. (5)). This process results in strong connections between units that respond to the centers of their categories because they will be active for both central and more

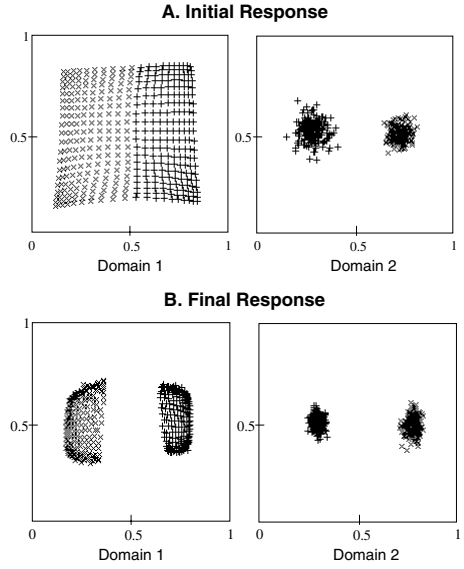

Figure 4: The initial (A.) and final (B.) response of the model to the data set in fig. 3.

peripheral inputs from a certain category. As a consequence, such central units will become more active than others when correlated inputs are presented. Their activation is a sum of the external activation caused by the inputs themselves, together with the activation mediated through the strengthened Hebbian weights from the other map (eq. (3)). Therefore, the response to peripheral stimuli will be pulled towards the center of each category. Fig. 4B shows the responses of the maps to the data after presentation of each data item and corresponding updating of the Hebbian connections. The continuous data in domain 1 has split into two clusters that correspond to the co-variance relations with the clusters in domain 2. Each cluster is based around a prototype determined by the central data item of each set. Similarly, the clusters in domain 2 have become very dense around their respective centers. Prototypes thus develop simultaneously in both domains, based on the interactions between the domain maps.

Categorical Perception

Categorical Perception (CP) is a phenomenon that occurs both innately and in learned categorization (see Harnad, 1987, for an overview): different stimuli within one category are perceived as more similar than stimuli with the same "distance" that straddle category boundaries. One example for innate CP is the perception of color, e.g. in a rainbow: although the light frequency changes continuously, we perceive separate bands of color. For example, within the red band we do not perceive differences between changing light frequencies, but an equally small change at the border of that band leads to the abrupt perception of orange.

It has been shown that CP also develops in learned categories such as phonemes of one's native language

(e.g. Kuhl *et al.*, 1992). More recently, CP has also been shown to arise for visual stimuli in categorization task experiments (Goldstone *et al.*, 1996). In these experiments, subjects had to group a set of continuously varied shapes into categories in a supervised learning task. After having learned the categories, they were better able to distinguish between two stimuli that were near a category boundary than between those that were within a category. Therefore, CP can be said to constitute a warping of similarity space in which the sensitivity to differences of (relevant) stimuli is enhanced near category boundaries and is decreased within categories.

Guenther and Gjaja (1996) modeled categorical perception for phonemes in an unsupervised model. They argued that the firing preferences of neurons in the auditory map reflect the distribution of sounds in the language, and due to the non-uniform distribution of these sounds CP arose in the model in a self-organizing process. While this model accounts well for CP in phoneme perception, it relies on a non-uniform distribution of the data. CP that arises for uniform stimuli as a result of explicit categorization has been modeled in supervised radial basis (Goldstone *et al.*, 1996) or backpropagation (Tijsseling and S.Harnad, 1997) networks. It therefore seems that CP can arise from different causes (data distribution or explicit teaching), and in the model presented here a third route is taken: it is studied how CP can arise in a homogeneously distributed data set that is correlated with non-uniform data in another domain, without the explicit teaching of category labels. Instead, categories form in an unsupervised way based on the correlational structures between the two domains.

In the present experiments, the x-coordinate of the data is the relevant dimension for determining category membership (with the categories defined by the correlations across domains). To establish whether CP did occur in the model, after training the map of domain 1 was presented with a sequence of data points from (0.0, 0.5) to (1.0, 0.5) in steps of 0.01, i.e., a walk from the left to the right side of the map. The difference between the responses of the model to every pair of adjacent data points is shown in fig. 5. There is a marked peak of sensitivity at the category boundary (0.5) where a difference of 0.01 in the input data is perceived as a difference of 0.08 in the responses. By contrast, at a distance from the category boundary, the sensitivity of the model to differences between stimuli is decreased.

This result models the basic characteristics of CP: an increased sensitivity to differences at the category boundary, and a diminished sensitivity within the categories.

