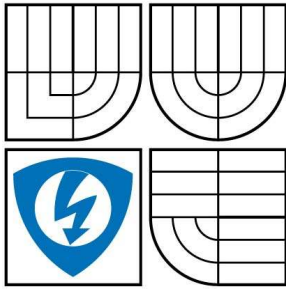


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

VÍCEKANÁLOVÉ METODY ZVÝRAZŇOVÁNÍ ŘEČI

MULTI-CHANNEL METHODS OF SPEECH ENHANCEMENT

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. ADAM ZITKA

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. ZDENĚK SMÉKAL, CSc.

BRNO 2008

LICENČNÍ SMLOUVA POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Bc. Adam Zitka
Bytem: Benedettiho 543, 53003, Pardubice
Narozen/a (datum a místo): 26.2. 1984, Pardubice

(dále jen „autor“)

a

2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií
se sídlem Údolní 244/53, 602 00, Brno
jejímž jménem jedná na základě písemného pověření děkanem fakulty:
prof. Ing. Kamil Vrba, CSc.
(dále jen „nabyvatel“)

Čl. 1

Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):

- disertační práce
- diplomová práce
- bakalářská práce
- jiná práce, jejíž druh je specifikován jako

.....
(dále jen VŠKP nebo dílo)

Název VŠKP: Vícekanálové metody zvýrazňování řeči
Vedoucí/ školitel VŠKP: prof. Ing. Zdeněk Smékal, CSc.
Ústav: Ústav telekomunikací
Datum obhajoby VŠKP:

VŠKP odevzdal autor nabyvateli v*:

- tištěné formě – počet exemplářů 2
- elektronické formě – počet exemplářů 2

* hodící se zaškrtněte

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.
3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.
4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

Článek 2

Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užít, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
 - ihned po uzavření této smlouvy
 - 1 rok po uzavření této smlouvy
 - 3 roky po uzavření této smlouvy
 - 5 let po uzavření této smlouvy
 - 10 let po uzavření této smlouvy(z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/ 1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3

Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísní a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne:

.....
Nabyvatel

.....
Autor

Anotace

Diplomová práce se zabývá vícekanálovými metodami zvýrazňování řeči. Vícekanálové metody zvýrazňování řeči používají pro snímání signálů více mikrofonů. Ze směsí signálů můžeme pomocí neuronových sítí oddělit například jednotlivé mluvčí, potlačit hluk v pozadí či šum. Tato úloha bývá nazývána cocktail-party efekt. Hlavní metoda řešící tento problém se nazývá analýza nezávislých komponent. V práci jsou nejdříve popsány její teoretické základy a jsou představeny podmínky a požadavky k její aplikaci. Jednotlivé metody ICA se snaží směsi rozdělovat pomocí hledání co nejmenších gaussovských vlastností signálů. Pro analýzu nezávislých komponent se používají různé matematické vlastnosti signálů jako je např. špičatost a entropie. Signály, které byly smíšeny uměle v počítači, lze poměrně dobře oddělovat např. pomocí FastICA či algoritmu rostoucího gradientu. Složitější situace nastává, chceme-li oddělit signály pořizené v reálném nahrávacím prostředí, protože separaci řeči osob současně mluvících v reálném prostředí ovlivňují různé další okolnosti jako akustické vlastnosti místnosti, šum, zpoždění, odrazy od stěn, pozice či typ mikrofonů apod. Práce představuje postup analýzy nezávislých komponent ve frekvenční oblasti, který dokáže úspěšně oddělit i nahrávky pořizené v reálném prostředí.

Klíčová slova

analýza nezávislých komponent, ICA, slepá separace zdrojů, cocktail-party efekt, zvýrazňování řeči

Abstract

This thesis deals with multi-channel methods of speech enhancement. Multichannel methods of speech enhancement use a few microphones for recording signals. From mixtures of signals, for example, individual speakers can be separated, noise should be reduced etc. with using neural networks. The task of separating speakers is known as a cocktail-party effect. The main method of solving this problem is called independent component analysis. At first there are described its theoretical foundation and presented conditions and requirements for its application. Methods of ICA try to separate the mixtures with help of searching the minimal gaussian properties of signals. For the analysis of independent components are used different mathematical properties of signals such as kurtosis and entropy. Signals, which were mixed artificially on a computer, can be relatively well separated using, for example, FastICA algorithm or ICA gradient ascent. However, difficult is situation, if we want to separate the signals created in the real recording environment, because the separation of speech people speaking at the same time in the real environment affects other various factors such as acoustic properties of the room, noise, delays, reflections from the walls, the position or the type of microphones, etc. Work presents approach of independent component analysis in the frequency domain, which can successfully separate also recordings made in the real environment.

Key words

independent component analysis, ICA, blind source separation, cocktail-party effect, speech enhancement

Prohlášení

Prohlašuji, že svou diplomovou práci na téma Vícekanálové metody zvýrazňování řeči jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....
podpis autora

Poděkování

Děkuji vedoucímu diplomové práce prof. Ing. Zdeňku Smékalovi, CSc. za odbornou pomoc a cenné rady při zpracování diplomové práce.

V Brně dne

.....

podpis autora

Obsah

Seznam proměnných, symbolů a zkratk	9
1 Úvod	10
2 Analýza nezávislých komponent (ICA)	11
2.1 Úvod	11
2.2 Definice	13
2.3 Nejednoznačnost analýzy nezávislých komponent	14
2.4 Nezávislost a nekorelovanost	14
2.5 Gaussovské proměnné	15
3 Odhady nezávislých komponent.....	16
3.1 Míra špičatosti	16
3.2 Negativní entropie	17
3.3 Aproximace negativní entropie	18
4 Předzpracování pro ICA	20
4.1 Centrování	20
4.2 Bělení.....	20
5 Algoritmus FastICA	22
5.1 FastICA pro jednu jednotku	22
5.2 FastICA pro několik jednotek.....	23
5.3 Separace signálů pomocí FastICA.....	24
6 Algoritmus rostoucího gradientu	29
6.1 Algoritmus rostoucího gradientu pro obecný případ.....	29
6.2 Algoritmus rostoucího gradientu pro supergaussovské signály	32
6.3 Separace signálů pomocí algoritmu rostoucího gradientu.....	32
7 Separace konvolutorních směsí řečových signálů	35
7.1 Algoritmus Fixed-point ICA ve frekvenční oblasti.....	36
7.2 Permutační a měřítkový problém	39
7.3 Separace signálů nahraných v reálném prostředí	41
8 Závěr.....	46
Seznam literatury	48
Seznam příloh.....	49

Seznam proměnných, symbolů a zkratek

A	směšovací matice
a_{ij}	prvky směšovací matice A
a	zesílení
γ_2	koeficient špičatosti
$E\{x^r\}$	moment r -tého řádu veličiny x
F	směrový diagram
η	iterační krok
G	nekvadratická funkce
g	derivace nekvadratické funkce G , distribuční funkce
H	entropie
h	funkce změny entropie
h_{ji}	impulsní odezva prostředí
∇h	gradient h
θ	směr přístupu
J	negativní entropie
L	Lagrangeova funkce
λ	Lagrangeův koeficient
M	počet signálů zachycených mikrofony
m	vektor střední hodnoty
P	permutační matice
p	hustota pravděpodobnosti
R	počet zdrojových signálů
\hat{S}	odhadovaný zdrojový signál ve frekvenční oblasti
s_i	zdrojové (originální) signály
s	vektor zdrojových signálů
tgh	hyperbolický tangens
x_i	smíšené signály
x	vektor smíšených signálů
y_i	výstupní nezávislé komponenty
W	separační matice (inverzní matice ke směšovací matici A)
w	váhový vektor neuronu

BSS	Blind Source Separation - slepá separace zdrojů
DFT	diskrétní Fourierova transformace
EEG	elektroencefalogram
FastICA	Fast Independent Component Analysis - rychlý algoritmus analýzy nezávislých komponent
ICA	Independent Component Analysis - analýza nezávislých komponent
SD	směrový diagram
SP	směr přístupu
TFSS	Time-Frequency Series of Speech - časově-frekvenční soubory řeči

1 Úvod

Diplomová práce se zabývá vícekanálovými metodami zvýrazňování řeči. Úkolem bylo prozkoumat oblast zpracování signálů vzniklých tzv. cocktail-party efektem. Jedná se o situace, kdy máme více zvukových nahrávek z určitého prostředí a snažíme se od sebe oddělit např. promluvy jednotlivých mluvčích, potlačit rušení atp. K danému problému se v této práci přistupuje z hlediska statistiky - prostor je věnován metodě nazývané analýza nezávislých komponent (ICA).

Ve druhé kapitole je úvodní popis řešení problémů pomocí metody ICA. Je předložena její definice a jsou stanoveny konkrétní předpoklady a podmínky pro její práci. Základním požadavkem metody ICA je, že zdrojové signály (např. jednotlivé promluvy mluvčích) jsou nezávislé a nemají Gaussovo (normální) rozdělení pravděpodobnosti. Na grafech je ukázán hlavní cíl metody, je definována nezávislost proměnných a je podáno vysvětlení, proč nelze používat proměnné s Gaussovým rozdělením pravděpodobnosti.

Třetí kapitola představuje různé možnosti výpočtu ICA. První přístup využívá špičatost. Špičatost je statistická vlastnost signálu a metoda vychází ze skutečnosti, že proměnné s Gaussovým rozdělením, v ICA zakázané, mají koeficient špičatosti nulový. Další v práci uvedený přístup využívá entropii, základní veličinu teorie informace. Největší entropii mají signály s Gaussovým rozdělením. Metoda vycházející z negativní entropie je založena na porovnání entropie sledovaného signálu a entropie gaussovského signálu se stejnou kovarianční maticí. Aproximace negativní entropie pak přináší praktičtější zjednodušení metody negativní entropie.

Čtvrtá kapitola ukazuje předzpracování signálu, která jsou velmi vhodná provést před samotným začátkem zpracování. Centrováním získáme veličinu s nulovou střední hodnotou, bělením zajistíme, že komponenty budou nekorelované a jejich rozptyly budou jednotkové.

V páté a šesté kapitole je ukázáno konkrétní použití metody ICA. Algoritmus FastICA je založen na aproximaci negativní entropie. Algoritmus FastICA je popsán pro jednu i více jednotek, v textu je uveden používaný iterační algoritmus. Na umělých a reálných signálech náhodným způsobem smíšených je tato metoda předvedena a jsou diskutovány její vlastnosti. Šestá kapitola pak ukazuje jiný přístup ICA - algoritmus rostoucího gradientu. Po podrobném matematickém popisu je tato metoda na reálných signálech vyzkoušena a výsledkově zhodnocena.

Sedmá kapitola se věnuje zpracování reálných nahrávek pořízených pomocí několika mikrofonů v reálném akustickém prostředí, kde se uplatňují další vlivy jako jsou akustické vlastnosti místnosti, šum, zpoždění, odrazy od stěn, pozice či typ mikrofonů a další. Hledání původních signálů je prováděno pomocí analýzy nezávislých komponent ve frekvenční oblasti. Matematický princip spočívá ve zpracování časově-frekvenčních souborů řeči s využitím mj. Fourierovy transformace a negativní entropie. Na nahrávkách pořízených v reálných prostředích je tato metoda vyzkoušena.

2 Analýza nezávislých komponent (ICA)

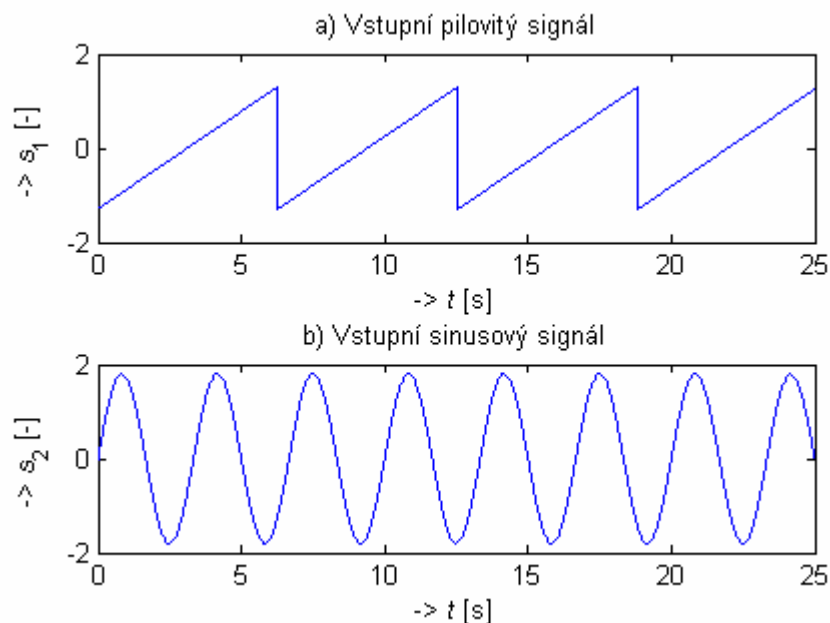
2.1 Úvod

Představme si situaci, že v určité místnosti mluví dvě osoby najednou, na odlišných místech jsou dva mikrofony a nahrává se zvuk v místnosti. Mikrofony nám dají dva signály, které mohou být označeny $x_1(t)$ a $x_2(t)$, kde x_1 a x_2 jsou okamžité hodnoty signálu a t je čas. Každý z těchto zaznamenaných signálů je vážená suma řečových signálů od dvou mluvčích, které můžeme označit $s_1(t)$ a $s_2(t)$ - říkáme jim zdrojové (originální) signály. Situaci můžeme vyjádřit pomocí lineárních rovnic [4]:

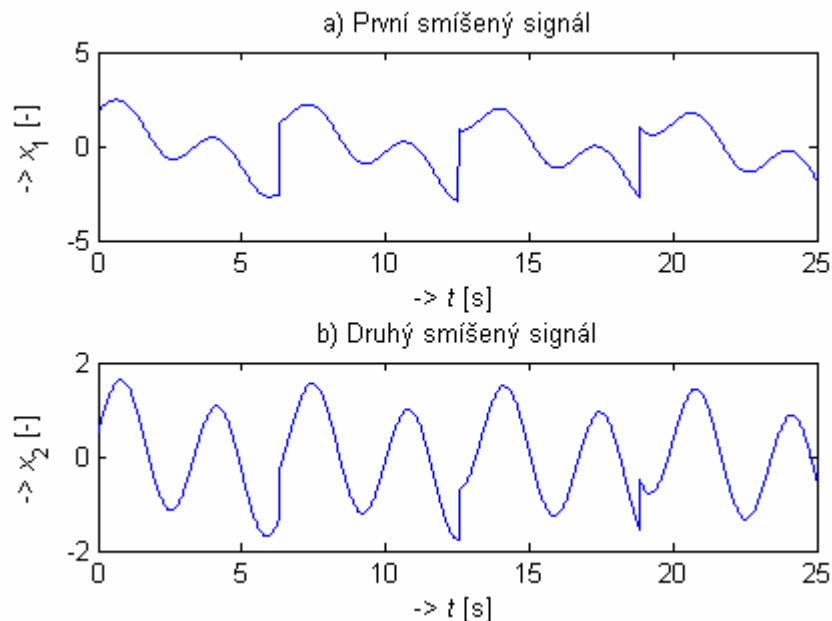
$$\begin{aligned}x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t),\end{aligned}\tag{2.1}$$

kde a_{11} , a_{12} , a_{21} a a_{22} jsou parametry závislé na vzdálenosti mikrofonů od mluvčích. Bylo by nyní velmi užitečné, kdybychom uměli získat dva oddělené zdrojové řečové signály $s_1(t)$ a $s_2(t)$ s použitím pouze nahraných (smíšených) signálů $x_1(t)$ a $x_2(t)$. Tento úkol se nazývá cocktail-party problém. Pro začátek neuvažujeme časové zpoždění či další faktory našeho zjednodušeného směšovacího modelu.

