Bimodal Integration of Phonemes and Letters: an Application of Multimodal Self-Organizing Networks

Lennart Gustafsson

Computer Science and Electrical Engineering
Luleå University of Technology, Sweden
Email: Lennart.Gustafsson@sm.luth.se

Andrew P. Papliński

Clayton School of Information Technology
Monash University, Australia
Email: app@csse.monash.edu.au

Abstract—Multimodal integration of sensory information has clear advantages for survival: events that can be sensed in more than one modality are detected more quickly and accurately, and if the sensory information is corrupted by noise the classification of the event is more robust in multimodal percepts than in the unisensory information. It is shown that using a Multimodal Self-Organizing Network (MuSON), consisting of several interconnected Kohonen Self-Organizing Maps (SOM), bimodal integration of phonemes, auditory elements of language, and letters, visual elements of language, can be simulated. Robustness of the bimodal percepts against noise in both the auditory and visual modalities is clearly demonstrated.

I. INTRODUCTION

Bimodal and more generally multimodal integration of sensory information is important for perception since many phenomena have qualities in more than one modality. We can both see and hear a car crash to take a drastic example. If we are close to the crash we might even detect an ominous smell of gasoline, causing us to take appropriate action. Multimodal integration has been studied extensively, for a comprehensive review, see [1]. Recently, functional magnetic resonance imaging (fMRI) has made non-invasive studies of multimodal identification and recognition by human subjects increasingly tractable and important, for a review see [2].

It has long been known that multimodal stimuli from the same phenomenon makes it possible for us to detect the phenomenon quicker [3], a characteristic of obvious survival value. Multimodal convergence occurs early in sensory processing in unimodal (thus rendering them not exclusively unimodal) sensory cortices [4], [5] with very little time delay from stimulus onset [6]. It has also been well documented in subcortical nuclei [7], [8].

Another important advantage of multimodal integration is that detection and identification of events is more robust under circumstances when disturbances are present [9]. Individual neurons with multimodal inputs in e.g. Superior Colliculus (SC) have been found to have superadditive response, i.e. the response to multimodal stimuli is stronger than the added responses to unimodal stimuli [8]. Such superadditive enhancement often follows the principle of inverse efficiency, i.e. the enhancement is strongest when one modality is weak [8]. Superadditive response obeying the principle of inverse

efficiency has also been documented on the behavioural level in humans [10].

Given the importance of language for human communication it is not surprising that we have an extensive language system with processing resources in several different cortical regions, for a review see [11]. Language has properties in two modalities, we hear speech and we see speech in mouth movements through lipreading. We have areas for processing auditory speech in sensory-specific auditory cortex and for visual speech in visual cortex [12], [4], [13]. The processing of speech sounds for phoneme perception, as opposed to “sine wave speech” with the same acoustical properties, has distinct neural resources in the left posterior Superior Temporal Sulcus (STSp) see e.g. [14], [15].

Through fMRI studies, bimodal integration of audiovisual speech has been found in the multimodal association area in the Superior Temporal Sulcus (STS) and the Superior Temporal Gyrus (STG) [13]. This area is suitably located between the sensory-specific auditory and visual areas. A number of studies have shown that audiovisual speech is more robust against noise in both modalities than auditory speech alone, see e.g. [16]. It has also been found that audiovisual speech is more rapidly identified than auditory speech alone, see e.g. [17].

Language has visual properties also in the form of written language. Perhaps surprisingly we acquire, through learning, neural resources in the STS and in other areas of cortex, dedicated to bimodal processing of spoken language and written language. The existence of neural resources in or close to the left fusiform gyrus, active in processing of letters as opposed to processing of visually similar digits, has been documented in several studies [18], [19], [20]. The integration of phonemes and letters takes place in the STS [21], [22].

Since written language is a recent human invention it is not possible to explain this as an evolutionary development. Rather it is proof of the plasticity of cortical organization, reading is important to us, we spend considerable time learning to read and to read, and dedicated neural resources enable us to do this efficiently.