Domain Integration: The McGurk Effect

Many experiments have shown that visual information can enhance the understanding of speech, suggesting an integration of the visual with the auditory signal in this task (see Massaro, 1998, for an overview). Striking evidence for the strength of this integration comes from the

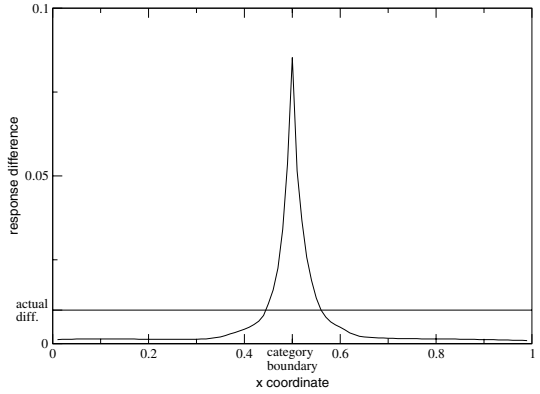

Figure 5: CP in the model: sensitivity to differences is increased around the category boundary.

McGurk effect (McGurk and MacDonald, 1976): when subjects are presented with conflicting visual and auditory data, their perception of what is said can be different from both the visual and the auditory signal. For example, when a face articulating /ga/ is dubbed with the auditory /da/, subjects perceive /ba/. This effect is highly pervasive and not subject to volitional control. It is not restricted to vision and auditory integration, but has also been found for touch and audition (Fowler and Dekle, 1991).

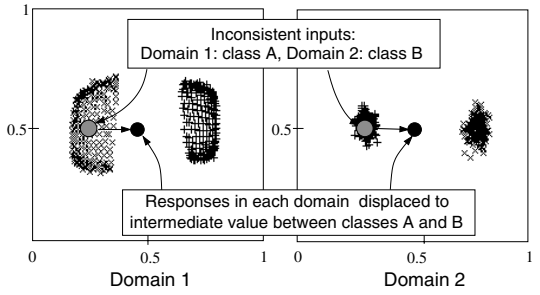

Figure 6: Exemplary response of the model to a data pair that does not correspond to the learned correlational structure. The previously learned responses are denoted by x and +, the data pair that does not correspond to the learned correlational structure by grey circles, and the response of the model to this data pair by black circles.

To test whether the model displayed a response similar to the McGurk effect in humans, data pairs were presented that did not correspond to the previously learned correlation structure. While during training the "left" half of the data set for domain 1 co-occurred with the "right" cluster in domain 2, now data from the "left" half in domain 1 was presented together with that from the "left" cluster of domain 2. Conceptually this corresponds to presenting e.g., an auditory /da/ together with a visual /ga/. The model integrated these conflicting inputs to a

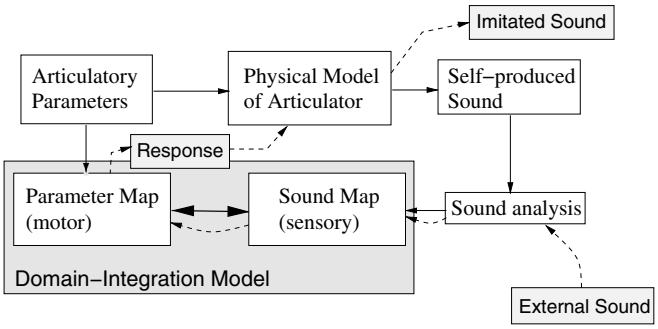

Figure 7: The model for sensori-motor integration and the imitation of sounds. Solid arrows indicate the "babbling phase" where the mapping from motor to sensory parameters is learned. The dashed arrows show the pathway for the subsequent imitation of sounds.

response that was a blend between the responses to each individual input (fig. 6).

While the McGurk effect has been studied in great detail and has revealed many results that are much more subtle than a simple blend between the auditory and visual information, the present model can give a principled account of how the domain integration that lies at the basis of this effect can arise. The details of the McGurk effect cannot be modeled with the artificial data set used here to investigate the general functionality of the model, but experiments are planned to use a more realistic set of auditory and visual signals that will give more detailed results.

In summary, the domain integration model displays, without the explicit teaching and labeling of categories, several of the effects that are generally supposed to rely on such labeling, namely, the formation of prototypes as attractors in the stimulus space, categorical perception in an evenly distributed set of stimuli, and an integration of stimuli from different domains to form a unified percept that forms a "compromise" when conflicting data is presented in the domains simultaneously.

The Model in Sensori-Motor Integration

In the previous sections it was described how the domain integration model integrates between two sensory domains, leading to psychologically observed phenomena such as prototype formation, categorical perception, and the McGurk effect. In this section, an extension to this model is proposed to account for sensori-motor integration (fig. 7). This extension works by presenting in one domain a representation of an action (e.g., motor parameters), and in the other, a representation of the sensory consequences of that action. The model then learns the associations between the motor commands and their sensory consequences, developing simultaneously in both domains prototypes of actions and consequences of these actions, based on a reliable correlation between them.