Pro názornost uvažujme křivky na obr. 2.1 a 2.2. Nejsou to samozřejmě reálné řečové signály, slouží jen pro příklad. Zdrojové řečové signály by mohly být takové jako jsou na obr. 2.1 a smíšené signály by mohly vypadat jako ty na obr. 2.2. Úkolem je obnovit data z obr. 2.1 za použití pouze informací z obr. 2.2. Když známe parametry a_{ij} , můžeme řešit lineární rovnice (2.1) klasickými metodami. Potíž je v tom, že když neznáme a_{ij} , problém je mnohem složitější.



Obr. 2.1: Časové průběhy zdrojových signálů



Obr. 2.2: Časové průběhy smíšených signálů

Jedním přístupem k řešení tohoto problému by mohlo být využití některých informací ze statistických vlastností signálů $s_i(t)$ k odhadu a_{ij} . Ve skutečnosti to možná překvapivě vede k tomu, že je nutné předpokládat, že $s_1(t)$ a $s_2(t)$ jsou v každém časovém okamžiku statisticky nezávislé. V mnoha případech to není tak nereálný předpoklad a v praxi nemusí být ani přesně pravdivý. Často používané techniky analýzy nezávislých komponent (ICA), které mohou být použity k odhadu a_{ij} , jsou založené na informacích z nezávislosti, které dovolují separovat dva zdrojové signály $s_1(t)$ a $s_2(t)$ z jejich směsi $x_1(t)$ a $x_2(t)$.

Metoda ICA byla původně vyvinuta k řešení jevů velmi úzce spojených s cocktail-party efektem. V době většího zájmu o ICA se zjistilo, že tento princip má také hodně jiných zajímavých využití.

Uvažujme například elektrický záznam mozkové aktivity (EEG). Data EEG se skládají ze záznamů elektrických potenciálů na mnoha různých místech hlavy. Výstupy ze snímačů na hlavě jsou pravděpodobně vytvářeny smícháním různých složek mozkové aktivity. Tato situace je docela podobná cocktail-party problému: chtěli bychom najít původní složky mozkové aktivity, ale můžeme pozorovat pouze směs složek. ICA může odhalit zajímavé informace mozkové aktivity prostřednictvím nezávislých složek.

Jiný velmi odlišný příklad využití metody ICA je extrakce důležitých rysů signálu (např. obličejových rysů). Základní problém ve zpracování číselných signálů je najít vhodné reprezentace pro obraz, audio nebo jiný druh dat pro úlohy jako je komprese a odšumování.

Výše popsané aplikace mohou být skutečně formulovány a sjednoceny matematickou konstrukcí ICA, tj. velmi obecnou a účelovou metodou zpracování signálu a analýzy dat.

2.2 Definice

Pro přesnou definici metody ICA můžeme použít statistický model „skrytých proměnných“ [4]. Předpokládejme, že sledujeme n lineárních směsí x_1, \dots, x_n pro n nezávislých složek:

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n, \text{ pro všechna } j. \quad (2.2)$$

V rovnici (2.2) je nyní vypuštěn čas t . V modelu ICA předpokládáme, že každá směs x_j stejně jako nezávislá složka s_k je náhodná proměnná nehledě na konkrétní časový signál. Sledované hodnoty $x_j(t)$, např. signály z mikrofonů v cocktail-party efektu, jsou pak vzorky této náhodné proměnné. Můžeme předpokládat, že obě (smíšené proměnné a nezávislé složky) mají nulovou střední hodnotu. Pokud to pravda není, mohou být sledované proměnné centrovány odečítáním střední hodnoty vzorků, čímž se vytvoří model s nulovou střední hodnotou.

Je výhodné používat vektorově-maticový popis namísto součtů jako v předchozí rovnici. Označme \mathbf{x} jako náhodný sloupcový vektor, jehož dílčí složky jsou směsi x_1, \dots, x_n a podobně \mathbf{s} jako náhodný sloupcový vektor s dílčími složkami s_1, \dots, s_n . Zaveďme matici \mathbf{A} se složkami a_{ij} . Matice \mathbf{A} má tolik řádků, kolik je smíšených signálů, a tolik sloupců, kolik je zdrojových signálů. Použitím tohoto vektorově-maticového popisu je výše popsán směšovací model zapsán jako:

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \quad (2.3)$$

Někdy potřebujeme sloupce matice \mathbf{A} . Je-li \mathbf{a}_i sloupec matice \mathbf{A} , tak předchozí rovnice může být zapsána i takto:

$$\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i. \quad (2.4)$$

Statistický model v rovnici (2.3) se nazývá analýza nezávislých komponent (ICA). ICA model je vytvářecím modelem, což znamená, že popisuje, jak jsou sledovaná (zaznamenávaná) data vytvářena procesem mísení složek (zdrojových signálů) s_i . Nezávislé komponenty jsou skryté proměnné, což znamená, že nemohou být sledovány přímo. Také směšovací matice je předpokládána jako neznámá. Vše, co dostáváme, je náhodný vektor \mathbf{x} a musíme odhadnout jak \mathbf{A} , tak \mathbf{s} . To je možné za určitých podmínek.

Základním bodem pro ICA je velmi jednoduchý předpoklad, že komponenty s_i jsou statisticky nezávislé. Statistická nezávislost bude přesně definována ještě dále v podkapitole 2.4. Bude také ukázáno, že musíme předpokládat, že nezávislé komponenty nesmí mít Gaussovo (normální) rozdělení pravděpodobnosti. Bohužel v základním modelu nemůžeme předpokládat, že rozdělení známe (pokud rozdělení známe, problém se značně zjednoduší). Pro jednoduchost dále předpokládejme, že neznámá směšovací matice \mathbf{A} je čtvercová. To znamená, že pořizujeme mikrofonem tolik nahrávek, kolik je zdrojů. Pokud čtvercová nebude a budeme zaznamenávat méně signálů než je zdrojů, všechny zdroje se odhadnout nepodaří. Potom po odhadu matice \mathbf{A} , můžeme dopočítat její inverzní matici, pojmenujme ji \mathbf{W} , a dostat nezávislou komponentu jednoduše pomocí:

$$\mathbf{s} = \mathbf{W}\mathbf{x}. \quad (2.5)$$

ICA je velmi úzce spojena s metodou nazývanou slepá separace zdrojů (BSS). „Zdrojem“ je originální signál, tj. nezávislá komponenta, jako je mluvčí v cocktail-party problému. „Slepá“ znamená, že víme velmi málo o směšovací matici. ICA je jednou z metod, pravděpodobně nejvíce rozšířenou, pro přípravu slepé separace zdrojů.

V mnoha aplikacích by bylo realističtější předpokládat, že v měření je určitý šum přidaný k modelu. Pro jednoduchost žádný šum nyní neuvažujeme, pro mnoho aplikací je bezšumový model využitelný.

2.3 Nejednoznačnost analýzy nezávislých komponent

V ICA modelu (2.3) je vidět následující nejednoznačnost (dvojznačnost):

1. Nemůžeme určit energie nezávislých komponent.

Protože \mathbf{s} a \mathbf{A} jsou neznámé, jakékoli skalární násobení v jednom ze zdrojů s_i může být vždy zrušeno dělením stejným skalárním číslem odpovídajícího sloupce \mathbf{a}_i matice \mathbf{A} - viz rovnice (2.4). V důsledku toho můžeme velice dobře určit moduly nezávislých komponent. Protože jsou to náhodné proměnné, nejpřirozenější cestou k získání modulu je předpokládat, že každá proměnná má jednotkový rozptyl $E\{s_i^2\} = 1$. Potom matice \mathbf{A} bude upravena v metodách řešení ICA tak, aby počítala s touto podmínkou. I tak stále zůstává dvojznačnost znaménka, ale můžeme klidně násobit nezávislou komponentu -1 bez vlivu na model. Tato nejednoznačnost je naštěstí pro většinu aplikací bezvýznamná.

2. Nemůžeme určit pořadí nezávislých komponent.

Důvodem je opět neznalost \mathbf{s} i \mathbf{A} . Můžeme volně měnit pořadí členů sumy v rovnici (2.4) a označit libovolnou nezávislou komponentu jako první. Pokud zavedeme permutační matici \mathbf{P} , která je stejného rozměru jako \mathbf{A} , a její inverzní matici \mathbf{P}^{-1} , můžeme s jejich využitím model napsat ve tvaru $\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$. Členy $\mathbf{P}\mathbf{s}$ jsou originální nezávislé proměnné s_j , ale v jiném pořadí. Matice $\mathbf{A}\mathbf{P}^{-1}$ je jen novou neznámou směšovací maticí k řešení algoritmu ICA.

2.4 Nezávislost a nekorelovanost

Pro definování nezávislosti uvažujeme dvě skalární náhodné proměnné y_1 a y_2 . Proměnné y_1 a y_2 jsou nezávislé, když znalost hodnoty y_1 nedává žádnou informaci o hodnotě proměnné y_2 a naopak. Výše bylo uvedeno, že toto je případ proměnných s_1 a s_2 , ale ne směsí x_1 a x_2 .

Formálně může být nezávislost definována hustotou pravděpodobnosti. Označme si $p(y_1, y_2)$ jako vzájemnou dvojrozměrnou hustotu pravděpodobnosti y_1 a y_2 . Dále označme $p_1(y_1)$ mezní hustotou pravděpodobnosti y_1 , tj. když uvažujeme:

$$p_1(y_1) = \int p(y_1, y_2) dy_2 \quad (2.6)$$

a podobně pro y_2 .

Pak definujme, že y_1 a y_2 jsou nezávislé právě tehdy, když vzájemná hustota pravděpodobnosti je rozložitelná následovně:

$$p(y_1, y_2) = p_1(y_1) p_2(y_2). \quad (2.7)$$

Tato definice platí pro libovolný počet n náhodných proměnných, v tom případě musí být vzájemná hustota pravděpodobnosti výsledkem n rozměrů.

Definice může být použita k odvození nejdůležitější vlastnosti nezávislých proměnných. Pro dvě funkce h_1 a h_2 vždy dostaneme:

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\}. \quad (2.8)$$

Slabší formou nezávislosti je nekorelovanost. Dvě náhodné proměnné y_1 a y_2 jsou nekorelované, když jejich kovariance je nulová:

$$E\{y_1 y_2\} - E\{y_1\}E\{y_2\} = 0. \quad (2.9)$$

Když jsou proměnné nezávislé, jsou nekorelované, což vyplývá z rovnice (2.9), vezmeme-li $h_1(y_1) = y_1$ a $h_2(y_2) = y_2$. Na druhé straně z nekorelovanosti nevyplývá nezávislost. Tím, že nezávislost implikuje nekorelovanost, mnoho metod ICA omezuje odhad tak, že vždy dává nekorelované odhady nezávislých komponent. To zmenšuje počet volných parametrů a zjednodušuje problém.

2.5 Gaussovské proměnné

Základním požadavkem metody ICA je, že nezávislé komponenty musí být negaussovské. Gaussovské náhodné proměnné mají symetrické rozdělení okolo střední hodnoty a to je pro všechny proměnné shodné. Proto je nelze oddělit - není informace o tvaru rozdělení. Přesněji řečeno, nejvýše jedna náhodná proměnná Gaussovo rozdělení mít může.

3 Odhady nezávislých komponent

Klíčem k odhadování modelu ICA je hledání signálů s minimální podobností s gaussovským signálem. Centrální limitní věta, klasický výsledek teorie pravděpodobnosti, říká, že rozdělení součtu nezávislých náhodných proměnných směřuje ke Gaussovu (náhodnému) rozdělení. Součet dvou nezávislých náhodných proměnných má obvykle rozdělení bližší Gaussovu než dvě originální náhodné proměnné. Snahou je tedy najít signál s minimálními gaussovskými vlastnostmi. Takový signál bude jedním z nezávislých signálů.

Pro použití negaussovských vlastností v odhadu ICA musíme mít kvantitativní měření negaussovských vlastností náhodné proměnné y . Pro zjednodušení předpokládejme, že y je centrován (nulová střední hodnota) a rozptyl je roven jedné. Metody předzpracování ICA (bude vysvětleno v kapitole 4), tato zjednodušení uskutečňují.

3.1 Míra špičatosti

Klasickou mírou negaussovských vlastností je špičatost (kurtosis). Koeficient špičatosti γ_2 je klasicky definován [4]:

$$\gamma_2(y) = \frac{E\{y^4\}}{(E\{y^2\})^2} - 3. \quad (3.1)$$

Ve skutečnosti, pokud předpokládáme, že y je jednotkového rozptylu, se pravá strana zjednoduší na $E\{y^4\} - 3$. To ukazuje, že špičatost je jednoduše normalizovaná verze momentu 4. řádu $E\{y^4\}$. Pro gaussovský y je moment 4. řádu roven $3(E\{y^2\})^2$. Takže gaussovská náhodná proměnná má nulový koeficient špičatosti, a proto je pro ICA nevhodná. Pro většinu (ne všechny ale skoro všechny) negaussovské náhodné proměnné je špičatost nenulová. Špičatost může být kladná i záporná. Náhodná proměnná se zápornou špičatostí se nazývá subgaussovská (nebo platykurtická), s kladnou špičatostí se nazývá supergaussovská (nebo leptokurtická). Supergaussovská náhodná proměnná má typicky „špičatou“ hustotu pravděpodobnosti se silnými chvosty, tzn. že hustota pravděpodobnosti je relativně velká v nule a ve vysokých hodnotách proměnné, zatímco malá je pro střední hodnoty. Typickým příkladem je Laplaceovo rozdělení, jehož hustota pravděpodobnosti (normalizovaná na jednotkový rozptyl) je dána:

$$p(y) = \frac{1}{\sqrt{2}} \exp(\sqrt{2}|y|). \quad (3.2)$$

Na druhé straně subgaussovské náhodné proměnné mají typicky „hladkou“ hustotu pravděpodobnosti, která je poměrně konstantní blízko nule a velmi malá pro vysoké hodnoty proměnné. Typickým příkladem je rovnoměrné rozdělení.

Míra negaussovských vlastností je typicky měřena pomocí absolutní hodnoty špičatosti. Mohou se použít také druhé mocniny špičatostí. Ty jsou nulové pro gaussovskou proměnnou a větší než nula pro většinu negaussovských náhodných

proměnných. Jsou i negaussovské náhodné proměnné, které mají nulovou špičatost, ale jsou velmi vzácné.

Špičatost nebo spíše její absolutní hodnota, je hojně využívána jako míra negaussovských vlastností v ICA. Hlavním důvodem je její jednoduchost - výpočetní i teoretická. Výpočetně může být špičatost odhadnuta jednoduše použitím momentu 4. řádu vzorků. Teoretická analýza je zjednodušena díky následující lineární vlastnosti: když x_1 a x_2 jsou dvě nezávislé náhodné proměnné, platí [4]:

$$\gamma_2(x_1+x_2) = \gamma_2(x_1) + \gamma_2(x_2) \quad (3.3)$$

a

$$\gamma_2(\alpha x_1) = \alpha^4 \gamma_2(x_1),$$

kde α je skalár.

Míra špičatosti má jako metoda ICA i svoje nevýhody. Hlavním problémem je, že špičatost je velmi citlivá na odlehlé hodnoty (výkyvy). Její hodnota může záviset na jen málo pozorováních na krajích rozdělení, které mohou být náchylné na chyby nebo mít nepodstatné vzorky.

