Transferring knowledge between learning systems

Andrew P Paplinski
Monash University
Melbourne, Australia
andrew.paplinski@monash.edu

William M Mount
University of New South Wales
Canberra, Australia
w.mount@adfa.edu.au

***Abstract*—We consider a pair of multimodal integration systems, a teacher and a learner and demonstrate how the knowledge is transferred from the teacher to the learner in an incremental fashion. By combining effectors to a perceptual integration network influenced by hypothesised endogenous "action thoughts", we present a simplified model of how a universal or common coding scheme can be used to represent dual auditory and visual sensory pathways and integrate these with articulatory and gestural forms of motor control. As an example the system learns visual representation of Chinese characters and their related Mandarin pronunciation.**

***Index Terms*—Learning Systems; Multimodal Integration; Visual Input; Auditory Input; Dual Coding; Common Coding; Self-Organizing Networks;**

I. Introduction

We postulate that the process of learning involves integration of information presented in different modalities, most typically auditory and visual. In this context the term multimedia learning has been recently used [1].

From classic psychological studies and more contemporary experiments in multimedia learning, it is well established that memory of verbal information is enhanced if relevant visual information is presented or imagined at the same time [2], [3]. This is also supported by fMRI (functional Magnetic Resonance Imaging) and PET (Positron Emission Tomography) studies, which have shown that different regions of the brain are activated to varying degrees during language processing, when spoken words heard by the subject are either paired or not readily paired with associated visual imagery [4], [5].

This type of learning originates from the dual theory of cognition [6], [7] that postulates that both visual and auditory information is used to represent information. Dual-coding theories complement a dual-route theory of reading, originally developed in the 1970's, which explain how readers simultaneously access orthographic and phonological information in order to recognize written words [8], [9]. According to this and related theories of speech and writing, the bimodal representation of language is fundamental to effective language comprehension, generation and acquisition.

Dual coding theory has it that different types of neuronal code are used for storage of visual and auditory information, based on the differences between the widespread, partially lateralised brain networks associated with each. These networks correspond to the visuospatial sketchpad and the phonological loop as described within Baddeley's model of working memory [10].

Visual information is usually taken to be represented by *analogue* codes which retain the main perceptual features of the underlying phenomenon being represented, while auditory and especially verbal information by *symbolic* codes, which may be more arbitrary and conceptual in nature [7]. The mental codes corresponding to these representations are used to organize incoming information that can be stored, acted upon and retrieved for subsequent use.

In our recent work we consider two models of learning systems [11], [12] that can integrate such bimodal information and multimodal information more generally. While to date we have not incorporated functional differences supposed for the visuospatial sketchpad and the phonological loop in these or the current model, the dual coding regime and dual stream reading framework [6], [9] is essentially retained.

In particular, the system presented in [12] describes a model of binding written words to mental objects, whereas the system from [11] integrates visual information, namely, rendering of Chinese characters with auditory information, in this case, Mandarin articulation of the related characters.

In this paper we present a model of transferring knowledge between such learning systems, a teacher and a learner and draw conclusion about the quality of such a process.

II. The structure of the learning systems

Our work on multimodal integration models has been strongly influenced by the following two models from the neuroanatomy area.

Firstly we refer to the dual-stream model presented in [13]–[16]. This model identifies seven general networks of processing speech information [13]. It starts at the *spectrotemporal analysis* module followed by the *phonological network* from which the processing diverges into two broad streams: the *articulatory stream* and the *lexical stream*. These two streams are interconnected by the *combinatorial network* integrating lexical and articulatory processing, and are also connected at the higher level to the widely distributed *conceptual network*.

Secondly, the model of neuronal circuitry for reading as presented in [9] includes thirteen interconnected cortical areas, arranged in five groups: visual input, visual word form, access to meaning, access to pronunciation and articulation, and top-down attention and serial reading.

In our work we use a much simplified model consisting of just five 'cortical' areas.

One of the basic premises of our modelling framework is the concept of a ubiquitous neuronal code, which implies a unified way of representing information exchanged by modules of the network.

