Does Distributional Input Improve the Categorization of Speech Sounds? Neurobiological Aspects and Computer Simulations.

Master Thesis Linguistics
University of Amsterdam

Supervisor:	Prof. dr. Paul Boersma
Second reader:	Dr. David Weenink
Student:	Karin Wanrooij
Student number:	0512133
Date:	July 22, 2009

TABLE OF CONTENTS

0. Introduction

In the first year of their lives infants change from listeners that can discriminate both native and non-native phonemes into listeners who are particularly sensitive to phonemes in their mother tongues. This thesis is an attempt to clarify the role of distributional learning in this process. Distributional learning refers to the creation of categories on the basis of statistical information in the input. To study this topic I explored neurobiological findings and I tried to obtain insights from a replicated computer model (Guenther and Gjaja, 1996).

Part I of this thesis explains the relevance of studying distributional learning and specifies the research questions. Parts II and III describe the neurobiological aspects and the computer simulations respectively. Note that Part II on neurobiological aspects explains some basic neuroscientific terms and findings that could be relevant for linguists interested in the coding of auditory information along the auditory pathways and the primary auditory cortex (during processing) and on the development of representations in the auditory cortex (during learning).

Also notice that a list of abbreviations is included at the end of the thesis. It refers to the sections that discuss the terms.

PART I: SETTING

Chapter 1: The relevance of studying distributional learning

1. The relevance of studying distributional learning

1.1 Introduction

Small infants have a remarkable ability to discriminate both native and non-native phonemes. After a few months, however, their behavior in discrimination tasks changes. More precisely, after about six months for vowels and eight to twelve months for consonants, the susceptibility to non-native contrasts declines strongly, while the sensitivity to contrasts in the mother tongue increases (Kuhl, 1992; Werker and Tees, 1984/2002; Kuhl, 2004). An important mechanism that may contribute to this developmental change is distributional learning, i.e., the use of 'distributional input' for the formation of 'categories'. I will first explain what is meant by 'distributional input' and 'categories'. Also, I will address the relevance of studying the nature and role of distributional input for category creation. Then I will specify the research questions and the methodology.

1.2 Distributional input

What is distributional input? Figure 1 shows an example of a unimodal distribution of speech sounds. The x-axis represents a continuum of speech sounds that vary in a certain aspect, in this case the first formant value (F1) of a hypothetical vowel phoneme /a/. The y-axis stands for the number of times that each sound occurs in the environment. The graph shows that the formant values vary according to a Gaussian-like distribution. Most tokens of /a/ have F1 values of around 800 mels (represented by the peak). There are also tokens that have slightly higher or lower F1 values, but they occur less often (as represented by the left and right slopes).

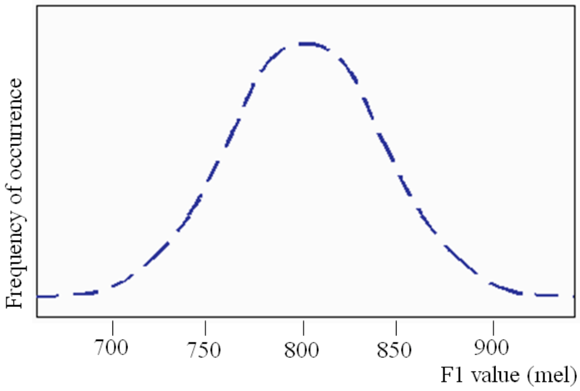


Figure 1: Unimodal distribution.
Hypothetical example for the vowel phoneme /a/ that varies in the first formant value

Distributions like this one can be drawn for other features in speech sounds as well (for instance for other formant values or for voice onset times in stops). Thus, in general the graph shows that speech sounds tend to vary in their features according to Gaussian-like distributions. Still, all sounds in the distribution belong together: they form a cluster and represent a phonemic category.

The grey curve (or yellow curve in a color print) in figure 2 is an example of a bimodal distribution, which represents two clusters of speech sounds. For example, if the feature that varies along the x-axis is again the F1 value, the clusters could stand for an open vowel /a/ (right peak) and a slightly more closed vowel /ɑ/ (left peak).

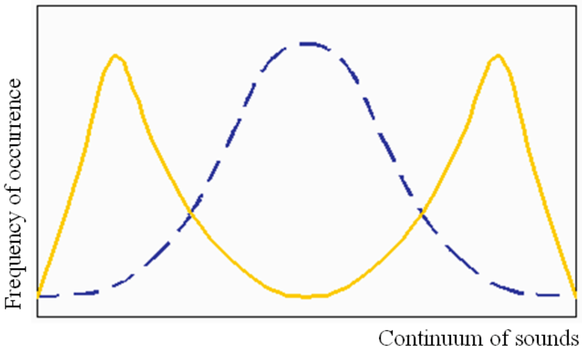


Figure 2: Unimodal and bimodal distributions (y-axis) of speech sounds varying in a particular feature along a continuum (x-axis).
(Unimodal = broken curve, bimodal = grey/yellow curve)

Maye and colleagues (Maye, Weiss & Aslin, 2008; Maye, Werker & Gerken, 2002) demonstrate that less than three minutes of exposure to a unimodal or bimodal statistical distribution of speech sounds suffice to make infants sensitive to this distribution in a subsequent phonetic discrimination task. Both studies use a series of tokens that differ in Voice Onset Times (VOTs) and which range from voiced unaspirated [da] to voiceless unaspirated [ta], although the actual stimuli and also the VOT values of the stimuli differ in both studies.[1]

Figure 3 shows the eight stimuli used by Maye, Weiss and Aslin (2008). VOT ranges from -100 milliseconds (i.e., 100 milliseconds voicing before the release) to +21 milliseconds (i.e., 21 milliseconds between the release and the start of voicing for the vowel). Tokens in the two distributions represent a contrast that had appeared *difficult* for 8-month-old infants in an

[1] For example, Maye, Werker and Gerken (2002) change the formant transitions in such a way that in going from voiced unaspirated [da] to voiceless unaspirated [ta] "the difference between formant frequencies at vowel onset vs. vowel center" (p. B105) diminishes. Maye, Weiss and Aslin (2008) do not report changes in formant transitions.

earlier study by Aslin, Pisoni, Hennessy and Perey (1981). In Maye, Weiss and Aslin (2008) infants were also 8 months old. In contrast, Maye, Werker and Gerken (2002) present [da] and [ta] tokens that had been shown to be *discriminable* for 6-to-8-month-old infants in an earlier study (Pegg & Werker, 1997). Infants in their study were also 6 to 8 months of age. The authors do not specify the VOT values.

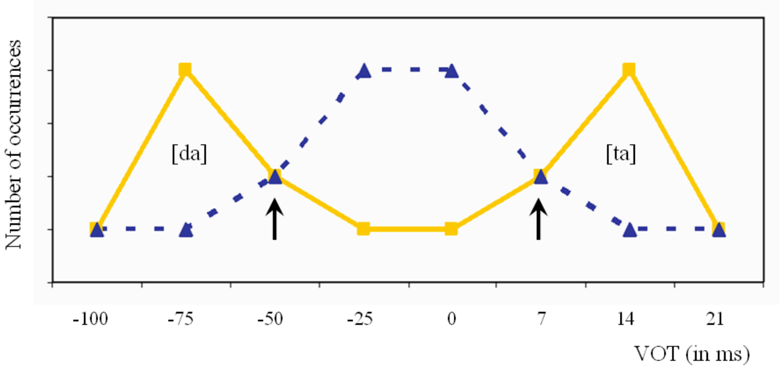


Figure 3: Unimodal and bimodal distributions of eight speech sounds varying in Voice Onset Times (in ms).
(Unimodal = broken curve, bimodal = grey/yellow curve)
Adapted from Maye, Weiss and Aslin (2008: 125)

After exposing one group of infants to a unimodal distribution and another group to a bimodal distribution, the infants in both studies were tested on their discrimination abilities with tokens that both groups had encountered equally often during the exposure. In figure 3 these tokens are pointed out by the arrows. Whereas infants that had been exposed to a bimodal distribution displayed discrimination of these two tokens, the infants that had been exposed to a unimodal distribution did not. Thus, exposure to a bimodal distribution of speech sounds that are difficult to discriminate before exposure can lead to enhanced discrimination (Maye, Weiss and Aslin, 2008), while exposure to a unimodal distribution of speech sounds that encompasses an initially discriminable contrast, can reduce discrimination performance (Maye, Werker and Gerken, 2002). Moreover, these changes occur very quickly (i.e., in these studies within less than three minutes). It is not clear, however, to what extent the discrimination performance reflects the incipient creation of *phonemic* categories. Nor is it evident, whether the process that underlies the discrimination performance corresponds to, for instance, the storage of exemplars or the 'warping' of the acoustic space. Warping refers to the phenomenon that in the process of learning some speech sounds come to be perceived as closer to one another in the acoustic space than other speech sounds, even when the acoustic distance between all sounds is identical (Kuhl, 1991).

Maye, Weiss and Aslin (2008) also tested whether infants were able to generalize the newly learned contrast to another, untrained place of articulation. More precisely, they tested whether the infants in the bimodal group would hear a difference between a velar [ga] versus [ka] on the basis of their exposure to [da] versus [ta]. (Another group of infants was first exposed to the velar continuum and was subsequently tested on the alveolar contrast. Of course, the velar continuum was constructed in the same way as the alveolar continuum). The results demonstrate that, indeed, "exposure to a bimodal distribution at one place of articulation facilitates discrimination at a second, untrained place of articulation." (Maye, Weiss and Aslin, 2008: 129). However, it is not transparent what process underlies the generalization performance. For example, did infants neglect the formant transitions that mark the place of articulation in this specific task and did they only pay attention to acoustic differences in VOT? (In that case they did not 'generalize'). Or did they hear the acoustic differences in VOT, while being aware of the distinct places of articulation? (If this is true, they generalized an acoustic difference). Still another option is that they even managed to extract a *phonological* feature 'voice', so that they could characterize the stimuli as either 'voiced' or 'unvoiced'. (Thus, in that case, they generalized a phonological feature). Also, the possibility remains that the new contrast was, after all, not so difficult for the infants as expected on the basis of the earlier study by Aslin, Pisoni, Hennessy and Perey (1981; see above in this section), so that the discrimination behavior did not reflect generalization abilities, but rather discrimination abilities that existed already before the exposure to the bimodal distribution.[2]

Distributional learning also seems to benefit adult second language (L2) learners, for whom discrimination and categorization of certain L2 phonemes are notoriously difficult. Studies on training methods show mixed results due to factors such as differences in tasks, training duration and phonemic contrasts. Interestingly, a recent study highlights the possible importance of training based on distributional learning as opposed to training based on explicit instructions: Gulian, Escudero and Boersma (2007) found that adults' discrimination performance progressed after being exposed to a bimodal statistical distribution, but was hampered if participants received simultaneous explicit instruction. Earlier, Maye and Gerken

[2] Unfortunately it is almost impossible to test discrimination performance before the training, since this would be time-consuming and longer tests with infants are prone to failure.

(2000, 2001) had also found an improvement in adult discrimination after distributional learning.

Due to mixed outcomes of studies it is uncertain whether adults can generalize the newly acquired discrimination abilities to novel stimuli. For example, adult participants in Maye and Gerken (2001) did not display enhanced discrimination of a newly learned VOT contrast at another place of articulation. Gulian, Escudero and Boersma (2007) found that the Bulgarian participants, who had been exposed to bimodal distributions of *synthetic* Dutch vowels, showed increased discrimination of corresponding *natural* stimuli. The authors conclude that the participants generalized distributional learning to new tokens. Again, as just discussed for infants, it is unsure what sort of process is represented by improved discrimination performance after distributional learning and what level of representation is affected (see section 1.3).

In the light of the potential importance of distributional learning for the creation of 'categories' of some sort (see section 1.3), it seems valuable to consider the role of Infant-Directed Speech (IDS) or 'motherese'. IDS is a style of speech used to address infants and toddlers. Apart from being relatively slow and high-pitched, it is characterized by enhanced phonetic differences, which could facilitate categorization. The presence of enhanced cues has been established for vowels in IDS spoken by English, Russian and Swedish mothers (Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg and Lacerda, 1997; Burnham, Kitamura & Vollmer-Conna, 2002) and also for stop consonants in Swedish IDS (Sundberg & Lacerda, 1998; Sundberg, 2001). The black curve (or red curve in a color print) in figure 4 shows an enhanced bimodal distribution as compared to the grey (or yellow) curve of the 'normal' bimodal distribution.

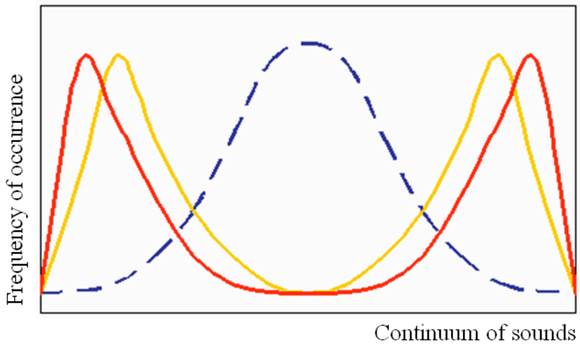


Figure 4: Unimodal, bimodal and enhanced bimodal distributions (y-axis) of speech sounds varying in a particular feature along a continuum (x-axis). (Unimodal = broken curve, bimodal = grey/yellow curve and enhanced bimodal = black/red curve)

It should be noted that the precise *form* of enhanced distributions in IDS is not transparent. There are indications that the variability of speech sounds in IDS is larger than that in Adult-Directed Speech or ADS (Kuhl, 2000; Kuhl et al., 1997). However, authors do not specify the cause of the variability. For example, Kuhl and colleagues mention vaguely that the "range" ((Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg and Lacerda, 1997: 685) of vowel formant values is larger; in Kuhl (2000) the "variety" (p. 11855) is enhanced in comparison to ADS; and still others only address the means of cue values in IDS, without touching on a measure of dispersion (e.g., Burnham, Kitamura and Vollmer-Conna, 2002). Because of the uncertainty as to the nature of the variability, we do not know if, for instance, the two peaks in the enhanced distribution (black/red curve) in figure 4 should actually be wider. Relatedly, we do not know if an enhanced distribution in IDS may facilitate or hamper the perception of certain ADS tokens. To understand this, consider again the enhanced distribution (black/red curve) in figure 4. It is drawn in such a way that some tokens in ADS (grey/yellow curve) could be difficult to discriminate for infants, in particular tokens that occur at the deepest points in the valley between the enhanced clusters and which occur more frequently in the ADS distributions. However, if the peaks in the enhanced distributions are actually wider, the tokens in the valley could occur slightly more often than is drawn in figure 4, so that they could become more discriminable (see the discussion of the studies by Maye et al., 2008, 2002, 2001 in this section).

Also, there is no clarity as to the exact *function* of enhanced distributions in IDS. There are many indications that cue enhancement facilitates perception in adults (Hazan and Simpson, 2000), but none of the studies involved is based on distributional learning. For infants, there is only one study hinting at a relation between enhanced differences in IDS and better speech discrimination abilities (Liu, Kuhl, and Tsao, 2003). All in all, the precise role in category formation remains unclear.

1.3 Categorization

What do I mean by the formation of 'categories'? Figure 5 depicts a small part of Boersma's model of parallel Bidirectional Phonology and Phonetics (see for example Boersma, 2007). It contains several levels of representation.

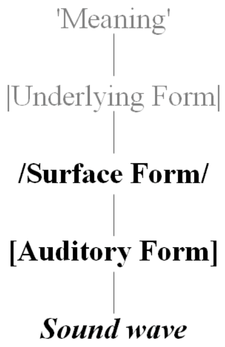


Figure 5: Part of Boersma's model of Bidirectional Phonology and Phonetics (e.g., Boersma, 2007)

To get familiar with the names of the representations, let us look at the comprehension process. In adult comprehension the acoustic waveform is transformed into the Auditory Form, which is a phonetic representation consisting of continuous auditory information such as formants and pitches. This Auditory Form is then partitioned into discrete known elements such as features and syllables, represented by phonological Surface Forms. Surface Forms, in turn, are mapped onto Underlying Forms, the phonological representations in the lexicon, which also contains the 'meaning'. In production the process is reversed and contains an additional step, which is left out in the figure: before turning into a sound wave the Auditory Form is mapped onto the Articulatory Form, which is a phonetic representation of articulatory gestures. Note that although this description suggests that processing proceeds serially, in fact the model is parallel: it assumes that the steps from Auditory Form to meaning (in comprehension) as well as the steps from meaning to Articulatory Form (in production) can occur simultaneously.

In addition to describing *processing* the model can be used to describe *learning*, i.e., infants' initial construction and the later adaptation of representations. Although the topic of this thesis is learning rather than processing, I will also address processing (chapters 2 and 3) in an attempt to get a better understanding of what has to be achieved by learning. Also, the studies discussed in the previous section show that distributional learning in infants (Maye et al, 2008; Maye et al., 2002) and adults (Gulian, Escudero and Boersma, 2007; Maye and Gerken, 2000, 2001) may occur within minutes, so that it is hard to define a precise boundary between processing and learning. Further, because the topic of this thesis is distributional learning (section 1.2), I will concentrate on the *bottom-up* influence of auditory input on the 'lower-level representations' (explained below) in the model. As a consequence, I ignore the top-down influence of higher representations on the formation of lower-level representations as well as the impact of representations relating to production.

The discussion in the previous section showed that it is uncertain what level of representation is affected primarily by distributional learning. The studies by Maye et al. (2002, 2008) leave open the possibility that it has a bearing on Surface Forms (for example if infants managed to extract a phonological feature 'voice' after distributional learning; see previous section). At the same time, it is possible that distributional learning influenced Auditory Forms (for instance if the infants focused on VOT differences; see previous section) or even lower levels of representation not illustrated in the picture (for example, if distributional input causes changes in non-cortical processing at one of the levels between the inner ear and the cortex (section 2.2). In view of the uncertainty as to the level of representation that is affected by distributional input, I will concentrate vaguely on 'categories', which in this thesis denote any representation that can arise at the lower levels, or more precisely, in between the sound wave and the Surface Form. This is highlighted in the figure.

Finally, it should be noted that looking at the process of distributional learning at these lower levels may also improve our understanding of higher-level categorizations, since there are indications that the effect of distributional learning extends beyond the discrimination of speech sounds to the extraction of words and word order patterns from a stream of syllables with certain transitional probabilities (Saffran, Aslin and Newport, 1996; Saffran and Wilson, 2003).

1.4 Questions

The discussion above showed that exposure to (enhanced) distributional input raises discrimination performance and may impact the creation of categories of some imprecise level. So far, the exact nature and role of distributional input in the learning of categories remain unclear. This thesis is an attempt to shed light on this issue by looking at a set of research questions that elaborate on the main question: does distributional input improve the categorization of speech sounds? The research questions can be grouped as follows: (1) What precisely *is* distributional learning? Can we define it in terms of 'necessary and sufficient' elements? And can we relate it to neural processes? (2) Relatedly, is distributional learning in adults and infants an identical process? Can it impact categorization in infants only or also in adult L2 learners? And (3) In what way(s) are (enhanced) distributions advantageous for categorization? For example, do they result in better, more stable categories? Do they speed up the categorization process? Do enhanced distributions in IDS facilitate the discrimination of tokens pronounced in ADS or only the IDS tokens?

1.5 Methodology

I took two approaches. First, because ultimately any linguistic theory of categorization must be consistent with neurobiological processes and because neuroscience is a quickly developing field, it seemed useful to look over the fence of linguistics into the garden of neuroscience to see what is known in this field that could help in solving the research questions. Although *distributional* learning appeared to be an as of yet unexplored topic in neuroscience, there turned out to be various findings that look relevant in the light of this study.

I include research on animals. Although animals do not process and learn speech sounds in the same way as humans do, animal studies can give valuable insights in lower levels of representation, which are presumably similar. Invasive measurements ensure that a particular area is examined and enable the researcher to look at the behavior of single neurons or specific groups of neurons.

As a second approach, I used a replication of a computer model (Guenther and Gjaja, 1996; replication Boersma 2009) to examine more specifically the effects on categorization of enhanced distributions in Infant-Directed Speech. Also, because simulations can be based on precise parameters and procedures, I studied the model in an attempt to sort out the 'necessary and sufficient' processes that can lead to self-organized clustering (or 'warping'; section 1.2) of the perceptual space.

PART II: NEUROBIOLOGICAL ASPECTS

Chapter 2: Coding mechanisms along the auditory pathways

Chapter 3: Auditory maps in adults

Chapter 4: Plasticity in the auditory cortex

2. Coding mechanisms along the auditory pathways

2.1 Introduction

In order to get a picture of (distributional) learning in the auditory cortex it is helpful to first understand in what way auditory information reaches the cortex and what it looks like when it gets there. Therefore, this and the next chapter will sketch how auditory cues in the input are coded for and transported along the auditory pathways (this chapter) and what coding maps in the primary auditory cortex look like (next chapter). Also, this chapter specifies some terms that will be used in the remainder of the thesis. An additional reason for looking at the pathways (and not just at the cortex) is that distributional learning may affect lower-level processes which subsequently influence cortical functioning.