Spoken language can be segmented on different levels,

such as phonemes, syllables, words and phrases. In many written alphabetical languages, such as Italian, Finnish and Russian, there is a good correspondence between letters or graphemes and phonemes. The correspondence on the syllable, word and phrase levels is then obvious. In English the correspondence between letters and phonemes is somewhat weaker and in other written languages such as Chinese, the written characters carry no phonetic information. It is not surprising that fMRI recordings show that when reading words out aloud Italian readers of Italian texts have a larger activation in the STS than English readers of English texts. The latter rely more on the left frontal cortex, higher up in the language processing system [23].

In this paper we study auditory and visual language on the lowest level, phonemes and letters. For this purpose we employ a multimodal self-organizing network (MuSON) consisting of two unimodal maps, receiving phonetic and graphic inputs respectively, forwarding their results to an integrating bimodal map. This architecture broadly mirrors the cortical architecture for the phonetic processing in the STSp, the visual processing of letters in the fusiform area and the bimodal integration in the STS. Multimodal integration is known to be mediated by both feedforward and feedback connections, see e.g. [24], [25]. This is also the case for bimodal speech processing, see [22]. Feedback is not represented in this study, it is the topic for further research.

We show how the templates for phonemes and letters resulting from self-organization of the unimodal maps integrate into templates for the bimodal percepts in the bimodal map. We furthermore demonstrate the robustness of the bimodal map against additive noise in the inputs to the unimodal maps.

II. THE MULTIMODAL SELF-ORGANIZING NETWORKS

Self-organizing neural networks have been inspired by the possibility of achieving information processing in ways that resemble those of biological neural systems. In particular, pattern associators based on Hebbian learning [26] and self-organizing maps [27] show similarities with biological neural systems. Pattern associators have been employed to simulate the multimodal sensory processing in cortex [28].

Kohonen Self-Organizing Maps (SOMs) are well-recognized and intensively researched tools for mapping multidimensional stimuli onto a low dimensionality (typically 2) neuronal lattice. A significant number of recent work are devoted to a variety of applications using a variant of the basic Kohonen algorithm for a single SOM. In [27] Kohonen discussed possible variations of SOMs that include: maps of varying topology that has been studied, for example, in [29], [30], and tree-structured SOMs to improve the winner search procedure, e.g. [31]. An interesting modification of a SOM in which multiple winners are employed and local intramap connections are trained using a temporally asymmetric Hebbian learning rule has been presented in [32].

However, to our knowledge, the network of interconnected SOMs, designated “for future research” in [27] p.194, has been only considered in our recent work [33]. In our current

paper we demonstrate that such a network of interconnected SOMs, referred to as Multimodal Self-Organizing Networks (MuSONs) may serve as a good simulation model for biological sensory processing.

As an example of a Multimodal Self-Organizing Network (MuSON) we consider interconnections of two-levels of SOMs as in Figure 1. The first level SOMs receive pre-

Fig. 1. A two-level Multimodal Self-Organizing Network (MuSON) for processing auditory and visual stimuli. The auditory stimuli are processed in SOM₁₂, and the visual stimuli in SOM₁₁. Bimodal integration then takes place in SOM₂₁

processed auditory x_{12} and visual x_{11} stimuli and generate the position/activity signals y_{12} and y_{11} , respectively. These signals are fed to the second-level bimodal map as $x_{21} = \{y_{11}, y_{12}\}$. The position/activity signals consists of the L -dimensional location of the winner and its intensity measure. Dimensionality of each map is, in general, equal to $l_{1,k}$, but is, most typically, equal to 2 for the ease of visualisation. In such a case dimensionality of the x_{21} is six. Each SOM performs mapping according to the following formula:

$$y(n) = g(x(n); W, V); \quad x \in \mathcal{R}^N, \quad y \in \mathcal{R}^L \quad (1)$$

where each W represents the weight matrix, V describes the positions of neurons, and $g(\cdot)$ represents an input-output mapping of a given SOM.

