A MULTIMODAL SELF-ORGANIZING NETWORK FOR SENSORY INTEGRATION OF LETTERS AND PHONEMES

ABSTRACT
Integration of signals from sensory receptors of different modalities is known to enhance perception. Integration takes place in bimodal and multimodal association areas of neocortex and results in robust bimodal or multimodal percepts as well as in feedback mediated enhanced unimodal perception. A Multimodal Self-Organizing Network, MuSON, is presented as a tool for simulating sensory integration. This MuSON is a development of an earlier version in that it takes into account also the degree of recognition of stimuli in the various maps of the network. The simulation results show the same characteristics as corresponding results from psychology and neuroscience.

KEY WORDS
self-organization, neural networks, letters, phonemes, sensory integration

1 Introduction

Many phenomena are manifest in two or more sensory modalities. Such is, e.g., the case of speech which can be seen in lip movements and heard. Bimodal or multimodal integration of sensory information is advantageous in such cases because the perception of the phenomena becomes more robust against noise in one or more modalities. This has been established in the case of audiovisual speech in e.g. [1]. For an extensive review of studies in bimodal and multimodal integration, see [2].

It has long been known that bimodal sensory integration takes place in association areas, e.g. in the superior temporal polysensory area (STP), see e.g. [3, 4], but multimodal convergence also occurs earlier in cortical sensory processing see e.g. [5, 6].

There are different mechanisms for integrating signals conveying auditory and visual information onto a neural structure. Both feed forward (bottom up) connections from lower levels to higher levels in the neural hierarchy and feedback (top down) connections going in the opposite direction serve to integrate information from different sensory modalities, see e.g. [7, 8].

Both feed forward and feedback in neural processing have been extensively studied. A presented stimulus will cause a rapid feed forward sweep of activity with a short delay at each hierarchical level [9]. This activity is subsequently modulated by feedback.

Feedback plays an important role in the processing of audiovisual speech. Speech is processed in a network of cortical regions, see [10] for a review, with early processing taking place in sensory specific cortices [11, 5, 12]. Processing for phoneme perception takes place in the left posterior Superior Temporal Sulcus (STSp), see e.g. [13, 14]. Integration of the two modalities of audiovisual speech takes place in the multimodal association area in the Superior Temporal Sulcus (STS) and the Superior Temporal Gyrus (STG) [12], located between the sensory-specific auditory and visual areas.

Audiovisual speech exists in two forms, lip reading and hearing and reading letters and hearing. In both forms the auditory perception is enhanced compared to purely auditory speech, see [15, 16]. The activity in unisensory auditory cortex is increased due to feedback from the bimodal area in the STS to auditory cortex [17, 18]. Letters are processed in unisensory visual cortex in or close to the left fusiform gyrus, see [19, 20, 21]. Bimodal integration of phonemes and letters takes place through feed forward processing in the STS [22, 18].

We have earlier [23, 24, 25, 26] modelled the processing of phonemes and letters in the sensory-specific areas and in the bimodal association area. We used a multimodal self-organizing network (MuSON), consisting of maps with phonetic and graphic inputs respectively, and an integrating bimodal map, corresponding to the cortical architecture described above. Feedback from the bimodal association area to the auditory cortex was also modelled in the auditory module.

With this model we have demonstrated [24, 25] that bimodal percepts are robust against additive noise in the letters and phonemes and that this robustness of the bimodal percepts is “transferred” down the auditory processing stream by feedback. The results from simulations with this model suggest that we hear a noisy phoneme better when we see the corresponding uncorrupted letter.

The feedback from the bimodal area to the auditory area should cause activity there even in the absence of auditory stimuli, provided a visual stimulus is present. It has been shown that there is activation in auditory cortex during lip reading [5], even though the sound has been eliminated. In this paper we will show that our model also exhibits this property. To do this we extend our model further with more versatile modules than we have previously been using [24, 25].