2.2 Anatomical overview of the auditory pathways

Figure 6 pictures an overview of the central auditory pathways. Ascending[3] acoustic information travels from the cochlear *inner hair cells* on the *basilar membrane* in the inner ear to neurons in the *spiral ganglion* (or cochlear ganglion, as the cells are strung around the bony core of the cochlea).[4] Subsequently, it is transmitted through the axons of spiral ganglion cells, which are bundled together in the *auditory nerve* (or cochlear nerve) of the eighth cranial nerve, to the *cochlear nuclei* in the *brainstem*. The brainstem consists of the medulla, the pons and the midbrain. The midbrain is the location of the *inferior colliculus* (IC). The final relay and processing centre for all ascending information destined for the (auditory) cortex is the *medial geniculate complex* (MGC) in the *thalamus*. A brief discussion of the role of these stages follows in the next section. (For a detailed discussion on the inner ear and other parts of the pathway, see e.g., Kandel, Schwartz and Jessel, 2000; Purves, Augustine, Fitzpatrick, Hall, LaMantia, McNamara and White, 2008).

[3] Ascending ('afferent') signals travel towards the cortex, while descending ('efferent') signals refer to top-down activity travelling from the cortex to the periphery.

[4] There is one spiral ganglion at each side of the head (and therefore in total two spiral ganglia per person). Each spiral ganglion consists of several spiral ganglion cells (section 2.3). Also, *efferent* spiral ganglion axons mainly innervate *outer* hair cells.

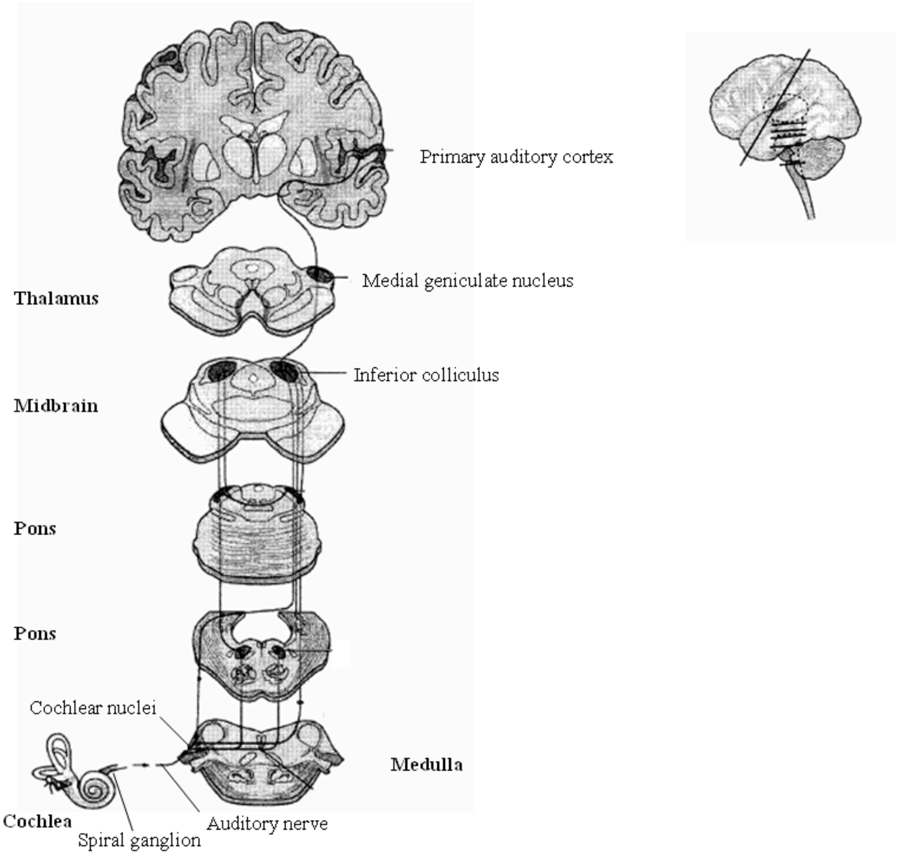


Figure 6: Overview of the central auditory pathways.
The diagram in the top right corner shows the orientation of the brain slices.
Further explanation: see text.
Figure copied and adapted from Kandel, Schwartz and Jessell, 2000: 604.

2.3 Sketch of coding along the pathways.

Coding along the pathways and in the auditory cortex is complex. This section gives a general impression (see e.g., Kandel et al., 2000; Purves et al., 2008). In sections 2.4 through 2.8 I will elaborate on some coding properties. At the lowest levels of the pathway each inner hair cell codes for a combination of stimulus frequency and intensity. (Each inner hair cell is connected to an average of ten spiral ganglion cells, which each have different sensitivities to intensity levels. There are roughly 3,000 inner hair cells per ear. The total number of spiral ganglion cells per ear is about 30,000). Processing of the frequency and intensity information begins in the cochlear nuclei in the brainstem, where different sorts of cell types have their specializations. For example, chopper cells fire at very regular rates for the duration of a stimulus despite small variations in stimulus frequency, and bushy cells mark stimulus onsets by firing once upon stimulation. (For a more elaborate discussion of different cell types in the cochlear nuclei see e.g., Langner, 1992; Kandel et al., 2000).

Signals entering the cochlear nuclei are split into different parallel pathways that reunite in the IC (section 2.2) in the midbrain and in the MGC (section 2.2) in the thalamus. This separation of information has been associated with sound localization. Its potential role in speech sound analysis is still an unexplored field (but see section 2.8). What is known, is that the IC plays a role in the analysis of complex temporal patterns, such as speech sounds. It contains neurons that are responsive to specific durations of sounds and neurons that are sensitive to specific frequency modulations. The latter property suggests that the IC is important for the analysis of pitch (e.g., Langner, 1992; Purves et al, 2008).

Apart from being separated and integrated along the pathways, information is transformed due to (excitatory and inhibitory) interactions between nerve cells as well as interfering efferent[5] activity. The net result of separation, integration and interaction is that along the pathway selectivity for specific properties in the signal increases (i.e., tuning sharpens) and that the information is coded in representations that are increasingly non-linear (e.g., Langner, 1992; Diesch and Luce, 1997; Eggermont 2001), i.e., the output at a level of processing cannot be described as a linear sum of the input components. For example, presenting the first and second vowel formants simultaneously results in a different activation pattern in the auditory cortex than summing the activation patterns caused by presenting the identical formants separately (Diesch and Luce, 1997; section 3.5). In the auditory cortex, single cells code for multiple dimensions (chapter 3).

2.4 Information transport by means of electrical activity

Neurons generate different types of electrical activity. For the purpose of this thesis (and for understanding neuroscientific articles in general) it is useful to know of at least two types: *action potentials* (APs), which are also called '*spikes*' (or 'discharges' or 'firings') and *postsynaptic potentials* (PSPs). APs in one cell can cause PSPs in a second cell in the following way. APs are generated in the cell body and travel down the axon to the axon terminal, which releases neurotransmitter into the synaptic cleft between the neuron's axon and the dendrites of the postsynaptic neuron. PSPs arise when the neurotransmitter, which was released by the presynaptic cell, binds to receptors on the postsynaptic cell. PSPs can, in

[5] See note 3.

turn, lead to the elicitation of an AP in the cell body of the second cell. APs are short. They last for about 1 millisecond. PSPs can last for tens or hundreds of milliseconds.

A single neuron's behavior in response to a stimulus is characterized by the temporal pattern of APs, i.e., its *firing pattern*.[6] However, APs can only be picked up by single-and multiple-unit recordings[7], i.e., by inserting electrodes into the brain, and not by for example ElectroEncephaloGram (EEG) recordings. This is because EEG can only pick up signals that occur *simultaneously* (so that the summed activities are large enough to be detected) and which arise from sources that have *similar orientations*. That is, the neurons have to be aligned perpendicularly to the scalp (again, so that the summed activities does not cancel). This is shown in figure 7. Most cells that match this requirement are pyramidal cells (named after the characteristic shape of the cell body, as is visible in figure 7).

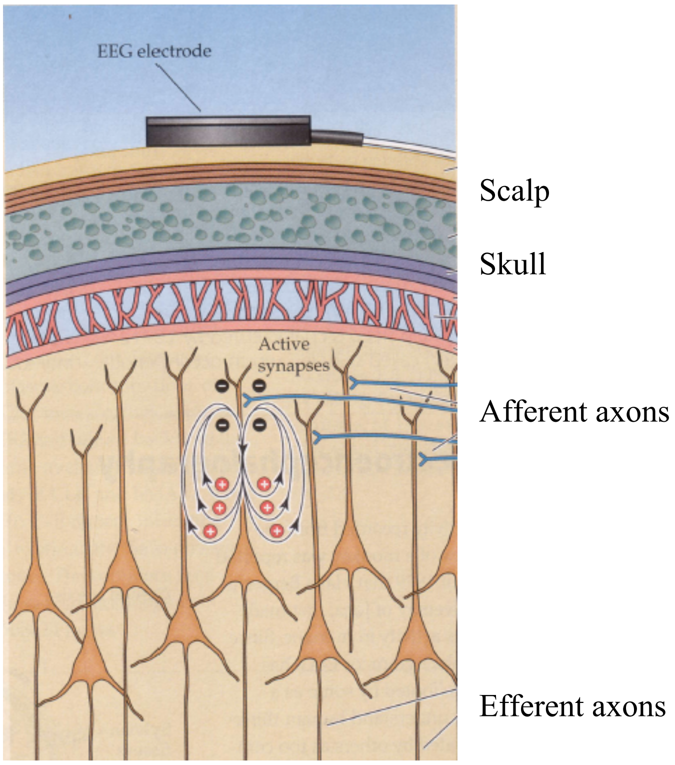


Figure 7: An EEG electrode on the scalp measures concerted PSPs of a large group of pyramidal neurons that result from active synapses (which arise due to APs from afferent neurons). The PSPs make the more superficial extracellular space negative compared to the deeper extracellular space.
Figure copied from Purves et al., 2008: 716.

[6] The minor differences in e.g., amplitude or duration are ignored.
[7] For a slightly more detailed description of the recording of Multiple Unit Activity (MUA), see section 3.5.1.

APs are far too short to occur simultaneously and the alignment of axons varies too much. Thus, the signals that are picked up by EEG recordings are concerted PSPs from large groups of neurons, which are aligned perpendicularly to the scalp. The same holds for Frequency-Following Responses (FFRs), which record concerted activity in the *brainstem*. It is measured on the scalp, just as EEG measurements.

Another technique which complements EEG is MagnetoEncephaloGraphy (MEG). This is because electrical fields are accompanied by magnetic fields, which are oriented perpendicularly to the electrical fields. Accordingly, MEG is most sensitive to electrical activity that runs *parallel* to the scalp. (More details on MEG can be found in section 3.5.2; Reference works that treat neural activity are e.g., Purves et al., 2008; on EEG and also to a lesser extent on MEG see Luck, 2005. On FFR see Abrams and Kraus, 2009).

2.5 Coding by single neurons

How does a neuron code for (speech) stimuli? And what stimulus properties are coded for? These questions are difficult and are still far from being understood fully. In this section I discuss some basic concepts.

2.5.1 Characteristic Frequency

Each neuron is particularly sensitive to a certain frequency. This frequency is usually expressed as the *characteristic frequency* (CF), i.e., the frequency that the neuron is most sensitive to at the lowest intensity level (the 'threshold intensity'). Figure 8 (Purves et al., 2008: 331) clarifies this. It shows frequency *tuning curves* of six different neurons in the auditory nerve. For a range of frequencies (x-axis) the graph shows the weakest sound intensity level that is required to increase the firing rate noticeably above the spontaneous level (y-axis). For example, the lowest intensity level that arouses the neuron in the utmost bottom picture is just below 10.000 Hertz, which is therefore that neuron's CF. Apart from revealing the CF, a frequency tuning curve also illustrates that each neuron is sensitive to a distribution of frequencies, so that it will not only fire for its CF, but also for frequencies close to its CF, albeit less strongly. Auditory nerve neurons that have high CFs have larger Receptive Fields (RFs) measured in Hertz than those with relatively low CFs. A neuron's RF refers to the range of stimulus parameter values (e.g. in this case: stimulus frequencies in Hertz) that affect the neuron's firing rate.

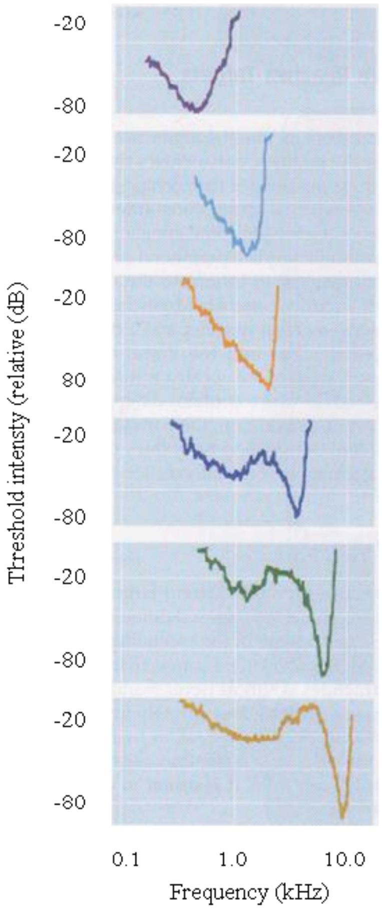


Figure 8: Frequency tunings curves for six neurons in the auditory nerve.
Figure copied and adapted from Purves et al, 2008: 331.

Sometimes other measures are used to denote the sensitive frequency of a neuron. For example, fMRI studies tend to report on *best frequencies* (BFs), because noise levels are usually too high to record proper CFs. The BF of a neuron is the frequency that elicits the highest response at a certain intensity level (Eggermont, 2008).[8]

2.5.2 Tonotopic order and 'place code'

At each stage along the auditory pathways (but at birth less so in the cortex) each neuron does not only have a particular tuning curve with a particular CF, but it also has a particular place in a tonotopic order. This order reflects the spatial arrangement of inner hair cells on the basilar membrane. (In going from the base to the apex of the basilar membrane the inner hair cells code for increasingly lower frequencies). In the auditory nerve this tonotopic order is

[8] A neuron's CF and BF do not have to be the same frequency value. This indicates that different intensity levels impact the frequency that a neuron is most sensitive to (Eggermont, 2008).

preserved, because the spiral ganglion cells "associate with the inner hair cells in approximately a one-to-one ratio" (Purves et al., 2008: 330; also recall section 2.3).[9]

This tonotopicity, which is preserved at higher stages along the pathways (section 2.2), is 'hard-wired' (i.e., determined by genetic codes) and present at birth up to but not including the auditory cortex (section 4.5). To what extent *cortical* tonotopicity is determined by genes is not clear. Most probably the development hinges on a combination of genetic coding, which would account for the similarity of tonotopic orientations across individuals, and auditory experience, which can explain the variability across individuals (section 3.3).

Since the CF and its location in the tonotopical order are intertwined along the pathways, it is difficult to say whether the place determines the CF or vice versa. As a consequence of the intertwined nature of CFs and their locations in the tonotopic order at each stage along the pathways, a common way of characterizing one way of coding is to say that a neuron codes for stimulus properties in terms of its place in the tonotopic order. This way of coding is commonly called 'place coding' (e.g., Brosch and Scheich, 2003; Sachs, 1984) or 'labeled-line coding' (e.g., Brosch and Scheich, 2003; Purves et al., 2008).

Importantly, features of speech sounds, such as formants (in particular the first two formants F1 and F2) are visible in cortical tonotopic maps (section 3.5). However, in particular in the cortex it is uncertain to what extent the place of neurons in the tonotopic order *directly* codes for stimulus properties, such as formants. For example, Eggermont (2001) questions whether the brain relies on place coding:

> "Do topographic maps participate directly in neural coding? (…) Maps can be interpreted or `read out' only by an external observer; they cannot be utilized internally by the brain, which `knows' only neuronal activity or the lack thereof. This leads to the conclusion that information about different stimuli and their functional significance is not conveyed directly by the map structure but only indirectly by patterns of activity resulting therefrom." (Eggermont, 1991: 9-10)

Not all authors agree with this viewpoint. For example, Brosch and Scheich (2003) say:

> "One possible solution [for interpreting firing rates; KW] for the external observer, *as well as for the brain* [italics: KW], is to integrate within the same topographic region

[9] Each spiral ganglion cell is connected to one inner hair cell. Each inner hair cell is connected to about 10 spiral ganglion cells (section 2.3).

> the activity of many neurons that respond to a given stimulus."(Brosch and Scheich, 2003: 157)

2.5.3 Firing pattern, 'rate code' and 'temporal code'

Regardless of the existence of a place code, it is clear that firing patterns code for stimulus properties. Firing patterns can be examined in several ways. Two important ways are by looking at the *firing rate* (in spikes per second) and the *spike time interval* (in milliseconds) (e.g., Brosch and Scheich, 2003; Eggermont, 2001; Sachs, 1984; Sachs and Young, 1979; Young and Sachs, 1979).[10] The former (firing rate) is an example of a measure that denotes the strength of the activity ('how strong'), while the latter (spike time interval) is a measure that reflects the precise timing of the spikes ('when'). Thus, in order to decode firing patterns to discover what stimulus properties are coded for, we can look at how strong a neuron fires and when it fires in response to a stimulus. A neuron tends to fire *vigorously* (firing rate) if the stimulus contains frequency components that fall within the neuron's RF (section 2.5.1). Similarly, a neuron tends to fire *in phase* (spike timing) with a frequency component in the stimulus, if this component falls within the neuron's RF (Sachs, 1984). Firing 'in phase' (or 'phase-locking' or 'time-locking') occurs when a neuron fires at the same points in the cycle of the stimulus waveform.[11] When authors talk about a '*temporal code*', they usually refer to phase-locking (and thus not to the firing rate, even though the firing rate pertains to temporal properties of firing as well). The term '*rate code*' denotes the representation of stimulus properties in terms of the firing rate (e.g., Brosch and Scheich, 2003; Eggermont, 2001; Sachs, 1984; Sachs and Young, 1979; Young and Sachs, 1979).

Looking at a neuron's firing rate to discover stimulus properties is problematic at higher intensity levels, because at these levels more and more neurons start firing at their saturation rates, so that the differences in firing rates between the neurons disappear (Sachs and Young, 1979; Sachs, 1984). In contrast, temporal coding is robust across intensity levels (at least at the level of the auditory nerve, as evidenced in single-unit population studies with animals; e.g., Young and Sachs, 1979). At the same time, looking at phase-locking of a neuron to stimulus frequencies is problematic at higher levels of processing along the pathways. This is because phase-locking to the stimulus tends to diminish as the signal travels

[10] Other possibilities that can typify firing patterns include "spike times", "particular spike sequences" and "spike time coincidences" (Eggermont, 2001: 6).

[11] A neuron does not have to fire at every cycle (Hayward, 2000). Also, a neuron can phase-lock to higher frequencies than those associated with its CFs, by firing at every so many cycles of the stimulus frequency (Phillips, 1993; Kandel et al., 2000).

up to the cortex: auditory nerve fibers can phase-lock to stimulus frequencies up to around 3,000 Hertz (Purves et al., 2008) to 3,500 Hertz (Phillips, 1993)[12], whereas in the brainstem this value has reduced to around 1,000 Hertz and further in the auditory cortex to approximately 100 Hz (Abrams & Kraus, 2009). The reduction may be partly caused by the fact that temporal precision is lost with every synapse along the pathways (Langner, 1992), although this is not certain. Some authors (e.g., Eggermont, 2001; Langner, 1992) reason that the reduction of phase-locking along the pathways must indicate a transformation of phase-locked information into a place code and/or higher-level representations. In particular, it is proposed that the neurons in the IC play a role in this transformation (Eggermont, 2001; Langner, 1992), since the IC receives converging input from lower levels and contains neurons that are particularly sensitive to temporal properties (section 2.3).

Neurons phase-lock to speech formants too. This was shown for humans (e.g., Krishnan, 2002; Aiken and Picton, 2008; both studies use FFR measurements that reflect brainstem activity) and also for animals (e.g., Young and Sachs, 1979, in the auditory nerve).

2.6 Neural ensemble coding

So far we looked at the way in which *single* neurons code for stimulus features. However, single-neuron firing patterns tend to be noisy and therefore unreliable. This noisiness is reduced by combining the firing patterns of several neurons coding for (more or less) the same stimulus properties. For example, each inner hair cell is connected to about ten spiral ganglion cells (section 2.3), which have overlapping tuning curves (section 2.5.1).[13] Similarly, at the level of the auditory cortex, it is unlikely that a single cell is responsible for coding a particular behaviorally important sound pattern, such as a phoneme or a specific animal vocalization (Brosch and Scheich, 2003; see also section 2.5), even though a neuron can be responsive to such patterns (Wang, Kadia, Lu et al., 2003; see also section 3.5). Thus, it seems that a representation is encoded by ensembles of nerve cells.

The question is, however, how the brain combines the information provided by these cells. Similarly, *researchers* need ways to interpret ensemble firing patterns, so as to be able to predict behavior on the basis of the observed patterns. Several solutions have been put forward for cortical ensemble *decoding*. The following discussion relies on Brosch and

[12] There are also other estimates. For example, Ohgushi (1983) finds that in the auditory nerve the upper limit of phase-locking to stimulus properties is 5,000 Hz.

[13] Also, a stimulus frequency excites several inner hair cells.

Scheich (2003), although information from other sources is added (references are mentioned below). Roughly speaking, there are three groups of approaches.