In a MuSON as in Figure 1 learning takes place concurrently for all SOMs, according to the well-known Kohonen learning law. In the practical example presented below we work with normalised stimuli and activity data located on respective hyper-spheres, therefore we use the simple “dot-product” learning law [27], [34]. In this case the update of the weight vector w_j for the j th neuron is described by the following expression:

$$\Delta w_j = \eta \cdot \Lambda_j \cdot (x^T - d_j \cdot w_j); \quad d_j = w_j \cdot x \quad (2)$$

where Λ_j is a neighbourhood function, Gaussian in our case, centred on the position of the winning neuron, and d_j is

the **post-synaptic activity** of the j th neuron. It is easy to show that for the above learning law, if the stimuli $\mathbf{x}$ are on a unity hyper-sphere, the resulting weight vectors $\mathbf{w}$ are located on, or close to such a sphere. For consistency, we also map neuron positions on a 3-D sphere. Consequently, the efferent signals $\mathbf{y}$ are being kept on a 4-D hyper-sphere.

III. THE VISUAL MAP FOR LETTERS

The data for the letter map has been obtained from the representation of letters in the 12-point Times New Roman font. Each letter is represented by a binary 21×25 image. These images are scanned vertically so that each letter is represented by a 525-element binary stimulus.

In order to reduce the dimensionality of the letter stimuli we perform principal component analysis thus reducing the dimensionality of the letter stimuli to $N = 22$, the total number of letters being equal to $K = 23$. In the final pre-processing step we project all letter vectors up on the unity hyper-sphere of dimensionality $N + 1$.

After self-organization according to equation (2) is complete, the properties of the visual letter map SOM_{11} may be summarized as in Figure 2. Some details of such a map

Fig. 2. Letter map. Patches of highest activity for labeled letters after self-organization on a map of 36×36 neurons.

formation are presented in sec. VI.

When an k -th letter represented by a vector $\mathbf{x}_{11}(k)$ is presented to the self-organized map $\text{SOM}_{11}(k)$, the output activity $d(\mathbf{x}_{11}(k))$ is highest from a population of neurons in a patch. This patch can be visualized by thresholding the activity at a level close to the maximum output level from the neurons. The resulting visual map for all the $K = 23$ letters in our material is shown in Figure 2. Neuronal populations in patches have developed as letter detectors.

Looking at the letter map of Figure 2 we can note that visually similar letters are placed in close proximity. Note for example the cluster of the letters **a**, **ä** and **ä**.

IV. THE AUDITORY MAP FOR PHONEMES

Phonotopic maps have a prominent place in the development of self-organizing neural networks. In 1988 Kohonen presented the “phonetic typewriter”, a SOM that learned to identify Finnish phonemes [35]. For a discussion of this and further developments, see [27]. In our study we let ten Swedish speakers, five male and five female, read a number of words, from which we parsed the initial phoneme, $K = 23$ phonemes in all. These phonemes constitute our auditory material — the automatic speech recognition task is deliberately kept simple in this study. A set of $N = 36$ melcepstral coefficients was determined for each phoneme by each speaker. These feature vectors were averaged over the speakers, yielding one thirty-six element feature vector $\mathbf{x}_{12}(k)$ for each phoneme. This averaged set of vectors constituted the inputs to the auditory map. The melcepstral representation of speech is a standard representation, one advantage being that the pitch of the speaker is filtered away which makes the averaging of phonemes as spoken by male and female speakers meaningful. For a discussion of the melcepstrum and its use for representation of speech see [36].

As in the letter map, neuronal populations in patches have developed during self-organization as phoneme detectors. The resulting auditory map SOM_{12} is shown in Figure 3. Note that the plosives **g**, **k**, **t** and **p**, the fricatives **s**, **S** (this

Fig. 3. Phoneme map. Patches of highest activity for labeled phonemes after self-organization on a map of 36×36 neurons

is our symbol for the sh-sound as in English ‘she’) and **f** and the nasal consonants **m** and **n** form three close groups on the map. Vowels with similar spectral properties are placed close

to each other. The back vowels **a**, **â** and **o** are in one group, the front vowels **u**, **ö**, **ä**, **e**, **y** and **i** in another group with the tremulant **r** in-between. The exact placing of the groups vary from one self-organization to another, but the existence of these groups is certain.