2 The multimodal self-organizing network

Self-organizing neural networks have been inspired by biological neural systems. Kohonen Self-Organizing Maps (SOMs) are well-recognized and much researched tools for mapping multidimensional stimuli onto a low dimensionality (typically 2) neuronal lattice, for an introduction and a review, see [27]. In this paper we will employ a network of interconnected modules, referred to as a Multimodal Self-Organizing Network (MuSON), see [23, 24, 25], consisting of SOMs and SumSOMs (Summing Self-Organized Maps). The SumSOM is introduced in this paper to enable the MuSON to fuse signals from different modules while taking their adherent activity levels (i.e. intensities) into account.

We first consider our previous feedforward Multimodal Self-Organizing Network (MuSON) as depicted in Figure 1. The pre-processed, sensory stimuli, $\mathbf{x}_{lt}$ and $\mathbf{x}_{ph}$ form the inputs to their respective unisensory maps, SOM$_{lt}$ and SOM$_{ph}$. Two-dimensional outputs from these maps, $\mathbf{y}_{lt}$ and $\mathbf{y}_{ph}$, are combined together to form a four-dimensional stimulus for the higher-level bimodal map, SOM$_{bm}$. The training of these maps is done in sequential order with the sensory maps being trained first whereupon the bimodal map is trained with concatenations of the winner neuron positions, in the two former maps, for corresponding sensory inputs. All maps are trained according to the well-known Kohonen learning law, see [23, 24] for details. After self-organizations each map performs a mapping of the form: $\mathbf{y}(k) = F(\mathbf{x}(k); W, V)$, where $\mathbf{x}(k)$ represents the k^{th} stimulus for a given map, W is the weight map, and V describes the structure of the neuronal grid. The 2-D output signal $\mathbf{y}(k)$ gives the 2-D position of the winner.

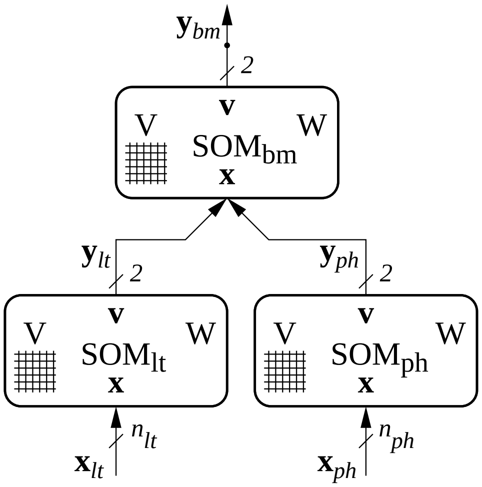

Figure 1. A two-level feedfoward-only Multimodal Self-Organizing Network (MuSON) processing auditory and visual stimuli consisting of Kohonen maps (SOMs). The auditory stimuli are processed in SOM_{ph}, and the visual stimuli in SOM_{lt}. Bimodal integration then takes place in SOM_{bm}.

We will now extend this MuSON with the intention of incorporating feedback, from the bimodal to the auditory processing unit, and enhancing its ability to exchange information between different processing units. This latter enhancement makes use of the activity levels generated by the processing units. Since the activity level of a neuron reflects how well its input agrees with previous training data, the response grading thus correlates with how well the submitting unit recognizes its input. Thus, as a consequence of this enhancement the units can communicate graded responses and these gradings can then be used to perform weighted response fusion. To carry out this fusion while also taking the degree of congruency between the responses into account, relations between the signals pending fusion have to exist, and these relations must be exploited using a technique that yields a fused signal that is meaningful.

3 A MuSON with SumSOMs

In this extended architecture we utilize SOMs whose output signals do not merely consist of the position $\mathbf{v}$ of a winner neuron, resolved via the structural description V as before, but also the activity level a of this neuron. To fuse two outputs of this kind in the desired way we employ signal transformation and postsynaptic activity combination. More specifically, we introduce a neuronal network configuration SumSOM which, using the two mentioned concepts, integrates a pair of outputs coming from Kohonen SOMs and/or SumSOMs and classifies this integrated signal. The transformation is applied to at least one of the SumSOMs incoming signals, prior to integration, to permit the integration step to be straight-forward.