The first possibility is to look at the *size of the activities* of single neurons. An example of this approach is the calculation of the *population vector*, which was proposed by Georgopoulos, Schwartz and Kettner (1986). The authors demonstrate that an arm movement (of a rhesus monkey) can be predicted accurately by a weighted average of preferred arm movement directions of single neurons.[14] The average is weighted by the activities of the neurons. Other authors applied the concept to different parts of the cortex, including the auditory cortex (e.g., Guenther and Gjaja, 1996; see the description of the computer model in chapter 5). The calculation of the population vector relies on the assumptions that "each neuron represents a particular property" (Brosch and Scheich, 2003: 166) of a stimulus (for example a characteristic frequency for which it fires most fiercely) and that "the information of different neurons is independent from each other" (Idem: 166). Based on these assumptions, the information represented by the different neurons can be added and averaged to yield the combined response.

An alternative to the population vector, which still focuses on the size of neuronal activities, is to select the cell with the largest activity. In this so-called *winner-takes-all* approach, a small number of neurons determines the code rather than a large number as in the calculation of the population vector (Gilbert, 1994). This approach has been associated with local and global inhibitory processes, which are shown to occur in different parts of the cortex including the auditory cortex (Salzman and Newsome, 1994, for visual processing; Kurt, Deutscher, Crook, Ohl, Budinger, Moeller, Scheich and Schulze, 2008, for processing in the auditory cortex).

Approaches to decoding that emphasize the magnitude of activities ignore *fine-grained temporal aspects* of firing, such as the issue whether the neurons fire in synchrony or not. These aspects are the focus of attention in the second group of solutions to the decoding problem. It is certain that correlated firing (or 'synchronization') occurs in the auditory cortex. Brosch and Scheich (2003) state that the prime role seems to be the *integration* of different stimulus properties rather than the coding of these separate properties. In this respect "It is interesting that the predominant time scale of neural synchrony in the auditory cortex is similar to the critical temporal window for the detection and processing of individual acoustic

[14] Similar to the nerve cells in the auditory cortex that represent certain favourite frequencies, as was described in section 2.5, each motor neuron has a preferred direction for which it fires most fiercely. If the direction deviates slightly from the preferred direction, the cell will still fire, but less fiercely.

events for a complex sound" (Brosch and Scheich, 2003: 167). Other researchers have demonstrated that this window is around 15 to 20 milliseconds (e.g., Hirsch, 1959). At the same time, further information (see the discussion of Steinschneider, Arezzo and Vaughan, 1990, in section 3.5.1) shows that precisely the integration of the information may adjust the *content* of the code. This occurs, because the summation of precise firing patterns (i.e., the combined waveform that can be constructed on the basis of many waveforms that reflect the precise single-neuron firing patterns) enhances certain components of the stimulus and suppresses other components (see section 3.5.1).

Thirdly, researchers (e.g., Pouget, Dayan and Zemel, 2000) also use *Bayesian decoding* to interpret ensemble firing patterns. Although this method can be more reliable than for example the calculation of the population vector, it is complicated in that the researcher must have a model of cell activity patterns. He or she then calculates the probability of getting the observed data, under the assumption that the model is correct.

2.7 Acoustic resolution at different levels for different speech sounds

Due to the reliance on different cues that are coded for differently along the pathways and at different places in the cortex, different speech sounds are 'resolved' at different levels and places along the pathways. For example, while already at the level of the brainstem mismatch responses (MMRs) can be recorded for differences between /ba/ and /wa/, differences between /ba/ and /ga/ only appear at the level of the cortex. (Kraus, McGee and Koch, 1998, who report this for guinea pigs). An MMR is a physiological response that is elicited when a deviant stimulus occurs in a train of identical stimuli. Thus, the relatively complex contrast between /ba/ and /ga/ can only be resolved at the level of the cortex.

Also, there are indications that in animals as well as in people spectral information that extends over comparatively long durations, such as that in vowels, is processed in the primary auditory cortex of the right hemisphere, while complex temporal information is handled on the left (for humans: e.g., Warrier, Wong, Penhune, Zatorre, Parrish, Abrams and Kraus, 2009; for animals: Kraus, McGee and Koch, 1998).

2.8 Different pathways for auditory 'who', 'what' and 'where' information

It has been noticed that different parts of the speech stimulus are coded for in separate pathways. In particular, there are indications that information relating to the source ('periodicity information') is analyzed separately from information pertaining to the filter ('spectral information'). The 'source' refers to the vocal fold vibrations and thus to the fundamental frequency. The 'filter' refers to the resonances in the vocal tract and thus to the formants. For example, Kraus and Nicol (2005) show that source and filter aspects can be detected in separate brainstem response components. The source responses in the FFR revealed the fundamental frequency, while the filter aspects disclosed the first formant, onsets and offsets. Further, certain cells in the IC in the brainstem seem to be specialized in periodicity analysis (e.g., Abrams and Kraus, 2009; Langer, Sams, Heil and Schulze (1997); Eggermont, 2001; see also Purves et al., 2008 and section 2.3). Interestingly, in some persons the representation of the fundamental frequency can be disrupted while the representation of formants and onsets remains intact, at least at the level of the brainstem (as measured with FFR; Abrams and Kraus, 2009). Also, Langer, Sams, Heil and Schulze (1997) find evidence that in the cortex spectral information and periodicity information are projected in orthogonal maps (section 3.4). In view of the fact that the relatively intensely studied analysis for sound localization along the auditory pathways could be labeled a 'where' pathway, it is thus tempting to label the filter analysis a 'what' pathway and the source analysis a 'who' pathway. Indeed, Wang, Kadia, Lu, Liang and Agamaite (2003) suggest such a division.[15]

2.9 Summary

The coding of auditory cues relies on the combined contributions of many neurons. Single neurons code for stimulus properties by means of their firing patterns, which depend on their CFs, and perhaps also by their place in the tonotopic order, although the latter point is not certain. The particular sensitivities of neurons reveal properties of speech sounds that neurons can code for. They include frequency (of formants and the fundamental frequency), intensity, duration, onset and offset. Along the auditory pathways the separation and integration of the neuronal contributions yields increasingly specific representations.

As different cues are coded for differently, the acoustic analysis of different phonemes is resolved at different levels along the pathways. For example, the integration of information that is necessary for the resolution of transient complex spectro-temporal patterns (as for example in formant transitions of stop consonants) only occurs at the level of the cortex, while longer-lasting spectral components (such as the spectra of vowels) can be resolved at lower levels.

[15] Kraus and Nicol (2005) associate the 'where' pathway with a 'where in frequency' and thus with spectral motion and linguistic content (the filter aspects) and the 'what' pathway, a bit surprisingly, with the identity of the speaker (the source aspects). They do so in an attempt to find correlates of the visual 'what' and 'where' pathways (as identified by Mishkin, Ungerleider and Macko, 1983) in the auditory system.

3. Auditory maps in adults

3.1 Introduction

This chapter discusses characteristics of the *adult* auditory cortex, in particular the primary auditory cortex (A1). The objective is to understand better how auditory input is processed in A1 and if there are signs of any 'categories' for speech sounds. Also, knowledge of the adult cortex makes possible a comparison with the infant cortex (discussed in the next chapter) and thus a more complete understanding of what infants must acquire.

3.2 Anatomical sketch of the auditory cortex

The mammalian auditory cortex comprises several regions that can coarsely be divided into primary or 'core' regions, secondary or 'belt' areas and tertiary or 'parabelt' areas (Hackett, 2003), which stand for increasingly higher levels of processing. In addition, each of these groups represents "a distinct level of cortical processing. At each successive level of processing, inputs appear to be processed in parallel by subdivisions within that region." (Hackett, 2003: 213).

Note that the core does not consist of a single area, although it is usually referred to as 'the primary auditory cortex'. Studies of primates show that there are at least two primary areas (e.g., Kaas, Hackett & Tramo, 1999 for macaque monkeys; Rauschecker, Tian, Pons and Mishkin, 1997 for rhesus monkeys), which each receive direct parallel input from the ventral division of the MGC. An fMRI study with human participants suggests that this may hold for the human primary auditory cortex as well (Formisano, Kim, Di Salle, van de Moortele, Ugurbil and Goebel, 2003). The role of parallel input into multiple primary maps is unclear. However, there are indications that the two maps represent the first cortical processing stages of the auditory 'what' and 'where' pathways (Formisano et al., 2003). Another point of note is that the belt and parabelt areas also receive direct input from the MGC, albeit from the dorsal rather than the ventral division. This fact implies that distributional input may directly impact higher-level processing, rather than only indirectly via the core.

The human auditory cortex lies on the superior temporal gyrus (one in each hemisphere). The core areas, for which I will use the common label 'A1'[16] from time to time, are situated on the transverse gyrus of Heschl (again one in each hemisphere), typically in the depth of the lateral sulcus. This latter fact complicates neurophysiological measurements such

[16] In animal studies A1 is the label of one of the core areas.

as EEG. (For plain anatomical descriptions of the human auditory cortex see Kandel et al. (2000) and Purves et al. (2008). More details of the primate auditory cortex are described by Hackett (2003). He includes a list of references, which compares auditory regions in human and non-human primates).

3.3 Tonotopicity as a central organizational principle

The most evident aspect of the *adult* mammalian auditory cortex is its tonotopic organization: along one surface axis neuronal CFs change systematically from low CFs to high CFs. Orthogonally to this frequency gradient, neurons lie in 'isofrequency bands', i.e., within these bands they have the same CFs (e.g., Merzenich and Brugge, 1973; Schreiner, 1998). (Realize that the tonotopic arrangement is thus expressed as a *two-dimensional map* with tonotopicity on one axis and isofrequency on the other axis).

The tonotopic arrangement of nerve cells is most prominent in the primary regions (Merzenich and Brugge, 1973). The secondary areas are also tonotopically organized, but less consistently. This difference has been attributed to the fact that A1 receives point-to-point input from the thalamus, while the belt areas get more diffuse inputs from A1 and secondary areas in the thalamus (Purves et al., 2008). Accordingly, A1 appears most responsive to pure tones, whereas the belt areas seem more responsive to complex stimuli (Kaas et al., 1994; Rauschecker et al., 1997).

As for the tonotopic *orientation* in the human A1, experimental findings are not consistent, due to variability among individuals and to differences in the techniques and experimental details. The predominant view is that the low frequencies are situated more laterally and frontally, while the high frequencies lie more medially and posteriorly (e.g., Wessinger, Buonocore, Kussmaul and Mangun, 1997, with fMRI; Romani, Williamson and Kaufman, 1982, with MEG; Verkindt, Bertrand, Perrin, Echallier and Pernier, 1995, with EEG[17]; for a slightly different perspective see Pantev, Bertrand, Eulitz, Verkindt, Hampson, Schuierer and Elbert, 1995, with MEG and EEG).

At the same time, there is growing evidence that the measurements reported in these studies reflect just one of two tonotopic maps in the primary auditory cortex. As just

[17] Verkindt et al. (1995) is often mentioned in relation to tonotopicity in the human cortex. However, EEG is not suitable for localization. It is questionable to what extent EEG can pick up signals coming from A1, as this area tends to be located in the depth of the lateral sulcus. In addition the orientation of the nerve cells in this area is not tangential to the scalp, as is required for the signal to be picked up by EEG measurements. Also, the results obtained by the chosen localization technique (dipole modelling) are heavily dependent on choices made by the researcher. It is therefore not surprising that the authors found localizations that are consistent with earlier studies.

mentioned in section 3.2, Formisano et al. (2003) demonstrate two primary maps in an fMRI study. In the rostral map the high to low frequencies lie along a rostral-caudal axis, while in the caudal region the frequencies run from low to high along the same axis. The low-frequency areas are thus adjacent to one another. This is consistent with the tonotopic orientation of the two core areas in macaque monkeys (Merzenich and Brugge, 1973).

3.4 Overlapping maps for other parameters

The stability of CFs across isofrequency bands triggers the question what is coded for along the isofrequency axis. Animal studies have explored this question by means of single- and multiple-unit recordings (e.g., Schreiner, 1998, for cats; Recanzone, Schreiner, Sutter, Beitel and Merzenich, 1999, for owl monkeys; for single- and multiple-unit recordings see section 2.4). It appears that several Receptive Field (RF; section 2.5) parameters vary systematically along the isofrequency axis. Schreiner (1998) and Recanzone et al. (1999) discern four groups of parameters, which are illustrated in figure 9 (Schreiner, 1998: 108) and explained below. The RF properties within each group co-vary, while the properties of different groups do not. Note that the figure illustrates an *idealized* version of the RF parameter maps in A1 and the map resulting from superimposing them. Also note that because the maps illustrate RF properties, they reflect properties of neurons rather than properties of the sound stimulus. (Of course the RF properties are important for understanding the stimulus properties that can be coded for). Finally notice that the figure is appreciated best in a color print.

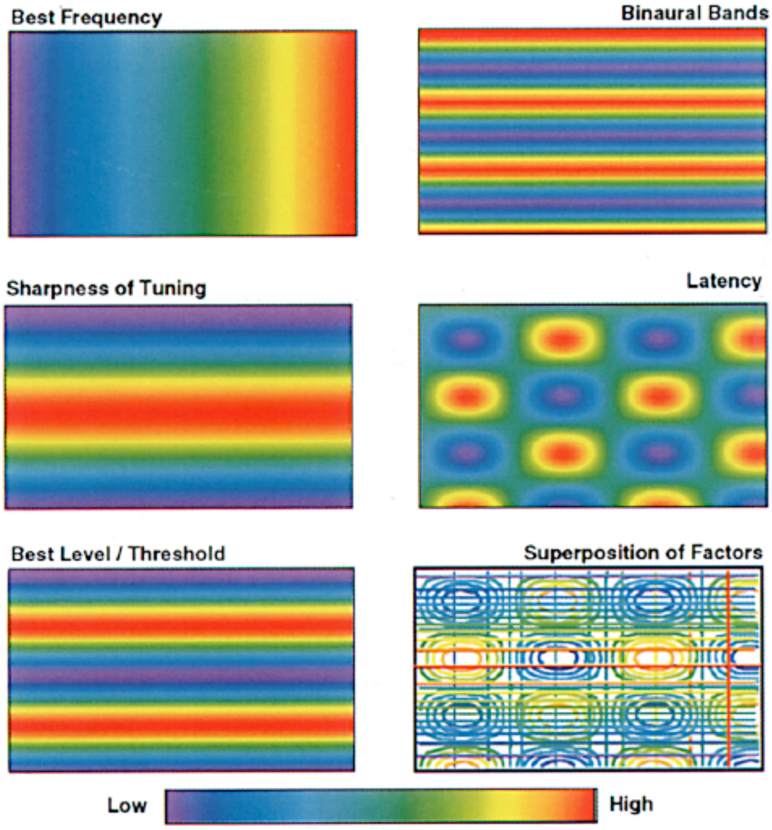


Figure 9: Idealized maps for different parameters in A1 and a schematic of the combined maps (bottom right). The top left picture ("best frequency") depicts the tonotopic axis from left to right. The isofrequency bands are visible from top to bottom. The parameters "sharpness of tuning", "best level/threshold" and "binaural bands" are organized orthogonally to the tonotopic axis and thus vary along the isofrequency axis. "Latency" is represented in patches.
Further explanation: see text.
Figure copied from Schreiner (1998: 108). The text is adapted.[18]

The first group contains frequency coding measures, such as the CF or BF (top left picture in figure 9). These measures vary along the tonotopic axis, which runs from low frequencies (on the left in the picture) to high frequencies (on the right). The second group consists of spectral distribution coding measures, such as sharpness of tuning (middle left picture in figure 9). Spectral distribution coding measures vary along the isofrequency axis,

[18] The adaptation is the elimination of the terms "ripple density" (as a second example of spectral distribution measures; middle left picture) and "local connectivity" (as a second example of parameters that may show a patchy distribution; right middle picture). The terms were not explained.

(which runs from top to bottom in figure 9 and is thus orthogonal to the tonotopic axis). The neurons that are most sharply tuned (at threshold intensity; section 2.5) reside in the *centre* of A1 (i.e., in the centre band of the middle left picture). The third group comprises absolute or relative intensity coding measures, i.e., measures that reflect sensitivity to intensity or changes in intensity. Examples are threshold intensity (section 2.5) and 'best level', which refers to the intensity level that makes the neuron fire most fiercely (bottom left picture in figure 9). The shape of the rate/level functions represents the cell's dynamic range and describes the firing rate at each intensity level. Intensity coding measures vary along the isofrequency axis, just as the measures in the second group. There is also evidence for an intensity gradient along the isofrequency axis in the human auditory cortex (Pantev, Hoke, Lehnertz and Lütkenhöner, 1989). The fourth group of parameters, which vary along the isofrequency axis as well, contains measures that relate to binaurality and sound location (top right picture in figure 9). These measures will not be discussed in this thesis.

There are indications that apart from these four groups that were identified by Schreiner (1998) and Recanzone et al. (1999), other RF properties vary along the isofrequency axis. In a MEG study Langner et al. (1997) revealed a periodotopic gradient in the human auditory cortex, which is also oriented orthogonally to the tonotopic map. Thus, in addition to their spectral tuning properties (as reflected in the CF or BF), neurons have temporal tuning properties (as reflected in for example the Best Modulation Frequency BMF). The authors propose that the periodotopic gradient reflects "the temporal periodicity analysis in the brainstem and corresponds to the perceptual quality of pitch" (Langner et al., 1997: 675) as opposed to the tonotopic gradient which represents "the cochlear frequency analysis and corresponds to the perceptual quality of timbre" (Idem).

Schreiner (1998) and Recanzone et al. (1999) discern still other RF properties that do not change along the isofrequency axis, but vary in a patchy manner across A1. This holds for, for example, onset latencies (middle right picture in figure 9), i.e., the time between the presentation of a stimulus and the onset of firing (see also Mendelson, Schreiner and Sutter, 1997). All parameter maps together result in a "multidimensional 'parameter space' projected onto two spatial dimensions (the dorsoventral and the rostrocaudal axis of A1)" (Schreiner, 1998: 107). Thus, neurons in A1 tend to code for combinations of features. This is illustrated in the bottom right picture in figure 9.

Schreiner (1998) also notes patterns in the degree of correspondence between the observed maps and idealized maps: at certain locations maps tend to be much more consistent with idealized maps than at other locations. He suggests that this may be due to advantages of local efficiency: "If specific local map configurations enhance local computational capacity, then differences in local parameter coherence may reflect computational strategies, such as integration, differentiation, and correlation capacities" (Schreiner, 1998: 109).

3.5 The coding of speech sounds in A1

The question is, of course, in what way the processing of *speech sounds* is related to the multidimensional maps in A1 and other auditory areas. (The central issue of the relation between speech sounds and the *formation* of auditory maps will be taken up in the next chapter). The processing of speech sounds has been studied in animals (section 3.5.1) and in humans (3.5.2).

3.5.1 Animal studies

Schreiner (1998 for cats) shows that activity patterns in the tonotopic map in A1 reveal the first, second and third formants (F1, F2 and F3), both in the steady-state vowel portions and in the formant transitions of the syllables /be/ (voiced) and /pe/ (unvoiced). In addition, the activity patterns provide onset information: for /pe/ the onset of the stimulus and the onset of voicing are marked separately in time, while for /be/ there is only one strong activity pattern at the beginning of the syllable, reflecting coinciding stimulus onsets and voicing onsets.

Similar results had been obtained earlier by Steinschneider, Arezzo and Vaughan (1990), who studied multiple-unit activity (MUA) in an old world monkey's A1. The results are more representative of human processing than Schreiner's findings (1998), because the overall and cytoarchitectural organizations of the auditory cortex of old world monkeys are similar to those of humans and old world monkeys also show similar discrimination performance of CV syllables. As will become apparent below, it is important to pay attention to the measurement technique used: MUA is recorded by inserting electrodes into the brain and summing the APs from multiple neurons that lie around the contacts. The result is a wave of combined responses, reflecting ensemble spike activity.

Steinschneider et al. presented synthetic syllables /ba/, /da/ and /ta/ to the monkey, as well as isolated formants and tones. Just as in Schreiner (1998) the responses to the syllables were "meaningfully related to the tonotopic organization of A1" (Steinschneider et al., 1990:.167). The same was true for the responses to isolated formants. The syllables elicited excitatory activation at sites that represent lower frequencies up to about 1200 Hz. Consequently, F1 and F2 with frequencies falling below 1200 Hz were reflected in the tonotopic map. In contrast, the syllables tended to inhibit sites representing higher frequencies.

As a result the F3 did not appear in the tonotopic map. At the same time, the high frequency regions revealed temporal characteristics better than the regions representing lower frequencies. For example, the interval between the onset of the aperiodic portion and the onset of the periodic portion of /ta/ corresponding to voice onset time (VOT) was reflected more accurately. Also, there was better phase-locking in response to high tones. In the lower-frequency regions temporal characteristics were also represented, but less precisely.