In Figure 1 we present a general structure of two learning systems considered in this paper, a teacher and a learner. Each

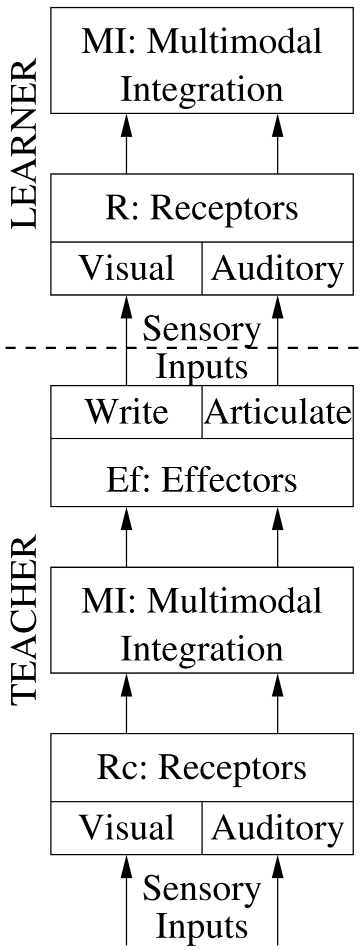

Fig. 1. Two learning systems: a teacher and a learner. Note three main building blocks: receptors, multimodal integration and effectors

learning system presented in Figure 1 consists of the following main parts:

- Rc — Receptors that receive the external sensory information, auditory and visual in our case,
- MI — Multimodal Integration part that interprets the sensory information and incorporates it within the internal knowledge structure of self-organizing modules
- Ef — Effectors that produce an external representation of knowledge, articulation and writing effectors in our case.

Receptors and the multimodal integration part of the overall system are described in some detail below and in greater detail in [11]. As an example we consider the system that learns Chinese characters and related Mandarin utterances.

For the visual receptor channel we use an angular integral of Radon Transform (aniRT) as 1-D descriptor of 2-D images. Details are discussed in [17]. For the auditory receptor channel, we follow our previous work [18], [19] and use melcepstral coefficients to represent related utterances.

The phonological information stored directly within the unimodal auditory maps and more indirectly within the fused bimodal map encodes a definite temporal relationship, however ultimately the bimodal words are represented as single conceptual entities with both spatial/orthographic and temporal/phonological properties.

All of the information stored within our model can be said to more closely resemble analogue codes in a dual-coding regime, since for both auditory and visual information this is represented as a universal encoding of the spatial and structural relationships between anIRT and melcepstral coefficients for characters and utterances, respectively. On the hand, we have previously considered the problem of binding of written or spoken names to mental objects [12] which we believe shows that the same system can be modified to represent a more symbolic form of encoding in which names or labels can be arbitrarily re-assigned.

The multimodal part of the learning systems has been considered in [11] and consists of (see Figure 2) five intercon-

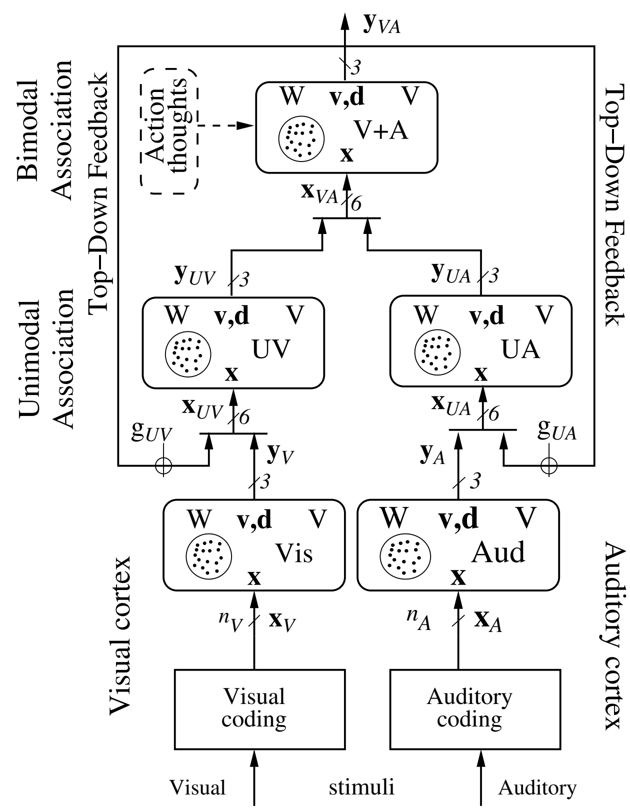

Fig. 2. The multimodal integration part of the learning system. Note four hierarchical levels: sensory coding, sensory processing, unimodal association and the bimodal integration/association.

nected self-organizing modules. Two **sensory** level modules, **Vis** and **Aud**, process visual and auditory stimuli, respectively, converting coded sensory information, $\mathbf{x}_V$ and $\mathbf{x}_A$ into the standard internal representation of signals $\mathbf{y}_V$ and $\mathbf{y}_A$. In the next hierarchical level, two **unimodal association** modules,

UV and **UA**, combine the signals from the sensory level, $\mathbf{y}_V$ and $\mathbf{y}_A$, with the modulating top-down feedback signals, $\mathbf{y}_{VA}$ [20], produced by the top level **bimodal association** module, **V+A**.