The SumSOM produces an output of the same kind that the standard SOMs do; a classification coded as a 2-D position of maximum activity and the activity level at that location. Therefore, the output from a SumSOM can be interpreted the same way as the output of the standard SOM. Our intention is for these SumSOMs to be seen as extensions to Kohonen's SOMs, allowing the processing of more than one input signal while replicating the SOMs behaviour during its application phase.

We employ this configuration with two variations; one in which the output from one SOM is transformed in order to enable modulation of the postsynaptic activity field of another, and one in which the output signals from two SOMs are both transformed so that the fused response may lie in a space that differs from both of the SOMs output spaces. We call the former variation a SumSOM of type 1 and the latter a SumSOM of type 2.

An outline of the extended architecture, using the described configurations as building blocks, is depicted in Figure 2. Initializing this architecture requires that the three maps discussed in section 2 are already organized as described. The initialization essentially consists of automatically arranging the neuron weights in the SumSOMs, using data from the three self-organized maps, in such a way that congruent inputs to them produce correctly fused outputs. More detailed accounts of how this is done are presented in section 4. What is important to note is that the additional initialization steps needed here does not alter the organizations produced by the previous training. Nor does it need any manual interventions if the network is reset and the training process restarted.

The same sensory stimuli as previously are input to the sensory processing units here, via $\mathbf{x}_{lt}$ and $\mathbf{x}_{ph}$. The feedback signals, $\mathbf{v}_{bm}$ and a_{bm}, going from the bimodal to the auditory processing unit are undefined from the outset resulting in $\mathbf{x}_{ph}$ being fused with a weak noise signal. The two sensory units output the positions of their respective maximum activities, on $\mathbf{v}_{lt}$ and $\mathbf{v}_{ph}$, along with the magnitude of these activities, on a_{lt} and a_{ph}. The bimodal processing area then fuses these signals and feeds its classification contained in $\mathbf{v}_{bm}$ back to the auditory SumSOM together with the activity level a_{bm}. Processing in the auditory unit is modulated by this feedback, resulting in a new output, and this recurrent process continues until it either converges or reaches a maximum number of allowed iterations.

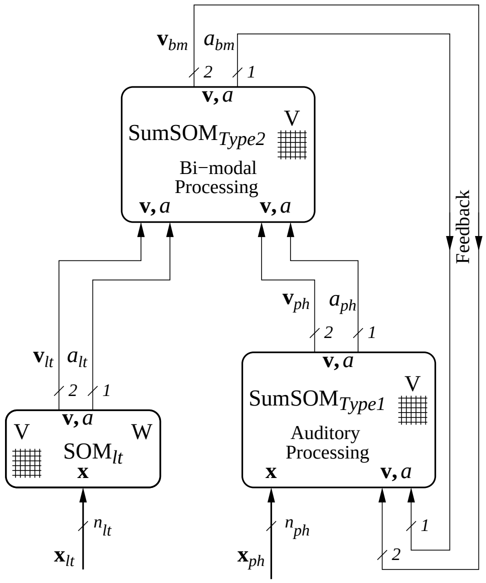

Figure 2. A two-level MuSON with feedback processing auditory and visual stimuli. The MuSON consists of SOMs and SumSOMs. SumSOMs combine signals that may come from SOMs and/or SumSOMs.

