Auditory processing and audiovisual integration revealed by combining psychophysical and fMRI experiments
Tomaskovic, Sonja

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2006

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 12-01-2020
Chapter 1

Introduction
Chapter 1

Functional magnetic resonance imaging was used as a tool to get the insight into brains’ function, and to come closer to the answer to the question: how does the brain work?

There are other modalities used to observe brains in action: single photon emission computed tomography (SPECT), positron emission tomography (PET), electroencephalography (EEG), or magnetoencephalography (MEG). The first two methods measure electric and magnetic fields caused by firing neurons, and the last two are based on detection of radio-active labeled molecules like water. They also differ in their spatial and temporal resolution. fMRI studies are capable of producing spatial resolutions as high as 1-3 mm (Baillet et al., 2001); however, temporal resolution is limited by the relatively slow hemodynamic response. MEG and EEG are measuring directly electrical brain activity and offer superior temporal resolution when compared to fMRI or PET. However, the spatial resolution of MEG or EEG mostly does not match that of fMRI or PET.

This thesis describes experiments conducted in order to investigate human perception and processing of the sound in the human brain. There are several stages in the sound processing. Firstly, in the ear, sound is recorded and transformed into the electrical signal, then information is transported to the brain stem and further to different parts of the brain. The main goal of this research is better insight and understanding of these processes in the brain. Combining psychophysical and fMRI experiments gives a better control of what subjects perceived and what was measured, since perception vary among subjects. Psychophysical experiments untangle subjects’ perception and neuroimaging, fMRI experiments describe the function of the brain.

The thesis is divided into the two major parts each describing perceptual phenomena investigated by combining psychophysical and fMRI experiments. Chapter 2 describes psychophysical experiment, that was carried out in order to examine effects of harmonic number of a band limited complex sound on perception. After detailed data analysis and observation that for different set
of auditory stimuli subjects’ responses vary, we tried to go step further and to observe brain at work. Using the fMRI we expected to detect differences on the brain level for different set of stimuli (chapter 3). The second part of the thesis, chapter 4, and chapter 5, describe audiovisual integration in speech perception. For certain combination of audiovisual speech sounds perception of subjects changes. These psychophysical results triggered performing the fMRI study and investigating if these perceptual effects can be also detected in the brain.

The introductory chapter briefly introduces concepts that are used in this thesis: physical principles of MRI, fMRI and Blood Oxygen Level Dependant (BOLD) signal, psychophysics, auditory and visual processing, and a brief description of brain anatomy.

1.1 Physical principles of MRI

The signal that is used in all magnetic resonance applications originates from the magnetic moment that is associated with the spin of an atomic nucleus. Every nucleus has a intrinsic angular momentum called nuclear spin. For fMRI the commonly used nucleus is the proton and in the absence of the external magnetic field these are randomly oriented (figure 1.1A). When they are placed in a magnetic field \( B_0 \) they will align either parallel or anti-parallel to the direction of the applied field \( B_0 \) (figure 1.1B). These orientations correspond to quantum mechanical energy states and parallel alignment is the energetically preferable state, therefore slightly more spins will align parallel to the magnetic field. That causes a small magnetic field of its own \( \vec{M} \) that is also aligned with \( B_0 \) and is called the longitudinal magnetization. Magnetization parallel to \( B_0 \) is usually represented as \( \vec{M}_z \) and if \( \vec{M}_z \) is at equilibrium this situation is indicated with \( \vec{M}_0 \).

Alignment of the magnetic moment of a single atom with the applied field \( (B_0) \) is not perfect. There are small angles between \( B_0 \) and the magnetic
Figure 1.1: (A) Hydrogen nuclei randomly oriented. (B) When external magnetic field $B_0$ is applied nuclei tend to align parallel to it.