At first sight, these results seem to fly in the face of statements by many authors (e.g., Abrams and Kraus, 2009) that phase-locking to high stimulus frequencies declines along the pathways in such a way that the highest stimulus frequencies that neurons can phase-lock to declines from around 3,500 Hz in the auditory nerve to 100 Hz in the auditory cortex (section 2.5.3). However, these statements are based on measurements of either multiple neuron *PSPs* or of *single* neuron APs (recall section 2.4). PSPs last too long to be able to extract APs. And indeed, in going along the pathways up to the cortex, it is increasingly difficult for individual neurons to accurately phase-lock to stimulus features. However, the waveform resulting from *combined AP patterns* (which is what is measured with MUA) reveals a *combined* phase-locked response. Even though individual neurons in the cortex cannot phase-lock to stimulus frequencies accurately, the auditory system can extract precise timing information from their combined AP patterns. Obviously the cortex needs this precise timing information to keep track of the ongoing speech signal.

Further, Steinschneider et al. (1990) found that the responses to the syllables revealed complex, non-linear summations of the responses to the isolated formants, both in A1 and in the thalamocortical afferents, which showed similar responses to all stimuli as the responses that were recorded in A1. The authors suggest that the non-linear summation is partly due to excitatory and inhibitory interactions between formants,[19] which occur in the MGC of the thalamus and even more so in A1, where multiple inputs converge. Also, referring to similar results in other animal studies, the authors point out that the interactions "may sharpen response differences that reflect consonant place of articulation" (p.166), leading to a higher specificity of cortical cell responses to complex stimuli as compared to the thalamic cell responses. As an interesting example, formant interactions with the F3s in particular appeared to sharpen differences between the syllables. Thus, although F3 did not show up as a separate representation in the tonotopic map, it enhanced the response differences between syllables.

[19] Note that the authors talk about interactions between *formants*, as measured in the combined response wave, not about interactions between neurons.

Another often cited animal study of vowel coding is an article by Ohl and Scheich (1997), who studied A1 in Mongolian gerbils. Gerbils are very sensitive to frequencies in the speech range and they can learn to discriminate vowels. Again F1 and F2 elicited tonotopic activation in the low-frequency areas, but importantly they showed up as a single area of activation and not as separate stripes. Also, the activated region was smaller for vowels with neighboring F1s and F2s and larger for vowels with a large distance between the formants in the frequency space. This finding was shown to be based on inhibitory interactions between neighboring formants: the "proportion of inhibitory formant interactions depended on both the formant distance F1-F2 and the spatial coordinates along isofrequency contours" (p.9442). These findings indicate that the topographic organization of the gerbil A1 emphasizes an F1-F2 relationship, which is crucial for vowel identification.

3.5.2 Studies on human perception

To what extent is the description of animal speech sound coding in A1 valid for humans? Although lower-level processing may be similar in humans and animals, there could also be important differences, as speech sounds are not meaningful to animals. Because obviously studies on human subjects cannot use invasive measurements, they tend to resort to MEG (section 2.4) when studying the representation of human speech in the auditory cortex. Stimuli are usually vowels and the main component that is looked at is typically the N100m (see below). Example studies include Diesch and Luce (1997) who compared responses to composite stimuli (both a tonal stimulus and a vowel stimulus) and their individual components (respectively pure tones and formants); Obleser, Elbert, Lahiri and Eulitz (2003) who studied German vowels; Shestakova, Brattico, Soloviev, Klucharev and Huotilainen (2004) who report on Russian vowels (see for these examples also Abrams and Kraus, 2009) and Mäkelä, Alku and Tiitinen (2003) for Finnish vowels.

The problem with MEG is, however, that the technique cannot provide accurate spatial results, so that it cannot shed light on questions such as where vowels are coded for in a specific region of the auditory cortex. Although the spatial resolution of MEG is better than that of EEG (which is partly because the signals are not blurred as they travel through the skull), ERMF (Event Related Magnetic Field) measurements still suffer from the non-uniqueness problem just as ERP (Event Related Potential) measurements do. That is, for every measurement there are an infinite number of possible source generators. Although this number can be reduced on the basis of assumptions, the conclusions on localization have limited reliability. According to Näätänen and Picton the N100m component has at least six subcomponents and is affected by several parameters such as stimulus onset, stimulus change and attention (Näätänen and Picton, 1987). Not surprisingly, the specification of the area in the auditory cortex that is targeted is usually lacking. However, there are suggestions that one source of the N100m component is on the gyri of Heschl (e.g., Diesch et al., 1996) and could therefore represent A1.

In spite of these shortcomings the MEG studies disclose results that seem meaningful in the light of the findings for animals. For example, they provide evidence for a mapping of F1 and F2 in a cortical map: Obleser et al. (2003) find for German vowels that the "acoustically most dissimilar vowels [a] and [i] showed more distant source locations than the more similar vowels [e] and [i]" (p. 207), implying that vowels that are close to one another in the F1-F2 vowel space are also close to one another in a cortical map. Similar results for vowels in other languages are obtained by Shestakova et al. (2004; for Russian vowels) and Mäkelä et al. (2003; for Finnish vowels). In addition, Diesch and Luce (1997) find evidence for interactions between F1 and F2: in their study the response to composite vowel stimuli is different in location and strength[20] than that of the combined response to the component-formant stimuli (F1 and F2).

The interactions reported in the MEG studies tend to be interpreted as support for the existence of a 'phonemotopic map' that is orthogonal to the tonotopic map, i.e., a map that spatially represents vowel phonemes rather than lower-level acoustic information, although the authors in the different studies stress that their results do not prove the existence of such a map. On the one hand, the animal studies show that at least in the case of vowels it is not strictly necessary to hypothesize a separate map of phonemes to account for the data. On the other hand, three observations leave open the possibility that A1 contains phonemic representations.

First, Shestakova et al. (2004) report their results (clear differences in cortical map location between the vowels; see above) despite the fact that they used highly variable stimuli: for each of the Russian vowels /a/, /u/ and /i/ they presented participants with 150 natural stimuli (450 in total), each of which was pronounced by a different male speaker (450 speakers in total). Unfortunately, as the authors remark, "without having a nonspeech control, it is impossible to conclude whether it is a [sic] phonological information which is used in the category encoding or the frequency differences as such." (Shestakova et al, 2004: 349)

[20] Note that the strength of the response in MEG influences localization.

Second, Diesch and Luce (1997) notice that the response to the composite vowel (see above in this section) is also faster. The phenomenon that speech sounds are processed faster than non-meaningful sounds is confirmed in other studies (e.g., Dehaene-Lambertz, Pallier, Serniclaes, Sprenger-Charolles, Jobert and Dehaene, 2005; Dehaene-Lambertz and Gliga, 2004; Mäkelä, Alku, Makinen, Valtonen, May, Tiitinen, 2002). At the same time, it seems more likely that the faster processing of speech sounds results from top-down influences rather than from separate representations in A1 (Dehaene-Lambertz et al., 2005).

Finally, animal studies reveal that a small population of cells in A1 responds very selectively to certain species-specific vocalizations (Wang, Kadia, Lu et al., 2003 for marmoset monkeys). The authors suggest that the population "serves as a signal classifier" (p.291) of familiar sounds, whereas the larger population of other cells in A1 "serves as a general sound analyzer that provides the brain with a detailed description" (p. 291) of unfamiliar sounds. The observations retain the possibility that the human A1 contains cells that are sensitive to specific phonemes.

3.6 Summary

A1 contains ordered maps of several parameters, which are spatially overlapping but can be separated functionally (Schreiner, 1998; section 3.4). Cells thus code for multiple features. The most widely studied parameter is tonotopicity.

As for speech sounds, spectral information seems to be represented in the tonotopic map. This holds in particular for lower-frequency information, encompassing the frequencies of F1 and F2, as was shown in animal studies and also in several MEG studies with human participants (section 3.5). Higher-frequency information, such as that related to higher formants, is not separately visible in the tonotopic map (as was demonstrated in the animal studies in section 3.5). The role of higher formants in speech sound processing appears when looking at the *summed* response waves of multiple neurons: the higher formants, in particular F3, contribute to sharpening the differences (Steinschneider et al., 1990; section 3.5.1).

Apart from reflecting acoustic analysis, A1 could play a role in the representation of phonemes or of some intermediate level between acoustics and phonemes. Although phoneme-sensitive cells have not been detected as of yet, there are indications that A1 neurons can develop specialized sensitivities to species-specific sounds (Wang, Kadia, Lu et al., 2003; section 3.5.2).

4. Plasticity in the auditory cortex

4.1 Introduction

The multidimensional maps in A1, which were discussed in the previous chapter, are not yet present at birth. This means that infants must form them as well as adapt them in response to changes later in life. In this chapter I attempt to unravel some of the maturational and experiential factors that determine the initial formation and later adaptation of the maps and the cortical representations of speech sounds therein. This is important for understanding the possible scope of distributional learning, in particular for understanding if distributional learning can help adult L2 learners as well.

In the following I will frequently use the term 'plasticity'. In general it denotes the capacity of being shaped. More specifically, it refers to different forms of plasticity, which are caused by different factors and considered at different levels of analysis. For example, adaptive plasticity refers to changes caused by experience (sections 4.2 and 4.3). This contrasts with plastic changes as a result of maturational factors (section 4.5). Also, plasticity can be studied at the level of the cortical map and the neurons therein (sections 4.2 and 4.3) or at the level of cellular mechanisms (section 4.4).

4.2 Adaptive plasticity in animal infants' A1

4.2.1 Studies on tonotopic maps in rat pups

Animal studies, in particular studies on rat pups, confirm that early experience has a profound influence on the development of A1. Both normal development and the exposure to a specific form of input have been studied, in particular for tonotopic maps. In this section I will discuss these findings. (For other parameters: see e.g., Eggermont, 1991 for the cat and Chang and Merzenich, 2003 for rat pups). As will become apparent in the discussion below, none of the studies to date have used controlled *distributional* inputs, such as bimodal distributions. Still, the results are interesting, as they provide insights into how input affects brain processes in infants. These insights have implications for distributional learning.

4.2.2 Normal development

In normal conditions the development of a rat pup's A1 is characterized by two distinct developments (Zhang, Bao and Merzenich, 2001). First, there is increasing tonotopicity in addition to narrowing RFs and shorter-latency responses of individual neurons in a zone that was initially responsive to low and middle frequencies. And second, there is a progressive deterioration of the zone with neurons that were initially responsive to all frequencies including high frequencies. The function of this shift in the tonotopic map is not clear. Only the first zone seems to develop into the adult rat A1. Further, the total area responding to tones is initially larger than that in adult rats[21], although the rat pup's brain is much smaller than the adult rat brain.

It is uncertain to what extent these developments also occur in the human infant brain. What is known, is that most probably human infants do not have tonotopic maps at birth yet (section 4.5). Also, it has been shown that latencies in response to pure tones and phonemes shorten with age (Cheour, Alho, Čeponiené, Reinikainen, Sainio, Pohjavuori, Aaltonen and Näätänen, 1998).[22]

4.2.3 Abnormal development triggered by specific inputs

If we focus on passive exposure (rather than conditioned exposure[23]), five main conditions have been studied. They consist of exposure to (1) continuous tones, (2) continuous noise, (3) pulsed tones, (4) pulsed noise and (5) silence.

Exposing rat pups to either *continuous tones* consisting of one frequency (Zhou, Nagarajan, Mossop and Merzenich, 2008) or *continuous white noise* (Chang and Merzenich, 2003) blocks the formation of a tonotopic map and the refinement of tuning, and thereby extends the closure of the critical period. (The existence of a critical period beyond which passive exposure does not induce substantial changes in tonotopicity and tuning properties any more, is discussed in many studies; e.g., for rats see De Villers-Sidani, Chang, Bao and Merzenich, 2007[24]). A parallel effect results from early sound deprivation, which also delays the maturation of the cortex, as evidenced by the fact that congenitally deaf animals as well as human children can attain normalized hearing after receiving cochlear implants (e.g., Kral, Hartmann, Tillein, Heid and Klinke, 2001, for cats and humans; Nicholas & Geers, 2007, for humans; Kral, 2009, for humans). Thus, the length of the critical period depends on the nature of the input: it is extended when animals are exposed to continuous sounds and when animals

[21] Judging from the pictures in the article (Zhang et al, 2001), the area responding to tones is about 5 mm^2 in rat pups versus less than 1 mm^2 in adult rats.

[22] The shortening of response latencies was measured with EEG. The paradigm was a MisMatch Negativity (MMN) paradigm. In these paradigms a series of identical sounds with occasional deviant sounds are presented to participants. The deviants elicit MMN responses in the brain.

[23] Conditioned exposure refers to exposure to sounds that is paired with, for example, a mild electric shock.

[24] For rats the critical period is very short: it extends from postnatal day 11 to postnatal day 13 (De Villers-Sidani, Chang, Bao and Merzenich 2007).

and humans do not hear sounds. As a consequence it is difficult to pinpoint the length of the critical period without defining the exact input.[25]

The effect caused by continuous sounds contrasts with the effect of exposure to either *pulsed tones*[26] consisting of one frequency (Zhang, Bao and Merzenich, 2001) or *pulsed white noise* (Zhang, Bao and Merzenich, 2002), which accelerate the consolidation of the cortical map. These findings indicate that temporal patterns in the input may be functional in ending the critical period.

Also, *continuous tones* reduce the area with neurons that selectively respond to this tone (Zhou et al., 2008), while *pulsed tones* lead to an enlarged area and thus to an over-representation of the presented tones (Zhang et al., 2001). This indicates that patterning enhances the salience of the input.

Further, *pulsed tones* do not only bring about larger areas, but also an overall deterioration of tonotopicity and neuronal response selectivity (i.e., the RFs were broad). Interestingly, in the study by Zhang and colleagues (2001) the latter effects (i.e., the deterioration of tonotopicity and neuronal response selectivity) could not have been caused by the lack of sufficient input or to other environmental factors, because tonotopic maps in other rat pups, which were raised in the same sound-attenuated room without tone pulses, developed normally. The disrupted tonotopicity and the broad RFs thus seem to stem from either the temporal pattern of the input (i.e., the pulsing) or from the over-representation of the input tones or from both factors.

Presenting *pulsed noise* instead of pulsed tones also yields disrupted tonotopicity and degraded RFs, but in addition the RFs are incomplete and may have multiple peaks (Zhang et al., 2002). In comparison, *continuous noise* causes the RFs to be complete and single-peaked (but broad) (Chang and Merzenich, 2003).

There are other studies than the ones mentioned on pulsed exposure that highlight the importance of temporal aspects. In particular, Nakahara, Zhang and Merzenich (2004) show critical temporal-order effects. Presenting rat pups with two tone sequences consisting of a sequence of three relatively low frequencies (2.8, 5.6 and 4 kHz) followed by a sequence of relatively high frequencies (15, 21 and 30 kHz) disrupted map formation, which could not be

[25] The length of the critical period for the formation of tonotopic maps in humans is unknown. Many articles referring to a 'critical period' are based on children who are born deaf and who receive cochlear implants later in life. Due to different criteria and measurement techniques the critical period differs in different studies. For example, Kral (2009) states that in order to develop more or less *normal hearing* a human child must get auditory input before the age of 4. Nicholas et al. (2007) mention that for *normal production* to develop the child must get input before the age of 2.

[26] Pulsed tones consisted of 25-ms tones of one frequency with 5-ms ramps. "No noticeable distortion of tonal stimuli was detected from sound spectrum [sic] recorded inside the test chamber." (: Zhang et al., 2001: 1129).

undone. Apart from a weakly tuned middle zone representing the middle frequencies (which had not been part of the stimuli), other more surprising results include an under-representation instead of an over-representation of one of the tones (viz., the third tone in the first sequence; this effect was attributed to temporal inhibition by one or both of the preceding stimuli) and a shifted representation of the first and the second tones in the first sequence as compared to the representations of other rats exposed to pure-tone stimuli (viz., the first lower-frequency tone was represented lower and the second higher-frequency tone was represented higher than expected). Because the third tone had a spectral content in between the first and the second tone, these effects were attributed to "the spatial and temporal interactions, among those inputs, and to an integration of various plasticity effects." (Nakahara et al., 2004: 7173)

4.2.4 Conclusion

It is clear that auditory experience shapes the parameter maps in infants' A1, even in the absence of behavioral relevance. Normal and thus distributional input supports a balanced development of tonotopicity and leads to sharper RFs and shorter response latencies. Exposure to atypical inputs illustrates that apart from stimulus content, temporal aspects such as continuous versus pulsed exposure or the order of presentation play a crucial role. Frequently recurring sounds (as opposed to sounds that are continuously present) tend to be represented by a larger area. The fact that non-distributional input disrupts or halts map formation suggests that only distributional input can lead to balanced map formation. This may be due to more balanced interactions between nerve cells in response to distributed input as opposed to a distortion of neuronal interactions triggered by non-distributional input.

4.3 Adaptive plasticity in adults' A1

4.3.1 Continuous plasticity

The continuing plasticity of the adult brain is clear in everyday life and has also been demonstrated in numerous studies with both animal and human participants. As evident from animal studies, the basic layout of parameter maps remains untouched, but RFs can change and there can be changes in the sizes of the areas that code for certain features. Representative is a study on owl monkeys by Recanzone, Schreiner and Merzenich (1993), showing that after a discrimination training the A1 area representing the trained frequencies had expanded and that the RFs in the targeted region had sharpened. This "relationship between receptive field size and the cortical area of representation (…) [which is found after discrimination training in many studies; KW] is thought to arise by competitive interactions of excitatory and inhibitory inputs between cortical neurons comprising a horizontally oriented network." (Recanzone, Schreiner and Merzenich, 1993: 100).

A possibly related result was obtained in an fMRI study with human participants, who displayed an increase in activation in A1 after a discrimination training (Guenther, Nieto-Castanon, Ghosh and Tourville, 2004). However, as the authors acknowledge, more activation as measured by fMRI does not necessarily result from an expanded area and changes in RFs, but could also follow from, for example, longer activation of the same cells. In this respect it is noteworthy that Recanzone et al. (1993) also found longer response latencies for the trained stimuli.[27] (Recall that response latencies tend to shorten during development in infants; section 4.2.2).

Changes in tuning of adult cortical cells can occur very rapidly. Although the length of training across experiments is very diverse, ranging up to several months, some studies show long-lasting adaptations in tuning after only a few minutes of training (e.g., Buonomano and Merzenich, 1998).[28] Further, once neurons in A1 have acquired new RFs, an animal can switch between RFs in accordance with task demands. This was shown in studies of the ferret A1 (Fritz, Elhilali and Shamma, 2005). This result suggests that the human A1 may contain overlapping representations for both Auditory and Surface Forms (section 1.3).

[27] The authors hypothesize that "The longer latencies in the trained monkeys may have resulted from increased neural processing at subcortical levels, or perhaps from a suppression of normally shorter latency inputs that are involved in processing other aspects of the acoustic signal."(Recanzone et al., 1993: 99)

[28] The measurements were done on slice preparations and not recorded in vivo. It is not clear to what extent long-lasting changes would also occur in vivo.

4.3.2 Training versus passive exposure

Although it is clear that the adult brain remains plastic, there is an ongoing debate as to the issues whether explicit training is indispensable or not and whether adults can generalize newly learned features to other stimuli or not. The dominant opinion on the former issue is that in contrast to infants, both animal and human adults "must attend to (i.e., respond to) a stimulus in order for plasticity to be induced" (Keuroghlian & Knudsen, 2007: p. 109; see also the impressive list of references therein, supporting this conclusion for many animal species with several sorts of stimuli presented in different paradigms; Also recall section 4.2.3). Some studies directly compare passive exposure to trained exposure. For example, Polley, Steinberg and Merzenich (2006), who studied rats, report that whereas passive exposure did not lead to cortical changes, trained exposure doubled the A1 region representing the target frequency. Notably, an even larger expansion of the target area (viz., 2.5 times enlargement) was observed in secondary areas.

In contrast to these accounts there are the few studies that were mentioned in chapter 1, which indicate that adults can learn passively if exposed to *distributional* inputs (Gulian, Escudero and Boersma, 2007; Maye and Gerken, 2000, 2001). Interestingly, such inputs have not been applied in any of the studies representing the dominant opinion.

4.3.3 Generalization to novel stimuli

As to the second issue, the prevailing claim is that adults cannot generalize learned features to novel stimuli (Keuroghlian & Knudsen, 2007). However, this assertion does not hold either. In combination with training there are several accounts of generalization abilities (e.g., McClaskey, Pisoni and Carrell, 1983; Tremblay, Kraus, Carrell and McGee, 1997). As was discussed in chapter 1, it is not clear if passive exposure to distributional inputs facilitates generalization.

4.3.4 Conclusion

The adult brain retains a high level of plasticity. Just as with infants, plasticity can impact neuronal RFs and the cortical territory for the trained parameter within minutes. Interestingly, single neurons can acquire different RFs and switch between these RFs depending on the situation and thus on top-down influence (Fritz et al., 2005). It implies that humans may develop phonetic and phonemic representations in the same map and even in the same cells.

At the same time the adult brain is less plastic than that in infants. There is an ongoing discussion as to the conditions that induce plasticity in the adult brain. Both training and passive exposure to distributional (but not to non-distributional) inputs may elicit plastic changes at different levels of representations, but their respective contributions cannot be pinpointed on the basis of the discussed studies. I will come back to this issue in section 4.5.