4 SumSOM details

Figure 3 shows a schematic of a type 1 SumSOM which thus integrates the output of a SOM, coming in on $\mathbf{v}_{1\ in}$ and $a_{1\ in}$ (as a 2-D position of a neuron and its activity), with the postsynaptic activity field of the other SOM, labelled SOM_0. The main prerequisite for initializing the neuron network TM_1 in a way that enables the SumSOM to achieve a meaningful integration of its training data is that there is a one-to-one correspondence between the $\mathbf{x}_{SOM_0}$ training data set and training data set of the SOM that generates the signals coming in on $\mathbf{v}_{1\ in}$. Assuming this prerequisite is fulfilled and that both SOMs are already organized the initialization of TM_1 aims at enabling it to become a transformation map that transforms its inputs into SOM_0's output space. The initialization begins by letting the neural lattice in TM_1 be of the exact same type as in SOM_0. The weights of TM_1's neurons W_1 are then modified in such a way that the winner in this network has the same position as the winner in SOM_0 for each corresponding pair $< \mathbf{x}_{SOM_0};\ \mathbf{v}_{1\ in} >$ in the training data. A neuron that did not get its weights modified by the previous procedure gets assigned with a varied version of the weight vector of its closest neighbouring neuron that did. This variation depends on the distance between the two neurons and the angle of the straight line connecting them. The aim is to create patches in the transformation network that become highly active for the training samples of $\mathbf{v}_{1\ in}$, as well as for small variations of those samples, that coincide with the patch of maximum activity that appears in SOM_0 for the corresponding training samples of $\mathbf{x}_{SOM_0}$. When all neurons have been assigned weights, the initialization phase is completed by nullifying the weights of all those neurons that are located on patch peripheries. This last step is merely done to ease visual interpretations of the map's postsynaptic activity surfaces.

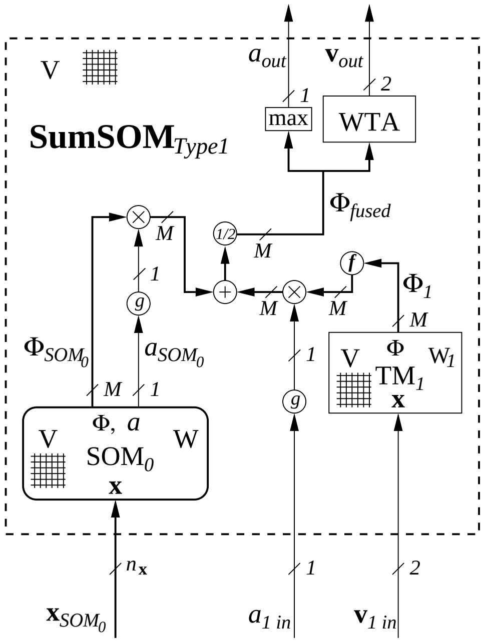

Figure 3. Outline of $SumSOM_{Type1}$ which fuses the input, $(\mathbf{v}_{1\ in}, a_{1\ in})$, coming from a SOM/SumSOM together with SOM_0's response.

Having trained and initialized SOM_0 and TM_1 respectively, SumSOM$_{Type1}$ fuses its input pairs together by transforming and superposing the induced activity fields of its neuron networks, treats the result as an integrated response field, and forms the fused output as location and intensity of the maximum activity in this field. In detail, the respective activity fields in SOM_0 and TM_1, caused by inputs on $< \mathbf{x}_{SOM_0};\ \mathbf{v}_{1\ in} >$, are forwarded as Φ_{SOM_0} and Φ_1. To bring about a combination of these fields that reflects how well the stimuli are recognized, Φ_{SOM_0} is multiplied with $g(a_{SOM_0})$ while Φ_1 is multiplied with $g(a_{1\ in})$, where the function g rescales activity levels in a way that adequately amplifies the difference between high and low activities. Prior to these multiplications the activity field Φ_1 is transformed using the function $\boldsymbol{f}$ which subtracts the minimal positive activity level from the field and then rescales all levels equally with the factor needed to restore the peak activity to its original magnitude. One way of viewing this operation is to interpret it as a kind of lateral inhibition, and it is aiming to make the maximal activities in the transformation map more prominent. Once these operations have been performed the activity fields are superimposed and rescaled with the constant $\frac{1}{2}$, forming a new field Φ_{fused} that is functionally comparable to those created by the two dimensional neuronal lattices that our Kohonen maps consist of, and the same methods of determining the maximum activity's position and the activity intensity can be used, thereby forming the unit's output $(\mathbf{v}_{out}, a_{out})$.