The bimodal association module is the central part of the learning system. We hypothesize that this module may be activated by endogenous "action thoughts", or "thought commands" that drive the effector systems, one for writing and one for articulation.

Information exchange between five learning modules, namely, the strength and position of the winner in the latent space, can be considered as an internal neuronal code. Such a code represents in a uniform way all internal signals transferred between modules.

Learning in the system can occur in an incremental fashion as described in [21]. The number of neuronal nodes (represented by dots in Figure 3 is generated to be statistically proportional to the current number of stimuli learned by the system.

III. The Transfer of Knowledge from the Teacher to the Learner

The teacher has its knowledge stored in the three modules, as described above. The bimodal map represents the top level of the hierarchy. An example of the knowledge structure represented by the bimodal module is given in Figure 3. The

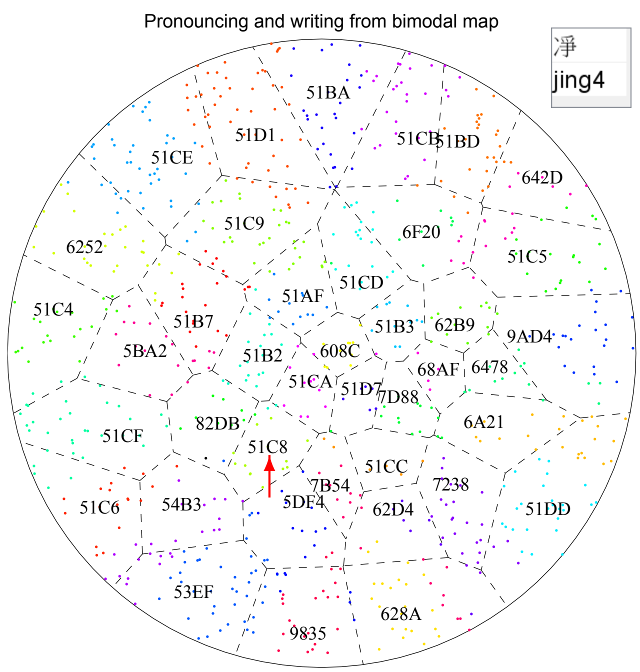

Fig. 3. Selecting an object labeled by a hexadecimal number 51C5 and represented by a Chinese character and related Mandarin articulation as indicated in the box in the upper right corner. Selected bimodal object will be used to drive the teacher's effectors and the learning process in the learner

hexadecimal annotation is used to emphasize the integrated representation of the audio-visual information. Neuronal units that respond strongly, that is, above the set threshold, to the given object/stimulus are marked in the same colour. In addition, for ease of orientation the tessellation of the neuronal space is marked by dashed lines. For each selected object we also show the Chinese character and its pronunciation in pinyin induced by the selection and produced by the relevant 'write' and 'articulate' effectors (see Figure 1).

The transfer of knowledge between the teacher and the learner can occur in one of the following three modes:

- Incrementally from the "fully learned" teacher.
- Concurrently with the teacher in the incremental way,
- All in one step (batch mode)

We will concentrate on the first mode in which the teacher's knowledge structure is already fixed. In this case the transfer of the knowledge occurs as follows. The teacher's exogenous action thought acts upon the bimodal map (see Figure 2). Such an action thought results in the effectors' articulation and writing action as in Figure 3. This pair of signals form the stimuli inputs for the learner. We can assume that the action thought is selected in accordance to the teacher's endogenous goal, which, in general, drives the learning process according to some performance measure.

Initially the learner is in the *tabula rasa* state so that its five maps are empty. Incrementally, for each teacher's action, the learning takes place as described in [21]. In Figure 4 we show an example of two unimodal association maps and the bimodal map after twelve learning/knowledge-transfer steps have been performed. The modules form their latent spaces according to information received from the sensory level maps and the modulating feedback from the bimodal map. We aim to maintain a statistically constant number of neuronal units per stimulus ϵ, say $\epsilon \approx 16$. Hence, after learning is completed for γ stimuli, the total number of neuronal units will be $\gamma \times \epsilon$, where γ is the total number of stimuli.