Figure 1.2: (A) A nucleus has a property called spin. The spinning positive charge causes a magnetic moment $\mu$. (B) Inside the magnetic field, $B_0$, the magnetic moment aligns with $B_0$ and precess around it with Larmor frequency.

moment of the individual spins (figure 1.2A). The spins of the nuclei precess around the main magnetic field (figure 1.2B). This is called Larmor precession, and it is described by:

$$\nu = \gamma_H B$$  \hspace{1cm} (1.1)

where $\gamma_H$ is gyromagnetic ratio for protons ($=42.58\text{MHz/T}$) and $B$ is the local magnetic field. In order to obtain information from the spins, use of radio frequency (RF) pulse (on the Larmor frequency) creates deflection of the net magnetization from its equilibrium. Once $\vec{M}$ has been deflected and RF field has been turned off, the individual magnetic moments of the protons will still precess around the $B_0$ (i.e., the z-axis) with the Larmor frequency. Therefore, their net result $\vec{M}$ will also show this rotational behavior. This rotating magnetic field is the basis of magnetic resonance imaging. The signal
produced because of rotation:

\[ s(t) = |\vec{M}| \cos(2\pi \nu t) \] (1.2)

After excitation, the nuclei return to equilibrium and this process is called relaxation. During the relaxation processes both longitudinal and transverse components of the net magnetization return to their equilibrium values. The longitudinal component decays by a \( T_1 \) (the spin-lattice) relaxation time. Nuclei that were flipped to high energy states by an excitation pulse, transfer energy to the lattice in the process of returning to the lower state. At the same time transverse magnetization decays, which occurs faster than \( T_1 \) relaxation only. This is due to spin-spin dephasing and is described by a \( T_2 \) relaxation time. For brain tissue \( T_1 \) is in the order of 1 s, and \( T_2 \) is in the order of 0.1 s (Haacke et al., 1999). Furthermore, spin coherence is affected by inhomogeneities in the applied magnetic field. The exponential decay in the signal resulting from the combination of \( T_2 \) relaxation and field inhomogeneities is referred to as \( T_2^* \) - the effective transverse relaxation time.

Bloch equations\(^1\) describes the behavior of the longitudinal and transverse magnetization vector:

\[ \frac{dM_z}{dt} = -\frac{M_z - M_{z0}}{T_1} \] (1.3)

\[ \frac{dM_{xy}}{dt} = -\frac{M_{xy}}{T_2^*} \] (1.4)

which results in:

\[ M_{xy}(t) = M_{xy}|_{t=0} e^{-t/T_2^*} \] (1.5)

\[ M_z(t) = M_0(1 - e^{-t/T_1}) + M_{z|t=0} e^{-t/T_1} \] (1.6)

\( T_1 \) is a property of a tissue, whereas \( T_2^* \) is also determined by local field inhomogeneities. \( M_{z|t=0} \) is the residual longitudinal magnetization immediately after the RF pulse. Including the transverse decay (formula 1.5) into

\(^1\)The rotational motion of the spins is ignored for now
signal $s$ (formula 1.2) gives:

$$s(t) = M_{xy|t=0}e^{-\frac{t}{T_2}}\cos(2\pi\omega t)$$ \hspace{1cm} (1.7)

### 1.1.1 Image formation

In order to reconstruct an image, it is necessary to encode the emitted signal and relate it to a spatial position of the nuclei contributing to the signal. By using the Larmor frequency we can use spin properties to create an image. The Larmor frequency depends on the local magnetic field strength. This field can be changed by applying a magnetic field gradient:

$$\omega = \gamma(B_0 + G_x \times x + G_y \times y + G_z \times z)$$ \hspace{1cm} (1.8)

$$\omega = \gamma(B_0 + G^T \mathbf{p})$$ \hspace{1cm} (1.9)

where $G_x$, $G_y$, and $G_z$ are the magnetic field gradients (in T/m) along the x-, y-, and z- axis, respectively.

The protons within the specific slice can selectively be excited along the longitudinal direction by applying the selective gradient. This gradient spatially varies the Larmor frequency and produces only limited band of frequencies. An RF pulse with selective bandwidth will only excite spins having a frequency within this bandwidth. This gradient is applied in the z-axes.

