

Replace the contents of this file with official assignment.
Místo tohoto souboru sem patří list se zadáním závěrečné práce.

Bachelor's thesis

NÁZEV PŘÍKLADNÉ ZÁVĚREČNÉ PRÁCE

Nikita Mortuzaiev

Faculty of Information Technology
Department of Applied Mathematics
Supervisor: Ing. Mgr. Ladislava Smítková Janků, Ph.D.
March 22, 2022

Czech Technical University in Prague
Faculty of Information Technology

© 2022 Nikita Mortuzaiev. Citation of this thesis.

This thesis is school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).

Citation of this thesis: Mortuzaiev Nikita. *Název příkladné závěrečné práce.* Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2022.

Contents

Acknowledgments	iv
Declaration	v
Abstrakt	vi
Acronyms	vii
1 Introduction	1
2 Physical Background	3
3 Biological Background	5
3.1 Outer and Middle Ear	5
3.2 Inner Ear	6
3.3 Auditory Scene Analysis	7
4 Mathematical Background	9
5 Computational ASA	11
6 Implementation	13
7 Experiments	15

Chtěl bych poděkovat především sit amet, consectetur adipiscing elit. Curabitur sagittis hendrerit ante. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue.

Declaration

FILL IN ACCORDING TO THE INSTRUCTIONS. VYPLŇTE V SOULADU S POKYNY.
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur sagittis hendrerit ante. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue. Donec ipsum massa, ullamcorper in, auctor et, scelerisque sed, est. In sem justo, commodo ut, suscipit at, pharetra vitae, orci. Pellentesque pretium lectus id turpis.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur sagittis hendrerit ante. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue. Donec ipsum massa, ullamcorper in, auctor et, scelerisque sed, est. In sem justo, commodo ut, suscipit at, pharetra vitae, orci. Pellentesque pretium lectus id turpis.

In Prague on March 22, 2022

.....

Abstrakt

Fill in abstract of this thesis in Czech language. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue.

Klíčová slova enter, comma, separated, list, of, keywords, in, CZECH

Abstract

Fill in abstract of this thesis in English language. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue.

Keywords enter, comma, separated, list, of, keywords, in, ENGLISH

Acronyms

DFA	Deterministic Finite Automaton
FA	Finite Automaton
LPS	Labelled Prüfer Sequence
NFA	Nondeterministic Finite Automaton
NPS	Numbered Prüfer Sequence
XML	Extensible Markup Language
XPath	XML Path Language
XSLT	eXtensible Stylesheet Language Transformations
W3C	World Wide Web Consortium

A horizontal row consisting of 28 small blue squares.

Introduction

Imagine a party. You can hear a variety of sounds: music in the background, conversations between people, noises of somebody coughing, maybe even a dog barking outside. . . . These sounds merge into a single stream that approaches your ears by vibrations in the air and then goes through different physical, biological and psychoacoustical processes to finally come in a form of electrical impulses to the brain. Despite all these sounds from different sources are mixed on the way to your ears, the brain can segregate one (or several) of them. You can focus your hearing on these “target” sounds and separate them from the complex mixture, leaving other sounds in the background. This phenomenon has been described as a “cocktail party effect”, and the process of integrating separate sounds into meaningful streams, or “auditory objects” – auditory scene analysis, or ASA.

In machine perception — specifically in machine hearing — a related concept is referred to as Computational ASA (CASA) and is tightly connected to the fields of sound recognition and digital signal processing. CASA systems indeed aim to separate sounds from mixtures, but they differ from BSS (blind source separation) systems in that they try to do this in a way a human ear does. Being based on and trying to combine works from different fields of science, CASA systems can bring new solutions and insights to the complex problem of signal separation.

The main objective for this thesis is to describe the principles and goals of CASA, existing applications and approaches. Another objective is to practically apply the theoretical knowledge and implement a simple CASA system to separate monophonic music from noise. But before all of this, since this thesis is made for an IT-oriented audience, it is needed to make a brief introduction to the underlying physics and biology.

Thus, the thesis is structured as follows:

Firstly, physical background theory will be provided, including an introduction to what a sound is. Since the implemented system from the practical part aims to segregate music from noise, a special focus in this part will be made on describing harmonic sounds and pitch perception.

Secondly, having in mind that CASA tries to mimic the human auditory system, a brief introduction to the biological structure of the human ear will be made. Here, auditory scene analysis according to Bregman will be introduced too.

Next, to cover the math in the implementation part, the basics of digital sound processing will be described. The related mathematical principles and functions used during the implementation will also be given some attention.