3.2 Negativní entropie

Druhým velmi důležitým měřením negaussovských vlastností je negativní entropie. Metoda negativní entropie je založena na množství informace - entropii. Entropie je základní kámen teorie informace. Entropie náhodné proměnné může být vysvětlena jako stupeň informace, který nám dá pozorování proměnné. Čím více „náhodná“, tj. nepředvídatelná a nestruturovaná, proměnná je, tím větší je její entropie.

Entropie H je definována pro diskrétní náhodnou proměnnou Y takto [4]:

$$H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i), \quad (3.4)$$

kde a_i jsou možné hodnoty Y . Tato velmi dobře známá definice může být zobecněna pro spojitou náhodnou proměnnou a vektory. V těchto případech se často nazývá diferenciální entropií. Diferenciální entropie H náhodného vektoru \mathbf{y} s hustotou $f(\mathbf{y})$ je definována jako:

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}. \quad (3.5)$$

Základním důsledkem teorie informace je, že gaussovská proměnná má největší entropii mezi všemi náhodnými proměnnými stejného rozptylu. To znamená, že entropie může být využita jako míra negaussovských vlastností. V podstatě to ukazuje, že Gaussovo rozdělení je „nejvíce náhodné“ neboli nejméně strukturované ze všech rozdělení. Entropie je malá pro rozdělení, která jsou zřetelně koncentrovaná na určité hodnoty, tj. když je proměnná zjevně koncentrovaná nebo má hustotu pravděpodobnosti, která je hodně špičatá.

Pro získání míry negaussovských vlastností, která bude nulová pro gaussovskou proměnnou a vždy nezáporná, někdo často používá mírně modifikovanou verzi definice

diferenciální entropie, nazývanou negativní entropie. Negativní entropie J je definována následovně:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y}), \quad (3.6)$$

kde $\mathbf{y}_{\text{gauss}}$ je gaussovská náhodná proměnná stejné kovarianční matice jako \mathbf{y} . Kvůli výše zmíněným vlastnostem je negativní entropie vždy nezáporná a je nulová právě a jen tehdy když má Gaussovo rozdělení.

Výhodou použití negativní entropie, resp. diferenciální entropie, jako míry negaussovských vlastností je, že je dobře podložená teorií statistiky. Problém využití negativní entropie je, že je velmi výpočtově složitá. Odhadování negativní entropie užitím definice by mohla vyžadovat odhad (třeba neparametrický) hustoty pravděpodobnosti. Proto jsou užitečné jednodušší aproximace negativní entropie.

3.3 Aproximace negativní entropie

Odhad negativní entropie je složitý, a proto tedy zůstává spíše jen teorií. V praxi jsou použitelné některé aproximace, které mají velmi slibné vlastnosti a které budou dále základem odvození efektivní metody ICA. Klasickou aproximací negativní entropie je využití momentů vyšších řádů, například [4]:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \gamma_2(y)^2. \quad (3.7)$$

Náhodná proměnná y se předpokládá s nulovou střední hodnotou a jednotkovým rozptylem. Platnost takových aproximací může být trochu omezená. Jiná aproximace je založena na principu maximální entropie. Obecně získáváme následující aproximaci:

$$J(y) \approx \sum_{i=1}^p k_i [E\{G_i(y)\} - E\{G_i(v)\}]^2, \quad (3.8)$$

kde k_i jsou kladné konstanty a v je gaussovská proměnná s nulovou střední hodnotou a jednotkovým rozptylem. Proměnná y má také nulovou střední hodnotu a jednotkový rozptyl a funkce G_i jsou nekvadratické funkce.

V případě, kdy se používá jedna nekvadratická funkce G , je aproximace:

$$J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2 \quad (3.9)$$

pro prakticky jakoukoli nekvadratickou funkci G . To je zjevně zobecnění aproximace založené na momentu, když y je symetrická.

Důležitým bodem je rozumný výběr G . Lze získat aproximaci negativní entropie, která je mnohem lepší než aproximace daná rovnicí (3.7). Následující výběr G se ukázal jako velmi užitečný:

$$G_1(u) = \frac{1}{a_1} \log \cosh a_1 u , \quad (3.10)$$

$$G_2(u) = -\exp\left(-\frac{u^2}{2}\right) ,$$

kde $1 \leq a_1 \leq 2$ je vhodná konstanta.

Takto dostaneme aproximaci negativní entropie, která dává dobrý kompromis mezi vlastnostmi dvou klasických negaussovských mír špičatosti a negativní entropie. Aproximace jsou koncepčně jednoduché, rychlé na výpočet, mají i zajímavé statistické vlastnosti. Proto se kontrastní funkce používají v metodách ICA. Praktický algoritmus založený na kontrastních funkcích bude představen v kapitole 5.

4 Předzpracování pro ICA

V předchozí části jsme diskutovali statistické principy základních metod ICA. Praktický algoritmus založený na těchto principech bude představen v následující části. Před použitím algoritmu ICA na datech je ale většinou velmi užitečné určitě předzpracování. V této části budou ukázány techniky, které dělají problém odhadu ICA jednodušší.

4.1 Centrování

Nejzákladnějším a nutným předzpracováním je centrovat (vystředit) \mathbf{x} , tzn. odečítat jeho vektor střední hodnoty $\mathbf{m} = E\{\mathbf{x}\}$ tak, abychom udělali \mathbf{x} proměnnou s nulovou střední hodnotou. Délka vektoru \mathbf{m} je dána počtem smíšených signálů \mathbf{x} . To naznačuje, že \mathbf{s} je také s nulovou střední hodnotou, jak může být vidět z předpokladů rovnice (2.3). Toto předzpracování je prováděno výhradně pro zjednodušení algoritmu ICA. Neznamena to, že střední hodnota nemůže být odhadnuta. Po odhadu směšovací matice \mathbf{A} s centrovanými daty můžeme dokončit odhad přidáním vektoru střední hodnoty pro \mathbf{s} zpátky do centrovaného odhadu \mathbf{s} . Vektor střední hodnoty veličiny \mathbf{s} je dán $\mathbf{A}^{-1}\mathbf{m}$, kde \mathbf{m} je střední hodnota, která byla odečtena v předzpracování.

4.2 Bělení

Jiným užitečným předzpracováním v ICA je nejdříve vybělit sledované proměnné. To znamená před aplikací algoritmu ICA (a po centrování), transformujeme sledovaný vektor \mathbf{x} lineárně tak, že získáme nový vektor $\tilde{\mathbf{x}}$, který je bílý, tj. komponenty jsou nekorelované a jejich rozptyly jsou jednotkové. Jinými slovy je kovarianční matice $\tilde{\mathbf{x}}$ rovna jednotkové matici [4]:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I}. \quad (4.1)$$

Transformace bělení je vždy možná. Jedna populární metoda bělení je použití rozkladu vlastní hodnoty kovarianční matice $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, kde \mathbf{E} je ortogonální matice vlastních vektorů $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$ a \mathbf{D} je diagonální matice jejích vlastních hodnot, $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$. $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$ může být odhadnuto standardním způsobem z dostupných vzorků $\mathbf{x}(1), \dots, \mathbf{x}(T)$. Bělení může být nyní provedeno:

$$\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{x}, \quad (4.2)$$

kde matice $\mathbf{D}^{-1/2}$ je počítána $\mathbf{D}^{-1/2} = \text{diag}(d_1^{-1/2}, \dots, d_n^{-1/2})$. Nyní se dá ověřit, že $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I}$.

Bělení transformuje směšovací matici do nové $\tilde{\mathbf{A}}$, využitím rovnice (2.3) a předchozí rovnice:

$$\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{A}\mathbf{s} = \tilde{\mathbf{A}}\mathbf{s}. \quad (4.3)$$

Užitečnost bělení spočívá ve faktu, že nová směšovací matice $\tilde{\mathbf{A}}$ je ortogonální. To může být vidět z:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \tilde{\mathbf{A}}E\{\mathbf{s}\mathbf{s}^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I}. \quad (4.4)$$

Vidíme, že bělení snižuje počet odhadovaných parametrů. Namísto odhadování n^2 parametrů, které jsou členy původní matice \mathbf{A} , potřebujeme pouze odhad nové, ortogonální směšovací matice $\tilde{\mathbf{A}}$. Ortogonální matice obsahuje $n(n-1)/2$ stupňů volnosti. Například ve dvourozměrném prostoru je ortogonální transformace určena jediným úhlovým parametrem. Ve vyšších dimenzích obsahuje ortogonální matice jen asi polovinu parametrů libovolné matice. Bělení je velmi jednoduchou procedurou, jednodušší než jakýkoli algoritmus ICA, a je tedy vhodné ke zjednodušení problémů v této oblasti.

5 Algoritmus FastICA

V kapitole 3 byly představeny různé míry negaussovských vlastností - účelových funkcí pro odhad ICA. V této části představíme velmi efektivní metodu pro realizaci ICA. Předpokládá se, že data jsou předzpracována centrováním a bělením.

5.1 FastICA pro jednu jednotku

Pro začátek ukažme verzi algoritmu FastICA s jednou jednotkovou. „Jednotkou“ rozumíme výpočetní jednotku - v podstatě umělý neuron, který má váhový vektor \mathbf{w} a je schopný se aktualizovat podle učicího pravidla. Učící pravidlo FastICA hledá směr, tj. jednotkový vektor takový, že průmět $\mathbf{w}^T \mathbf{x}$ maximalizuje negaussovské vlastnosti. Negaussovské vlastnosti jsou zde měřeny aproximací negativní entropie $J(\mathbf{w}^T \mathbf{x})$, rozptyl $\mathbf{w}^T \mathbf{x}$ musí být omezen na jednotkový.

Algoritmus FastICA [4] je založen na iteraci s pevnou řádovou čárkou pro nalezení maxima negaussovských vlastností $\mathbf{w}^T \mathbf{x}$. Může být také odvozen aproximační Newtonovou iterací. Označme g jako derivaci nekvadratické funkce G (3.8), například derivace funkcí (3.9) jsou:

$$\begin{aligned} g_1(u) &= \operatorname{tgh}(a_1 u) \\ g_2(u) &= u \exp\left(\frac{-u^2}{2}\right) \end{aligned} \tag{5.1}$$

kde $1 \leq a_1 \leq 2$ je vhodná konstanta, často se bere $a_1 = 1$. Základní forma algoritmu FastICA je následující:

1. Vybereme počáteční (např. náhodný) váhový vektor \mathbf{w} .
2. Necht' $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - E\{g'(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$.
3. Necht' $\mathbf{w} = \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$.
4. Pokud nekonverguje, zpátky k bodu 2.

Norma ve třetím kroku je jakákoli běžná maticová norma. Poznamenejme, že konvergence znamená, že stará a nová hodnota \mathbf{w} jsou ve stejném směru. Není nezbytné, aby vektor konvergoval do jednoho bodu. Je to z toho důvodu, že nezávislé komponenty mohou být definovány až na znaménka. Dále uveďme, že se zde předpokládá, že data jsou předzpracována bělením.

Nyní k odvození FastICA. Zprvce poznamenejme, že maxima aproximace negativní entropie $\mathbf{w}^T \mathbf{x}$ jsou získána z určitého optima $E\{G(\mathbf{w}^T \mathbf{x})\}$. Podle Kuhn-Tuckerových podmínek optima $E\{G(\mathbf{w}^T \mathbf{x})\}$ za omezení $E\{(\mathbf{w}^T \mathbf{x})^2\} = \|\mathbf{w}\|^2 = 1$ jsou získány v bodech, kde platí:

$$E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{w} = 0. \tag{5.2}$$

Řešme tuto rovnici Newtonovou metodou. Označením funkce na levé straně (5.2) jako F získáme její Jacobiho matici $JF(\mathbf{w})$:

$$JF(\mathbf{w}) = E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{I}. \quad (5.3)$$

Pro zjednodušení inverze této matice se rozhodneme aproximovat první výraz v (5.3). Racionální aproximací se zdá být $E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{w}^T \mathbf{x})\} \approx E\{\mathbf{x}\mathbf{x}^T\} E\{g'(\mathbf{w}^T \mathbf{x})\} = E\{g'(\mathbf{w}^T \mathbf{x})\} \mathbf{I}$. Potom se Jacobiho matice stane diagonální a může být snadněji invertována. Tak dostaneme následující aproximační Newtonovu iteraci:

$$\mathbf{w}^+ = \mathbf{w} - \frac{[E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{w}]}{[E\{g'(\mathbf{w}^T \mathbf{x})\} - \beta]}. \quad (5.4)$$

Tento algoritmus může být zjednodušen násobením obou stran členem $\beta - E\{g'(\mathbf{w}^T \mathbf{x})\}$. To dá (po algebraickém zjednodušení) iteraci FastICA.

V praxi musí být předpoklady ve FastICA nahrazeny jejich odhady. Přirozené odhady jsou samozřejmě odpovídající průměry vzorků. Ideálně by měla být použita všechna dostupná data, což ale není většinou dobrý nápad, protože výpočet může být příliš náročný. Průměry mohou být odhadnuty použitím menšího vzorku, jehož velikost může mít značný efekt na přesnost výsledného odhadu. Vzorové body by měly být vybrány zvláště v každé iteraci. Když konvergence není dostatečná, může se zvětšit velikost vzorků.

5.2 FastICA pro několik jednotek

Jedno-jednotkový algoritmus v předchozí části odhaduje jen jednu nezávislou komponentu. Pro odhad několika nezávislých komponent potřebujeme spustit jedno-jednotkový algoritmus FastICA s použitím několika jednotek (např. neuronů) s váhovými vektory $\mathbf{w}_1, \dots, \mathbf{w}_n$. Abychom předešli konvergování různých vektorů do stejného maxima, musíme dekorelovat výstupy $\mathbf{w}_1^T \mathbf{x}, \dots, \mathbf{w}_n^T \mathbf{x}$ po každé iteraci. Toho se dosahuje třemi metodami.

Jednoduchou cestou zajištění dekorelace je schéma založené na Gram-Schmidtově dekorelaci. To znamená, že odhadneme nezávislé komponenty jednu po druhé. Když máme odhadnout p nezávislých komponent nebo p vektorů $\mathbf{w}_1, \dots, \mathbf{w}_p$, necháme proběhnout jedno-jednotkový algoritmus s pevnou řádkovou čárkou pro \mathbf{w}_{p+1} , a po každém iteračním kroku odečteme od \mathbf{w}_{p+1} projekci $\mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j, j = 1, \dots, p$ předchozích odhadnutých p vektorů a normalizujeme \mathbf{w}_{p+1} :

1. Necht' $\mathbf{w}_{p+1} = \mathbf{w}_{p+1} - \sum_{j=1}^p \mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j$

2. Necht' $\mathbf{w}_{p+1} = \frac{\mathbf{w}_{p+1}}{\sqrt{\mathbf{w}_{p+1}^T \mathbf{w}_{p+1}}}$.

V určitých aplikacích bohužel může být vyžadováno použití symetrické dekorelace, ve které nejsou žádné vektory „upřednostňované“ před jinými.