Two other dimensions of the slice are also encoded by use of the gradients. After the slice selective gradient $G_z$ and the RF pulse are turned off, the Larmor frequency can spatially be varied again by applying the appropriate gradients and starting data acquisition. While the phase-encoding gradient, $G_y$, is on, nuclei at different y- positions precess at different Larmor frequencies; spins in the higher field will precess faster and get ahead of the other spins which result in a phase difference. When the gradient pulse is turned off, each reverts to precessing at its original frequency (the same Larmor frequency), but each position within a given y- direction has a slightly different phase angle. While the readout gradient, $G_x$, is turned on, frequencies of
the protons also change according to their spatial location. This gradient is ‘on’ during the data collection, so the protons do not have a chance to return to the same Larmor frequency as they do when the gradient is turned off. Instead a distribution of Larmor frequencies is created along the x-axis. The measured signal from position \( \vec{p} \) now becomes:

\[
s(\vec{p}, t) = A(\vec{p}) \cos (\gamma G_x(p_x)t) \tag{1.10}
\]

where \( p_x \) is x-coordinate of \( \vec{p} \), \( t = 0 \) at the start of \( G_x \), and \( A(\vec{p}) \) is the amplitude at location \( p \). The overall measured signal is a summation of the signal over all locations:

\[
S(t) = \sum_{\vec{p}} s(\vec{p}, t) \tag{1.11}
\]

\( S(t) \) is not measured continuously, but sampled in time. Thus, \( t \) is replaced by \( t_n \):

\[
t_n = n \delta t, n \epsilon N \tag{1.12}
\]

and \( \delta t \) is the time between two sampled points. An inverse Fourier transform separates this combined signal into the individual frequencies, that corresponds to the positions, and their relative amplitudes, often defined as grey value in the image. The amplitude depends on the local \( T1 \) and \( T2^* \) (or \( T2 \)) properties and on the number of protons from a region contributing to the signal.

The recorded signal for a specific location depends on the gradient strength \( G \), and the time \( t \) while the gradient was turned on (see equation 1.10). We can write:

\[
k_x = \gamma G_x t_n \tag{1.13}
\]

yielding:

\[
s(\vec{p}, t_n) = A(\vec{p}) \cos (p_x k_x) \tag{1.14}
\]

Correspondingly, if the y-gradient (\( G_y \)) is turned on for a time \( t_y \), \( k_y \) will be:

\[
k_y = \gamma G_y t_y \tag{1.15}
\]
Now, the signal from the position $\vec{p}$, for a certain strength and duration of $G_x$ and $G_y$ is described by:

$$s(\vec{p}, \vec{k}) = I(\vec{p}) \cos (\vec{k}^T \vec{p})$$  \hspace{1cm} (1.16)

where $\vec{k}$ is the column-vector containing the elements $k_x$ and $k_y$. The sum of the contributions for each position (see equation 1.11) will be:

$$S(\vec{k}) = \sum_{\vec{p}} s(\vec{k}, \vec{p})$$  \hspace{1cm} (1.17)

Many combinations of $k_x$ and $k_y$ can be filled into a k-space. Once k-space is adequately sampled, a 2-dimensional discrete Fourier transformation can construct the 2D image. Finally, by gathering individual slices a 3D image is created. Echo planar imaging (EPI), technique introduced by Mansfield (1977), is used to define the path of filling in the k-space. The most condensed approach to MR imaging would encode all spatial information into the MR signal after a single excitation. Although this technique is considered to be fast, however it produces a lot of noise which might mask the stimuli and furthermore it can interfere with brain function.