The other variation of our SumSOM allows us to perform weighted signal fusion of outputs coming from two SOMs (and/or SumSOMs) while replicating the responses of a third (a template SOM). A schematic is shown in figure 4. As with $SumSOM_{Type1}$, setting up meaningful integration of the inputs of a $SumSOM_{Type2}$ requires a one-to-one correspondence of the training data set that generates the signals on $\mathbf{v}_{1\ in}$ and $\mathbf{v}_{2\ in}$. Additionally, the setup phase of a type 2 SumSOM also needs an external template that dictates the resolution of the SumSOM's response field and supplies a predetermined winner position for each corresponding pair in the training data. Here, the template consists of a Kohonen SOM that has been trained with the concatenations of the SOMs' signals $\mathbf{v}_{1\ in}$ and $\mathbf{v}_{2\ in}$ generated when these are presented with corresponding samples of the training data (i.e. as described in section 2). Set up thus begins with letting the neuron lattices in both TM_1 and TM_2 be of the exact same type as in the template SOM. The weights of these networks are then assigned so that corresponding training signals on $\mathbf{v}_{1\ in}$ and $\mathbf{v}_{2\ in}$ yield the same positions of winner neurons in both networks and that these positions agree with the winner positions generated by the template SOM. Initialization of the both maps concludes as with TM_1 in the type 1 SumSOM as described earlier.

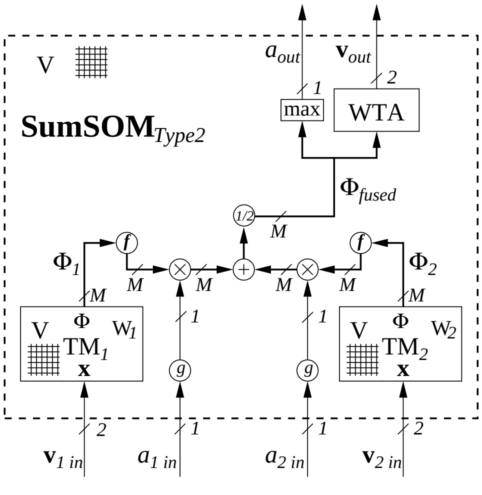

Figure 4. Outline of $SumSOM_{Type2}$ which fuses together the outputs, $(\mathbf{v}_{1\ in}, a_{1\ in})$ and $(\mathbf{v}_{2\ in}, a_{2\ in})$, of two SOMs/SumSOMs.

The operation of a $SumSOM_{Type2}$ after it has been initialized is similar to that of a $SumSOM_{Type1}$. Input signals received via $< \mathbf{v}_{1\ in};\ \ \mathbf{v}_{2\ in} >$ induces activity fields in the two transformation maps TM_1 and TM_2. Before these activity fields are superposed they are both transformed, through the application of the function $\boldsymbol{f}$, to increase the numerical significance of the maximum activity levels. The altered fields $\boldsymbol{f}(\Phi_1)$ and $\boldsymbol{f}(\Phi_2)$ are also rescaled with $g(a_{1\ in})$ and $g(a_{2\ in})$, respectively, where g has the same purpose as previously described. After superposition the combined field is element-wise multiplied with the constant $\frac{1}{2}$ and the result Φ_{fused} is the unit's response field. The position of the peak activity and its magnitude are respectively output via $\mathbf{v}_{out}$ and a_{out}.