4.4 Underlying mechanisms of cortical plasticity

4.4.1 Hebbian and other forms of synaptic plasticity

In the previous sections we saw that cortical map changes resulted from changes in RFs, both for infants and for adults. The cellular mechanisms that drive this adaptive cortical plasticity (i.e., the experience-dependent changes in neuronal RFs which affect the layout of parameter maps) are still largely a mystery, in spite of steady advances in research. The common assumption is that the underlying mechanism is *synaptic* plasticity, in particular Long Term Potentiation (LTP) of synapses on the basis of Hebbian learning (Buonomano and Merzenich, 1998; the discussion in this section is based on this article, unless stated otherwise). Hebbian learning (Hebb, 1949) refers to the strengthening of a synapse between cells if the cells are active simultaneously. Hence the commonly used slogan "fire together, wire together" to explain Hebbian learning. More precisely, Hebbian learning occurs when the firing of a presynaptic cell is paired with depolarization in the postsynaptic cell, so that the synapse between them is strengthened and remains strengthened in the longer run. Because the change in synaptic strength is long-lasting it is labeled LTP. When Hebbian learning occurs, synapses are excitatory, because the presynaptic activity increases the probability that the postsynaptic cell will fire.

There is considerable evidence that Hebbian-like synaptic plasticity occurs in the cortex, including the auditory cortex, although the process has not been registered firmly in living animals. It is shown that LTP can occur after just a few minutes of stimulation and that NMDA receptors (which are a type of glutamate receptors) play a role in detecting the co-activity of cells. In addition, there are indications that blocking these receptors hampers cortical reorganization.

However, Hebbian plasticity is not the only form of synaptic plasticity affecting cortical reorganization. For example, some synapses may induce *inhibitory* rather than *excitatory* activity when they are strengthened. In more technical terms, they may invoke Inhibitory PostSynaptic Potentials (IPSPs) rather than Excitatory PostSynaptic Potentials (EPSPs).[29] In addition, there are factors that *weaken* rather than *strengthen* synapses and which thus lead to Long Term Depression (LTD) rather than to LTP. One of these factors is low-frequency stimulation. Thus, while the pairing of a presynaptic input and subsequent

[29] Inhibitory and excitatory synapses (evoking IPSPs and EPSPs respectively) depend on different neurotransmitters (section 2.4). Most excitatory synapses use glutamate as a transmitter. There are several types of glutamate receptors, such as AMPA (alpha-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate) receptors and NMDA (N-methyl-D-aspartate) receptors. Most inhibitory synapses use GABA (gamma-aminobutyric acid) or glycine as transmitters.

postsynaptic depolarization leads to LTP, activity in the postsynaptic cell that is not temporally related to presynaptic input will produce LTD. Consequently, it is not primarily the amount of activity that determines whether LTP or LTD occurs, as is assumed in many computer models that implement Hebbian learning, but rather the relative timing of pre- and postsynaptic activity. Although the importance of this so-called "Spike Timing-Dependent Plasticity" (STDP) for understanding plasticity has been known for some time (see the list of references in Abbott and Nelson, 2000, which contains for example Levy and Steward, 1983) more recent research has boosted interest in the topic and has led to a rapidly rising number of articles (see e.g., the references in Abbott and Nelson, 2000; and Dan and Poo, 2004).

4.4.2 Synaptic plasticity and synaptogenesis during development

If synaptic plasticity is an important underlying cellular mechanism driving cortical plasticity, as was suggested in the previous section, we expect differences in synaptic plasticity between infants and adults. Indeed, genetically programmed cellular mechanisms in infants facilitate changes in synaptic strength, for instance by stretching the duration of PSPs, which are shorter in adults (Kral, 2009). In addition, other innate cellular mechanisms promote rapid synaptogenesis (i.e., the formation of synapses), a process which continues until approximately the first birthday. The subsequent maturation and elimination of synapses appears largely guided by experience and lasts until about 12 years of age (Kral, 2001).

4.4.3 Sites of plasticity

Do the cortical changes arise in direct response to the input or are they affected indirectly via plastic changes at lower levels? That these changes exist at every step along the pathways is undebated (e.g., Buonomano and Merzenich, 1998 in general; Song, Skoe, Wong and Kraus, 2008 for the brainstem). Also, it is known that changes at lower levels induce cortical changes. For instance, when adults get deaf due to impaired afferents the cortex reorganizes so as to over-represent unimpaired frequencies (e.g., Keuroghlian and Knudsen, 2007). Nonetheless, the common impression is that the cortex is the *primary* site of plasticity in response to changes in the environment, so that mostly lower-level (i.e., subcortical and peripheral) changes arise as a result of cortical reorganization (Buonomano and Merzenich, 1998). This is also more in accordance with the fact that in contrast to the cortical maps the organization of lower levels is hard-wired to a large extent. (Recall section 2.5). Thus, it seems that the cortex is the first and foremost site affected by (distributional) input.

4.4.4 Summary

Different forms of synaptic plasticity underlie the changes in the RFs, which in turn affect cortical map organization. Apart from the intensity of cell activities, the timing of the activities is crucial in determining whether a synapse is strengthened or weakened. In addition, strong synapses may either excite or inhibit the firing of post-synaptic cells.

Maturational factors cause an explosion of new synapses and boost synaptic plasticity during development, in particular in the first year of life, i.e., precisely when infants form categories for speech sounds. Still, these maturational mechanisms do not settle the issue whether adults can or cannot learn from passive exposure.

4.5 Maturational plasticity in human infants' A1

4.5.1 Introduction

Synaptogenesis and enhanced synaptic plasticity are not the only cards that the developing brain can play to boost plasticity. In this section I discuss the genetically programmed gross anatomical and cytoarchitectural development of the human auditory cortex, in particular A1. Although the data do not give definite answers on issues such as the differences in distributional learning between infants and adults, they add pieces to the puzzle.

4.5.2 Nature versus nurture

Jean Moore and colleagues (Moore and Guan, 2001; Moore, 2002; Moore and Linthicum, 2007) spent years studying the histological structure of the auditory cortex. They looked at post-mortem tissue, obtained from human fetuses, infants, children and adults. Their results challenge the common opinion that a newborn infant's cortex is mature apart from undeveloped myelination, as expressed in for example Gazzaniga, Ivry and Mangun (2002):

> "(…) virtually the entire adult pattern of gross and cellular anatomical features is present at birth. With the exception of complete myelination of axons in the brain, the newborn has a well-developed cortex that includes the cortical layers and areas characterized in adults." (Gazzaniga, Ivry and Mangun, 2002: 630)

Moore and colleagues show that the newborn's auditory cortex is far from mature. Although it is likely that the majority (or even all) cells are present at birth and also that they have attained their approximate relative positions, it is not the case that the only remaining development is myelination. The nerve cells need to (1) differentiate, (2) grow and, indeed, (3) develop myelin sheaths around the axons to acquire rapid conduction velocity. The first aspect (i.e., differentiation and the accompanying laminar organization) is not acquired until around the first birthday. The second aspect, 'growing', involves different aspects such as dendrite development and the growth of axons in length and diameter. The third aspect, i.e., axonal maturation, continues until around 12 years of age.

Importantly, it is not the case that the primary areas in the auditory cortex (section 3.2) develop first. The architectural development as expressed by aspects 1, 2 and 3 proceeds in the same order and in the same amount of time throughout the auditory cortex, both in the primary and the higher areas (section 3.2) across individuals. This finding by Moore and colleagues demonstrates that this development is not affected by (distributional) input, but is genetically programmed. Of course, input may determine the *representations* that are captured in the architecture. Indeed, this seems very likely in view of the fact that the input shapes neuronal RFs (sections 4.2 and 4.3). The distinction between on the one hand architectural development, which is probably genetically coded for, and on the other hand representational development, which depends heavily on input, has been put forward by others (e.g., Karmiloff-Smith, 2006 and references therein). The following sketches the cytoarchitectural development of the auditory cortex. Unless stated otherwise, the description relies on Moore and Guan, 2001; Moore, 2002 and Moore and Linthicum, 2007.

4.5.3 Gestation: the emergence of layer I and the development of the pathways

Of all six layers in the adult cortex (see figure 10) layer I develops first and turns out to play a special role in cortical development. It arises early in gestation (around 8 weeks), when a primordial layer splits into a marginal layer (turning into layer I later) and a subplate (afterwards developing into layers II through VI). Already around the 22nd fetal week axonal maturation occurs in this future layer I, while the other layers are not even detectable. The role of layer I will become apparent below.

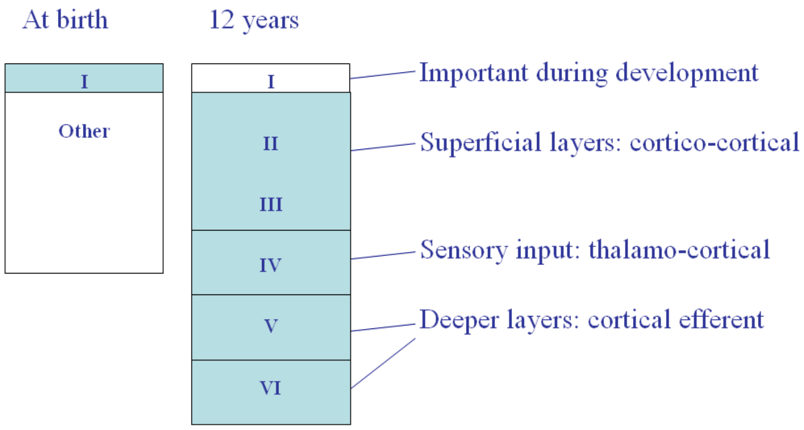

At birth
12 years
I
Other
I
II
III
IV
V
VI
Important during development
Superficial layers: cortico-cortical
Sensory input: thalamo-cortical
Deeper layers: cortical efferent

Figure 10: Simplified view of cortical layers at birth (left) and from the age of 12 (right)
Dark = myelinated
Figure made on the basis of data and microphotographs in Moore and colleagues (2001, 2002, 2007).

Around the beginning of the third trimester of gestation all elements in the auditory pathways extending from the cochlea through the brainstem up to the thalamus are present (including the efferent pathways) and start functioning. Consistent responses to sound in the form of facial and body movements arise around the 28th to 29th week of gestation. However, these responses result mainly from brainstem reflexes (i.e., the sound does not reach the cortex, but travels up to the brainstem, where it elicits a motor reflex). The influence from the sparse afferent axons entering the marginal layer from the thalamus is still limited.

4.5.4 Around birth: mature pathways and immature cortex

Around the time of birth the cortex is 1.2 mm deep (as compared to around 2.0 mm in adults) and still immature, whereas the auditory pathways up to and including the thalamic relays have attained an adult-like appearance. Apart from layer I, layers are not detectable, although a relative concentration of immature neurons hints at incipient lamination of future layers II and IV. (Layer II will develop into a layer that receives input from other cortical areas. IV is the main thalamic input layer; see figure 10).

Layer I contains myelinated axons, which are absent from the rest of the cortex. Notably, the axons run "tangentially for long distances, up to several millimeters, across the cortical surface. As they traverse the layer, they contact and synapse with large numbers of apical dendritic tufts of deeper-lying pyramidal cells. It has been suggested (Marin-Padilla and Marin-Padilla 1982) that non-specific stimulation of cortical plate neurons through their apical dendrites plays a vital role in their structural and functional maturation." (Moore and Guan, 2001: 307) This interpretation is supported by the fact that the number of myelinated axons in this layer is at its maximum from the late fetal period into the first year of life. The number decreases from about 4.5 months of age. Layer I thus seems crucial for the initial development of the cortex and it seems feasible that it serves a role in map formation.

4.5.5 Until around 4.5 months: prominent role for layer I

Around 4.5 months the cortex is 1.4 to 1.6 mm deep. Layers II and IV appear somewhat more clearly and differentiation into the layers II through VI has begun. All these layers are still immature. Layer I has developed into a two-tiered band. The lower tier contains "closely packed axons" (Moore and Guan, 2001: 302) belonging to neurons that originate from the marginal layer ('intrinsic neurons') and which are most numerous during gestation. They disappear after this time (i.e., after 4 to 5 months of age), which emphasizes their role in the development of the cortex rather than in processing. The upper tier contains dispersed axons and represents neurons that receive direct input from the thalamus ('extrinsic neurons'). Animal studies show that this thalamic input probably originates from the medial division of the MGC, which projects directly onto layer I, and from collateral branches of axons from the ventral division of the MGC, which project to layer IV.[30] The number of the extrinsic neurons also decreases after around 4 to 5 months, but the layer retains a limited number into adulthood. The role of these axons in adulthood is unclear, but could relate to the maintenance of plasticity.

4.5.6 Further maturation until 12 years of age

Around the first birthday the cortex is 1.8 mm to 2.0 mm deep, which is adult-like. It has finally attained a mature laminar organization. In layer I the number of axons has decreased substantially, particularly in the lower tier. After one year of life, there are no prominent changes in depth and organization any more, but neurons keep maturing (section 4.5.2).

The maturation process starts in the deeper layers (IV, V and VI), already before the first birthday in the second half of the first year. The neurons in these layers represent thalamo-cortical afferents (layer IV) and cortical efferents (layers V and VI; see figure 8). Subsequently the more superficial layers (II and III) ripen. They represent (part of the) cortico-cortical projections. This process is most evident after the age of 5. Axons in all layers keep maturing until the child is about 12 years old.

4.5.7 Discussion

The most notable factors in the maturation process of the infant's cortex are (1) the initial prominent role of mature axons in layer I, which disappear between 4 to 5 months of age and the first birthday and (2) the development that starts with a lack of other layers than layer I at birth to an adult-like differentiation into six layers around the first birthday. In particular the first factor must play a role in cell tuning and concomitant map formation. Also, it is tempting to relate the period of 4 to 5 months to the age at which infants become sensitive to native vowel phonemes in discrimination tasks (around 6 months; Kuhl 2004). Similarly, it is remarkable that infants complete basic phoneme learning around the time that the prominence of layer I reduces substantially and laminar differentiation is accomplished.

Further, the early prominence of layer I and the thalamo-cortical layer IV in the absence of axonal maturation in the superficial cortico-cortical layers suggests that in infants (distributional) input reaches the cortex primarily via layers I and IV. Both layers receive direct input from the thalamus. It can be expected, therefore, that auditory input in small children, in particular in the first year of life, shapes neuronal RFs in these layers, in particular

[30] Animal studies also show inputs from adjacent auditory areas. See the references in Moore and Guan, 2001, p. 307.

in layer IV (because most of the axons in layer I disappear). The lack of differentiation and the immature axons in the superficial layers prevent large-scale influence from higher-level cortical areas.

In contrast, the adult auditory cortex contains prominent superficial layers that receive input from other cortical areas. Because top-down information is processed faster than the lower-level information (section 3.5.2), it tends to dominate information coming in from the thalamus. Interestingly, a number of studies show that the new RFs that adults develop when they are learning arise precisely in superficial layers representing cortico-cortical input rather than in the thalamic-input layer IV (Keuroghlian and Knudsen, 2007). This observation has consequences for the differential results that have been described for adult learning during passive exposure: it could be that in the studies that report learning during passive exposure distributional learning somehow bypasses the strong impact of information coming in from the superficial layers. It is unknown whether the scarce remaining axons in layer I play a role in this process.

PART III: COMPUTER SIMULATIONS

5. Description of the model by Guenther and Gjaja (1996)

5.1 Introduction

To study the 'necessary and sufficient' aspects of warping, and the effects of *enhanced* distributions I used the model proposed by Guenther and Gjaja (1996), because it provides an explicit account of the way in which mere exposure to statistical properties of input sounds leads to the creation of perceptual clusters. Specifically, the authors state to model the 'perceptual magnet effect' (Kuhl, 1991), i.e., the perceptual warping of the acoustic space near category centers in such a way that the same acoustic differences are perceived as smaller near 'best exemplars' of a category and as wider near 'poor exemplars'.

In the following I describe a replication of the model (script Boersma, 2009) and two simulations (chapter 6). Finally I will discuss the neurobiological validity of the model (chapter 7). Unless stated otherwise the parameters and procedures in the description below are taken from Guenther and Gjaja (1996).

5.2 The model in short

Figure 11 presents an overview of the model. It contains two layers of artificial neurons. The input layer is called the 'formant representation'. Guenther and Bohland (2002) labeled this layer the 'thalamic map', which was stated to represent neurons in the thalamus. The output layer is the 'auditory map' and symbolizes neurons in the auditory cortex. The figure shows only part of the connections or 'synapses' between the two layers. In reality all thalamic map nodes are connected to all cortical map nodes. In accordance with our definition of distributional learning the connections are unidirectional ('bottom up'), so that there is no ('top down') feedback from the cortical layer to the thalamic layer. The connections are also *adaptive*: their weights change during the learning process in such a way that they gradually come to reflect the input distributions. I will explain how this is accomplished in more detail below.

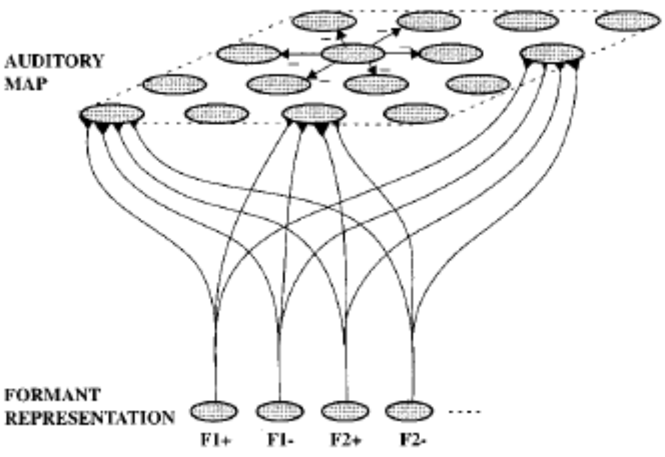


Figure 11: Overview of the model proposed by Guenther and Gjaja (1996)
The input layer ("formant representation") and the output layer ("auditory map") are connected by unidirectional adaptive weights. The arrows between the cells in the auditory layer reflect inhibition. Further explanation: see text.
(Figure taken from Guenther and Gjaja, 1996: 1112).

5.3 Input vectors (layer 1)

The input into the thalamic map consists of formant values expressed in mels. In all simulations we present two formant values, either the first and the second formant (F1 and F2) for vowels or the second and the third formant (F3) for the American English phonemic categories /r/ and /l/. In order to avoid that the high formant values invoke higher summed thalamic node activities than low formants, each formant value is normalized and represented by antagonistically paired values x_i^+ and x_i^- (explanation below) according to the formulas in (1).

(1) $$x_i^+ = \frac{F_i - F_{iMIN}}{\sqrt{(F_i - F_{iMIN})^2 + (F_{iMAX} - F_i)^2}}$$

$$x_i^- = \frac{F_{iMAX} - F_i}{\sqrt{(F_i - F_{iMIN})^2 + (F_{iMAX} - F_i)^2}}$$

where the values x_i^+ and x_i^- symbolize the thalamic node activities, resulting from the i^{th} formant input; F_i is the value of the i^{th} formant input; and F_{iMIN} and F_{iMAX} correspond to respectively the minimum and the maximum values of the i^{th} input formant. The following values were used: F_{1MIN} = 100 mels, F_{1MAX} = 1100 mels, F_{2MIN} = 200 mels, F_{2MAX} = 2200 mels F_{3MIN} = 300 mels and F_{3MAX} = 3300 mels.

Thus, each input formant is represented by an input vector consisting of the values x_i^+ and x_i^-. These values are *normalized*: they vary between 0 and 1 (see table 1). Also, they are *antagonistically paired*: they are paired in such a way that that *the sum of their squared values is always 1*:

(2) $(x_i^+)^2 + (x_i^-)^2 = 1$

Table 1 shows that this is true for example input values. As a consequence of equation (2) and the Pythagorean theorem the length of any input vector (x_i^+, x_i^-) is always 1.

Table 1: Normalization and antagonistic pairing of input formant values.
The values of x^+ and x^- vary between 0 and 1, i.e., they are normalized.
The sum of the squared input vector values x^+ and x^- is always 1, i.e., x^+ and x^- are antagonistically paired.
(Calculation of x+ and x- is according to equation 1)

Example inputs *(in mels)*	**x+**	**x-**	**$(x+)^2 + (x-)^2$**
F1			
100	**0.00**	**1.00**	**1.00**
300	0.24	0.97	**1.00**
500	0.55	0.83	**1.00**
700	0.83	0.55	**1.00**
900	0.97	0.24	**1.00**
1100	**1.00**	**0.00**	**1.00**
F2			
200	**0.00**	**1.00**	**1.00**
600	0.24	0.97	**1.00**
1000	0.55	0.83	**1.00**
1200	0.71	0.71	**1.00**
1400	0.83	0.55	**1.00**
1800	0.97	0.24	**1.00**
2200	**1.00**	**0.00**	**1.00**
F3			
300	**0.00**	**1.00**	**1.00**
800	0.20	0.98	**1.00**
1300	0.45	0.89	**1.00**
1800	0.71	0.71	**1.00**
2300	0.89	0.45	**1.00**
2800	0.98	0.20	**1.00**
3300	**1.00**	**0.00**	**1.00**

5.4 Weight vectors (connections)

All simulations start with an untrained model. That is, the connections have not yet been affected by any inputs. They have randomly chosen weights, which are indicative of the initial preferred tunings or Best Frequencies (BFs, see section 2.5) of the auditory map cells. Guenther and Gjaja do not specify how they assign random values. We use the following procedure (script Boersma, 2009). We take uniformly random values of F1, F2 and F3 and tune each auditory map cell to these values according to the equations in (3). The formulas show that the weights are represented by an input vector consisting of the values z_{ij}^{+} and z_{ij}^{-}. These values are normalized and antagonistically paired in the same way as the inputs.