Necht' $\mathbf{W} = (\mathbf{W}\mathbf{W}^T)^{-1/2} \mathbf{W}$,

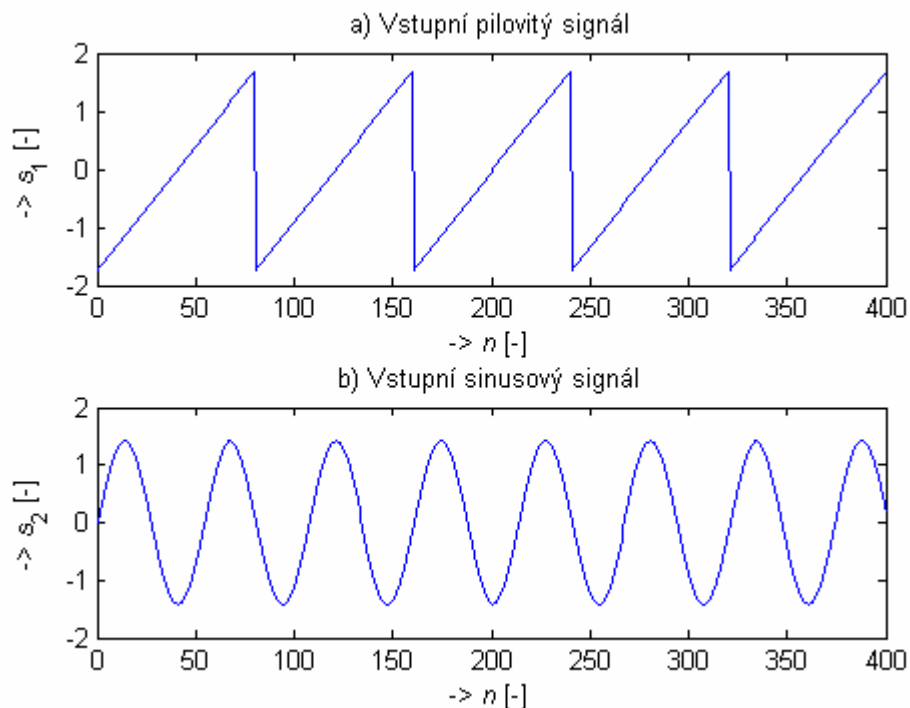
kde \mathbf{W} je matice $(\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ vektorů a inverzní čtvercová matice $(\mathbf{W}\mathbf{W}^T)^{-1/2}\mathbf{w}$ je získána z dekompozice vlastních hodnot $\mathbf{W}\mathbf{W}^T = \mathbf{F}\mathbf{D}\mathbf{F}^T$ jako $(\mathbf{W}\mathbf{W}^T)^{-1/2} = \mathbf{F}\mathbf{D}^{-1/2}\mathbf{F}^T$. Jednodušší variantou je následující iterační algoritmus:

1. Necht' $\mathbf{W} = \frac{\mathbf{W}}{\sqrt{\|\mathbf{W}\mathbf{W}^T\|}}$.
2. Necht' $\mathbf{W} = \frac{3}{2}\mathbf{W} - \frac{1}{2}\mathbf{W}\mathbf{W}^T\mathbf{W}$.
3. Opakovat krok 2 do konvergence.

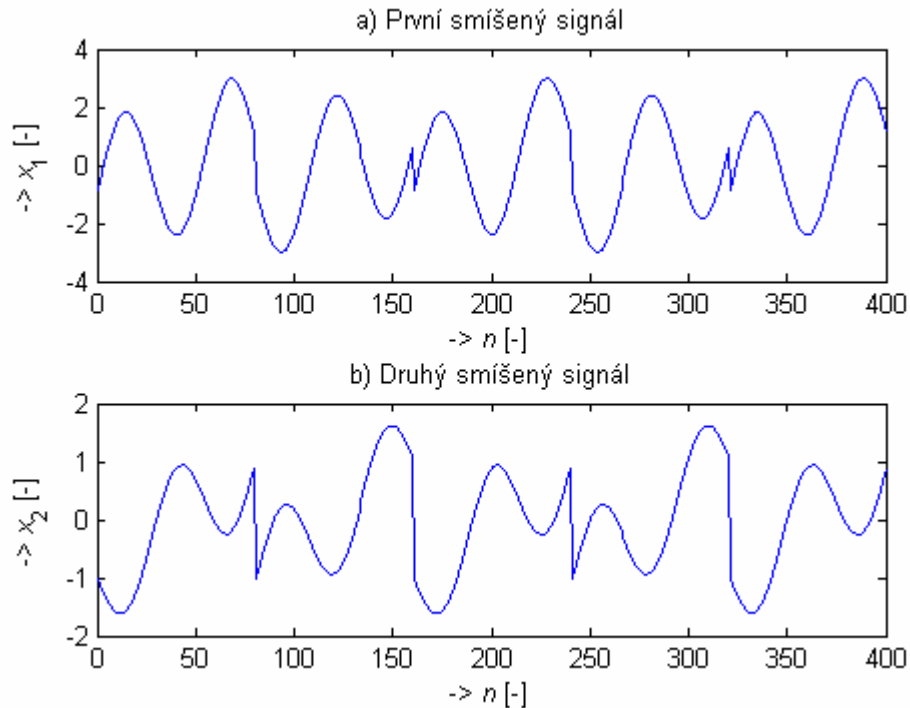
5.3 Separace signálů pomocí FastICA

Algoritmus FastICA pro oddělení signálů bude předveden na dvou příkladech. První příklad bude od sebe oddělovat směs dvou umělých signálů, druhý bude oddělovat směs dvou reálných signálů. V obou případech budou signály smíšeny uměle (v počítači) náhodnou směšovací maticí. Pro algoritmus FastICA je vytvořen v Matlabu programový balík The FastICA package for Matlab [6]. Pomocí tohoto programového balíku můžeme nastavovat nejrůznější parametry zpracování. Vybíráme především počet odhadovaných (separovaných) komponent, nelineární funkci g , bělení a stabilizaci.

Ukažme si práci algoritmu nejdříve na směsi dvou umělých signálů. Jedním signálem je pilovitý signál (v Matlabu zadaný příkazem sawtooth), druhým je sinusový signál (příkaz sin). Zdrojové signály jsou vykresleny na obr. 5.1. Tyto signály byly sečteny pomocí náhodně vygenerované čtvercové matice \mathbf{A} . Získali jsme tak dva smíšené signály, jejichž průběh je na obr. 5.2. V obou směsích jsou v různém poměru namíchány oba zdrojové signály z obr. 5.1.



Obr. 5.1: Průběhy zdrojových umělých signálů



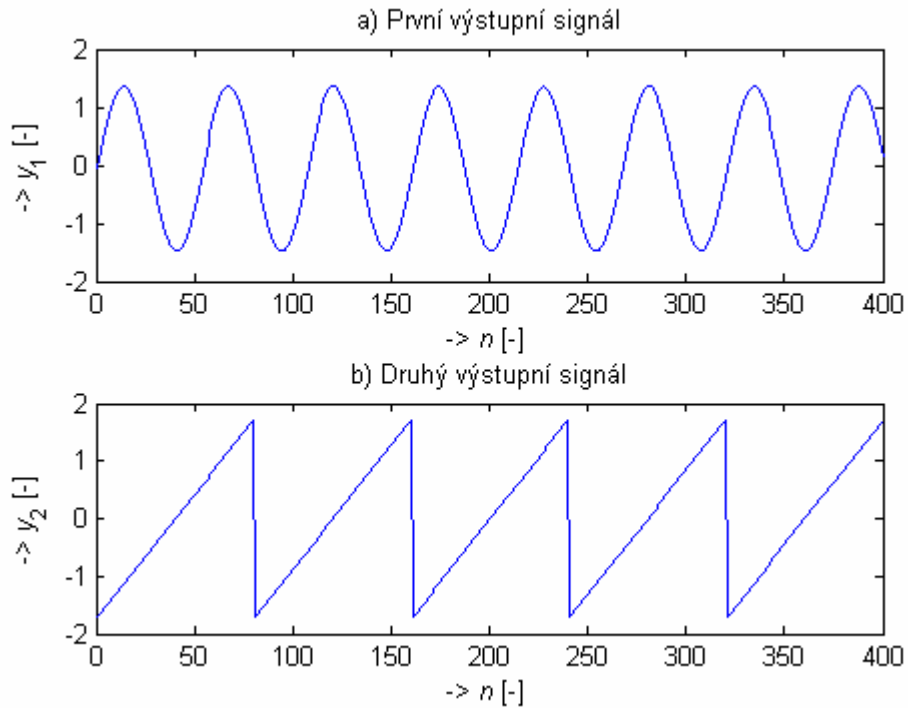
Obr. 5.2: Průběhy smíšených umělých signálů

Smíšené signály x se zadají jako vstupní parametry funkce `fastica` souboru `fastica.m` programového balíku `FastICA package`. Nastavení funkce pro prezentovaný příklad vypadá takto:

```
[y, A, W] = fastica(x, 'approach', 'defl', 'numOfIC', '2', 'g', 'tanh', 'stabilization', 'on');
```

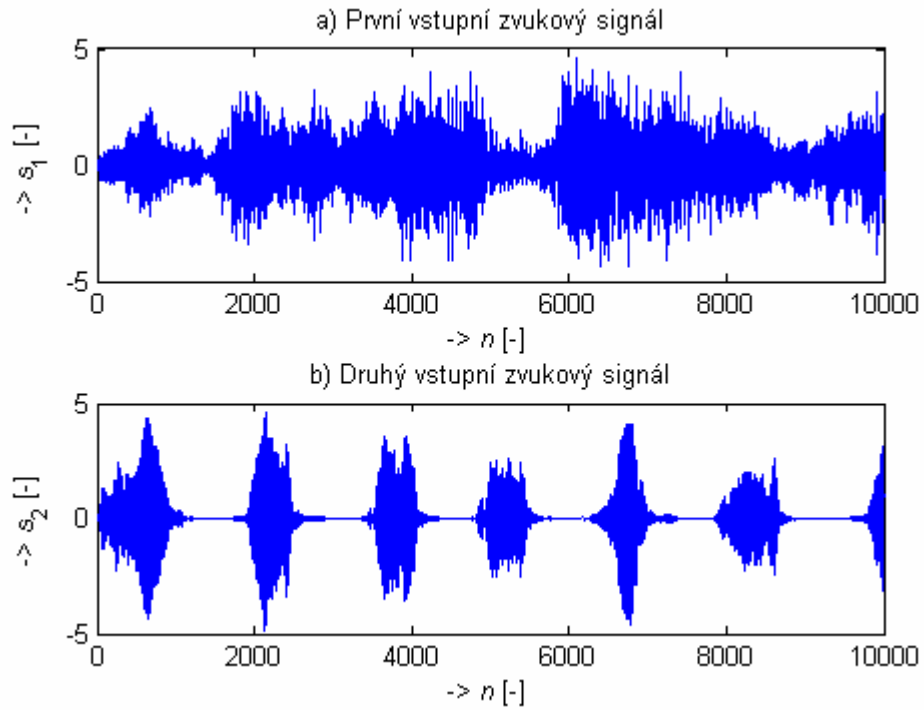
Volba `'approach'` na `'defl'` znamená, že výstupní komponenty y jsou separovány jedna po druhé, `'numOfIC'` s hodnotou `'2'` říká, že výstupem mají být dvě komponenty, nelineární funkce `'g'` je nastavena na `'tanh'`, tedy podle rovnice (5.1) a je zapnuta stabilizace pro iterační krok.

Oddělené signály (nezávislé komponenty) po provedení algoritmu `FastICA` jsou vidět na obr. 5.3. Porovnáním s průběhy zdrojových signálů na obr. 5.1 je vidět, že pro uvedené nastavení se podařilo umělé signály rekonstruovat téměř dokonale.

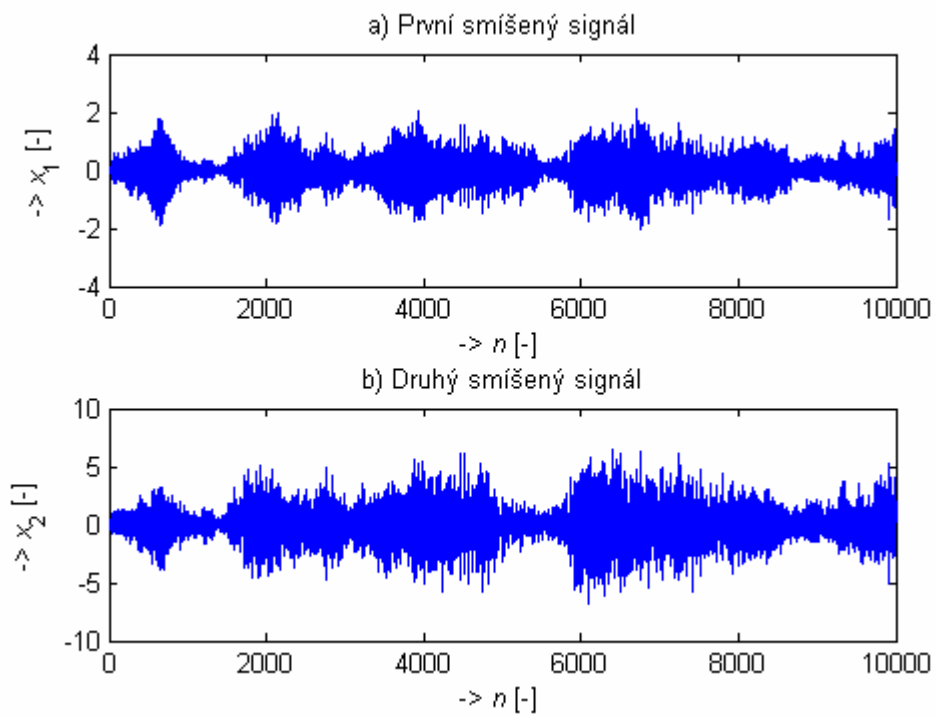


Obr. 5.3: Průběhy nezávislých komponent umělých signálů

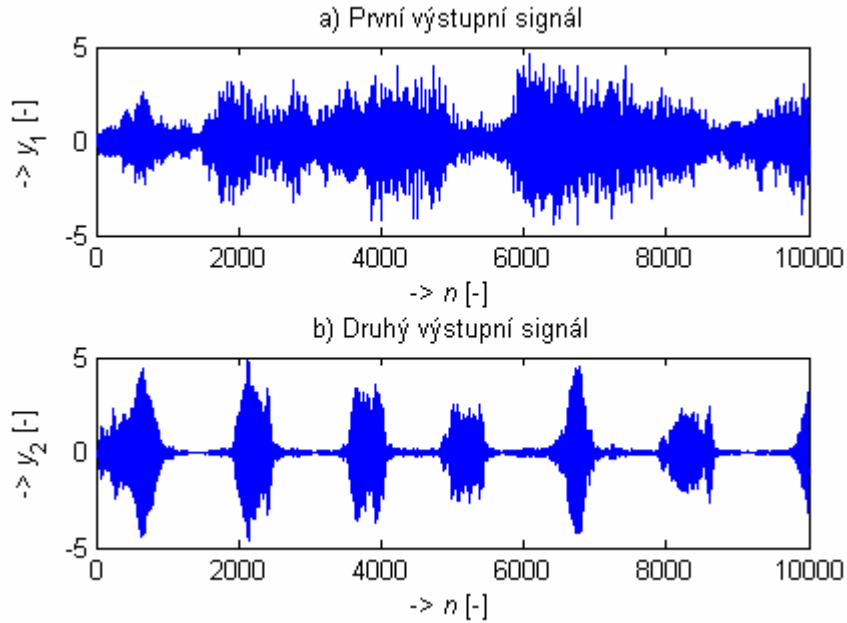
V praxi je potřeba od sebe oddělovat reálné signály - např. oddělit projev mluvčího od ostatních zvuků v místnosti (řeč jiných mluvčích, hluk) z nahrávek mikrofonů umístěných na různých místech. Druhá ukázka práce algoritmu FastICA je zaměřena na reálné zvukové signály. V počítači byly smíšeny dva zvukové signály, které můžete vidět na obr. 5.4. Jedná se o matlabovské signály laughter a chirp. Pokud provedeme opět separaci pomocí FastICA package, dostaneme nezávislé komponenty jako jsou na obr. 5.5. I v tomto případě došlo k dobré rekonstrukci původních zdrojových signálů, což se v Matlabu dá ověřit i poslechem. Smíšené signály jsou při poslechu velmi nezřetelnou směsí obou zvukových signálů, zatímco výstupní nezávislé komponenty jsou každá od druhého signálu téměř očištěna.