Another way of looking at the behaviour of the modules is to plot the surface, or the equivalent image of postsynaptic excitations of all neuronal units for a given stimulus. Such an image is shown in Figure 5. We can see the strength of postsynaptic excitation for the bimodal map of Figure 4 for the stimulus 'leng3'. Comparing with the bimodal map from Figure 4, we can notice that for the stimulus 'leng3', there is a significant excitation equivalent to stimulus 'liang2' because of the joint visual and auditory similarities as perceived by the bimodal map.

If the process of learning is continued until all the knowledge from the teacher system is transferred to the learner system, the final result might look like the one in Figure 6. For convenience the maps are annotated using the pinyin representation of Chinese characters. The teacher and the learner maps are different thus emphasising the fundamental fact that the teacher and the learner are different individuals in the sense that they have formed different bimodal associations between the written and spoken language components, or more generally, that they created different views of their limited "worlds" due to the history of the learning process.

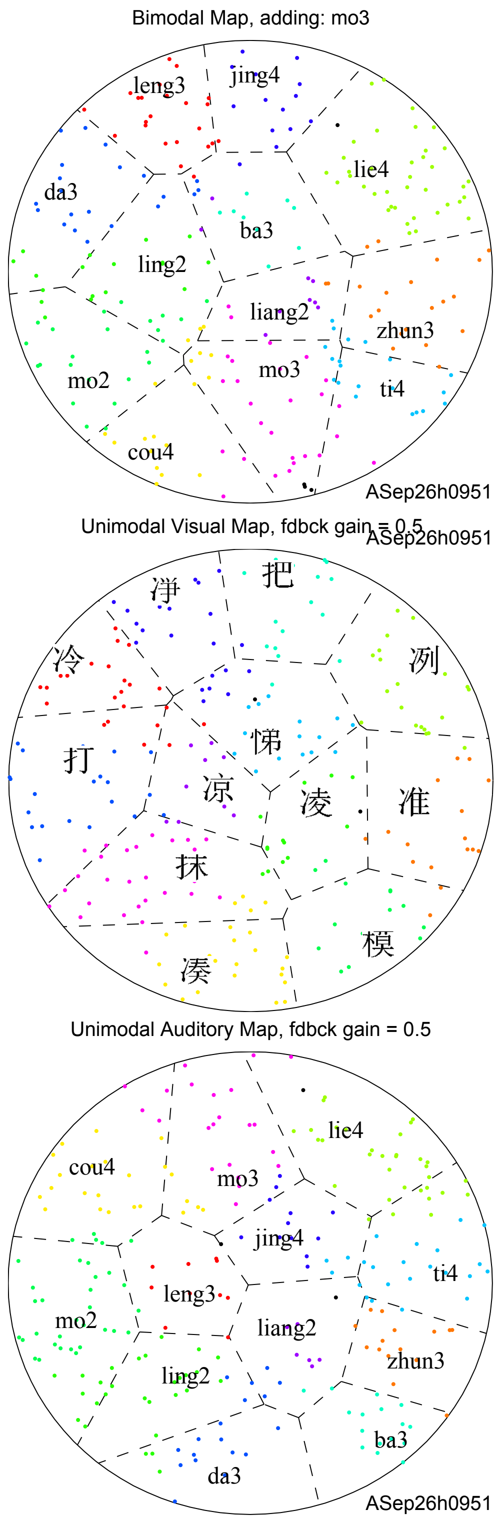

Fig. 4. Two unimodal association maps and the bimodal map after 12 knowledge transfer steps. The objects in the bimodal map are, for convenience, labeled with pinyin rather than with the equivalent hexadecimal codes for characters as in Figure 3