### 1.2 Functional magnetic resonance imaging

Functional MRI is an indirect method for measuring brain activity, and it relies on metabolic changes that follows the neuronal activity (Buxton, 2002; Jezzard et al., 2001). There is a certain time delay between actual neuronal firing and maximum of the BOLD response. Approximately, 2 to 3 seconds after stimulus onset, an increase in signal intensity can be measured. After the stimulus is turned off, the BOLD response will decay to the baseline. Blood consists of erythrocytes that contain hemoglobin with bound oxygen or deoxyhemoglobin without oxygen. Hemoglobin contains a ferrous core that changes with respect to its magnetic properties when it binds to oxygen. That leads to a change in the magnetic susceptibility of the blood.
tissue. Deoxyhemoglobin and hemoglobin have different magnetic properties; deoxyhemoglobin is paramagnetic and locally introduces inhomogeneity of the main magnetic field, and hemoglobin is weakly diamagnetic and has little effect. A stimulation produces rapid neuronal activation, which increases cerebral blood flow (CBF), cerebral blood volume (CBV), and oxygen delivery. As CBF increases more than CBV, oxygen delivery quickly exceeds in local oxygen needs owing to the activation. As the delivered oxygen exceeds local demand, the capillary and venous beds fill with a larger ratio of oxygenated to deoxygenated hemoglobin (compared to when the cortex was at rest). Deformations due to the paramagnetic properties of deoxyhemoglobin will lead to a quick dephasing of the nuclear magnetic moments, reducing the net transverse magnetization. Thus, regions of the brain that are active will respond with an increase of local blood oxygenation levels (due to increase of CBF and CBV). This will effect the local field inhomogeneities (and rise of $T2^*$), which will increase due to the increase of oxygenation levels. In a $T2^*$ sensitive EPI sequence this will show up as a local increase in image intensity called the hemodynamic response signal. By acquiring the data for both the baseline condition\(^2\) and the stimuli condition, the resulting differences in image intensities can be detected.

### 1.3 Psychophysics

Psychophysical experiments try to define the relation between physical stimuli and their perception. In auditory experiments the aim is to investigate mechanisms involved in sound perception and processing. The most common way of doing it is two (or more) alternative forced-choice experiments, where stimuli are presented repeatedly. Subjects have to compare those stimuli, and answer if the stimuli are the same. A psychometric curve describes the relationship between a parameter of a physical stimulus and the subjects’

\(^2\)Condition with no stimuli.
responses. The psychometric function usually has the shape of a sigmoid function. That can be described by:

\[ p(x) = \frac{1}{1 + e^{-x}} \quad (1.18) \]

Detection probability as a function of strength are represented on the y-axes, and the physical parameters (stimuli) on the x-axes (figure 1.3). The 50% point is the point at which function reaches the middle between chance level responses and 100% correct response.

### 1.4 Auditory processing

The ear (figure 1.4) can be divided in three parts: the external, middle and inner ear.

Sound waves first reach the external ear, and through the external ear canal arrive to the middle ear and cause the tympanic membrane (eardrum) to oscillate. The three bones in the ear (malleus, incus, and stapes) pass these vibrations to the cochlea. The cochlea is a snail-shaped, fluid-filled structure in the temporal bone, consisting of three adjacent compartments (scala
Figure 1.4: A schematic drawing of the human ear. It is divided into a three main parts: outer ear, middle ear, and inner ear.

vestibule, scala media, and scala tympani) separated by two membranes: basilar membrane and Reissners membrane. The basilar membrane plays a crucial role in sound processing. Sound causes oscillations of the basilar membrane, and different frequencies excite different parts of the membrane. High frequencies excite the basal part of the basilar membrane, while low frequencies ‘travel’ to the other end of the cochlea and have their response at the apical part of the basilar membrane. Normally, the human auditory system can detect frequencies from 20 Hz to 20 kHz. The basilar membrane contains hair cells. Two different type of hair cells exist which differ in their organization and function. A single row of inner hair cells along the basilar membrane detects the resonant motion that causes bending of the hairs (stereocilia) of the hair cells. Outcome of this bending is that ion channels in the stereocilia open and close eliciting an action potential in auditory nerve. The frequency decomposition that is present on the basilar membrane is also present in the auditory nerve. Nerve fibers at different locations, carry information about different frequencies. The frequency to which the nerve fiber is the most sensitive is called its characteristic frequency. A nerve fiber will not only carry the information of its characteristic frequency, but also of neighboring frequencies. Two frequencies are resolved if they are more than
Figure 1.5: A schematic drawing of the human eye. Indications on the figure are for different parts of the eye.

one third octave apart. From the hair cells impulses are send through the cochlear nucleus (CN), that transmits signals further via at least two pathways. The binaural pathway contain connections to the bilateral superior olivary nuclei (SO) where binaural comparison of time-delay and sound intensity are made on behalf of spatial localization. Signals are transmitted via the lateral lemniscus (LL) to the inferior colliculi (IC). In parallel, the monaural contralateral pathway directly relays spectral information from the cochlear nucleus via the contralateral LL to the IC. The IC is the center of auditory processing in the brain stem and has afferents that project to the bilateral medial geniculate (MG) nuclei in the thalamus. From here, signals are further transmitted to the auditory cortex in the brain.