Obr. 5.4: Průběhy zdrojových reálných signálů



Obr. 5.5: Průběhy smíšených reálných signálů



Obr. 5.6: Průběhy nezávislých komponent reálných signálů

Příklad konvergoval po 5 krocích. Algoritmus FastICA v porovnání s jinými algoritmy ICA má velmi rychlou konvergenci. Nevybírá se velikost kroku, algoritmus je tak snadnější k použití. S využitím nelinearity g hledá přímo nezávislé komponenty a výběrem vhodné nelinearity (5.1) můžeme získat požadované vlastnosti. FastICA má mnoho výhod v neuronových algoritmech - je paralelní, distribuovaná, výpočetně jednoduchá a nemá velké nároky na paměť.

6 Algoritmus rostoucího gradientu

Analýza nezávislých komponent má v praxi různé varianty. Další takovou je algoritmus rostoucího gradientu (ICA gradient ascent). Uvažujme, že M neznámých zdrojových signálů $\mathbf{s} = (s_1, s_2, \dots, s_M)^T$ mají společnou distribuční funkci g a hustotu pravděpodobnosti p_s . Matice \mathbf{W} (inverzní ke směšovací matici \mathbf{A}) je matice, která odděluje M signálů $\mathbf{y} = (y_1, y_2, \dots, y_M)^T$ ze souboru nahraných směsí signálů \mathbf{x} , entropie signálů $\mathbf{Y} = g(\mathbf{y})$ je [5]:

$$H(\mathbf{Y}) = H(\mathbf{x}) + E \left[\sum_{i=1}^M \ln p_s(y_i) \right] + \ln |\mathbf{W}|, \quad (6.1)$$

kde $y_i = \mathbf{w}_i^T \mathbf{x}$ je i -tý signál, který je vyjmutý pomocí i -tého řádku matice \mathbf{W} . Tato hodnota bude vypočítána použitím N vzorkovaných hodnot směsí \mathbf{x} .

Hustota pravděpodobnosti p_s je derivací distribuční funkce g :

$$p_s(y) = \frac{dg(y)}{dy}, \text{ resp. } g'(y) = p_s(y). \quad (6.2)$$

Můžeme tak psát:

$$H(\mathbf{Y}) = H(\mathbf{x}) + E \left[\sum_{i=1}^M \ln g'(y_i) \right] + \ln |\mathbf{W}|. \quad (6.3)$$

Hledáme matici \mathbf{W} , která maximalizuje entropii $H(\mathbf{Y})$. Protože entropie $H(\mathbf{x})$ smíšených signálů \mathbf{x} je konstanta a nemá vliv na $H(\mathbf{Y})$, nebudeme ji uvažovat. Můžeme proto při hledání \mathbf{W} pokračovat maximalizací funkce:

$$h(\mathbf{Y}) = E \left[\sum_{i=1}^M \ln g'(y_i) \right] + \ln |\mathbf{W}|, \quad (6.4)$$

kteřá je změnou v entropii spojenou se zobrazením z \mathbf{x} do \mathbf{Y} . Můžeme najít optimální matici \mathbf{W}^* použitím rostoucího gradientu na h iteračním přizpůsobením \mathbf{W} za účelem maximalizace funkce h . Pro efektivní provedení rostoucího gradientu vyžadujeme výraz pro gradient h vzhledem k matici \mathbf{W} .

6.1 Algoritmus rostoucího gradientu pro obecný případ

Pokračujme hledáním parciální derivace h vzhledem k jednomu skalárnímu prvku \mathbf{W}_{ij} z \mathbf{W} , kde \mathbf{W}_{ij} je prvek i -tého řádku a j -tého sloupce \mathbf{W} . Váha \mathbf{W}_{ij} určuje poměr j -té směsi v i -tém extrahovaném signálu y_i .

Protože platí, že $\mathbf{y} = \mathbf{W}\mathbf{x}$ a každý zdrojový signál má tu samou hustotu pravděpodobnosti g' , je parciální derivace h vzhledem k ij -tému prvku \mathbf{W} :

$$\frac{\partial h}{\partial \mathbf{W}_{ij}} = E \left[\sum_{i=1}^M \frac{\partial \ln g'(y_i)}{\partial \mathbf{W}_{ij}} \right] + \frac{\partial \ln |\mathbf{W}|}{\partial \mathbf{W}_{ij}}. \quad (6.5)$$

Nyní rozebereme dvě derivace na pravé straně předchozí rovnice. Vnitřek sumy prvního členu pravé strany můžeme přepsat následujícím způsobem:

$$\frac{\partial \ln g'(y_i)}{\partial \mathbf{W}_{ij}} = \frac{1}{g'(y_i)} \frac{\partial g'(y_i)}{\partial \mathbf{W}_{ij}}. \quad (6.6)$$

A dále upravit takto:

$$\frac{\partial g'(y_i)}{\partial \mathbf{W}_{ij}} = \frac{dg'(y_i)}{dy_i} \frac{\partial y_i}{\partial \mathbf{W}_{ij}}. \quad (6.7)$$

Derivace na pravé straně může být přepsána takto:

$$\frac{dg'(y_i)}{dy_i} = g''(y_i), \quad (6.8)$$

kde $g''(y_i)$ je druhá derivace g podle y_i , dále:

$$\frac{\partial y_i}{\partial \mathbf{W}_{ij}} = x_j. \quad (6.9)$$

Dosazení rovnic (6.8) a (6.9) do (6.7) dá rovnici:

$$\frac{\partial g'(y_i)}{\partial \mathbf{W}_{ij}} = g''(y_i)x_j. \quad (6.10)$$

Rovnici (6.10) dosadíme do (6.6):

$$\frac{\partial \ln g'(y_i)}{\partial \mathbf{W}_{ij}} = \frac{1}{g'(y_i)} g''(y_i)x_j. \quad (6.11)$$

Výraz v rovnici (6.11) pak můžeme vložit do sumy prvního členu na pravé straně rovnice (6.5), zavést označení:

$$\psi(y_i) = \frac{g''(y_i)}{g'(y_i)} \quad (6.12)$$

a psát tak:

$$E \left[\sum_{i=1}^M \frac{\partial \ln g'(y_i)}{\partial \mathbf{W}_{ij}} \right] = E \left[\sum_{i=1}^M \frac{g''(y_i)}{g'(y_i)} x_j \right] = E \left[\sum_{i=1}^M \psi(y_i) x_j \right]. \quad (6.13)$$

Pokud se podíváme na druhý člen pravé strany rovnice (6.5), pak platí [5]:

$$\frac{\partial \ln |\mathbf{W}|}{\partial \mathbf{W}_{ij}} = [\mathbf{W}^{-T}]_{ij}, \quad (6.14)$$

kde \mathbf{W}^{-T} je inverzní transponovaná matice \mathbf{W} : $\mathbf{W}^{-T} = [\mathbf{W}^T]^{-1}$.

Tedy $[\mathbf{W}^{-T}]_{ij}$ je ij -tý prvek inverzní transponované matice \mathbf{W} . Sloučením rovnic (6.14) a (6.13) do vztahu (6.5) vznikne:

$$\frac{\partial h}{\partial \mathbf{W}_{ij}} = [\mathbf{W}^{-T}]_{ij} + E \left[\sum_{i=1}^M \psi(y_i) x_j \right]. \quad (6.15)$$

Pokud uvažujeme všechny prvky \mathbf{W} , pak máme:

$$\nabla h = \mathbf{W}^{-T} + E[\psi(\mathbf{y})\mathbf{x}^T], \quad (6.16)$$

kde ∇h (nabla - grad h) je $M \times M$ matice (Jakobián) derivací, v kterých ij -tý prvek je $\partial h / \partial \mathbf{W}_{ij}$.

Máme-li konečný počet vzorků N zachycených směsí \mathbf{x}^t pro $t = 1, \dots, N$ a předpokládanou matici \mathbf{W} , předpoklad $E[\cdot]$ může být odhadnut jako střední hodnota:

$$E[\psi(\mathbf{y})\mathbf{x}^T] = \frac{1}{N} \sum_{t=1}^N \psi(\mathbf{y}^t) [\mathbf{x}^t]^T, \quad (6.17)$$

kde $\mathbf{y}^t = \mathbf{W}\mathbf{x}^t$.

Pravidlo rostoucího gradientu je tedy:

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \eta \nabla h, \quad (6.18)$$

resp.:

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \eta (\mathbf{W}^{-T} + \frac{1}{N} \sum_{t=1}^N \psi(\mathbf{y}^t) [\mathbf{x}^t]^T), \quad (6.19)$$

kde η je malá konstanta. Toto pravidlo maximalizuje entropii signálů $\mathbf{Y} = g(\mathbf{y})$.

6.2 Algoritmus rostoucího gradientu pro supergaussovské signály

Nyní budeme derivovat výraz pro ∇h pro specifickou distribuční funkci zdrojových signálů. Obvykle používanou distribuční funkci pro extrakci supergaussovských signálů je funkce tgh. Vezmeme-li distribuční funkci:

$$g(\mathbf{y}^t) = \text{tgh}(\mathbf{y}^t), \quad (6.20)$$

tak hustota pravděpodobnosti g' je dána první derivací tgh:

$$g'(\mathbf{y}^t) = 1 - \text{tgh}^2(\mathbf{y}^t) \quad (6.21)$$

a druhá derivace tgh je:

$$g''(\mathbf{y}^t) = -2\text{tgh}(\mathbf{y}^t)g'(\mathbf{y}^t), \quad (6.22)$$

a tak:

$$\psi(\mathbf{y}^t) = \frac{g''(\mathbf{y}^t)}{g'(\mathbf{y}^t)} = -2\text{tgh}(\mathbf{y}^t). \quad (6.23)$$

Vložením rovnice (6.23) do (6.16) je:

$$\nabla h = \mathbf{W}^{-T} + E[-2\text{tgh}(\mathbf{y}^t)[\mathbf{x}^t]^T]. \quad (6.24)$$

Máme-li konečný počet vzorků N zachycených směsí \mathbf{x}^t pro $t = 1, \dots, N$ a předpokládanou matici \mathbf{W} ($\mathbf{y} = \mathbf{W}\mathbf{x}$), předpoklad $E[.]$ může být odhadnut jako střední hodnota:

$$E[-2\text{tgh}(\mathbf{y})\mathbf{x}^T] = \frac{-2}{N} \sum_{t=1}^N \text{tgh}(\mathbf{y}^t)[\mathbf{x}^t]^T. \quad (6.25)$$

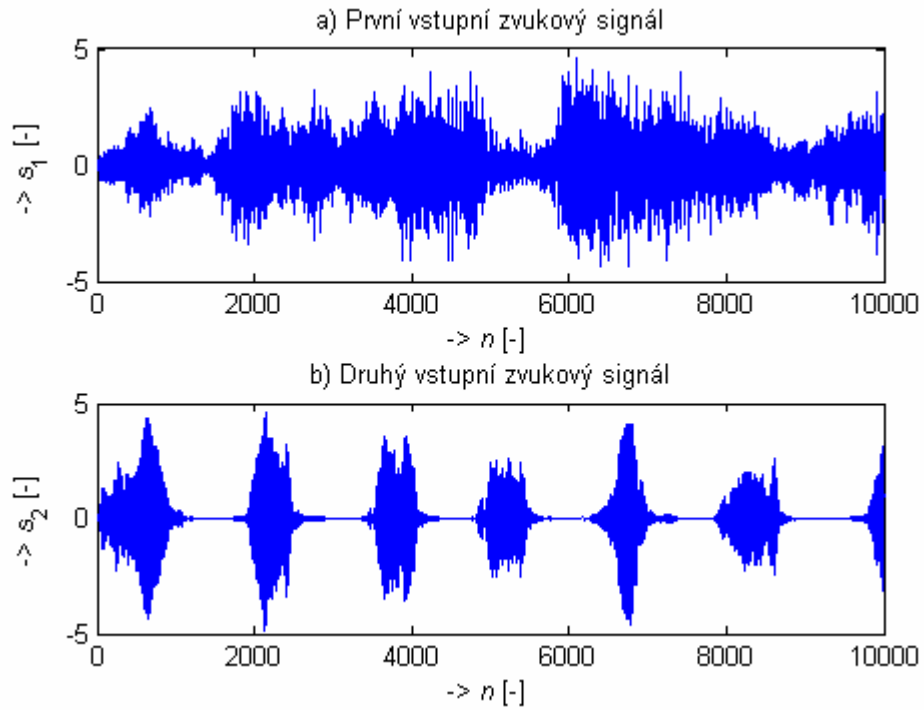
Pravidlo pro zpřesňování \mathbf{W} za účelem maximalizace entropie signálů $\mathbf{Y} = g(\mathbf{y})$ je tedy:

$$\mathbf{W}_{new} = \mathbf{W}_{old} + \eta(\mathbf{W}^{-T} - \frac{2}{N} \sum_{t=1}^N \text{tgh}(\mathbf{y}^t)[\mathbf{x}^t]^T), \quad (6.26)$$

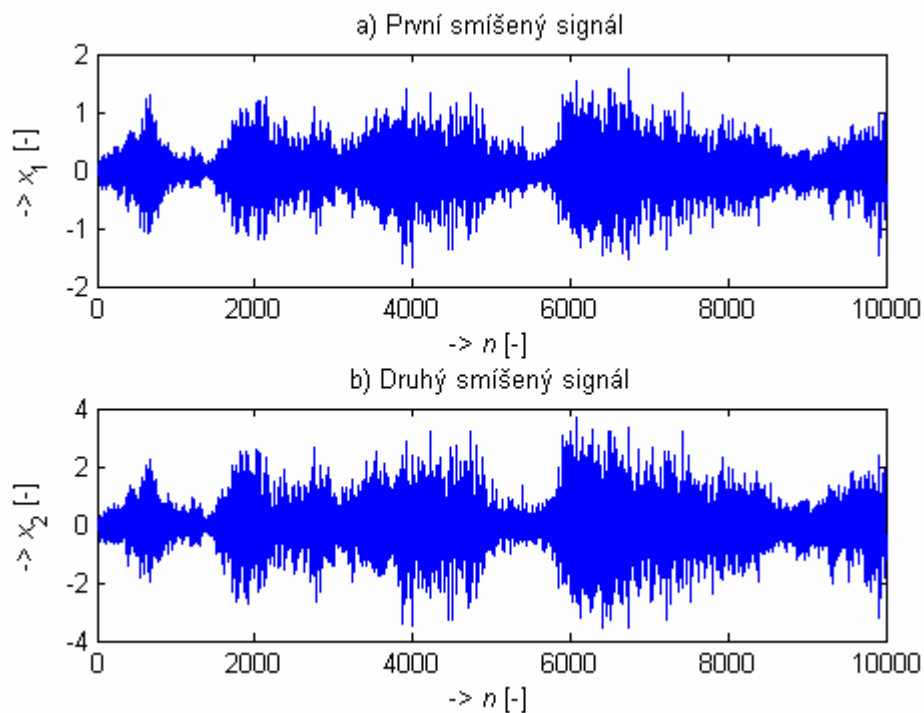
kde η je malá konstanta.