5 The maps for letters, phonemes and bimodal percepts

The pre-processed stimuli inputs $\mathbf{x}_{lt}$ and $\mathbf{x}_{ph}$ to the letter and auditory processing maps consist 22- and 36-element vectors respectively. The pre-processing of stimuli is described in [24]. Resulting self-organized maps for letters

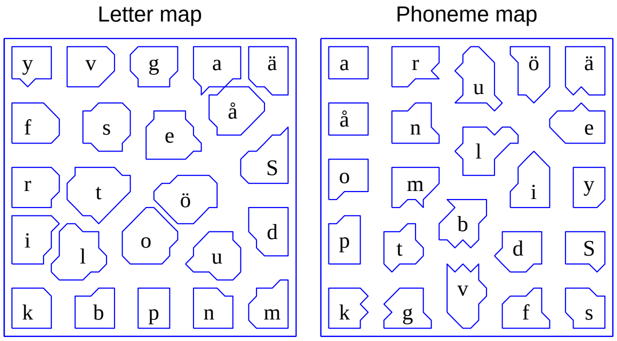

Figure 5. Patches of highest activity for labelled letters and phonemes after self-organization. In the letter map the response field consist of the output signals of 36 × 36 neurons. The phoneme map's response field is composed of $2 \times\ (36 \times 36)$ fused neuron signals.

and phonemes are shown in Figure 5. For a discussion of these maps and the phonetic typewriter from 1988 by Kohonen, see [24, 25, 28, 27].

The patches in Figure 5 cover populations of neurons which show the highest activity for their respective stimuli. These neuronal populations constitute the detectors of the respective stimuli.

The bimodal map integrates letters and phonemes as shown in Figure 6. The similarity characteristics of this map are derived from the placement of the patches in the letter and phoneme maps and thus only indirectly reflect the features of the letters and phonemes.

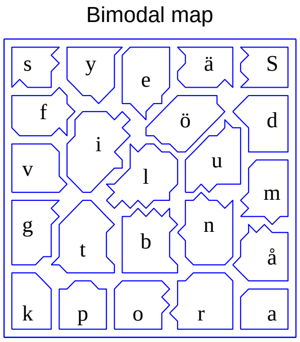

Figure 6. Bimodal map. Patches of highest activity for labelled letter/phoneme combinations after self-organization. Response field consists of $2 \times (36 \times 36)$ fused neuron signals.

The maps in Figures 5 and 6 have been obtained through self-organization of the original MuSON summarized in section 2 and have been retained in the extended MuSON by initialization in the expanded MuSON of Figure 2.

6 Robustness of the bimodal percepts and the recoded phoneme map against unimodal disturbances

When uncorrupted letters and corrupted phonemes are presented to the extended MuSON, the resulting bimodal percepts are very stable against the corruption of the phonemes as are the auditory processing results due to feedback from the bimodal map. As stated in the introduction, this is an important advantage of sensory integration.

We choose to illustrate this with the uncorrupted letter p and the corrupted phoneme p in Figure 7. It is seen that the auditory processing at first yields a borderline t. But after only one loop through feedback the classification changes to a p, and after six loops the activity peak has reached the ideal position; the position for an uncorrupted phoneme p. The initial bimodal percept is within the neuronal patch for p and reaches the ideal position for p after five loops. This very powerful corrective action against corruption is typical for all letter and phoneme pairs.

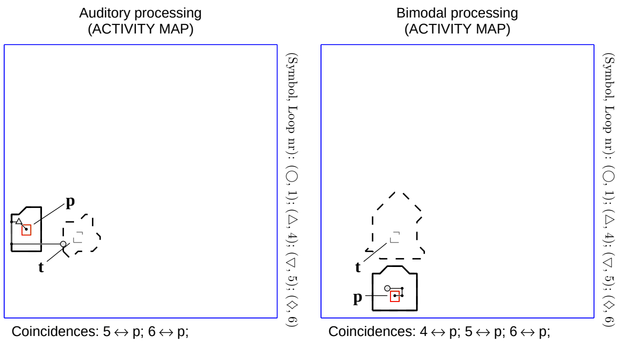

Figure 7. The peak activities through loops are shown as dots and symbols on the progression curves. The ∘-symbols show the initial positions of these activities. Phoneme corruption consists of mixing the uncorrupted versions p and t together.