1.5 Visual processing

The eye (figure 1.5) is the light-sensitive organ that is the first component of the visual system. All of the individual components through which light travels within the eye before reaching the retina are transparent, thereby minimizing dimming of the light. The cornea and lens help to focus (converge) light rays onto the retina. The visual pathway that mediates perception
and pathway that control eye movements originate in the retina. This light causes chemical changes in the photosensitive cells of the retina, the products of which trigger nerve impulses. Photoreceptor synapse on ganglion cells that are retinal projection neurons. The information leaves the eye by way of the optic nerve, and there is a partial crossing of axons at the optic chiasm, the route to the thalamus and the brain stem. After the chiasm, the axons are called the optic tract. The optic tract wraps around the midbrain to get to the lateral geniculate nucleus (LGN), where all the axons must synapse. This thalamic relay nucleus projects to the primary visual cortex via a pathway called the optic radiations. This projection is important for perception. Certain ganglion cells project to the midbrain directly, principally to two structures: superior colliculus and the pretectal nuclei. In humans the superior colliculus has a minimal role in perception but an important role in the control of rapid eye movements from one object to another (i.e., saccades).

1.6 Brain anatomy

The cerebral cortex consists of four lobes: frontal, parietal, temporal, and occipital lobe (figure 1.6).

The frontal lobe is mostly devoted to the planning and production on body and eye movements, speech, cognition, and emotions. The precentral gyrus contains the primary motor cortex, which participates in controlling the mechanical actions of movement, and adjacent to the primary motor cortex is premotor area which is important in motor decision making and planning of movements. In the left inferior frontal gyri most of the people contain Broca’s area which is essential for the articulation of speech. Much of the frontal lobe is association cortex, and is involved in complex processing of emotions, memories, thoughts, and behavior.

The parietal lobe is separated from the frontal lobe by central sulcus. In the postcentral gyrus, the primary somatic sensory cortex is located and it
Figure 1.6: Human brain. Indications on the figure for the cerebrum: frontal lobe, parietal lobe, occipital lobe, and temporal lobe; cerebellum, brain stem, and spinal cord.

is important for perception of touch, pain, and limb position. The remaining portions of parietal lobe consists of the inferior parietal lobe which is involved in integration of sensory information for speech and perception, and the superior parietal lobe, which mediates behavioral interaction with the world around us.

The occipital lobe is the posterior lobe of the brain and is most unique in its function. It is involved in visual processing, and perception.

The temporal lobe is located below the frontal and parietal lobes. The auditory cortex is located on the superior temporal gyrus and it is involved in perception and recognition of auditory stimuli and memory. The temporal lobe is very important for language. The area for phonological recognition is mostly in the superior temporal gyrus (for example Wernicke’s area), and adjacent portions of the temporal lobe seem to play a large role in managing lexical and semantic information (Martin, 2003).
1.6.1 The auditory cortex

The auditory cortex can be subdivided in numerous sub-areas on the basis of neurophysiological criteria. One of the criteria is assigned by Brodmann (Brodmann, 1909), and the brain was divided into various sub-regions by analyzing each area’s cellular structure. The auditory cortex consists of: Broadmann area (BA) 41 (primary auditory cortex), BA42 (secondary auditory cortex), and BA22 (auditory association cortex). Processing in auditory cortex seems to be performed in a hierarchical fashion. Sharp frequency tuning and short-latency responses are processed in primary regions, while more complex analysis of sounds like recognizing rhythm and melody elicit responses in secondary auditory areas (Wessinger et al., 2001). Furthermore, the hemispheric specialization with regard to the physical characteristic of sound are claimed. It has been suggested that the left hemisphere specializes in rapid temporal processing while the right hemisphere is involved in fine spectral analysis (Zatorre 2001; Sininger et al., 2004).

References

7) Haacke EM, Brown RW, Thompson MR, Venkatesan R. Magnetic reso-