(3)

$$z_{ij}^{+} = \frac{F_{ij\,BEST} - F_{iMIN}}{\sqrt{(F_{ij\,BEST} - F_{iMIN})^2 + (F_{iMAX} - F_{ij\,BEST})^2}}$$

$$z_{ij}^{-} = \frac{F_{iMAX} - F_{ij\,PREF}}{\sqrt{(F_{ij\,BEST} - F_{iMIN})^2 + (F_{iMAX} - F_{ij\,BEST})^2}}$$

where z_{ij}^{+} and z_{ij}^{-} represent the weight vector between the i^{th} formant input node in the thalamic layer and the j^{th} auditory layer node; $F_{ij\,BEST}$ is the BF of the j^{th} auditory layer node for the i^{th} formant. (See sections 5.7.3 and 5.8.2 for a further explanation of the BF); and F_{iMIN} and F_{iMAX} correspond to the minimum and maximum values for the i^{th} formant. These values are specified under the formulas in (1).

The weights *z* will change during learning (section 5.7). As a result, they will not stay normalized. However, the sum of the squared values will remain in the range from (about) 0.98 to 1. This range was measured after 4000 learning steps (section 5.7.1). Normalizing the *z*-values after every learning step does not change the results. Guenther and Gjaja do not mention this procedure either.

5.5 Activities of auditory map cells (layer 2)

The thalamic-layer activities *x* do not directly affect the weights. First they change the activities of the cells in the auditory layer. Subsequently, in the learning phase the auditory map cells with the largest activities are allowed to learn, i.e., to adapt the connections weights. Therefore, in order to make the model learn, we have to first calculate the cortical-map activities, then select the largest activities and finally change the weights associated with the auditory map cells with the largest activities.

To calculate the activities we use the following formula:

$$(4) \qquad m_j = \frac{\sum (x_i^+ z_{ij}^+ + x_i^- z_{ij}^-)}{M}$$

where m_j is the activity of the j^{th} auditory map cell; the x_i^+ and x_i^- represent the input vector for the i^{th} formant (see the formulas in 1); z_{ij}^+ and z_{ij}^- stand for the weight vector between the i^{th} formant input node in the thalamic layer and the j^{th} auditory layer node (see the formulas in 2); and M is the number of formant values. (Guenther and Gjaja state incorrectly that M is the number of map cells, i.e., 500; however, this number would reduce the activity values m too far below 1). Thus, the activity m of each auditory map cell is the dot product of the input vector and the cell's weight vector divided by the number of formant values M. In all simulations M is 2.

Because the denominator in (4) does not vary, the activity of a cell is largest when the dot product is maximal. This happens when the two vectors are parallel, i.e., "when the angle between them is zero" (Guenther and Gjaja, 1996: 1114). In addition, as the x-vector is normalized and the summed squared z-values remain close to 1, the largest dot product for one formant is always nearly 1. Consequently, the largest dot product for two formants is about 2. Since we divide the summed dot product by the number of formant values, the maximum activity m of a cell is always close to 1.

5.6 Inhibition

Once the activities of all auditory map cells have been calculated, the algorithm selects a limited number of cells with the largest activities. These cells remain active, while the activity values of all other cells revert to zero. According to the authors "This process approximates the effects of competitive interactions between map cells" (p. 1113) as described by for example Kohonen (1982). Note that the arrows in the auditory map layer in figure 1 only symbolize the inhibition process. In Guenther and Gjaja's model the cells do not have locations, so that the arrows do not show spatial neighbors (as they would do in Kohonen, 1982). Every auditory map cell is connected to (i.e., compared to) all other auditory map cells.

The number of cells that are allowed to remain active (L) declines from 35 to 1 during learning (section 5.7) and is 30 in the test phase (section 5.8). Guenther and Gjaja state that in the former case this number "is a monotonically decreasing function of time" (p.1113). However, they do not specify how much time is needed to decline from 35 to 1 or, to put it differently, in how many learning steps (section 5.7.1) this decline is accomplished. In the replication we use the function in (5), which decays more smoothly than a linear decline. This is to allow for a variable number of steps. (This number can be specified by the user of the model).

$$(5) \qquad L = L_{END} + L_{START} * e^{-step/400}$$

where L is the number of surviving cells; L_{END} is the number of surviving cells at the end of the training, i.e. 1; and L_{START} is the number of surviving cells at the beginning of the training, i.e. 35. The *step* refers to the number of steps in the training, which is determined by the user (see also section 5.7.1). The number 400 affects how fast the number of surviving cells declines. Table 2 shows that after around 1700 steps L approaches 1, i.e., from that moment on the number of surviving cells will not drop anymore, but will stay constant at 1.

Table 2: Gradual decline in the number of surviving cells *L* as learning proceeds (i.e., as the total number of learning steps mounts)

Steps	*L*
1	36
10	35
50	32
100	28
200	22
300	18
400	14
500	11
1000	4
1500	2
1600	2
1700	1
1800	1
1900	1
2000	1

5.7 Learning

5.7.1 Gaussian distributions

In the learning phase the model is trained with Gaussian distributions representing phonemic categories (see also section 1.2). More precisely, at each *learning step* one 'input combination' of formant values is presented to the model. Each input combination consists of two formant values.[31] Each formant value represents one token in a Gaussian distribution. For example, for the American English phonemic categories /r/ and /l/ Guenther and Gjaja present tokens from distributions with a centre F3 of 2075 mels for /l/ and of 1200 mels for /r/ (both with a standard deviation of 60 mels), while the centre F2 for both categories is 1000 mels (with a standard deviation of 40 mels). These values are based on perception data provided by Iverson and Kuhl (1996).

5.7.2 Computing the activities in layers 1 and 2

Each input formant value is converted into normalized and antagonistically paired inputs according to the equations in (1) and the activities of the auditory map cells are computed on the basis of formula (4). Also, inhibition is implemented in accordance with equation (5). For example, for the simulated warping of the American English categories /r/ and /l/, which consist of F2 and F3-values, the activity *m* of the *j*th auditory map cell is:

(6) $$m_j = \frac{x_2^+ z_{2j}^+ + x_2^- z_{2j}^- + x_3^+ z_{3j}^+ + x_3^- z_{3j}^-}{2}$$

5.7.3 Learning equations

Only the surviving auditory map cells are allowed to learn, i.e., to adapt the connection weights z_{ij}^+ and z_{ij}^-. Guenther and Gjaja use the differential learning equations in (7):

(7) $$\frac{dz_{ij}^+}{dt} = \alpha m_j (x_i^+ - z_{ij}^+)$$

$$\frac{dz_{ij}^-}{dt} = \alpha m_j (x_i^- - z_{ij}^-)$$

where alpha is the learning rate, which is set at 0.04. However, the authors do not specify the time step *dt*. If we assume that *dt* is 1, we can determine the new weights as in the formulas in (8):

[31] It is possible to do the simulations with only one formant value or on the basis of more than two formant values.

(8) $$z_{ij}^{+}(new) = z_{ij}^{+}(old) + \alpha m_j(x_i^{+} - z_{ij}^{+})$$

$$z_{ij}^{-}(new) = z_{ij}^{-}(old) + \alpha m_j(x_i^{-} - z_{ij}^{-})$$

where $z_{ij}^{+}(new)$ and $z_{ij}^{-}(new)$ refer to the new weight vector and $z_{ij}^{+}(old)$ and $z_{ij}^{-}(old)$ are the old weight vector. We choose the old values to represent z_{ij}^{+} and z_{ij}^{-}.

The formulas in (8) show that the modification of the weights is dependent on both the 'pre-synaptic' input-layer activity *x* and the 'post-synaptic' output-layer activity *m*. This coupling is meant to reflect Hebbian learning (see section 4.4.1). According to the formula an active cell (i.e., *m*>0) gradually rotates its weight vector towards the input vector. If a weight *z* is smaller than an input *x*, *αm(x-z)* will be positive and the weight will increase. If a weight *z* is larger than an input *x*, *αm(x-z)* will be negative and the new weight will be smaller than the old weight. In both situations, the new weight will be somewhat closer to the input than the old weight. Learning will stop when the input vector and the weight vector are equal. In that case *αm(x-z)* will be zero and the equations in (8) can be rewritten as in (9):

(9) $$z_{ij}^{+}(new) = z_{ij}^{+}(old)$$

$$z_{ij}^{-}(new) = z_{ij}^{-}(old)$$

Imagine that the input vector remains the same over some learning steps. In this case the activity *m* of a learning cell will grow with each step until it reaches the maximum value of 1, while the size of the change in the weights, i.e., $z_{ij}^{+}(new)$ - $z_{ij}^{+}(old)$, will decline until it is zero. At this point the weight vector is parallel to the input vector and learning stops. Since the cell fires maximally for this input, the cell can be said to be perfectly tuned to the input. Or, in other words, the input vector that is parallel to the cell's weight vector is the cell's *preferred stimulus* or BF. Importantly, because the weight vectors of learning cells change, their BFs also change. Warping results from the fact that the BFs of the cells come to reflect the input distributions.

5.8 Testing

When testing the model (either the untrained map or the map after learning), we present it with a series of formant combinations and plot the *perceptual* spacing between the tokens by calculating the population vector (recall section 2.6). Let us look at this in more detail.

5.8.1 Presenting evenly spaced formant combinations

The formant combinations that are presented to the model in the test phase are auditorily evenly spaced (see figure 12). In the simulation for the American English phonemes /r/ and /l/ we use values that we inferred from Iverson and Kuhl (1996)[32]: the perceptual distance was 200 mels between two adjacent F2 values and 175 mels between two adjacent F3 values.

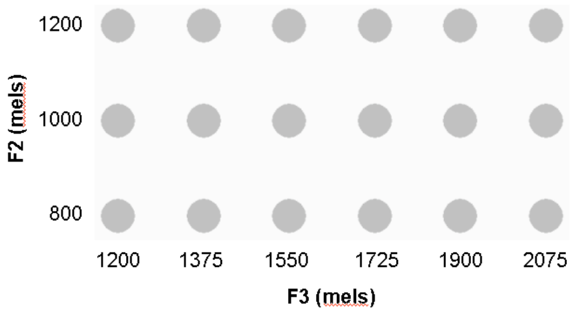


Figure 12: Formant frequencies (in mels) used in the test phase when simulating the perceptual warping of the American English phonemes /r/ and /l/. (Adapted from Iverson and Kuhl, 1996: 1132)

In the script[33] (Boersma, 2009) it is possible to adjust the grid, so that it contains more formant combinations. For example, instead of presenting 3 times 6 combinations, we could present 5 times 10 combinations. The distance remains evenly spaced. Also, notice that the grid pictures a frequency space, not a cell location space, as the cells do not have locations.

Each formant input is converted into an input vector according to the equations in (1). Then, for each input the activities are computed on the basis of equation (4). The 30 auditory map cells with the largest activities remain active, while the other cells are inhibited.

[32] Guenther and Gjaja refer to Iverson and Kuhl (1994) and Kuhl (1995). However, the stimuli that they use were specified in Iverson and Kuhl (1996). Also, Iverson and Kuhl's data are stated in Hertz. The Hertzes were transformed into mels according to the formula: 2595*log(1+F/700).

[33] The script (Boersma, 2009) provides for a pop-up window, which allows the user to adapt several parameters easily. One of the parameters that can be adapted is the number of formant values displayed in the grid. Thus, this number can be adjusted either in the script or in the pop-up window.

5.8.2 Computing each surviving cell's BF

Subsequently, for each input formant *i* we compute $F_{ij\,BEST}$, which is the BF of each surviving cell *j* for input formant *i*. As explained in section 5.7.3, the BF of an auditory map cell is the input formant vector that maximally activates it and this input vector is parallel to the afferent weight vector, as expressed in (10):

$$(10)\qquad \frac{x_i^+}{x_i^-} = \frac{z_{ij}^+}{z_{ij}^-}$$

Using the equations in (1), we infer (11):

$$(11)\qquad \frac{x_i^+}{x_i^-} = \frac{F_i - F_{iMIN}}{F_{iMAX} - F_i}$$

Since we are computing the BF ($F_{ij\,BEST}$), we can substitute F_i with $F_{ij\,BEST}$. On the basis of equations (10) and (11) we then deduce $F_{ij\,BEST}$ as in (12):

$$(12)\qquad F_{ij\,BEST} = \frac{z_{ij}^+ * F_{iMAX} + z_{ij}^- * F_{iMIN}}{z_{ij}^+ + z_{ij}^-}$$

5.8.3 Computing the population vector

The model uses a population vector (section 2.6) to represent neuronal ensemble coding (section 2.6), i.e., the perceptual result for each formant input is represented by a weighted average of BFs (sections 5.7.3 and 5.8.2). The average is weighted by map cell activities, as illustrated in formula (13).

$$(13)\qquad \vec{F}_{iPERC} = \frac{\sum m_j \vec{F}_{ij\,BEST}}{\sum m_j}$$

where $\vec{F}_{iPERC}$ is the perceived frequency, i.e. the perceptual result, for the i^{th} formant; m_j is the activity of the j^{th} auditory map cell and $\vec{F}_{ij\,BEST}$ is the BF or the best i^{th} formant stimulus of the j^{th} auditory map cell.

The perceived formants are plotted in a grid as the grid in figure 12 (section 5.8.1). Figure 13 shows the result for an untrained network with random BFs. The silver circles are part of the grid and show the *presented* formant combinations, while the dark circles stand for the formant combinations as they are *perceived* by the network.

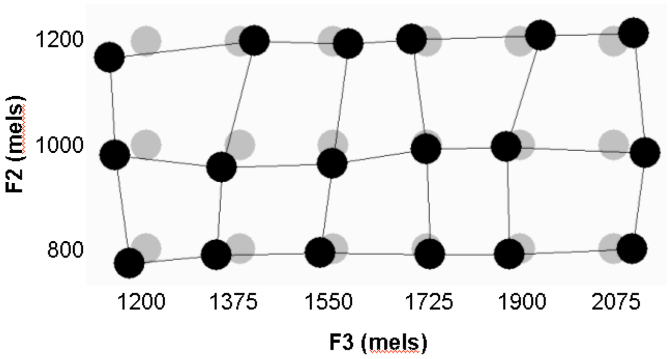


Figure 13: Perception of F2-F3 formant combinations by an untrained network

6. Simulation

6.1 Replication of the perceptual warping of /r/ versus /l/

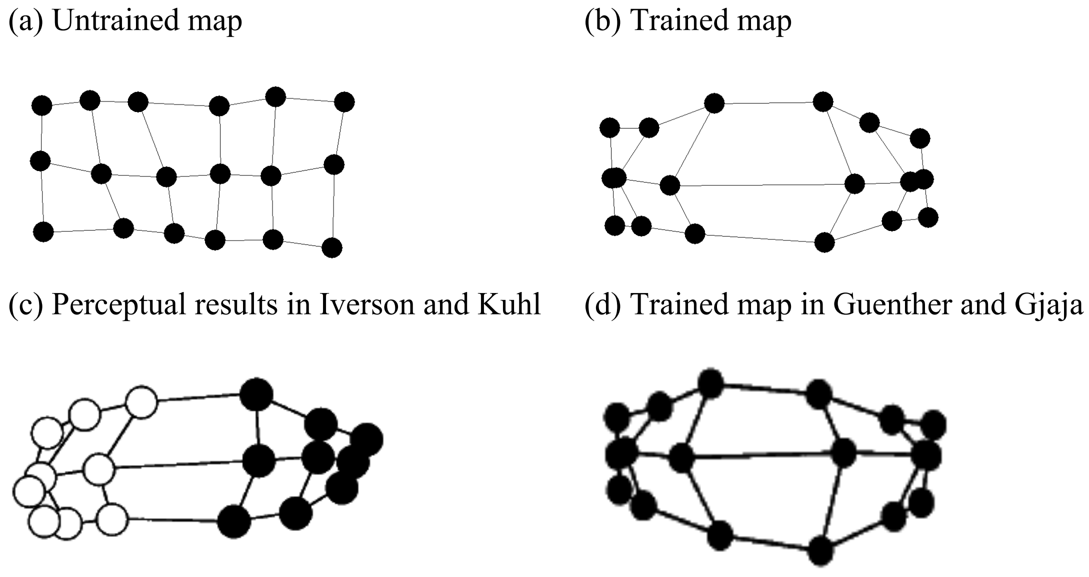
The purpose of the simulation was to show that the replicated results are very similar to those obtained in Guenther and Gjaja's model and that they can reflect a real warping process as revealed in experiments on human perception: the simulation demonstrates the perceptual warping for the American English phonemic categories /r/ versus /l/, as reported for humans by Iverson and Kuhl (1996) and as modeled by Guenther and Gjaja (1996).[34] All parameters for the simulations are listed in table 3. Figure 12 (section 5.8.1) showed all formant combinations used when testing the model.

Table 3: Simulation parameters
Italics = values not reported by Guenther and Gjaja (1996)

Centre frequencies of the formant distributions (in mels)	
F_2 for l	1000
F_2 for r	1000
F_3 for l	2075
F_3 for r	1200
Standard deviations of the formant distributions (in mels)	
SD for F1	20
SD for F2	40
SD for F3	60
Minimum and maximum formant values presented to the model (in mels)	
F_{1MIN}	100
F_{1MAX}	1100
F_{2MIN}	200
F_{2MAX}	2200
F_{3MIN}	300
F_{3MAX}	3300
Lowest and highest formant values shown in the grid (in mels)	
F_{2LOW}	*800*
F_{2HIGH}	*1200*
F_{3LOW}	*1200*
F_{3HIGH}	*2075*
Number of map cells	
	500
Number of surviving cells (see equation 5)	
L_{START}	35
L_{END}	1

[34] See note 32.

Figure 14 (a) depicts an untrained map at the start of the simulation. After feeding the network with F2 and F3 distributions representative for /r/ and /l/, the warping result in figure (b) appeared. The warping is very similar to the results obtained by Guenther and Gjaja (d) and to Iverson and Kuhl's findings (c). The cluster for /r/ is on the left. It has warped towards the centre frequency of the F3 distribution for /r/, which was the lowest F3-value in the grid (1200 mels). The cluster for /l/ is on the right. It has warped towards the centre frequency of the F3 distribution for /l/, which was the highest F3-value in the grid (2075 mels). Also, both clusters have warped towards the centre frequencies of the F2 distributions, which correspond to the middle F2-value in the grid for both /r/ and /l/ (1000 mels).



(a) Untrained map

(b) Trained map

(c) Perceptual results in Iverson and Kuhl

(d) Trained map in Guenther and Gjaja

Figure 14: Comparison of (a) an untrained map in the replicated model, (b) a trained map in the replication, (c) the perceptual results as reported by Iverson and Kuhl (1996). The figure was copied from this article (p.1133) and (d) a trained map as reported by Guenther and Gjaja (1996). The figure was copied from this article (p. 1115). The horizontal axis is the F3-axis, running from low values to high values from left to right. The low to high F2-values are represented on the vertical axis from bottom to top.

Importantly, the simulation shows what Kohonen (1982) described in an analogous fashion for topologically ordered maps, viz., that warping is contingent on only two factors: (1)

a competitive process that selects a group of winning neurons in the map and (2) an adaptive change in the weights that are related to these winners.[35]

6.2 Model perceptual warping with realistic standard deviations

A point of concern is the narrow standard deviations that Guenther and Gjaja use in the simulation (as well as in the other simulations which are not reported here): for modeling /r/ versus /l/ the values are based on Iverson's and Kuhl's reported standard deviations *only* for stimuli that were perceived as 'best tokens' of /r/ and /l/. These standard deviations seem unrealistically narrow: they are of 40 mels for F2 and 60 mels for F3. If we enlarge the standard deviations for this contrast to 100 mels for F2 and 200 mels for F3 (which seem even on the low side; Kinnaird and Zapf, 2004) and if we keep all other parameters equal to the values reported in table 3, then the model fails to warp properly, as illustrated in figure 15. Note that the larger standard deviations do not result in strongly overlapping distributions: the centre values of the F3 distributions for /r/ versus /l/ are still 4.4 standard deviations apart (i.e., 2075 mels for the centre F3 of /l/ minus 1200 mels for the centre F3 for /r/ is 875 mels; 875 mels divided by a standard deviation of 200 mels is 4.4).

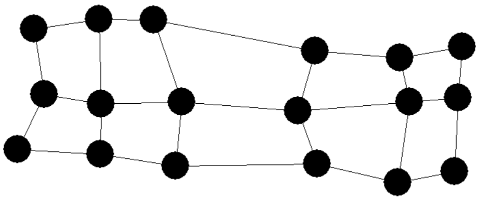


Figure 15: Model trained on /r/ and /l/ in 1000 learning steps
Parameters: see table 3. However, standard deviations were changed into 100 mels for F2 and 200 mels for F3.

In order to make the model warp with realistic and thus larger standard deviations, we have to increase *L* (i.e., the number of surviving cells; section 5.6). Strong warping occurs when *L* is kept at about 200 throughout the training. (This can be achieved by setting L_{START} at 1 and L_{END} at 200.[36] See equation 5 in section 5.6). This is demonstrated in figure 16.

[35] In Kohonen, 1982, the winners are spatial neighbors.