7 Activation of auditory cortex by visual speech alone

As stated in the Introduction there is activation of parts of auditory cortex during silent lip reading [5]. We will here show how this effect of sensory integration, based on the model given in [18], is also manifest when visual speech consists of letter reading. Silence in our experiment is characterized by very low initial activity in the auditory processing; this level has been chosen to be approximately 10% of the activity caused by an uncorrupted phoneme. Silence will have an initial winner position in the auditory processing map, but the coordinates of this position becomes unimportant in the subsequent processing since the adherent activity level is low. Figure 8 depicts the results from a simulation run of our extended MuSON when it has been presented with the letter i together with "silence": The maximum activity in the auditory processing unit is located at the ideal position after only one feedback loop, whereas the bimodal percept immediately manages to dampen the

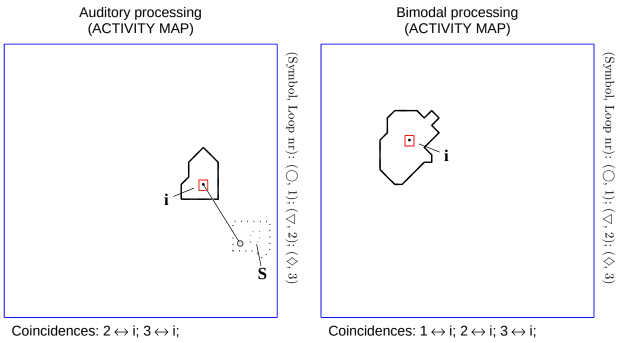

Figure 8. The MuSON received "silence" and the letter i. Initial peak activity in the auditory unit is indicated by the symbol ∘.

weak auditory signal sufficiently for it to not have an influence. The peak activity levels generated by both units are somewhat lower compared to the levels attained when the MuSON is presented with the uncorrupted phoneme i instead of silence.

8 Conclusion

We have shown that modelling bimodal integration of audiovisual speech consisting of phonemes and letters with our extended Multimodal Self-Organizing Network yields results that agree with known results from psychology and neuroscience.

Acknowledgment

The authors acknowledge financial support from the 2006 Monash University Small Grant Scheme, and the Kempe Foundation in Sweden.

References

[1] W. Sumby and I. Pollack, "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am.*, vol. 26, pp. 212–215, 1954.

[2] E. G. Calvert, C. Spence, and B. E. Stein, *The handbook of multisensory processes*, 1st ed. Cambridge, MA: MIT Press, 2004.

[3] C. Schroeder and J. Foxe, "The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex," *Cognitive Brain Research*, vol. 14, pp. 187–198, 2002.

[4] C. Schroeder, J. Smiley, K. Fu, T. McGinnis, M. O'Connell, and T. Hackett, "Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing," *Int. J. Psychophysiology*, vol. 50, pp. 5–17, 2003.

[5] G. A. Calvert, E. T. Bullmore, M. J. Brammer, R. Campbell, S. C. Williams, P. McGuire, P. W. Woodruff, S. D. Iversen, and A. S. David, "Activation of auditory cortex during silent lipreading," *Science*, vol. 276, pp. 593–596, 1997.

[6] J. Driver and C. Spence, "Crossmodal attention," *Curr. Opin. Neurobiol.*, vol. 8, pp. 245–253, 1998.

[7] G. A. Calvert and T. Thesen, "Multisensory integration: methodological approaches and emerging principles in the human brain," *J. Physiology Paris*, vol. 98, pp. 191–205, 2004.

[8] J. J. Foxe and C. E. Schroeder, "The case for feedforward multisensory convergence during early cortical processing," *Neuroreport*, vol. 16, no. 5, pp. 419–423, April 2005.

[9] V. Lamme and P. Roelfsema, "The distinct modes of vision offered by feedforward and recurrent processing," *Trends Neuroscience*, vol. 23, pp. 571–579, 2000.

[10] C. J. Price, "The anatomy of language: contributions from functional neuroimaging," *J. Anat.*, vol. 197, pp. 335–359, 2000.