V. THE BIMODAL MAP INTEGRATING PHONEMES AND LETTERS

The outputs from the auditory phoneme map and the visual letter map combine as inputs to the bimodal map SOM_{21} . Self-organization results in the map shown in Figure 4. The

Fig. 4. Bimodal map. Patches of highest activity for labeled letter/phoneme combinations after self-organization on a map of 36×36 neurons

similarity characteristics of this map are derived from the placement of the patches in the unimodal maps and thus only indirectly reflect the featural characteristics of the phonemes and letters. The fricative consonants **s**, **S** and **f** form a group in the combined map as do the nasal consonants **m** and **n**. Most, but not all, vowels form a group and those who are isolated have obviously been placed under influence from the visual letter map.

VI. SOME DETAILS OF THE ACTIVITY MAPS FORMATION

With reference to Figure 1, for each pair of letter/phoneme stimuli, $\{\mathbf{x}_{11}(k), \mathbf{x}_{12}(k)\}$ we can calculate relevant post-synaptic activities of all neurons in all three maps

$$d_{11}(\mathbf{v}_{11}, k) = \mathbf{w}_{11}(\mathbf{v}_{11}) \cdot \mathbf{x}_{11}(k), \quad (3)$$

$$d_{12}(\mathbf{v}_{12}, k) = \mathbf{w}_{12}(\mathbf{v}_{12}) \cdot \mathbf{x}_{12}(k), \quad (4)$$

$$d_{21}(\mathbf{v}_{21}, k) = \mathbf{w}_{21}(\mathbf{v}_{21}) \cdot \mathbf{x}_{21}(k) \quad (5)$$

where vectors $\mathbf{v}$ represent position of neurons on respective neuronal lattices.

As an example, in Figure 5 we plot the activity surface $d_{21}(\mathbf{v}_{21}, k)$ in a bi-modal map SOM_{21} for a pair k of

Fig. 5. The activity in the trained bimodal map when the letter and the phoneme **o** is the input to the sensory-specific maps. The patch of neurons representing **o** is clearly distinguished from other patches

unimodal stimuli representing phoneme/letter **o**. We note one winning patch with activity descending away from this patch as illustrated in Figure 5. Thresholding the activities in Figure 5 we obtain one patch of activities as presented in the complete bimodal map of Figure 4.

Similarly, in Figure 6 we present images of activities of the trained bimodal map for each phoneme/letter applied to the lower level maps. Note again that the bimodal map as in Figure 4 has been created by thresholding activities for each stimulus as presented in Figure 6.

VII. ROBUSTNESS OF THE BIMODAL PERCEPTS AGAINST UNIMODAL DISTURBANCES

An important advantage of integration of stimuli from sensory-specific cortices into multimodal percepts in multimodal association cortices is that even large disturbances in the stimuli may be eliminated in the multimodal percepts. Our model has the same advantage, as can easily be demonstrated.

In the letter map in Figure 2, we can observe that the letter **i** is given a semblance with **l**, the **ä** is given a semblance with **ä**, and the **u** a semblance with **n**. The first two disturbances are quite realistic, the third maybe less so. Now, if we corrupt **i** with **l**, **ä** with **ä** and **u** with **n**, the resulting activities move as shown in Figure 7. The solid lines represent activities for the base letters, whereas dash-dotted lines represent activities when the inputs consist of letters heavily corrupted by neighboring letters.

Similarly, we create corrupted phoneme stimuli. We use three unclearly pronounced phonemes as inputs to the auditory phoneme map as in Figure 3. An **i** is pronounced very close to a **y**, and an **â** is pronounced very close to an **o**. The corrupted stimuli are formed by linearly combining the corresponding melcepstral coefficient vectors and the resulting activity map is shown in Figure 8. The first two

Fig. 6. The activities in the trained bimodal map for each phoneme/letter stimulus applied to the lower level maps. An additional test stimulus is marked with ‘-’

disturbances, that is, **i** with **y** and **å** with **ø** are quite realistic but less realistic is the **u** pronounced with a large influence from an **n**. The activity on the phoneme map is lowered and the maximum activity moves away from the pure phoneme patches in the direction of the patch of the disturbing phoneme, as is shown in Figure 8.