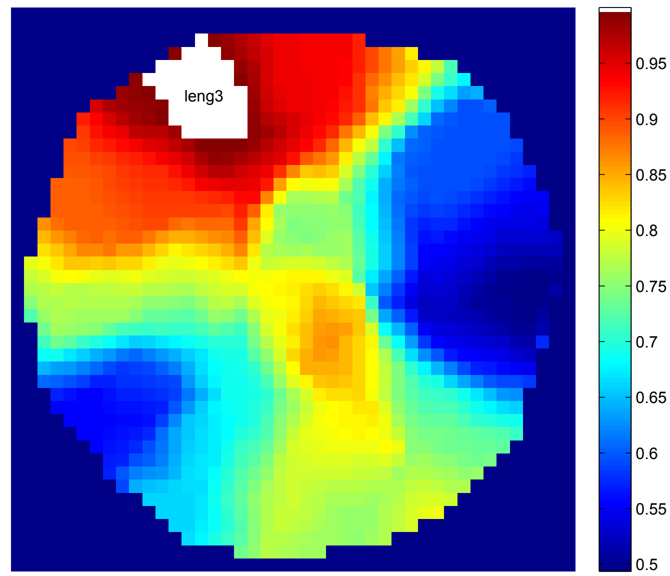

Fig. 5. The image of postsynaptic excitations in the bimodal map of Figure 4 for the stimulus ‘leng3’

IV. Concluding Remarks

The above experiment shows how it is possible to extend a simple self-organising processing and integration network based on dual stream reading and listening to incorporate an effector stage sharing similar bimodal perceptual fusion of words; in this case written Chinese characters and their spoken utterances.

Through this we may move from consideration of *dual* coding to a *common* coding scheme [22] in which actions are postulated to be represented in terms of their perceptual consequences. Perception of environmental stimuli and events, internal cognitive manipulation of these and generation of perception related effects can be seamlessly integrated within such a neuro-architectural framework, using a common and ubiquitous neuronal code.

It is worth noting that some known behaviour of *mirror neurons* which respond both to specific actions performed by an individual and perception of the same actions performed by others [23] can also be represented within such a model. In this simple teacher-learner scenario, the action is the expression by the teacher and perception the impression made upon the learner. As both of these are registered and may be initiated or influenced by endogenous thought commands, we can speculate that the bimodal integration layer of our model plays an analogous role to such mirror neuron systems.

The bimodal layer in this seven-module network incorporates simplistic endogenous action thought inputs which directly activate fused percepts and a mechanism to express these either via written (or gestural) or spoken (articulatory) effector pathways. By addition of a centralised attentional modulation and control subsystem with recurrent connections, however, it may be possible to extend this to implement a form of “working memory” which coordinates perceptual and action thought patterns.

In this regard, the type of complex activation patterns shown in Figure 5 are likely to be relevant, since these may be considered not only to indicate the current conscious percept or

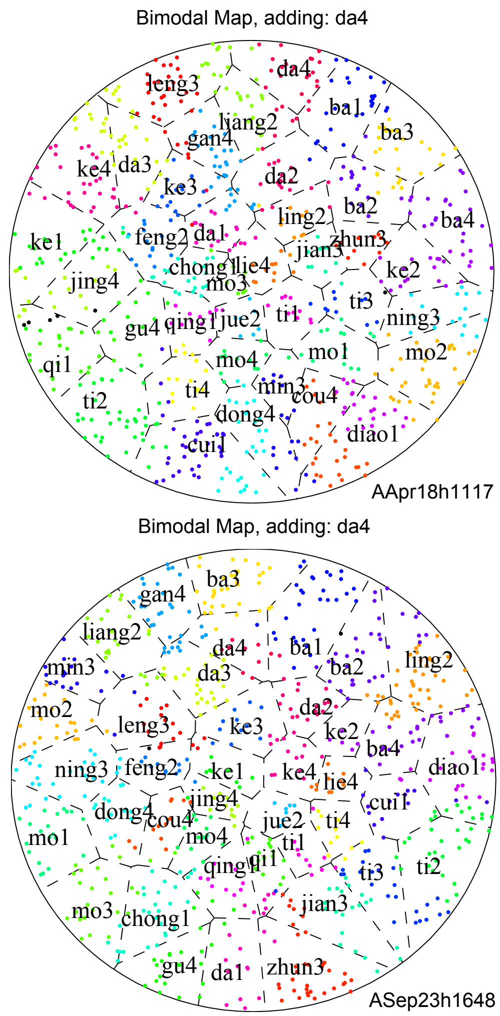

Fig. 6. Bimodal maps for the teacher (top) and the learner (bottom). As expected, bimodal associations between the written and spoken language components developed differently between the teacher and the learner.

active response, but also unconscious candidates of alternative perceptions and perception-influencing actions which might become amplied or attenuated through interaction with central excitatory and/or inhibitory regulation systems.