The sensori-motor variant of the model was tested on sound production. For this purpose, a physical model of a speech synthesizer (Boersma and Weenink, 1996) was used. In initial experiments, two parameters, jaw opening and the position of the styloglossus muscle (a muscle that controls the position of the back of the tongue) were varied continuously at 18 steps each, and the resulting sounds were analyzed with respect to their firsts two formant values. The model was trained on the resulting two-domain data set with 324 items. Fig. 8 shows the initial and final responses of the model. While the motor parameters are evenly distributed prior to training, after training prototypical parameter-sound pairs have formed in both domains due to their correlational structure.

The sensori-motor integration model corresponds to the ideomotor principle which postulates a tight coupling between perception and action. As such it can give an account of the imitation of sounds (fig. 7, fig. 8B): an external sound that is presented to the model evokes a response on the auditory map. This response is propagated through the developed Hebbian connections to the motor map where a motor response is evoked which can be used to articulate, i.e., imitate, the corresponding sound. However, the imitation of the heard sound is displaced towards a prototype that the model has developed during training (indicated by an arrow in the auditory map in fig. 8B). In this way, imitation is not merely a reproduction of an external stimulus, but a re-interpretation of that stimulus based on the developed structure of the model.

Discussion

The model described in this paper presents an algorithm to integrate sensory information between two domains to form a unified percept, thereby displaying phenomena also observed in human categorization. The model can be extended to also account for sensori-motor integration and the imitation of low level percepts. While the simple data sets used in this paper were used to demonstrate the principled functionality of the model, more realistic and

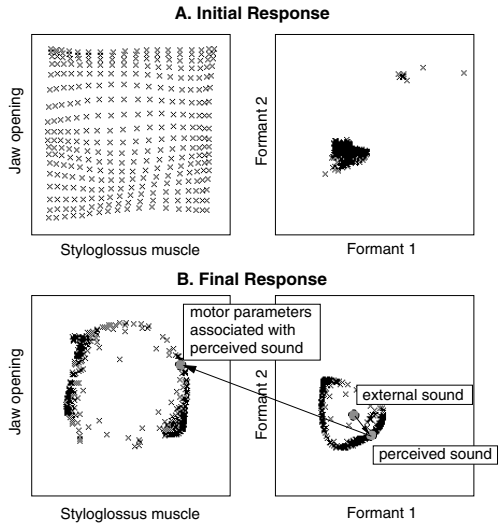

Figure 8: A. Initial and B. final responses of the sensori-motor integration model.

extensive experiments are necessary to establish whether it can account for more detailed results in these domains. We have now started to use higher-dimensional data for the learning of the articulation–perception mapping in sound production and imitation, and preliminary results look promising.

An important property of the model is that it shows a unified account of sensori-sensor and sensori-motor integration in a neurobiologically inspired framework.

An alternative view of this model could be as a variant of supervised category learning: when one map receives the inputs (i.e., object representations) and the other the targets (i.e., category labels), the model learns the mapping from the category members to their labels if there is a sufficient number of different categories. The domain integration model, however, adds an important aspect that is often neglected in supervised category learning models: not only category members, but also the concept of "category" has a topology and is changed by its members. For example, the "concepts" of the dog and cat categories will move closer together on the target map if their members share properties. In this way it becomes possible to measure the similarity between concepts by investigating the developed topology on the target map.

In its present form the model is simple, though it allows insights into how perception can change due to categorization. However, more realistic training data, as well as an extension of the model to be able to handle sequential and more complex data, are necessary. These will be the next steps in the described research.

Acknowledgments I would like to thank Eduardo Miranda for providing the data set for the sound imitation experiments.

References

Boersma, P. and Weenink, D. (1996). Praat, a system for doing phonetics by computer. Technical Report 132, Institute of Phonetic Sciences of the University of Amsterdam.

Calvert, G., Brammer, M., Bullmore, E., Campbell, R., Iversen, S., and David, A. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, **10**, 2619–2623.

de Sa, V. R. and Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation*, **10**, 1097–1117.

Fowler, C. and Dekle, D. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 816–828.

Georgopoulos, A. P., Kettner, R. E., and Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neural population. *Journal of Neuroscience*, **8**, 2928–2937.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, **6**, 298–304.

Goldstone, R. L., Steyvers, M., and Larimer, K. (1996). Categorical perception of novel dimensions. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, pages 243–248, Hillsdale, NJ. Erlbaum.

Guenther, F. H. and Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustic Society of America*, **100**, 1111–1121.

Harnad, S. (1987). *Categorical Perception*. Cambridge University Press, Cambridge.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–608.

Massaro, D. W. (1998). *Perceiving Talking Faces*. MIT Press, Cambridge, MA.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746–748.

Schyns, P. G., Goldstone, R. L., and Thibaut, J. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, **21**, 1–54.

Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, **4**, 303–312.

Tijsseling, A. and S.Harnad (1997). Warping similarity space in category learning by backprop nets. In M. Ramscar, U. Hahn, E. Cambouropolos, and H. Pain, editors, *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization*, pages 263 – 269. Department of Artificial Intelligence, Edinburgh University.