[11] J. R. Binder, J. A. Frost, T. A. Hammeke, P. S. F. Bellgowan, J. A. Springer, J. N. Kaufman, and E. T. Possing, "Human temporal lobe activation by speech and nonspeech sounds," *Cerebral Cortex*, vol. 10, pp. 512–528, 2000.

[12] G. A. Calvert and R. Campbell, "Reading speech from still and moving faces: The neural substrates of visual speech," *J. Cognitive Neuroscience*, vol. 15, no. 1, pp. 57–70, 2003.

[13] G. Dehaene-Lambetrz, C. Pallier, W. Serniclaes, L. Sprenger-Charolles, A. Jobert, and S. Dehaene, "Neural correlates of switching from auditory to speech perception," *NeuroImage*, vol. 24, pp. 21–33, 2005.

[14] R. Möttönen, G. A. Calvert, I. Jääskeläinen, P. M. Matthews, T. Thesen, J. Tuominen, and M. Sams, "Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus," *NeuroImage*, vol. 19, 2005.

[15] R. Frost, B. Repp, and L. Katz, "Can speech perception be influenced by simultaneous presentation of print?" *J. Mem. Lang.*, vol. 27, pp. 741–755, 1988.

[16] T. Dijkstra, U. H. Frauenfelder, and R. Schreuder, "Bidirectional grapheme-phoneme activation in a bimodal detection task," *J. Physiology Paris*, vol. 98, no. 3, pp. 191–205, 2004.

[17] G. Calvert, R. Campbell, and M. Brammer, "Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex," *Current Biology*, vol. 10, pp. 649–657, 2000.

[18] N. van Atteveldt, E. Formisano, R. Goebel, and L. Blomert, "Integration of letters and speech sounds in the human brain," *Neuron*, vol. 43, pp. 271–282, July 2004.

[19] I. Gauthier, M. J. Tarr, J. Moylan, P. Skudlarski, J. C. Gore, and A. W. Anderson, "The fusiform "face area" is part of a network that processes faces at the individual level," *J. Cognitive Neuroscience*, vol. 12, no. 3, pp. 495–504, 2000.

[20] T. A. Polk and M. J. Farah, "The neural development and organization of letter recognition: Evidence from functional neuroimaging, computational modeling, and behavioral studies," *PNAS*, vol. 98, pp. 847–852, February 1998.

[21] T. A. Polk, M. Stallcup, G. K. Aguire, D. C. Alsop, M. D'Esposito, J. A. Detre, and M. J. Farah, "Neural specialization for letter recognition," *J. Cognitive Neuroscience*, vol. 14, no. 2, pp. 145–159, 2002.

[22] T. Raij, K. Uutela, and R. Hari, "Audiovisual integration of letters in the human brain," *Neuron*, vol. 28, pp. 617–625, November 2000.

[23] A. P. Papliński and L. Gustafsson, "Multimodal feedforward self-organizing maps," in *Lect. Notes in Comp. Sci.*, vol. 3801. Springer, 2005, pp. 81–88.

[24] L. Gustafsson and A. P. Papliński, "Bimodal integration of phonemes and letters: an application of multimodal self-organizing networks," in *Proc. Int. Joint Conf. Neural Networks*, Vancouver, Canada, July 2006, pp. 704–710.

[25] A. P. Papliński and L. Gustafsson, "Feedback in multimodal self-organizing networks enhances perception of corrupted stimuli," in *Lect. Notes in Artif. Intell.*, vol. 4304. Springer, 2006, pp. 19–28.

[26] S. Chou, A. P. Papliński, and L. Gustafsson, "Speaker-dependent bimodal integration of Chinese phonemes and letters using multimodal self-organizing networks," in *Proc. Int. Joint Conf. Neural Networks*, Orlando, Florida, Aug. 2007.

[27] T. Kohonen, *Self-Organising Maps*, 3rd ed. Berlin: Springer-Verlag, 2001.

[28] ——, "The "neural" phonetic typewriter," *Computer*, pp. 11–22, March 1988.