6.3 Separace signálů pomocí algoritmu rostoucího gradientu

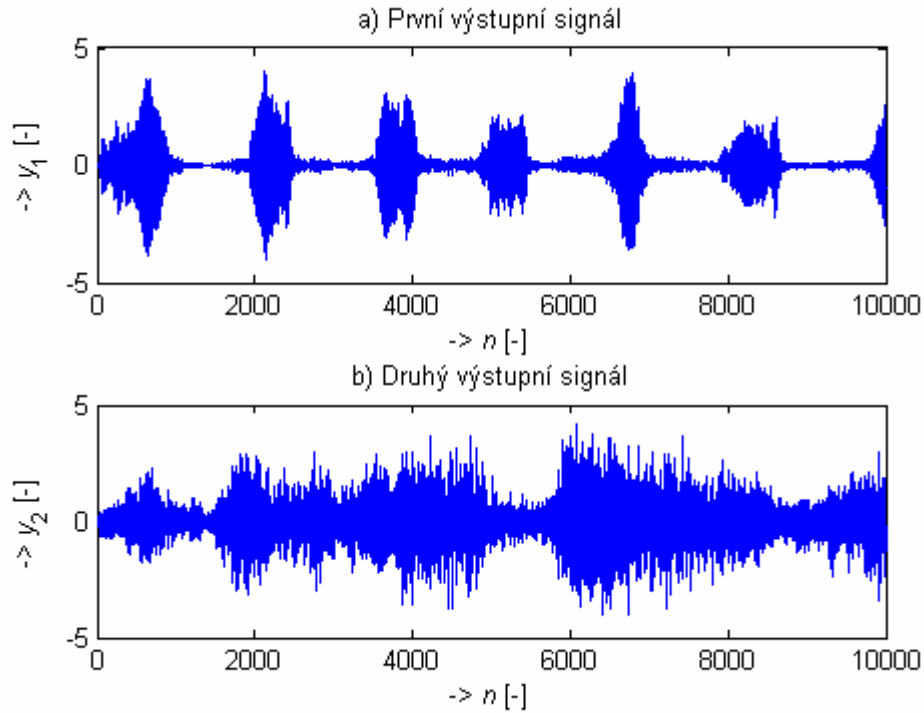
V této podkapitole se podíváme na konkrétní použití algoritmu rostoucího gradientu na oddělování dílčích složek smíšeného signálu. Jako testovací signály budou použity opět reálné zvukové signály (smíšené náhodnou směšovací maticí) z části 5.3, kde jsme je zpracovávali pomocí algoritmu FastICA. Zdrojové signály jsou vidět na obr. 6.1, smíšené signály na obr. 6.2.



Obr. 6.1: Průběhy zdrojových reálných signálů



Obr. 6.2: Průběhy smíšených reálných signálů



Obr. 6.3: Průběhy nezávislých komponent reálných signálů

Algoritmus rostoucího gradientu se v Matlabu realizuje výpočty podle rovnic uvedených v podkapitole 6.2. V tomto algoritmu se volí parametry: η (iterační krok), počet iterací, distribuční funkce. Maximální počet iterací byl nastaven na 100, distribuční funkce byla tgh. Výsledné nezávislé komponenty jsou na obr. 6.3. Z obr. 6.3 je porovnáním se zdrojovými signály na obr. 6.1 vidět, že se signály podařilo obnovit s poměrně dobrou úspěšností. Počet iterací ve srovnání s FastICA je větší, algoritmus tak trvá déle. Z uvedených grafů také můžeme vidět další vlastnost analýzy nezávislých komponent: nelze určit pořadí výstupních nezávislých komponent.

7 Separace konvolutorních směsí řečových signálů

Přístupy extrakce zdrojových signálů prezentované v předchozích kapitolách nejsou příliš efektivní pro použití v reálných aplikacích. Bohužel jsou limitovány na použití v situacích, ve kterých jsou signály smíšené bez časových zpoždění. Tato kapitola představuje rozšíření ICA z předchozích kapitol - její konvoluční verzi, která umožňuje separaci signálů smíšených i s časovými zpožděním.

V reálném nahrávacím prostředí se signály zachycené mikrofony skládají nejen z přímých signálů, ale i z jejich zpožděných (odražených) a tlumených verzí a z šumu. Řečový signál složený z M prvků lineárního mikrofonního pole proto může být modelován jako lineární konvolutorní směs R zdrojových signálů s_i tak, že M -rozměrný obdržený signál $x(n)=[x_1(n), x_2(n), \dots, x_M(n)]^T$ je dán [8]:

$$x_j(n) = \sum_{i=1}^R \sum_{p=1}^P h_{ji}(p) s_i(n-p+1) \quad (7.1)$$

pro $j = 1, 2, \dots, M$, kde $s(n)=[s_1(n), s_2(n), \dots, s_R(n)]^T$ reprezentuje originální zdrojový signál, h_{ji} je impulsní odezva délky P mezi zdrojem i a mikrofonem j . V této práci je uvažován případ, kdy máme dva mikrofony a dva zdroje, tj. $M = R = 2$. Podle toho jsou zachycené signály:

$$\begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \otimes \begin{bmatrix} s_1(n) \\ s_2(n) \end{bmatrix} = \begin{bmatrix} ref_{11} + ref_{12} \\ ref_{21} + ref_{22} \end{bmatrix}, \quad (7.2)$$

kde $ref_{11} = h_{11} \otimes s_1(n)$, $ref_{12} = h_{12} \otimes s_2(n)$, $ref_{21} = h_{21} \otimes s_1(n)$, $ref_{22} = h_{22} \otimes s_2(n)$ jsou tzv. referenční signály a \otimes znamená operaci konvoluce (odtud označení vzniklých signálů jako konvolutorní), která je obecně definována:

$$x(n) = \sum_{i=-\infty}^{\infty} h(i) s(n-i). \quad (7.3)$$

Ve frekvenční oblasti je stejný model (7.2) reprezentován pomocí krátkodobé Fourierovy transformace rovnice (7.1) a model v rovnici (7.2) může být vyjádřen:

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = H(f) S(f) = \begin{bmatrix} H_{11}(f) & H_{12}(f) \\ H_{21}(f) & H_{22}(f) \end{bmatrix} \cdot \begin{bmatrix} S_1(f) \\ S_2(f) \end{bmatrix}, \quad (7.4)$$

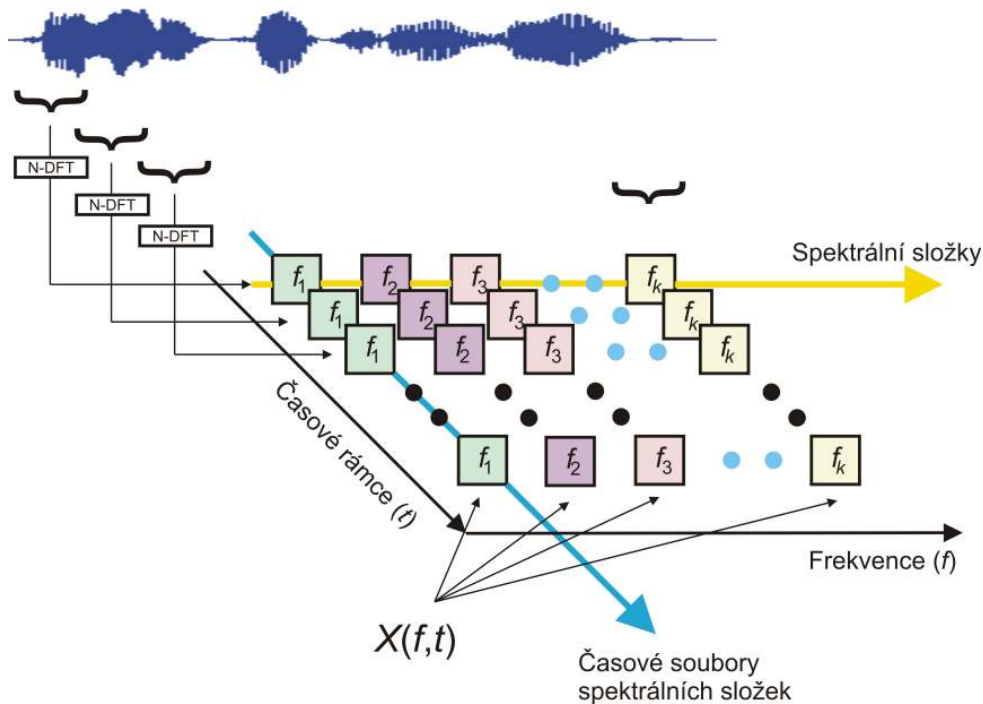
kde symboly velkými písmeny značí Fourierovu transformaci členů zapsaných v rovnici (7.2) malými písmeny. Analýza nezávislých komponent ve frekvenční oblasti odděluje signály na každé frekvenci nezávisle, přičemž separační proces je zapsán:

$$\begin{bmatrix} \hat{S}_1(f) \\ \hat{S}_2(f) \end{bmatrix} = \begin{bmatrix} Y_1(f) \\ Y_2(f) \end{bmatrix} = W(f) X(f) = \begin{bmatrix} W_{11}(f) & W_{12}(f) \\ W_{21}(f) & W_{22}(f) \end{bmatrix} \cdot \begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix}, \quad (7.5)$$

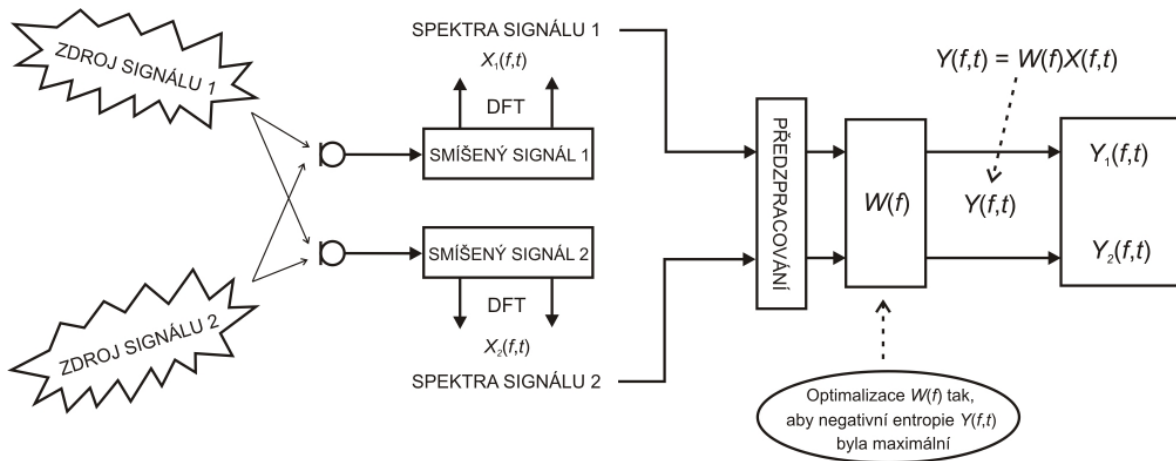
kde $[Y_1(f), Y_2(f)]^T$ jsou nezávislé komponenty a $W(f)$ je separační matice na frekvenci f .

7.1 Algoritmus Fixed-point ICA ve frekvenční oblasti

Algoritmus Fixed-point ICA ve frekvenční oblasti pracuje na časově-frekvenčních souborech řeči (TFSS) smíšených řečových dat s cílem získat časově-frekvenční soubory řeči nezávislých komponent na každé frekvenci. Celý proces vytváření časově-frekvenčních souborů pomocí analýzy krátkodobou Fourierovou transformací diskrétního signálu je naznačen na obr. 7.1. Časově-frekvenční soubory sestávají z řečových spektrálních komponent o stejné frekvenci ze všech analyzovaných rámců v časovém pořadí. Základní činnost fixed-point ICA ve frekvenční oblasti je ukázána na obr. 7.2. Algoritmus je založen na heuristickém předpokladu, že když se smísí negaussovské signály, vzniknou více gaussovské signály, a tak měření jejich negaussovských vlastností může poskytnout nezávislé komponenty. Směšovací model ve frekvenční oblasti pro řečový signál v rovnici (7.4) ukazuje, že časově-frekvenční soubor na jakékoli frekvenci je superpozicí spektrálních příspěvků každého zdroje. Časově-frekvenční soubor smíšených řečových signálů na jakékoli frekvenci je tedy více gaussovský než jakýkoli nezávislý zdroj.



Obr. 7.1: Proces vytváření časově-frekvenčních sérií řeči



Obr. 7.2: Funkce Fixed-point ICA algoritmu ve frekvenční oblasti

Proces měření negaussovských vlastností se skládá ze dvou kroků - předbělení a rotace pozorovaného vektoru. Předbělení je polovinou ICA úlohy a dává prostorově dekorelované signály. Bělení časově-frekvenčního souboru s nulovou střední hodnoty je provedeno podle Mahalanobisovy transformace. Vybělený signál $X_w(f,t)$ na frekvenci f je získán [9]:

$$\mathbf{X}_w(f,t) = Q(f)\mathbf{X}(f,t), \quad (7.6)$$

kde $Q(f) = \Lambda_x^{-0.5}V_x$ je tzv. bělicí matice. $\Lambda_x = \text{diag}\{1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, \dots, 1/\sqrt{\lambda_n}\}$ je diagonální matice s kladnými vlastními hodnotami $\lambda_1 > \lambda_2 > \dots > \lambda_n$ kovarianční matice $X(f_p,t)$ a V_x je ortogonální matice složená z vlastních vektorů.

Postup následující po bělení zahrnuje rotaci vyběleného vektoru $\mathbf{X}_w(f,t)$ separační maticí tak, že $\mathbf{Y}(f) = W(f)\mathbf{X}_w(f,t)$ se rovná nezávislým komponentám. Výpočetní funkce může být založena na různých mírách negaussovských vlastností - např. špičatosti či negativní entropii. Negativní entropie $J(\mathbf{Y})$ časově-frekvenčního souboru možné nezávislé komponenty $\mathbf{Y}(f,t)$ je dána vztahem:

$$J(\mathbf{Y}) = H(\mathbf{Y}_{\text{gauss}}) - H(\mathbf{Y}), \quad (7.7)$$

kde H je diferenciální entropie a $\mathbf{Y}_{\text{gauss}}$ je gaussovská náhodná proměnná se stejnou kovariancí jako \mathbf{Y} . Tato definice zajistí, že negativní entropie J bude nulová, když $\mathbf{Y}(f,t)$ je gaussovská, nebo bude rostoucí, když $\mathbf{Y}(f,t)$ směřuje k negaussovosti. Takže kontrastní funkce založená na negativní entropii může být maximalizována k získání optimální negaussovské komponenty.