In the following chapter, having all the related theory in mind, an introduction to the main principles and goals of CASA will be made, along with an overview of its applications and selected models.

Then, in the practical part, the focus will be made on describing the implementation of specific parts of the CASA system built for this thesis (see attached medium).

Finally, an overview of the experiments made to test the implemented system will be provided.

..... Chapter 2

Physical Background

Biological Background

Sounds... For sure, they are one of the most important sources of information in our everyday life. By listening to them, one can describe what is happening around, understand how to react to occurring situations, or even tell if a danger is approaching, and it is time to take action. It is hard to imagine human perception without hearing, but as easy as this may sound (no pun intended), the biology behind it is quite complicated. This chapter will introduce the reader to how sound as a mechanical phenomenon is converted to sound as perception and provide a basic overview of the structures in the human ear, along with the mechanical and neurobiological processes happening inside of them.

3.1 Outer and Middle Ear

At the beginning, sound approaches the ear by vibrations in the air (or any other elastic medium) and enters the outer ear, which consists of the visible part (called the auricle, or the pinna) and the ear canal. The auricle is a thin plate of elastic cartilage, covered with integument, and connected to the surrounding parts by ligaments and muscles; and to the beginning of the ear canal by fibrous tissue [Wikipedia citation – Outer ear]. The ear canal is a tube leading from the bottom of the auricle to the middle ear, separated from it by the eardrum (or tympanic membrane). The main purpose of the ear canal is to focus the sound energy gathered by the auricle on the eardrum. It also amplifies frequencies between 3 kHz and 12 kHz.

Being gathered on the eardrum, the mechanical vibrations propagate through the middle ear. Three bones (called the ossicles) are located inside of it. The malleus (also called the hammer) is connected to the eardrum and transfers the vibrations from it to the incus (the anvil). These vibrations are quite chaotic, but the malleus is connected to the eardrum in a linear manner, also helping the ear to respond more linearly and smoothly. The incus, in turn, connects to the stapes (the stirrup). The footplate of the stapes introduces pressure waves in the inner ear, which starts with the oval window of the cochlea. The structures of the middle ear can be seen on Figure 3.1a

It may sound redundant to have additional structures in the ear which propagate the vibrations even further, when they could travel just one centimeter more in a way like before, in the ear canal, but in reality, the pressure of these mechanical vibrations is too small to cause the waves of the same velocity in the cochlear fluids. The ossicles help to amplify the pressure of these vibrations. They are positioned to form a lever, and, because the oval window is about 14 times smaller than the eardrum, the pressure gain becomes quite significant in the end –

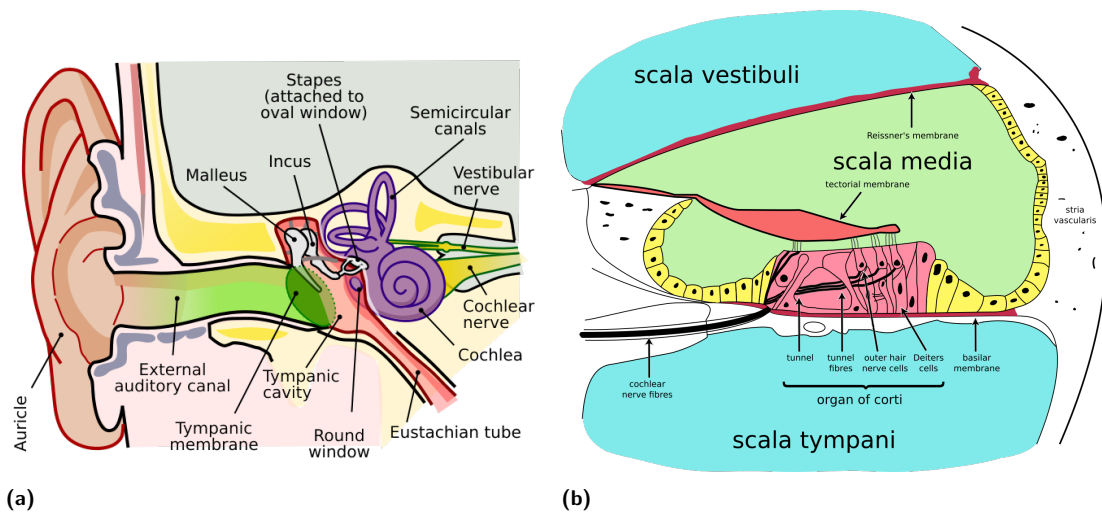


Figure 3.1 (a) Anatomy of the human ear. The ossicles of the middle ear are shown in white. The inner ear is shown in purple. (b) Cross-section of the cochlea showing the organ of Corti and three chambers filled with cochlear fluids. Both pictures were taken from <https://commons.wikimedia.org/>

at least 18.1 times [Wikipedia citation – Middle ear].