[36] More precisely, when setting L_{START} at 1 and L_{END} at 200, the number of surviving cells is 201 for the first 277 learning steps (section 5.7.1) and 200 afterwards. (See formula 5; section 5.6).

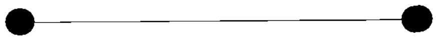


Figure 16: Model trained on /r/ and /l/ in 1000 learning steps
Parameters: see table 3. However, standard deviations were changed into 100 mels for F2 and 200 mels for F3. Also, the number of surviving cells *L* was approximately 200 throughout the training (see text).

However, plasticity diminishes with experience and age (sections 4.2.3 and 4.4.2). Therefore it seems unrealistic to keep *L* more or less constant over time (as we did in order to obtain figure 16). If we set L_{START} at 200 and L_{END} at 1 (equation 5 in section 5.6), *L* declines from 200 to 17 in 1000 learning steps. Figure 17 shows an example outcome.

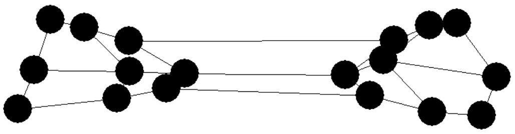


Figure 17: Model trained on /r/ and /l/ in 1000 learning steps
Parameters: see table 3. However, standard deviations were changed into 100 mels for F2 and 200 mels for F3. Also, the number of surviving cells *L* declined from 200 at the start of the training to 17 at the end of the training (see text).

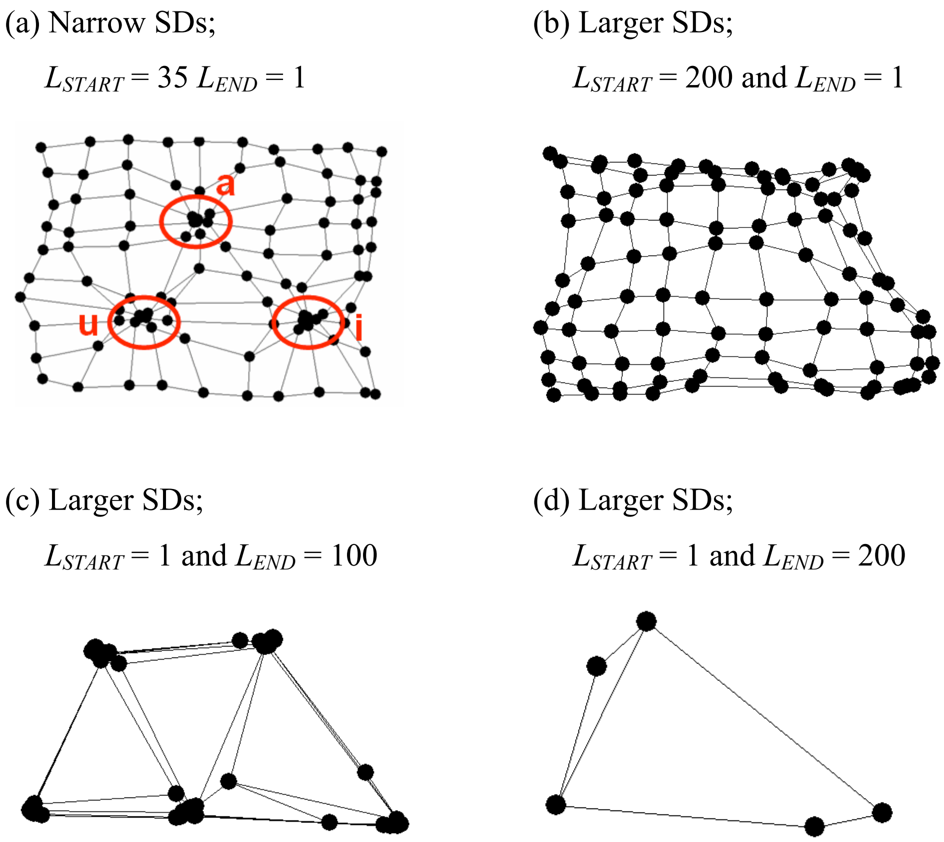
Thus, the model can show perceptual warping even with realistic standard deviations provided the number of surviving cells (*L*) is large enough. Warping is stronger if this number remains high over time. Unfortunately, however, we do not know what a realistic *L* is and whether and how precisely *L* changes over time. Nor can we discover *L* by simply adjusting it until we get a realistic simulated map, because we do not really know what a realistic map is. (The effects of sound on cortical map changes, which were discussed in chapter 4, are far too imprecise to predict how input distributions for phonemes like /r/ and /l/ would change the cortical map). All we can say at this point is that changes in the parameters L_{START} and L_{END} (and thus in *L*; see equation 5) have a substantial impact on the simulation results, as is illustrated by the differences between the figures 14b, 15, 16 and 17. Therefore, discovering the properties of the *L*-function (e.g., how it changes over time and what the roles of age and experience are) is an important issue for future research.

6.3 No conclusions on enhanced distributions

The intention was to look at the predictions of the model for two questions, viz., (1) whether Infant-Directed-Speech (IDS) leads to 'better', more stable categories or not and (2) how infants perceive tokens that fall 'in between' IDS distributions, but which are common in Adult-Directed Speech (ADS) (sections 1.2 and 1.4). Guenther and Gjaja do not simulate the effects of enhanced distributions. However, in view of the uncertainty as to the value(s) of *L* (previous section) it was impossible to predict the answers on these questions.

Figure 18 illustrates once more (cf., the previous section) the large differences between simulation results due to different choices for L_{START} and L_{END}, which define *L*. I fed an untrained network with production data on the Russian vowels /i/, /a/ and /u/, as depicted in graphs by Kuhl and colleagues for both Adult-Directed Speech (ADS) and Infant-Directed Speech (IDS) (Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg and Lacerda, 1997). Figure 18 only gives example outcomes for ADS.[37] The parameters are listed in table 4. In accordance with the description in the previous section, the model warps the perceptual space nicely when small standard deviations are chosen (as used by Guenther and Gjaja; figure 18a), but performs weakly with more realistic standard deviations of 100 to 250 mels for F1 and 200 to 300 mels for F2 (on the basis of Pols, Tromp and Plomp, 1973, for Dutch vowels). Again, as in the previous section, when increasing *L*, the model produces very dissimilar maps depending on the chosen values for L_{START} and L_{END}. This is illustrated in pictures 18b, 18c and 18d. Because we do not know what values of L_{START} and L_{END} could reflect reality, we cannot use the resulting maps for predictions on the effects of enhanced distributions.

[37] Pictures for IDS are similar. Applying the parameter settings of figure 18d to IDS yields a larger triangle than in figure 18d. This is expected, since IDS-values are enhanced (section 1.2).


(a) Narrow SDs; $L_{START} = 35$ $L_{END} = 1$

(b) Larger SDs; $L_{START} = 200$ and $L_{END} = 1$

(c) Larger SDs; $L_{START} = 1$ and $L_{END} = 100$

(d) Larger SDs; $L_{START} = 1$ and $L_{END} = 200$

Figure 18: Pictures **(a) and (b)** show outcomes after 1000 learning steps. They remain virtually the same when learning is continued. Pictures **(c) and (d)** show outcomes after 5000 learning steps. Due to the settings of L_{START} and L_{END} , which cause the number of learning cells to remain virtually constant at 100 (in c) and 200 (in d) the pictures in (c) and (d) still change substantially after 1000 learning steps and keep more susceptible to change even after 5000 learning steps.

(a) The standard deviations (SDs) for F1 and F2 are in accordance with those reported by Guenther and Gjaja (1996). However, The SDs are unrealistically narrow: SD_{F1} = 20 mels and SD_{F2} = 40 mels; **(b)(c)(d)** The SDs are larger than in (a): SD_{F1} = 120 mels and SD_{F2} = 240 mels (based on Pols, Tromp and Plomp, 1973, for Dutch vowels). Each picture in (b)(c) and (d) has different values for L_{START} and L_{END}: **(b)** When L_{START} is 200 and thus larger than in (a) and when L_{END} stays the same as in (a) there is no clear warping. A similar picture results when L_{START} is 500 and thus includes all map cells; **(c)** There is warping when the number of learning cells remains virtually constant at 100 throughout the training. However, warping yields five instead of the expected three clusters; **(d)** There is 'extreme' warping when the number of surviving cells remains virtually constant at 200 throughout the training: the picture stays the same even when the resolution of the grid is increased from 100 circles (10 times 10 circles) to 10,000 circles (100 times 100 circles). There are three to five clusters. It is not clear if the two topmost circles represent one cluster or two. The same holds for the two rightmost clusters.

Table 4: Parameters for training the model on the Russian vowels /i/, /a/ and /u/
Italics = values not reported by Guenther and Gjaja (1996)
Bold = different values from those reported by Guenther and Gjaja (1996)

Centre frequencies of the formant distributions (in mels) *	
F_1 for i	*570*
F_1 for u	*570*
F_1 for a	*960*
F_2 for i	*1810*
F_2 for u	*1100*
F_2 for a	*1350*
Standard deviations of the formant distributions (in mels) **	
SD for F1	20 / **120**
SD for F2	40 / **240**
Minimum and maximum formant values presented to the model (in mels)***	
F_{1MIN}	100
F_{1MAX}	**1450**
F_{2MIN}	200
F_{2MAX}	**2400**
Lowest and highest formant values shown in the grid (in mels)	
F_{1LOW}	*350*
F_{1HIGH}	*1220*
F_{2LOW}	*750*
F_{2HIGH}	*2120*
Number of map cells	
	500

*: The values are inferred from Kuhl, Andruski et al. (1997).

**: The larger standard deviations (120 mels for F1 and 240 mels for F2) are based on data for Dutch vowels, as reported by Pols, Tromp and Plomp (1973).

***: It was necessary to use maximum formant values that exceed those reported by Guenther and Gjaja (1996) to accommodate the Russian vowel distributions.

7. Neurobiological validity

Guenther and Gjaja's model (1996) incorporates several neurobiological elements. It contains cells representing the thalamus and cells representing the auditory cortex. They are connected via synapses that can vary in strength. And this adaptive process depends on both the activities of presynaptic and postsynaptic cells and thus reflects Hebbian learning (section 4.4.1). Also, the fact that all thalamic cells are connected to all cortical cells can be taken to parallel the broad activating influences that are sent out from layer I during infancy (section 4.5). Furthermore, the model implements some sort of competitive interactions between cortical map cells and it mimics the process of diminishing plasticity (section 4.4.2). Finally, a population of cells, rather than a single cell, determines perception (section 2.6).

At the same time the model is simplistic. First, the cells do not have locations and therefore the model cannot implement tonotopicity. On the one hand, it is interesting that the model yields warping *without* tonotopic ordering: in this way it illustrates that a place code is not strictly necessary for warping to occur and it confirms the viewpoint that place does not code directly for stimulus properties (section 2.5.2). This is an important finding. On the other hand, tonotopicity *is* omnipresent along the pathways and in the adult auditory cortex (sections 2.5 and 3.3). Also, the absence of tonotopicity has consequences for the way in which the model implements competition. Whereas in humans *neighboring* cells can influence each other's activities (sections 3.5, 4.2, 4.3), the competitive interactions in the model encompass a comparison of *all* cortical cells for any input.

Interestingly, Kohonen (1982) proposed a neural network that can create topologically ordered maps through self-organization. The generation of the topological order seems to hinge on precisely the introduction of spatial competition: one winner and its *neighbors* are allowed to learn. However, Kohonen did not model the warping of categories in a topological map.

Second, the model ignores temporal aspects. Although it is complex to integrate effects of timing, in reality these effects play a role both in the coding of information (as with e.g., phase-locking to stimulus features, section 2.5.3, and synchronic firing of neuronal ensembles in order to integrate auditory elements, section 2.6) and in influencing synaptic plasticity (which is more dependent on the timing of firing than on the intensity, section 4.4.1). As a consequence of disregarding place and temporal codes (sections 2.5.2 and 2.5.3), information is transmitted in the form of "activities" only (which supposedly reflects rate codes, section 2.5.3).

Third, the model cannot handle ongoing plasticity well. Although there is always one cell that is allowed to learn (section 5.6) and although it is possible to increase this number, there is no procedure that revives plasticity 'in adulthood' in response to new auditory input.

A fourth point, which is somewhat more difficult to judge properly, is that the cortical cells in the model have multiple BFs (section 2.5.1), one for each formant. This means that they could be phoneme-sensitive. Although this does not seem correct for most cells, there are indications that cells *can* have multiple peaks in their frequency tuning curves and that A1 in animals can contain cells that are sensitive to species-specific vocalizations (Wang, Kadia, Lu et al., 2003; section 3.5.2).

Relatedly, in the calculation of a perceived frequency (the population vector) each cell's contribution is considered to be independent from that of other cells (section 2.6). However, in reality the neurons interact (section 3.5). Also, the cell's contribution to the calculation of one formant is taken to be independent from its contribution to the calculation of the other formant, whereas in fact there are interactions between formants (section 3.5). A final point is that the model treats frequency as an isolated parameter, while frequency coding is actually influenced by intensity levels and other parameters (section 2.5.1).

In sum, the model simplifies many neurobiological aspects. The question is how bad this is. Unfortunately, it is difficult to determine what aspects we can abstract away from and what aspects are indispensable for a comprehensive model of distributional learning. The simplification was useful in that it revealed two essential aspects for warping to occur: a form of competition that yields a limited number of 'survivors' and an adaptive process in which the survivors learn. Accordingly, the simplification also showed that warping does not require tonotopicity (as this is not implemented in the model) or top-down influence (which is also not implemented), although it is probably affected by both factors.

8. Discussion and future research

The topic of this thesis was the impact of distributional input on speech sound categorization in the primary auditory cortex (A1). To study this topic I looked at neurobiological aspects and computer simulations. Each approach had its limitations: *distributional* learning (as opposed to learning on the basis of non-distributional input) has not been studied in controlled paradigms in neuroscience, while the computer model was oversimplified. Nonetheless, the combination of approaches yielded some outcomes that seem relevant. Let us consider what can be said on each of the three research questions (section 1.4).

8.1 What is distributional learning?

The first question was: 'what *is* distributional learning?' Studies with human participants (Maye et al, 2008; Maye et al. 2002; Maye and Gerken, 2001; Gulian, Escudero and Boersma, 2007; see chapter 1) showed that exposing infants or adults to bimodal distributional inputs of speech sounds can trigger better discrimination performance of two tokens that each belong to one of the distributions. Similarly, presenting the participants with a single distribution can weaken their discrimination performance for two tokens in that distribution that they could discriminate before exposure. Although it is still uncertain if this discrimination performance is based on warping (section 1.2) of the perceptual space, there are a couple of observations that support this possibility.

First, studies with infant animals (section 4.2) demonstrated a direct impact of auditory input on the receptive fields (RFs) of neurons in A1 and concomitantly on the formation of RF parameter maps in A1. Even though the authors of these studies did not test perception directly (since they looked at the behavior of cells rather than at the perception behavior of animals), obviously the RFs are related to perception. Furthermore, this relation emerged more clearly in the study by Recanzone, Schreiner and Merzenich (1993; section 4.3), who measured RFs in owl monkeys that received a discrimination training.

Second, *non*-distributional input was shown to distort or arrest map formation in infants' A1 and this distortion was attributed to unbalanced interactions between neurons (section 4.2). Consequently, distributional input seems to have a special function in eliciting *balanced* interactions that lead to stable parameter maps. In this respect a topic for future research is to examine the role of Infant-Directed Speech (IDS) in the formation of stable maps, since IDS is supposedly 'more distributed' (i.e., more variable) than ADS (section 1.2).

Thus, RFs (which are supposedly representative of acoustic perception) are shaped by input and interactions between neurons. Interestingly, the replication of Guenther and Gjaja's

(1996) computer model (script Boersma, 2009) revealed a very similar process and it showed that this process can be self-organizing, i.e., it can evolve in the absence of meaning or any other top-down influence. In addition, the model confirmed Kohonen's observation (1982) that this self-organized warping relies on two 'necessary and sufficient' factors: (1) a competitive process that yields a (group of) winning neurons in the map and (2) an adaptive process that allows these winners to become more responsive to the input. In neurobiological terms, these processes correspond to precisely the interactions between neurons and the adaptation of RFs in response to these interactions, which were both reported in the animal studies (see above).

The cortex is most probably the prime site of plasticity that is triggered by input. (Buonomano and Merzenich, 1998; section 4.4.3). Consequently, the prompt changes in discrimination behavior, which occur after exposure to distributional input, seem to arise here. In view of the development of the auditory pathways during gestation, which starts at the cochlea (Moore and Guan, 2001; section 4.5), it is likely that plastic changes proceed from the cortex in a retrograde fashion (Buonomano and Merzenich, 1998; section 4.4.3). Thus, the lower the level, the more it relies on genetic programming and the less it is susceptible to change on the basis of the input. At this point it is not clear what precisely the effect is of lower-level changes on clustering (sections 1.2 and 5.1) in the cortex and how much time is needed for these adaptations to occur.

In addition, input does not only directly impinge on A1, but also on higher-level auditory areas due to parallel direct input from the thalamus into both primary and secondary cortical regions. While it is not clear if this direct input also leads to warping at higher levels, it may account for part of the findings by Saffran and colleagues on sensitivity to statistical information in the order of syllables, which are probably processed in higher-level auditory regions (Dehaene-Lambertz et al., 2005).

All in all, the findings indicate that (1) distributional learning may indeed be characterized as a warping of the perceptual space, in particular a *balanced* warping as opposed to an unbalanced warping that results from non-distributional input or deprivation of input; that (2) this warping is based on changes in RFs, specifically cortical RFs and that (3) these changes arise from interactions between neurons, which are elicited by the input.

At this point it should be noted, however, that other interpretations of distributional learning cannot be excluded. This holds, for example, for exemplar approaches. In these approaches every input unit is said to be stored as an exemplar preserving phonetic detail. Proponents of this approach (e.g., Wedel, 2004) could claim that RFs can store many different 'exemplars', since we saw that a neuron can acquire more than one RF and that it can switch between RFs depending on the task. While it is difficult to see how a neuron would accumulate all the different RFs, label them and calculate the differences between them (as is required in exemplar approaches), the simplified warping process that was reflected in Guenther and Gjaja's model, cannot readily account for this behavior (i.e., the switch in tuning properties depending on the task) either.

8.2 Is distributional learning in adults and infants an identical process?

As to the second question ('Is distributional learning in adults and infants an identical process?'), the following can be said. The process is similar in that both in infants and in adults input affects RFs of cortical neurons (sections 4.2 and 4.3; see also section 8.1). In a further comparison there are mainly differences, which highlight that the roles of nature and nurture are intertwined. A notable disparity is that for infants input seems to influence mostly RFs in other layers of the cortex than for adults (section 4.5). Histological studies by Moore and colleagues (2001, 2002, 2007) suggest that input into infants' A1 must enter via layer I and simultaneously via the thalamic input layer IV. The input via layer I occurs predominantly in the first months of life, precisely when the parameter maps are formed. Interestingly, this period coincides with the period in which vowels are acquired. It thus looks as if vowel input plays a crucial role in map formation. Myelinated axons in layer I start disappearing after 4 to 5 months of age and have substantially reduced in number around the first birthday, precisely when the acquisition of basic phonemes is completed.

Inputs into the adult A1 do not seem to enter via layer I (or maybe in a restricted manner as a very small number of myelinated axons remain into adulthood). The input enters via the main thalamic input layer IV. However, this input seems quickly affected by top-down activity in the superficial layers (i.e., layers II and III), which only ripen in children after the age of five: there are several indications that the adult brain is organized in such a way that expectations on the input patterns are processed faster than the detailed bottom-up analysis (e.g., Dehaene-Lambertz, Pallier, Serniclaes, Sprenger-Charolles, Jobert and Dehaene, 2005; Dehaene-Lambertz and Gliga, 2004; Mäkelä, Alku, Makinen, Valtonen, May, Tiitinen, 2002; section 3.5.2) Although the precise workings of the vertical ('column') interactions between layers remain a puzzle, it appears that input into the adult A1 produces RF changes in the superficial layers rather than in layer IV, which is the prime site of RF changes in infants and small children (Keuroghlian and Knudsen, 2007; section 4.5.7). This difference may account for the dominant opinion that adults cannot learn without attention (section 4.3.2): if learning occurs in superficial layers, where top-down information from other cortical regions arrives, then focusing the attention, which is a top-down process, may help the learning process. A further difference between adult and infant learning is that the effect of input on RF changes (and concomitant changes in map layout) is tiny in adults as compared to infants due to several genetically programmed factors that boost synaptic plasticity (section 4.4.2). It is logical that in adults these mechanisms are not at work, since they would prevent stabilization, essential for recognizing previously learned patterns.

In view of these considerations it is not transparent what made fast distributional learning (within minutes) successful for the adult participants in the experiments by Gulian, Escudero and Boersma (2007). Future experiments should examine this issue further by, for example, looking at the effect of engaging participants in an unrelated task (e.g., drawing a picture), while exposing them to the sound distributions. Also, the role of the lingering axons in layer I deserves attention.