For example, such nascent activations may be evaluated by a separate system linked to internal drives and desired goal states, while the efferent activation signals relayed back via an attentional control system could influence the conscious sequence of perceptions and actions represented by bimodal word units.

Finally, by completion of the loop within our two network system, our future aim is to show how the interactions between learner and teacher and the repeated interpretation and expression of percepts by each, can result in a complex cycle of endogenous and exogenous influence representing a "conversation" guiding the flow of information from teacher to learner network within a integrated supervised and self-organised learning process.

REFERENCES

[1] R. E. Mayer, *Multimedia Learning*. Cambridge University Press, 2001.

[2] J. R. Anderson and G. H. Bower, *Human Associative Memory*. Washington, DC: Winston, 1973.

[3] T. T. Brunye *et al.*, "Repetition and dual coding in procedural multimedia presentations," *Applied Cognitive Psychology*, vol. 22, pp. 877–895, 2008.

[4] M. Denis and E. Mellet, "Neural bases of image and language interactions," *International Journal of Psychology*, vol. 37, no. 4, pp. 204–208, 2002.

[5] M. Just *et al.*, "Imagery in sentence comprehension: an fMRI study," *NeuroImage*, vol. 21, pp. 112–124, 2004.

[6] A. Paivio, *Mental Representations*. Oxford University Press, 1986.

[7] R. J. Sternberg, *Cognitive theory*, 3rd ed. Thomson/Wadsworth, 2003.

[8] M. Coltheart, "Modeling reading: The dual-route approach," in *The science of reading: A handbook*, M. Snowling and C. Hulme, Eds. Blackwell: Oxford University Press, 2005, pp. 6–23.

[9] S. Dehaene, *Reading in the Brain*. Viking, 2009, http://pagesperso-orange.fr/readinginthebrain/figures.htm.

[10] A. D. Baddeley, *Working memory*. Oxford: Oxford University Press, 1986.

[11] A. P. Papliński and W. M. Mount, "Bimodal incremental self-organizing network (BiSON) with application to learning Chinese characters," in *LNCS*, vol. 8226. Springer, 2013, pp. 121–128.

[12] A. P. Papliński, L. Gustafsson, and W. M. Mount, "A recurrent multimodal network for binding written words and sensory-based semantics into concepts," in *LNCS*, B.-L. Lu, L. Zhang, and J. Kwok, Eds., vol. 7062. Springer, 2011, pp. 413–422.

[13] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Rev. Neurosci.*, vol. 82, pp. 393–402, 2007.

[14] J. Binder, R. Desai, W. Graves, and L. Conan, "Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies," *Cerebral Cortex*, vol. 19, pp. 2767–2796, 2009.

[15] C. Price, "The anatomy of language: a review of 100 fMRI studies published in 2009," *Ann. N.Y. Acad. Sci.*, vol. 1191, pp. 62–88, 2010.

[16] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing," *Nature Neurosci.*, vol. 12, no. 6, pp. 718–724, 2009.

[17] A. P. Papliński, "The angular integral of the Radon transform (aniRT) as a feature vector in categorization of visual objects," in *ISNN, LNCS*, vol. 7951. Springer, 2013, pp. 523–531.

[18] A. P. Papliński, L. Gustafsson, and W. M. Mount, "A model of binding concepts to spoken names," *Aust. Journal of Intelligent Information Processing Systems*, vol. 11, no. 2, pp. 1–5, 2010.

[19] T. Jantvik, L. Gustafsson, and A. P. Papliński, "A self-organized artificial neural network architecture for sensory integration with applications to letter–phoneme integration," *Neural Computation*, vol. 23, pp. 2101–2139, 2011.

[20] A. Mechelli, C. Price, K. Friston, and A. Ishai, "Where bottom-up meets top-down: neuronal interactions during perception and imagery," *Cerebral Cortex*, vol. 13, pp. 1256–1265, Nov. 2004.

[21] A. P. Papliński, "Incremental self-organizing map (iSOM) in categorization of visual objects," in *ICONIP, LNCS*, vol. 7664. Springer, 2012, pp. 125–132.

[22] W. Prinz, "Modes of linkage between perception and action," in *Cognition and motor processes*, W. Prinz and A. F. Sanders, Eds. Berlin: Springer, 1984, pp. 185–193.

[23] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annual Review of Neuroscience*, vol. 27, pp. 169–192, 2004.