To regulate the middle ear and protect it from damage due to very loud sounds, two muscles are located inside of it: the stapedius muscle and the tensor tympani muscle. These muscles are controlled by unconscious reflexes and hold the ossicles when the vibrations become too intense.

To provide ventilation and drainage of the middle ear and to equalize pressures in this isolated environment, the middle ear is connected to the back of the throat by the eustachian tube [Auditory Neuroscience reference].

3.2 Inner Ear

The inner ear starts with the above-mentioned oval window, which is connected to the stapes of the middle ear. The oval window is a part of the cochlea — a structure of the inner ear dedicated to hearing. Along with the cochlea, the inner ear also contains the vestibular system, which is responsible for the sense of balance and spatial orientation and uses the same kinds of fluids and cells as the cochlea does. The vestibular system will not be covered in this thesis, but the fluids and cells will be described in more detail later in the section.

The cochlea itself is a spiral-shaped cavity made of bony tissue, which makes about 2.75 turns around its axis and is about 3 cm long [Wikipedia citation - Cochlea]. The core component of it is the basilar membrane, which runs along almost its entire length and separates two of the three chambers of the cochlea filled with different fluids: the tympanic duct filled with perilymph (scala tympani), and the cochlear duct filled with endolymph (scala media). The third chamber, the vestibular duct (scala vestibuli), is separated from the cochlear duct by the Reissner's membrane and is filled with perilymph (Figure 3.1b). When the footplate of the stapes of the middle ear introduces movements to the cochlear fluids, the basilar membrane is affected too, and the endolymph in the cochlear duct moves along.

The most interesting property of the basilar membrane is that its stiffness and width is different throughout its length – the membrane is narrow and stiff at the basal end of the cochlea, and

wide and floppy at the apical end [Auditory Neuroscience citation]. And here sound waves have two possible routes to take while propagating through the basilar membrane: a shorter path, which includes going through the stiffer parts of it, or a longer path, which means travelling along the membrane until it becomes easier to pass through, but pushing more fluid on the way. In fact, high-frequency waves tend to choose the shorter path, and low-frequency waves – the longer one.

Thus, the basilar membrane moves in different places depending on the frequencies of the vibrations. The organ of Corti, which sits on top of it and runs along its entire length, contains displacement cells able to respond to movements of the fluid nearby and send electrical impulses when this happens. Such cells are packed with a bunch of stereocilia (hair) that stick out of its top, and thus are called hair cells. These cells can be of two types: inner hair cells that are located closer to the center of the cochlea, and outer hair cells that sit closer to its outer side. Inner hair cells are less numerous than outer hair cells and form a single row along the organ of Corti, while outer hair cells usually form three rows [Auditory Neuroscience citation].

Now, it is important to mention that the endolymph in the cochlear duct contains high amounts of positively charged ions (primarily potassium and calcium). When it moves in response to the sound pressure, the stereocilia of the inner hair cells are deflected, and tiny ion channels open in them. This allows the charged ions from the endolymph to enter the stereocilia. The cell becomes depolarized, and a receptor potential is produced. This results in releasing the neurotransmitters at the basal end of the cell and then triggering action potentials in the nerve nearby [Wikipedia citation - Hair cells]. In this way, inner hair cells detect movements around them and convert mechanical sound waves to electrical nerve signals.

Outer hair cells, in turn, serve as amplifiers of the quiet sounds. Their receptor potentials are converted to cell body movements, thus increasing the sound pressure.

3.3 Auditory Scene Analysis

..... Chapter 4

Mathematical Background

Computational ASA

Now, having described all the underlying concepts from different fields of science in previous chapters, it is time to finally focus on computational auditory scene analysis. CASA is said to be a field of study that groups practical, programmable solutions for auditory scene analysis problems, and thus can introduce new discoveries and insights to it. CASA systems are used primarily for source separation, meaning that they are machine listening systems that aim to separate sounds from different sources in mixtures. However, they are not the same as systems for blind signal separation – the core difference is that CASA systems try to mimic (at least to some extent) the mechanisms inside the human ear, which were described in chapter 3. In this chapter, main principles of CASA systems will be described, along with a typical structure, desired outputs and applications. In the second part, major works that use computational auditory scene analysis for source separation will be reviewed and compared.

[illegible]

Implementation

■ ■

Experiments