8.3 In what way(s) are (enhanced) distributions advantageous for categorization?

The third question was: in what way(s) are (enhanced) distributions advantageous for categorization? (section 1.4) Unfortunately, the simulations, which were intended as a means to explore this question, could not be used for this purpose. The parameter settings provided by Guenther and Gjaja (1996) only led to proper warping when the model was exposed to distributions with unrealistically narrow standard deviations. (Guenther and Gjaja use these narrow standard deviations in all of their simulations). Although it is possible to make the model warp with larger, more reasonable standard deviations by raising *L* (i.e., the number of cells that survive the competitive process and which are allowed to learn; section 5.6), it is not clear what a reasonable number of learning cells is and whether, why and how this number changes over time. Because changes in the parameters defining *L* (viz., L_{START} and L_{END}) yield very different outcomes, the model could not be used to predict the effects of enhanced distributions on category formation in the auditory cortex.

At the same time the model was useful in that it revealed this sensitivity to *L*, or more generally to the competitive process (sections 5.6 and 8.1). This is analogous to Kohonen's (1982) observations for topologically ordered maps, which appeared to be sensitive to parameter settings defining the "lateral interaction" (p.68) of neighboring cells. Understanding neuronal interactions (and concomitantly competitive processes and the implications for *L* in the model) is an important topic for future research.

A related point that remains unclear is the form of variability in IDS distributions and more generally the limits of variability in effective distributions. On the one hand distributed input yields balanced maps due to balanced interactions (see 8.1). At the same time, too widely peaked distributions are likely to lead to non-solidified clusters. For understanding the precise impact of IDS on categorization, it is essential that we determine the shape of IDS distributions (Benders, in preparation).

Although it was impossible to determine the impact of enhanced distributions on the categorization process, an animal study (Fritz et al, 2005; section 4.3.1) presented an interesting possible side-effect of IDS. It showed that cells in adult animals can acquire new RFs and that subsequently the animal can switch between old and new RFs depending on the situation. Thus, if a child tunes his or her RFs first to IDS and subsequently to ADS, then later in life (s)he should be able to switch between IDS and ADS perception. As patterns that are acquired early in life tend to persist strongly, this mechanism could underlie the 'hyperspace phenomenon', i.e., that listeners tend to choose enhanced vowels rather than normally occurring vowels as representative vowel tokens (Johnson, Flemming and Wright, 1993). In other words, it is conceivable that listeners switch to first acquired clusters when performing this task. Although difficult to test, this observation predicts a correlation between hearing a lot of IDS early in life and prototypicality judgments that favor enhanced tokens later in life. Future research is needed to compare this option with existing explanations for the hyperspace effect, such as the proposals by Johnson, Flemming and Wright (1993) and Boersma (2006).

8.4 Does distributional input improve the categorization of speech sounds?

At this point, let us reconsider the main question of this thesis: does distributional input improve the categorization of speech sounds? As the discussion of the first sub-question indicated, distributional input leads to balanced interactions which seem to facilitate balanced clustering. However, it is still difficult to pinpoint what level of representation these clusters represent in Boersma's model. What seems relevant is the observation that the locations of speech sound representations in the auditory pathways (i.e., the places along the pathways that could be related to the creation of Auditory Forms and Surface Forms) may differ for different phonemes. Vowels and possibly other phonemes that rely on longer-duration formant cues (such as approximants), are probably resolved subcortically (Kraus, McGee and Koch, 1998; section 2.7) and turn up in A1 as characteristic F1-F2 combinations (Obleser et al., 2003; Shestakova et al., 2004; Mäkelä et al. 2003; section 3.5.2). In contrast, stop consonants are only resolved in the cortex, although it is unclear if this happens in A1 (Kraus, McGee and

Koch, 1998; section 2.7). Although obviously more research is needed to determine what representations are affected, these findings hint at the possibility that in A1 distributional input molds Surface Form-like representations of vowels and Auditory-Form-like representations of stop consonants.

9. References

Abbott, L.F. & Nelson, Sacha B. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience, 3*, 1178-1183.

Abrams, Daniel & Kraus, Nina. (2009). Auditory pathway representations of speech sounds in humans. In J. Katz, L. Hood, R. F. Burkard & L. Medwetsky (Eds.), *Handbook of Clinical Audiology* (pp. 611-626). Philadelphia: Lippincott Williams & Wilkins.

Adank, Patti; van Hout, Roeland & Smits, Roel. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America, 116*(3), 1729-1738.

Aiken, Steven J. & Picton, Terence W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*(1/2), 35-47.

Aslin, Richard N.; Pisoni, David B.; Hennessy, Beth L. & Perey, Alan J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development, 52*, 1135-1145.

Belin, Pascal & Zatorre, Robert J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience, 3*(10), 965-966.

Benders, Titia (in preparation). Unsupervised learning of cue-weighting in phoneme perception: human and computer learners. PhD project (2008-2012).

Boersma, Paul. (2007). Cue constraints and their interactions in phonological perception and production. *Rutgers Optimality Archive,* pp. 1-40.

Boersma, Paul (2006). Prototypicality judgments as inverted perception. In Fanselow, Gisbert; Féry, Caroline; Schlesewsky, Matthias & Vogel, Ralf (eds.): *Gradience in Grammar*, (pp. 167-184). Oxford: Oxford University Press.

Brosch, Michael & Scheich, Henning. (2003). Neural representation of sound patterns in the auditory cortex of monkeys. In A. A. Ghazanfar (Ed.), *Primate Audition. Ethology and Neurobiology.* (pp. 151-175). Boca Raton/London/New York/Washington D.C.: CRC Press.

Buonomano, Dean V. & Merzenich, Michael M. (1998). Cortical plasticity: from synapses to maps. *Annual Review of Neuroscience, 21*(1), 149-186.

Burnham, Denis; Kitamura, Christine & Vollmer-Conna, Uté. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435-1435.

Chang, Edward F. & Merzenich, Michael M. (2003). Environmental noise retards auditory cortical development. *Science, 300*(5618), 498-502.

Cheour, M.; Alho, K.; Čeponiené, R.; Reinikainen, K.; Sainio, K.; Pohjavuori, M.; Aaltonen, O. & Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology, 29*(2), 217-226.

Dan, Yang & Poo, Mu-ming. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron, 44*, 23-30.

De Villers-Sidani, Etienne; Chang, Edward F.; Shaowen, Bao & Merzenich, Michael M. (2007). Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *Journal of Neuroscience, 27*(1), 180-189.

Dehaene-Lambertz, G. & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience, 16*(8), 1375-1387.

Dehaene-Lambertz, G.; Pallier, C.; Serniclaes, W.; Sprenger-Charolles, L.; Jobert, A. & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24*(1), 21-33.

Dehaene-Lambertz, G. & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport : an international journal for the rapid communication of research in neuroscience, 12*(14), 3155-3158.

Diesch, Eugen & Luce, Thomas. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology, 34*(5), 501-510.

Eggermont, Jos J. (1991). Maturational aspects of periodicity coding in cat primary auditory cortex. *Hearing Research, 57*, 45-56.

Eggermont, Jos J. (2001). Between sound and perception: reviewing the search for a neural code. *Hearing Research, 157*, 1-42.

Eggermont, Jos J. (2008). The role of sound in adult and developmental auditory cortical plasticity. *Ear and Hearing, 29*(6), 819-829.

Formisano, Elia; Kim, Dae-Shik; Di Salle, Francesco; van de Moortele, Pierre-Francois; Ugurbil, Kamil & Goebel, Rainer. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron, 40*(4), 859-869.

Fritz, Jonathan B.; Elhilali, Mounya & Shamma, Shihab A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *Journal of Neuroscience, 25*(33), 7623-7635.

Gazzaniga, Michael S.; Ivry, Richard B. & Mangun, George R. (2002). *Cognitive Neuroscience. The Biology of the Mind. Second Edition.* New York/London: W.W. Norton and Company.

Georgopoulos, Apostolos P.; Schwartz, Andrew B. & Kettner, Ronald E. (1986). Neuronal population coding of movement direction. *Science, 233*(4771), 1416-1419.

Gilbert, Charles D. (1994). Neuronal dynamics and perceptual learning. *Current Biology, 4*(7), 627-629.

Guenther, Frank & Gjaja, Marin. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America, 100*(2), 1111-1121.

Guenther, Frank H. & Bohland, Jason W. (2002). Learning sound categories: a neural model and supporting experiments. *Acoustical Science and Technology, 23*(4), 213-221.

Guenther, Frank H.; Nieto-Castanon, Alfonso; Ghosh, Satrajit S. & Tourville, Jason A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech Language and Hearing Research, 47*(1), 46-57.

Gulian, Margarita; Escudero, Paola & Boersma, Paul. (2007). Supervision hampers distributional learning of vowel contrasts. *Proceedings of the International Congress of Phonetic Sciences, Saarbrücken, Germany, August 6-10*, 1893-1896.

Hackett, Troy A. (2003). The comparative anatomy of the primate auditory cortex. In A. A. Ghazanfar (Ed.), *Primate Audition. Ethology and Neurobiology.* (pp. 199-225). Boca Raton/London/New York/Washington D.C.: CRC Press.

Hayward, Katrina. (2000). *Experimental Phonetics*. Harlow, England: Pearson Education.

Hazan, Valerie & Simpson, Andrew. (2000). The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Language and Speech, 43*(3), 273.

Hebb, Donald O. (1949). *The Organization of Behavior*. New York: Wiley.

Hirsch, Ira J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America,31*(6), 759-767.

Iverson, Paul & Kuhl, Patricia K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99*(2), 1130-1140.

Johnson, Keith; Flemming, Edward & Wright, Richard. (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language, 69*(3), 505-528.

Kaas, Jon H.; Hackett, Troy A. & Tramo, Mark Jude. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology, 9*(2), 164-170.

Kandel, Eric R.; Schwartz, James H. & Jessell, Thomas M. (2000). *Principles of neural science*. New York, NY [etc.]: McGraw-Hill.

Karmiloff-Smith, Annette. (2006). The tortuous route from genes to behavior: a neuroconstructivist approach. *Cognitive, Affective and Behavioral Neuroscience, 6*(1), 9-17.

Keuroghlian, Alex S. & Knudsen, Eric I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology, 82*(3), 109-121.

Kinnaird, Susan Kuzniak & Zapf, Jennifer. (2004). An acoustical analysis of a Japanese speaker's production of English /r/ and /l/ [Electronic Version]. Retrieved June 28, 2009 from https://www.indiana.edu/~iulcwp/pdfs/04-kinnaird.pdf.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 59-69.

Kral, Andrej. (2009). Frühe Hörerfahrung und sensible Entwicklungsphasen. *HNO, 57*(1), 9-16.

Kral, Andrej; Hartmann, Rainer; Tillein, Jochen; Heid, Silvia & Klinke, Rainer. (2001). Delayed maturation and sensitive periods in the auditory cortex. *Audiology and Neuro-Otology, 6*(6), 346-362.

Kraus, Nina; McGee, Therese J. & Koch, Dawn Burton. (1998). Speech sound representation, perception, and plasticity: a neurophysiologic perspective. *Audiology & Neuro-Otology, 3*(3), 168-182.

Kraus, Nina & Nicol, Trent. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences, 28*(4), 176-181.

Krishnan, Ananthanarayan. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing Research, 166*(1/2), 192-201.

Kuhl, Patricia K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93-107.

Kuhl, Patricia K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences Colloquium "Auditory Neuroscience: Development, Transduction, and Integration", 97*, 11850-11857.

Kuhl, Patricia K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-843.

Kuhl, Patricia K.; Andruski, Jean E.; Chistovich, Inna A.; Chistovich, Ludmilla A.; Kozhevnikova, Elena V.; Ryskina, Viktoria L.; Stolyarova, Elvira L.; Sundberg, Ulla & Lacerda, Francisco. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 227*(5326), 684-686.

Kuhl, Patricia K.; Williams, Karen A.; Lacerda, Francisco; Stevens, Kenneth N. & Lindblom, Björn. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science. New Series., 255*(5044), 606-608.

Kurt, Simone; Deutscher, Anke; Crook, John M.; Ohl, Frank W.; Budinger, Eike; Moeller, Christoph K.; Scheich, Henning & Schulze, Holger. (2008). Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions. *PLoS ONE, 3*(3), e1735.

Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research, 60*(2), 115-142.

Langner, G.; Sams, M.; Heil, P. & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *Journal of Comparative Physiology, 181*(6), 665-676.

Levy, W.B. & Steward, D. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience, 8*, 791-797.

Liu, Huei-Mei; Kuhl, Patricia K. & Tsao, Feng-Ming. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), 1-10.

Lively, Scott E.; Logan, John S. & Pisoni, David B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*(3), 1242-1255.

Luck, Steven J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA [etc.]: The MIT Press.

Mäkelä, Anna Mari; Alku, Paavo & Tiitinen, Hannu. (2003). The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neuroscience Letters, 353*(2), 111-114.

Mäkelä, Anna Mari; Alku, Paavo; Mäkinen, Ville; Valtonen, Jussi; May, Patrick & Tiitinen, Hannu. (2002). Human cortical dynamics determined by speech fundamental frequency. *NeuroImage, 17*, 1300-1305.

Maye, Jessica & Gerken, LouAnn. (2000). Learning phonemes without minimal pairs. In C. Howell (Ed.), *BUCLD 24 Proceedings* (pp. 522-533). Somerville, MA: Cascadilla Press.

Maye, Jessica & Gerken, LouAnn. (2001). Learning phonemes: how far can the input take us? In A. H.-J. Do (Ed.), *BUCLD 25 Proceedings* (pp. 480-490). Somerville, MA: Cascadilla Press.

Maye, Jessica; Weiss, Daniel & Aslin, Richard. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science, 11*(1), 122-134.

Maye, Jessica; Werker, Janet F. & Gerken, LouAnn. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111.

McClaskey, Cynthia L.; Pisoni, David B. & Carrell, Thomas D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics, 34*(4), 323-330.

Mendelson, J.R.; Schreiner, C. E. & Sutter, M. L. (1997). Functional topography of cat primary auditory cortex: response latencies. *Journal of Comparative Physiology A, 181*, 615-633.

Merzenich, Michael, M. & Brugge, John F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Research, 50*(2), 275-296.

Mishkin, Mortimer; Ungerleider, Leslie, G. & Macko, Kathleen, A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences, 6*, 414-417.

Moore, Jean K. (2002). Maturation of human auditory cortex: implications for speech perception. *Annals of Otology, Rhinology and Laryngology, 111*, 7-10.

Moore, Jean K. & Guan, Yue-Ling. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *JARO, 2*(4), 297-311.

Moore, Jean K. & Linthicum, Fred H. (2007). The human auditory system: a timeline of development. *International Journal of Audiology, 46*(9), 460-478.

Näätänen, Risto & Picton, Terence. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Nakahara, Haruka; Zhang, Li I. & Merzenich, Michael M. (2004). Specialization of primary auditory cortex processing by sound exposure in the "critical period". *Proceedings of the National Academy of Sciences of the United States of America, 101*(18), 7170-7174.

Nicholas, Johanna Grant & Geers, Ann E. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *Journal of Speech, Language & Hearing Research, 50*(4), 1048-1062.

Obleser, Jonas; Elbert, Thomas; Lahiri, Aditi & Eulitz, Carsten. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research, 15*(3), 207-213.

Ohgushi, Kengo. (1983). The origin of tonality and a possible explanation of the octave enlargement phenomenon. *Journal of the Acoustical Society of America, 73*(5), 1694-1700.

Ohl, Frank W. & Scheich, Henning. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences of the United States of America, 94*(17), 9440-9444.

Pantev, C.; Bertrand, O.; Eulitz, C.; Verkindt, C.; Hampson, S.; Schuierer, G. & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology, 94*(1), 26-40.

Pantev, C.; Hoke, M.; Lehnertz, K. & Lütkenhöner, B. (1989). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology, 72*, 225-231.

Pegg, J.E. & Werker, Janet F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*, 3742-3753.

Phillips, Dennis P. (1993). Representation of acoustic events in the primary auditory cortex. *Journal of Experimental Psychology: Human Perception and Performance, 19(1)*, 203-216.

Polley, Daniel B.; Steinberg, Elizabeth E. & Merzenich, Michael M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of Neuroscience, 26*(18), 4970-4982.

Pouget, Alexandre; Dayan, Peter & Zemel, Richard. (2000). Information processing with population codes. *Nature Reviews Neuroscience, 1*, 125-132.

Pols, Louis C.W.; Tromp, H.R.C. & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America, 53*(4), 1093-1101.

Purves, Dale; Augustine, George J.; Fitzpatrick, David; Hall, William C.; LaMantia, Anthony-Samuel; McNamara, James O. & White, Leonard E. (Eds.). (2008). *Neuroscience*. Sunderland, Mass.: Sinauer Associates.

Rauschecker, Josef P.; Tian, Biao; Pons, Timothy & Mishkin, Mortimer. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *The Journal of Comparative Neurology, 382*(1), 89-103.

Recanzone, G. H.; Schreiner, C. E. & Merzenich, Michael M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience, 13*(1), 87-103.

Recanzone, Gregg H.; Schreiner, Christoph E.; Sutter, Mitchell L.; Beitel, Ralph E. & Merzenich, Michael M. (1999). Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *Journal of Comparative Neurology, 415*(4), 460-481.

Romani, Gian Luca; Williamson, Samuel J. & Kaufman, Lloyd. (1982). Tonotopic organization of the human auditory cortex. *Science, 216*(4552), 1339-1340.

Sachs, Murray B. (1984). Neural coding of complex sounds: speech. *Annual Review of Physiology, 46*, 261-273.

Sachs, Murray B. & Young, Eric D. (1979). Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *Journal of the Acoustical Society of America, 66*(2), 470-479.

Saffran, Jenny R.; Aslin, Richard N. & Newport, Elissa L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, Jenny R. & Wilson, Diana P. (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273.

Salzman, C. Daniel & Newsome, William T. (1994). Neural mechanisms for forming a perceptual decision. *Science, 264*(5156), 231-237.

Schreiner, Christoph E. (1998). Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiology and Neuro-Otology, 3*(2-3), 104-122.

Shestakova, Anna; Brattico, Elvira; Soloviev, Alexei; Klucharev, Vasily & Huotilainen, Minna. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Cognitive Brain Research, 21*(3), 342-350.

Song, Judy H.; Skoe, Erika; Wong, Patrick C.M. & Kraus, Nina. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience, 20*(10), 1892-1902.

Steinschneider, Mitchell; Arezzo, Joseph, C. & Vaughan Jr., Herbert, G. (1990). Tonotopic features of speech-evoked activity in primate auditory cortex. *Brain Research, 519*(1-2), 158-168.

Sundberg, Ulla. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one-year-olds. In *Working Papers 49.* (pp. 148-151): Dept. of Linguistics. Stockholm University.

Sundberg, Ulla & Lacerda, Francisco. (1999). Voice onset time in speech to infants and adults. *Phonetica, 56*(3-4), 186-199.

Tremblay, Kelly; Kraus, Nina; Carrell, Thomas D. & McGee, Therese. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America, 102*(6), 3762-3773.

Verkindt, Chantal; Bertrand, Olivier; Perrin, François; Echallier, Jean-François & Pernier, Jacques. (1995). Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and Clinical Neurophysiology, 96*(2), 143-156.

Wang, Xiaoqin; Kadia, Siddhartha C.; Lu, Thomas; Liang, Li & Agamaite, James A. (2003). Cortical processing of complex sounds and species-specific vocalizations in the marmoset monkey (Callithrix jacchus). In A. A. Ghazanfar (Ed.), *Primate Audition. Ethology and Neurobiology.* (pp. 279-299). Boca Raton/London/New York/Washington D.C.: CRC Press.

Warrier, Catherine; Wong, Patrick; Penhune, Virginia; Zatorre, Robert J.; Parrish, Todd; Abrams, Daniel A. & Kraus, Nina. (2009). Relating structure to function: Heschl's Gyrus and acoustic processing. *Journal of Neuroscience, 29*(1), 61-69.

Wedel, A. B. (2004). Self-Organization and Categorical Behavior in Phonology. *Dissertation Abstracts International, 65*(3), 914.

Werker, Janet F. & Tees, Richard C. (2002). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 25*(1), 121-133.

Wessinger, C. Mark; Buonocore, Michael H.; Kussmaul, Clif L. & Mangun, George R. (1997). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping, 5*(1), 18-25.

Young, Eric D. & Sachs, Murray B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *Journal of the Acoustical Society of America, 66*(5), 1381-1403.

Zhang, Li I.; Bao, Shaowen & Merzenich, Michael M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience, 4*(11), 1123-1130.

Zhang, Li I.; Bao, Shaowen & Merzenich, Michael M. (2002). Disruption of primary auditory cortex by synchronous auditory inputs during a critical period. *Proceedings of the National Academy of Sciences of the United States of America, 99*(4), 2309-2314.

Zhou, X.; Nagarajan, N.; Mossop, B. J. & Merzenich, Michael M. (2008). Influences of un-modulated acoustic inputs on functional maturation and critical-period plasticity of the primary auditory cortex. *Neuroscience, 154*(1), 390-396.

10. List of abbreviations

Abbreviation	Section that discusses the term
A1	3.2
AP	2.4
BF	2.5
BMF	3.4
CF	2.5
EEG	2.4
ERMF	3.5.2
ERP	3.5.2
FFR	2.4
IC	2.2
MEG	3.5.2
MGC	2.2
MMR	2.7
MUA	3.5.1
PSP	2.4
RF	2.5.1