The three pairs of corrupted stimuli **i**, **aa** and **u** are now applied to the second-level bimodal map. As expected, in the bimodal map illustrated in Figure 9, the activities for the disturbed **i** and **å** have moved very little as indicated by solid and dash-dotted lines, respectively. The recognition of these bimodal percepts is much less influenced by the disturbances than the recognition in the unimodal maps. In the case of the **u** there is no improvement, but no deterioration either. No case of deterioration of recognition in the multimodal map over recognition in the unimodal maps has been found. This clearly demonstrates the robustness of the multimodal representation.

VIII. CONCLUSION

It has been shown that multimodal integration of sensory information in different modalities, in this case bimodal integration of auditory information in phonemes and visual

Fig. 7. The maximum activities, shown by solid lines in the letter map when the inputs consist of different letters and by dash-dotted lines when the inputs consist of letters heavily corrupted by neighboring letters

Fig. 8. The maximum activities, shown by solid lines in the phoneme map when the inputs consist of different phonemes and by dash-dotted lines when the inputs consist of phonemes heavily corrupted by neighboring phonemes

information in letters, can be achieved through the use of Multimodal Self-Organizing Networks (MuSON). The robustness of the resulting bimodal percepts has been demonstrated.

Fig. 9. The maximum activities, shown by solid lines in the bimodal map when the auditory and visual inputs to the sensory-specific maps are perfect and by dash-dotted lines when the inputs to the unisensory maps are heavily corrupted. In two of the three cases the bimodal percept is seen to be very robust against corruption of the unimodal inputs

ACKNOWLEDGMENT

The authors acknowledge financial support from the 2005 Monash University Small Grant Scheme.