Zbývá odvodit učící pravidlo pro každý řádek separační matice \mathbf{W} . Negativní entropie může být aproximována nekvdatickou nelineární funkcí G takto:

$$J(y) = \sigma[E\{G(y)\} - E\{G(y_{\text{gauss}})\}]^2, \quad (7.8)$$

kde σ je kladná konstanta. Provedení algoritmu závisí na použité nekvadratické nelineární funkci G . Výběr funkce G závisí na funkci rozdělení pravděpodobnosti dat. Některé funkce používané pro separaci signálů jsou:

$$\begin{aligned} G_1(Y) &= \sqrt{a_1 + Y}; a_1 = 0,01 \\ G_2(Y) &= \log(a_2 + Y); a_2 = 0,01 \\ G_3(Y) &= \frac{Y}{|Y|}; \forall Y \neq 0. \end{aligned} \quad (7.9)$$

Algoritmus pro učení separační matice $\mathbf{W}(f)$ je získán maximalizací negativní entropie kontrastní funkce. Použitím kontrastní funkce můžeme dále napsat:

$$J(\mathbf{Y}) = E\left\{G(|\mathbf{w}^H \mathbf{X}_w|^2)\right\}, \quad (7.10)$$

kde \mathbf{w} je M -rozměrný komplexní vektor takový, že

$$E\left\{G(|\mathbf{w}^H \mathbf{X}_w|^2)\right\} = 1 \Rightarrow |\mathbf{w}| = 1. \quad (7.11)$$

Tato kontrastní funkce může mít M lokálních nebo globálních optimálních řešení \mathbf{w}_i ($i = 1, 2, \dots, M$) pro každý zdroj. Takže učení každé \mathbf{w} volá po maximalizaci (7.10) za podmínky dané (7.11). Maximum $J(\mathbf{Y})$ může být nalezeno řešením Lagrangeovy funkce $L(\mathbf{w}, \mathbf{w}^H, \lambda)$ výše uvedeného:

$$L(\mathbf{w}, \mathbf{w}^H, \lambda) = E\left\{G(|\mathbf{w}^H \mathbf{X}_w|^2)\right\} \pm \lambda [E\{\mathbf{w}^H \mathbf{X}_w\} - 1], \quad (7.12)$$

kde λ je Lagrangeův koeficient. Za účelem nalezení maxima kontrastní funkce musí být řešeny následující rovnice:

$$\frac{\partial L}{\partial \mathbf{w}} = 0; \quad \frac{\partial L}{\partial \mathbf{w}^H} = 0; \quad \frac{\partial L}{\partial \lambda} = 0 \quad (7.13)$$

Tyto rovnice mohou být získány z rovnice (7.10) takto:

$$\frac{\partial L}{\partial \mathbf{w}} = E\left\{g(|\mathbf{w}^H \mathbf{X}_w|^2) \mathbf{w}^H\right\} + \lambda \mathbf{w}^H = 0, \quad (7.14)$$

$$\frac{\partial L}{\partial \mathbf{w}^H} = E\left\{g(|\mathbf{w}^H \mathbf{X}_w|^2) \mathbf{X}_w^H \mathbf{w}\right\} + \lambda \mathbf{w} = 0, \quad (7.15)$$

$$\frac{\partial L}{\partial \lambda} = |\mathbf{w}|^2 - 1 = 0. \quad (7.16)$$

Řešení užitím Newtonovy iterační metody je dáno:

$$\mathbf{w}_{new} = \mathbf{w} - \left[\frac{\partial L}{\partial \mathbf{w}^H} \right] \left[\frac{\partial L}{\partial \mathbf{w}} \left(\frac{\partial L}{\partial \mathbf{w}^H} \right) \right]^{-1}, \quad (7.17)$$

$$\mathbf{w}_{new} = \mathbf{w} \left(E \left\{ g'(|\mathbf{w}^H \mathbf{X}_w|^2) \right\} \right) + (|\mathbf{w}^H \mathbf{X}_w|^2) g''(|\mathbf{w}^H \mathbf{X}_w|^2) - E \left\{ g'(|\mathbf{w}^H \mathbf{X}_w|^2) (\mathbf{X}_w^H \mathbf{w}) \mathbf{X}_w \right\} \quad (7.18)$$

Kritérium ukončení iterace je definováno číslem $\delta = (|\mathbf{w}_{old} - \mathbf{w}_{new}|)^2$, které bude velmi malé při konvergenci. Po každé iteraci je \mathbf{w} normováno.

Nezávislé zdroje jsou extrahovány v klesajícím pořadí negativní entropie ze smíšeného signálu. Po každé iteraci je také nezbytné dekorelovat \mathbf{w} , aby nekonvergovalo do předešlého bodu. K uskutečnění tohoto se dá využít Gram-Schmidtovy ortogonalizace, v které jsou odečteny komponenty všech předchozích získaných separačních vektorů spadající do směru aktuálního vektoru. Ortogonalizovaný separační vektor \mathbf{w}_i pro i -tý zdroj po j -té iteraci je dán:

$$\mathbf{w}_i = \mathbf{w}_i - \sum_{j=1}^{i-1} (\mathbf{w}_i^T \mathbf{w}_j) \mathbf{w}_j. \quad (7.19)$$

Aktualizační rovnice (7.18) je použita k odhadu separačního vektoru na každé frekvenci z vybělených časově-frekvenčních souborů smíšeného signálu pro každý zdroj a separační matice $\mathbf{W}(f)$ na frekvenci f je dána:

$$\mathbf{W}(f) = \begin{bmatrix} \mathbf{w}_1 \\ \cdot \\ \mathbf{w}_R \end{bmatrix} = \begin{bmatrix} W_{11}(f) & \cdots & W_{1M}(f) \\ \cdot & \cdots & \cdot \\ W_{R1}(f) & \cdots & W_{RM}(f) \end{bmatrix}. \quad (7.20)$$

Každý řádek této separační matice jednoznačně odpovídá separačnímu vektoru \mathbf{w} pro každý zdroj. Protože separační matice byla odvozena použitím vybělených signálů, její přednásobení s vybělenými signály ve frekvenční oblasti dává časově-frekvenční soubory $\mathbf{Y}(f,t) = [\mathbf{Y}_1(f,t), \mathbf{Y}_2(f,t), \dots, \mathbf{Y}_R(f,t)]^T$ separovaného signálu, tj.:

$$\hat{\mathbf{S}}(f,t) = \mathbf{Y}(f,t) = \mathbf{W}(f) \mathbf{X}_w(f,t). \quad (7.21)$$

7.2 Permutační a měřítkový problém

Pro úspěšné získání separovaných signálů musí být pořadí separačních vektorů (pozice řádků) ve $\mathbf{W}(f)$ stejné na každé frekvenci. Algoritmus separuje originální zdroje v klesajícím pořadí negativní entropie. Ale pořadí negativní entropie pro nezávislé zdroje nezůstává stejné na všech frekvencích, což postupně vede ke změně řádků $\mathbf{W}(f)$ v neznámém pořadí. Tento jev je nazýván jako permutační problém. Jiný problém je spojen s odlišným zesílením na každé frekvenci, které by pro přesnou rekonstrukci signálu mělo být stejné. To je nazýváno jako měřítkový problém. Pokud tyto dva problémy nejsou

řešeny, rovnice (7.21) dá jiné smíšené signály namísto separovaných komponent. Pro řešení těchto problémů bylo navrženo několik řešení. Jeden z postupů je směrový diagram (SD) vyžadující znalost směru přístupu (SP) každého zdroje. V naprosto slepém nastavení to nemůže být známé, takže je odhadnuto ze směrového diagramu separační matice. Směrový diagram $F_R(f, \theta)$ mikrofonového pole v R -tém zdroji je [9]:

$$F_R(f, \theta) = \sum_{k=1}^M W_{Rk}^{(ICA)}(f) \exp[j2\pi d_k \sin \theta / c], \quad (7.22)$$

kde $W_{Rk}^{(ICA)}$ je členem separační matice obdržené v rovnici (7.19), $R=1,2$. Směrový diagram separační matice obsahuje nuly na každém zdrojovém směru. Pozice nul se mění na každé frekvenci pro stejný zdrojový směr. Směr přístupu R -tého zdroje může být odhadnut jako:

$$\hat{\theta}_R = \frac{2}{N} \sum_{p=1}^{N/2} \theta_R(f_p), \quad (7.23)$$

kde $\theta_R(f_p)$ označuje směr nul na p -té frekvenci. Pro případ dvou zdrojů jsou dány rovnice:

$$\begin{aligned} \theta_1(f_p) &= \min \left[\arg \cdot \min_{\theta} |F_1(f_p, \theta)|, \arg \cdot \min_{\theta} |F_2(f_p, \theta)| \right], \\ \theta_2(f_p) &= \max \left[\arg \cdot \min_{\theta} |F_1(f_p, \theta)|, \arg \cdot \min_{\theta} |F_2(f_p, \theta)| \right], \end{aligned} \quad (7.24)$$

kde $\min[u, v]$ a $\max[u, v]$ jsou definovány k výběru minima a maxima. Pak separační matice na každé frekvenci je uspořádána v souladu se směry nulových hodnot, což vyřeší permutační problém. Po odhadu SP je zesílení na každé frekvenci normováno na každém zdrojovém směru. Zesílení na R -tém zdrojovém směru na p -té frekvenci je dáno:

$$\alpha_R(f_p) = \frac{1}{F_R(f_p, \hat{\theta}_R)}, \quad (7.25)$$

kde $\hat{\theta}_R$ je odhadovaný směr R -tého zdroje. Upravená separační matice má potom tvar:

$$\mathbf{W}(f_p) = \begin{bmatrix} \alpha_1(f_p) & 0 \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 \cdots & \alpha_R(f_p) \end{bmatrix} \begin{bmatrix} W_{11}(f_p) & \cdots & W_{1M}(f_p) \\ \cdot & \ddots & \cdot \\ W_{M1}(f_p) & \cdots & W_{RM}(f_p) \end{bmatrix}. \quad (7.26)$$

Tato matice je použita k separaci signálů na každé frekvenci. Užitím techniky overlap-add je signál v časové oblasti rekonstruován z časově-frekvenčního souboru každého zdroje. Pro použití $\mathbf{W}(f)$ z rovnice (7.20) v časové oblasti je dobré provést ještě opak bělení:

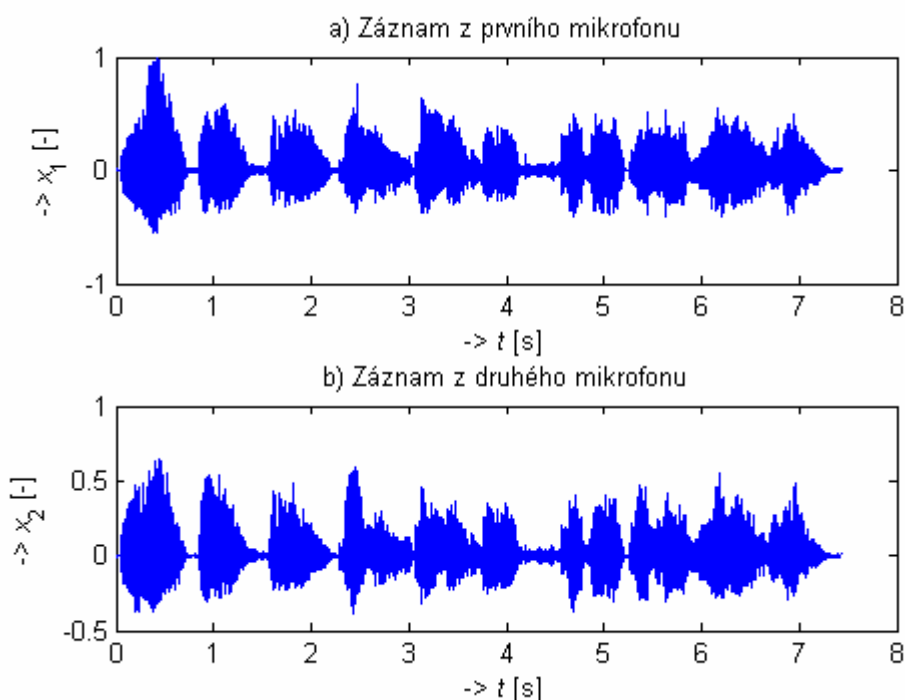
$$\mathbf{W}(f) = \mathbf{W}(f)(\mathbf{Q}(f))^{-1}. \quad (7.27)$$

Pak použitím $\mathbf{W}(f)$ může být odvozen filtr FIR délky P k separaci signálů přímo v časové oblasti takto:

$$y(n) = \sum_{r=0}^P w(r)x(n-r). \quad (7.28)$$

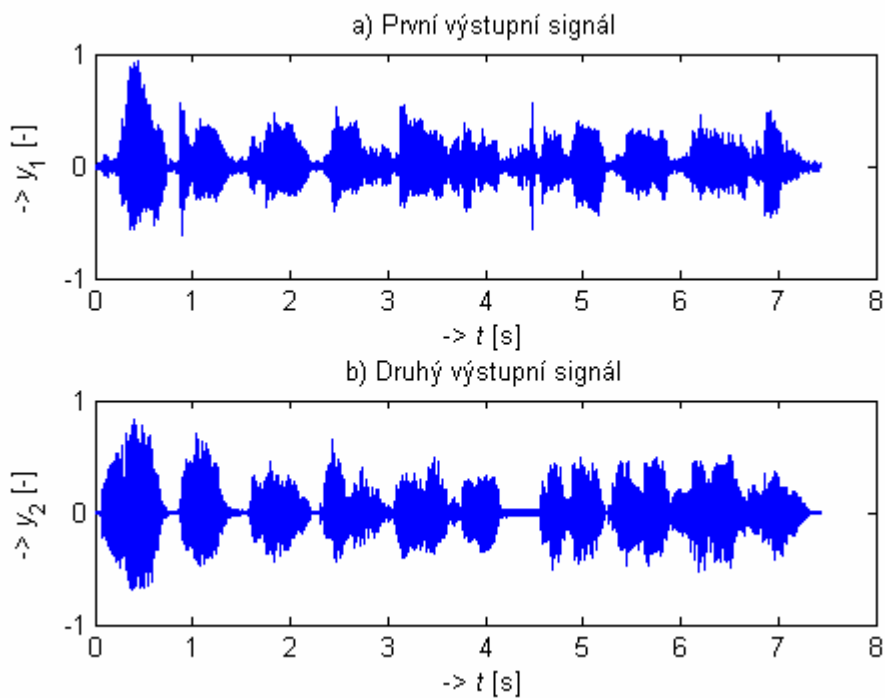
7.3 Separace signálů nahraných v reálném prostředí

Algoritmus Fixed-point ICA ve frekvenční oblasti navržený v předchozích podkapitolách demonstrujeme na nahrávkách pořízených v reálném prostředí pomocí dvou mikrofonů. Separaci provádí matlabovský toolbox BSS for Speech z japonského Nara Institutu pro vědu a techniku. Pomocí tohoto programu lze vytvářet i vlastní nahrávky ze dvou mikrofonů. Vybírat a měnit se dají např. parametry jako je počet vzorků DFT, velikost rámců signálu, nelineární funkce G , okénková funkce a další. Zvukové ukázky vstupních smíšených signálů i výsledky experimentů jsou k poslechu na přiloženém CD. Nejdříve byl tento postup vyzkoušen na směsi dvou řečových signálů (separace typu řeč-řeč). Zaznamenané signály jsou na obr. 7.3.