REFERENCES

- [1] (Eds:) Gemma Calvert, Charles Spence, and Barry E. Stein, *The handbook of multisensory processes*, MIT Press, Cambridge, MA, 1st edition, 2004.
- [2] A. Amedi, K. von Kriegstein, N. M. van Atteveldt, M. S. Beauchamp, and M. J. Naumer, "Functional imaging of human crossmodal identification and object recognition," *Exp. Brain Res.*, vol. 166, pp. 559–571, 2005.
- [3] M. Hershenson, "Reaction time as a measure of intersensory facilitation," *J. Exp. Psychol.*, vol. 11, no. 3, pp. 289–293, 1962.
- [4] G. A. Calvert, E. T. Bullmore, M. J. Brammer, R. Campbell, S. C. Williams, P. K. 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.
- [5] J. Driver and C. Spence, "Crossmodal attention," *Curr. Opin. Neurobiol.*, vol. 8, pp. 245–253.
- [6] A. Fort, C. Delpuech, J. Pernier, and Giard. M-H., "Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans," *Cerebral Cortex*, vol. 12, pp. 1031–1039, October 2002.
- [7] B. Gordon, "Receptive fields in deep layers of cat superior colliculus," *J. Neurophysiol.*, vol. 36, pp. 157–178, 1973.
- [8] M. T. Wallace, M. A. Meredith, and B. E. Stein, "Multisensory integration in the superior colliculus of the alert cat," *The American Physiological Society*, pp. 1006–1010, 1998.
- [9] D. E. Callan, A. M. Callan, C. Kroos, and E. Vatikiotis-Bateson, "Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep eeg case study," *Cognitive Brain Research*, vol. 10, pp. 349–353, 2001.
- [10] M. H. Giard and F. Peronnet, "Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study," *Journal of Cognitive Neuroscience*, vol. 11, no. 5, pp. 473–490, 1999.
- [11] C. J. Price, "The anatomy of language: contributions from functional neuroimaging," *J. Anat.*, vol. 197, pp. 335–359, 2000.
- [12] 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.
- [13] G. A. Calvert and R. Campbell, "Reading speech from still and moving faces: The neural substrates of visual speech," *Journal of Cognitive Neuroscience*, vol. 15, no. 1, pp. 57–70, 2003.
- [14] G. Dehaene-Lambertz, 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.
- [15] R. Möttönen, G. A. Calvert, I. P. 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.
- [16] K. W. Grant and P-F. Seitz, "The use of visible speech cues for improving auditory detection of spoken sentences," *J. Acoust. Soc. Am.*, vol. 108, no. 3, pp. 1197–1208, September 2000.
- [17] J. Besle, A. Fort, C. Delpuech, and M-H. Giard, "Bimodal speech: early suppressive visual effects in human auditory cortex," *European Journal of Neuroscience*, vol. 20, no. 3, pp. 2225–2234, 2004.
- [18] 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," *Journal of Cognitive Neuroscience*, vol. 12, no. 3, pp. 495–504, 2000.
- [19] 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.
- [20] 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," *Journal of Cognitive Neuroscience*, vol. 14, no. 2, pp. 145–159, 2002.
- [21] T. Raji, K. Uutela, and R. Hari, "Audiovisual integration of letters in the human brain," *Neuron*, vol. 28, pp. 617–625, November 2000.
- [22] 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.
- [23] J. A. Fiez, "Sound and meaning: how native language affects reading strategies," *Nature Neuroscience*, vol. 3, no. 1, pp. R731–R735, January.
- [24] G. A. Calvert and T. Thesen, "Multisensory integration: methodological approaches and emerging principles in the human brain," *Journal of Physiology Paris*, vol. 98, pp. 191–205, 2004.
- [25] 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.
- [26] J. Hopfield, "Neural networks and physical systems with emergent collective computational properties," *Proc. Nat. Academy of Sci. USA*, vol. 79, pp. 2554–2588, 1982.
- [27] Teuvo Kohonen, *Self-Organising Maps*, Springer-Verlag, Berlin, 3rd edition, 2001.
- [28] Edmund T. Rolls, "Multisensory neuronal convergence of taste, somatosensory, visual, and auditory inputs," in *The Handbook of multisensory processes*, Gemma Calvert, Charles Spencer, and Barry E. Stein, Eds., pp. 311–331. MIT Press, 2004.
- [29] Daminda Alahakoon, Saman K. Halgamuge, and Bala Srinivasan, "Dynamic self-organizing maps with controlled growth for knowledge discovery," *IEEE Trans. Neural Networks*, vol. 11, no. 3, pp. 601–614, May 2000.
- [30] M. Milano, P. Koumoutsakos, and J. Schmidhuber, "Self-organizing nets for optimization," *IEEE Trans. Neural Networks*, vol. 15, no. 3, pp. 758–765, May 2004.
- [31] Pengfei Xu, Chip-Hong Chang, and Andrew Papliński, "Self-organizing topological tree for on-line vector quantization and data clustering," *IEEE Tran. System, Man and Cybernetics, Part B: Cybernetics*, vol. 35, no. 3, pp. 515–526, June 2005.
- [32] Reiner Schulz and James A. Reggia, "Temporally asymmetric learning support sequence processing in multi-winner self-organizing maps," *Neural Computation*, vol. 16, pp. 535–561, 2004.
- [33] Andrew P. Papliński and Lennart Gustafsson, "Multimodal feedforward self-organizing maps," in *Proc. Int. Conf. Comp. Intell. Sec. (CIS05)*, Xian, China, Dec. 2005.

- [34] Andrew P. Papliński and Lennart Gustafsson, “Detailed learning in narrow fields – towards a neural network model of autism,” in *Lect. Notes in Comp. Sci.*, O. Kaynak, E. Alpaydin, and L. Xu, Eds. 2003, vol. 2714, pp. 830–838, Springer.
- [35] Teuvo Kohonen, “The “neural” phonetic typewriter,” *Computer*, pp. 11–22, March 1988.
- [36] Ben Gold and Nelson Morgan, *Speech and audio signal processing*, John Wiley & Sons, Inc., New York, 2000.