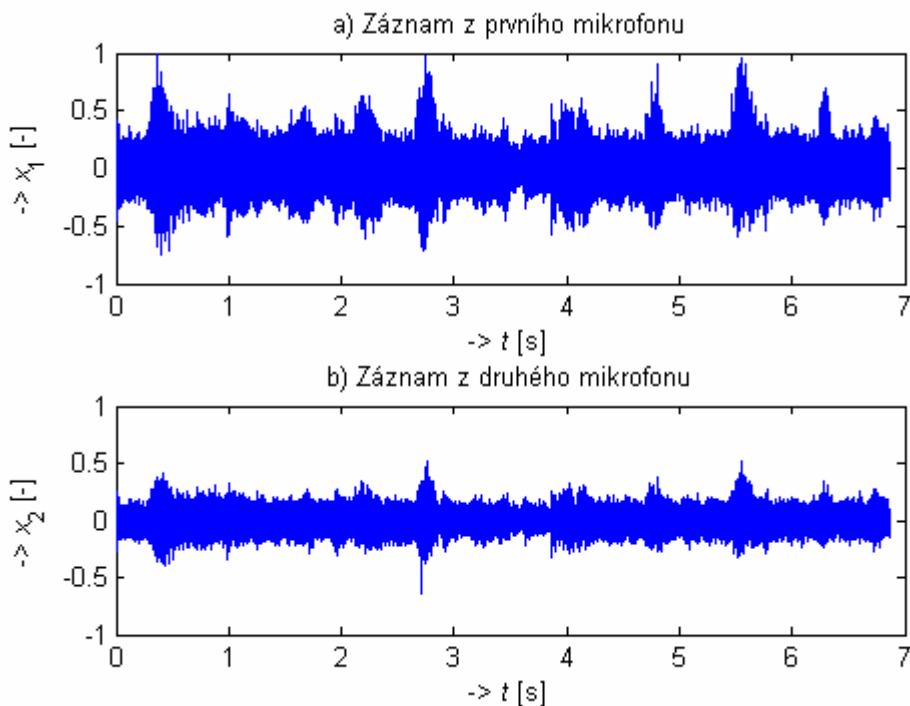
Obr. 7.3: Průběhy smíšených signálů při separaci řeč-řeč

Ze srovnání průběhů zachycených (obr. 7.3) a separovaných (obr. 7.4) signálů a zvláště pak z poslechu je zřejmé, že smíšené signály se podařilo velmi dobře oddělit.

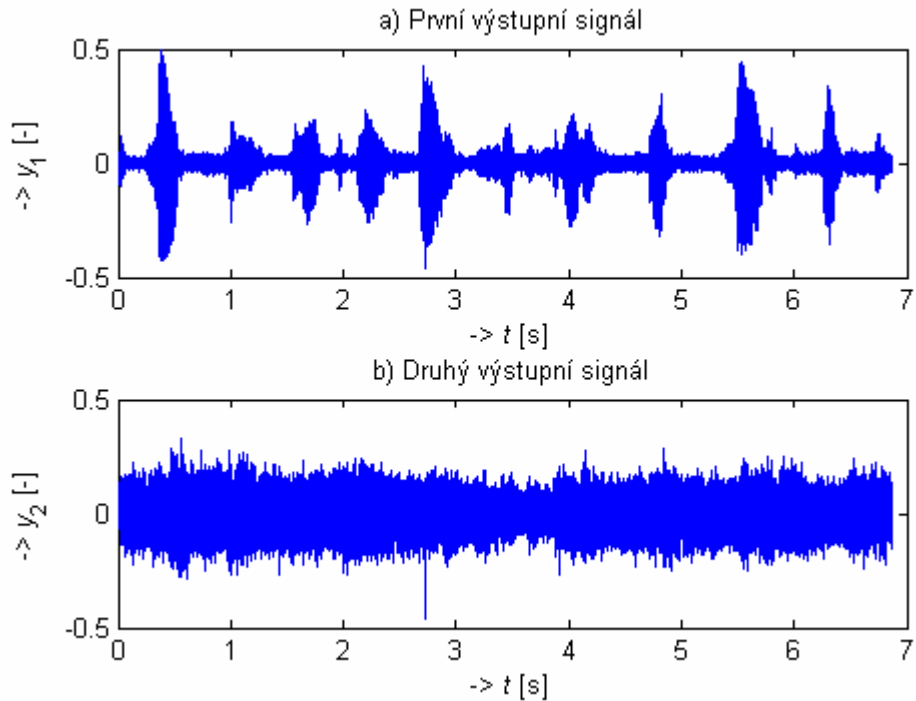


Obr. 7.4: Průběhy nezávislých komponent signálů při separaci řeč-řeč

Další testovanou situací byla nahrávka s řečí a hudbou. Smíšené signály jsou na obr. 7.5. I v tomto případě se podařilo provést úspěšnou separaci a získat tak signál, kde je v podstatě pouze řeč (a hudba je jen velmi slabá) - obr. 7.6a, a signál, kde je hudba zcela dominantní - obr. 7.6b.

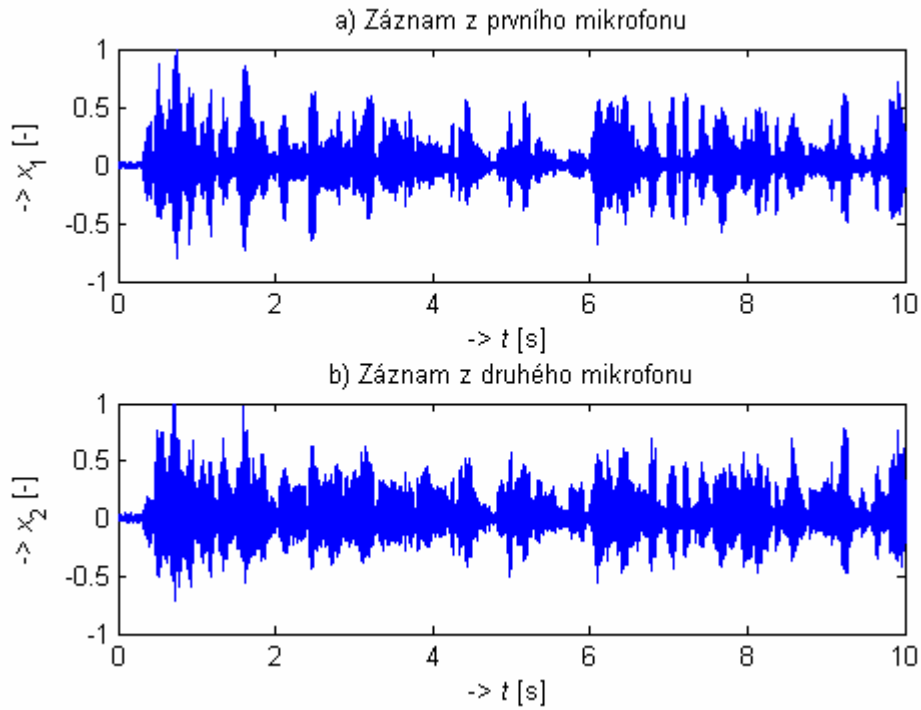


Obr. 7.5: Průběhy smíšených signálů při separaci řeč-hudba

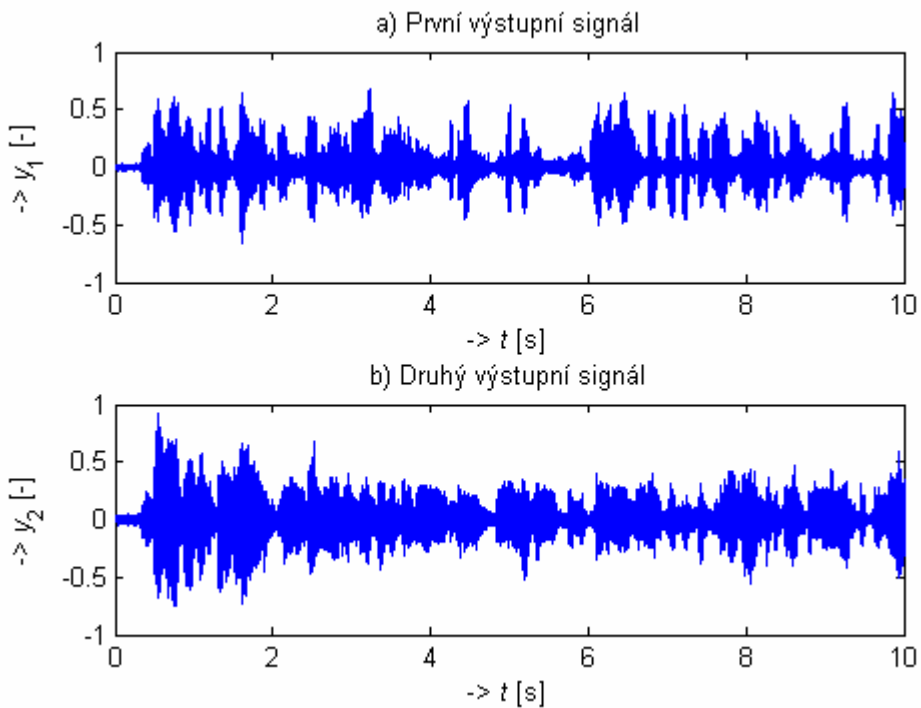


Obr. 7.6: Průběhy nezávislých komponent signálů při separaci řeč-hudba

Algoritmus ICA ve frekvenční oblasti však určitě není 100% úspěšný za všech okolností. Poslední příklad ukazuje, jak dopadla separace dvou řečových signálů nahraných ve velmi rušném prostředí (rušná kancelář, delší vzdálenost od mikrofonů). Smíšené signály jsou na obr. 7.7. Separace těchto signálů (obr. 7.8) již nedává tak výborné výsledky jako v předchozích případech, i když stále je zřejmé, že vždy jeden zdroj je v daném odhadu nezávislé komponenty výraznější. Každý výstupní signál však obsahuje i tlumenou verzi druhého signálu.



Obr. 7.7: Průběhy smíšených signálů při separaci řeč-řeč pořízených v náročnějších podmínkách



Obr. 7.8: Průběhy nezávislých komponent signálů při separaci řeč- řeč pořízených v náročnějších podmínkách

Algoritmus ICA ve frekvenční oblasti vykazuje velmi dobré výsledky při separaci konvolutorních směsí. Časová délka výpočtu tohoto algoritmu je poměrně krátká, ovlivňuje ji použité okno, délka DFT, délka řečové segmentu a další parametry. Nevýhodou algoritmu je, že je obtížně rozšiřitelný na separaci více než dvou zdrojů.

8 Závěr

Diplomová práce se věnovala vícekanálovým metodám zvýrazňování řeči, zvláště pak metodě analýzy nezávislých komponent (ICA). Cílem bylo nalézt úspěšnou metodu, která by oddělila zdrojové signály ze směsi reálných nahrávek. V práci je uvedena teorie i praxe k řešení cocktail-party efektu, tj. situace, kdy např. na večírku mluví více lidí, v místnosti se vyskytuje i další rušení a my chceme pomocí nahrávek z několika mikrofonů odlišit jednotlivé řečové projevy. Tato oblast zpracování signálů může mít i další navazující využití např. pro zvýrazňování zvuku pro lidi s poruchami sluchu nebo v systémech rozpoznávání řeči. Pro úspěšnost metody ICA je nutné předpokládat, že zdrojové signály jsou statisticky nezávislé a nemají Gaussovo rozdělení pravděpodobnosti. Před samotným zpracováním je vhodné signály upravit centrováním a bělením.

Jednotlivé metody ICA se snaží směsi rozdělovat pomocí hledání co nejmenších gaussovských vlastností signálů. Byla tak v práci vysvětlena analýza nezávislých komponent s využitím koeficientu špičatosti vycházející z toho, že gaussovská proměnná má nulový koeficient špičatosti. Metoda negativní entropie srovnává entropii nezávislých komponent a gaussovského signálu. Výpočetní náročnost metody zjednodušuje vhodná aproximace negativní entropie.

Na aproximaci negativní entropie je založen algoritmus FastICA. Tento algoritmus byl v práci vyzkoušen na umělých i reálných signálech. Uměle vytvořenými signály byly pilovitý a sinusový signál. Z grafů vyplývá, že takové signály lze pomocí FastICA separovat spolehlivě, pořadí nezávislých komponent však ovlivnit nelze. Reálné zvukové signály byly smíšeny náhodnou čtvercovou maticí v počítači. Za těchto zjednodušujících podmínek algoritmus dokáže komponenty dobře odhadnout. Algoritmus má velmi rychlou konvergenci, různými parametry můžeme také měnit způsob výpočtu nezávislých komponent a přizpůsobovat tak požadované vlastnosti. S nahrávkami pořízenými v reálném prostředí si však příliš poradit nedokáže.

Další praktickou metodou ICA je v diplomové práci algoritmus rostoucího gradientu. Na příkladu supergaussovských signálů je ukázána rekonstrukce původních signálů, signály uměle smíšené lze touto metodou úspěšně oddělit. Tento algoritmus má oproti algoritmu FastICA složitější výpočet, tím je pomalejší, provádí se zpravidla více iteračních kroků. I zde se dají nastavovat různé parametry s ohledem na charakter signálů, jež jsou zpracovávány. Ani tento algoritmus však není dostatečný pro separaci skutečných nahrávek.

Složitějším případem separace signálů je separace tzv. konvolutorních směsí - tedy nahrávek pořízených v reálném prostředí, které jsou ovlivněny dalšími vlivy jako jsou akustické vlastnosti místnosti, šum, zpoždění, odrazy od stěn, pozice či typ mikrofonů a další. Na takové složitější podmínky lze aplikovat algoritmus nazvaný Fixed-point ICA ve frekvenční oblasti, který zpracovává časově-frekvenční soubory řeči pomocí Fourierovy transformace a negativní entropie. V práci jsou ukázány příklady separace konvolutorních směsí signálů, které dokazují, že ICA ve frekvenční oblasti dokáže pracovat i s těmito složitějšími podmínkami. Nejdříve byl algoritmus testován na směsi dvou řečových signálů. Tyto signály se podařilo velmi dobře oddělit. Dále je v práci ukázáno úspěšné oddělení řeči a hudby. Velmi rušnou nahrávku dvou řečových promluv se podařilo rekonstruovat už jen částečně. V každém z výstupních signálů je druhý tlumeně obsažen.

Algoritmus ICA ve frekvenční oblasti je pro reálné situace určitě použitelný, nevýhodou ale je, že je velmi špatně rozšiřitelný na separaci více než dvou původních signálů.

Seznam literatury

- [1] HYVÄRINEN, A., KARHUNEN, J., OJA, E.: Independent Component Analysis. John Wiley and Sons, New York, 2001. ISBN 0-471-40540-X.
- [2] QUATIERI, T. F.: Discrete-Time Speech Signal Processing (Principle and Practice). Prentice-Hall, 2002 Upper Saddle River. ISBN 0-13-24292-X.
- [3] SMĚKAL, Z.: Číslicové zpracování signálů (MCSI). Elektronické texty VUT, ÚTKO, FEKT, VUT v Brně, 2007.
- [4] HYVÄRINEN, A., OJA, E.: Independent Component Analysis: Algorithms and Applications. Neural Networks Research Centre, Helsinki University of Technology, 2000.
- [5] STONE, J. V.: Independent Component Analysis: A Tutorial Introduction. MIT Press, 2004. ISBN 0262693151.
- [6] The FastICA package for MATLAB. Helsinki University of Technology, 2005, <http://www.cis.hut.fi/projects/ica/fastica/>
- [7] ZAPLATÍLEK, K., DOŇAR, B.: Matlab: začínáme se signály. BEN, Praha, 2006. ISBN 80-7300-200-0.
- [8] PRASAD, R., SARUWATARI, H., LEE, A., SHIKANO, K.: A Fixed-point ICA Algorithm for Convolved Speech Signal Separation. 4th International Symposium on ICA and BSS (ICA2003), 2003, Nara, Japan.
- [9] PRASAD, R., SARUWATARI, H., SHIKANO, K.: An ICA Algorithm for Separation of Convolutional Mixture of Speech Signals. International Journal of Information Technology Volume 2 Number 4, 2007, Nara, Japan.
- [10] KOVÁR, M.: Maticový a tenzorový počet (MMAT). Elektronické texty VUT, Ústav matematiky, FEKT, VUT v Brně, 2006.

Seznam příloh

Příloha 1: CD s textovými a demonstračními soubory